

The Neural Time Course of Semantic Ambiguity Resolution in Speech Comprehension

Lucy J. MacGregor¹, Jennifer M. Rodd², Rebecca A. Gilbert¹,
Olaf Hauk¹, Ediz Sohoglu^{1,3}, and Matthew H. Davis¹

Abstract

■ Semantically ambiguous words challenge speech comprehension, particularly when listeners must select a less frequent (subordinate) meaning at disambiguation. Using combined magnetoencephalography (MEG) and EEG, we measured neural responses associated with distinct cognitive operations during semantic ambiguity resolution in spoken sentences: (i) initial activation and selection of meanings in response to an ambiguous word and (ii) sentence reinterpretation in response to subsequent disambiguation to a subordinate meaning. Ambiguous words elicited an increased neural response approximately 400–800 msec after their acoustic offset compared with unambiguous control words in left frontotemporal MEG sensors, corresponding to sources in bilateral frontotemporal brain regions. This response may reflect increased demands on processes by which multiple alternative meanings are activated and maintained until later

selection. Disambiguating words heard after an ambiguous word were associated with marginally increased neural activity over bilateral temporal MEG sensors and a central cluster of EEG electrodes, which localized to similar bilateral frontal and left temporal regions. This later neural response may reflect effortful semantic integration or elicitation of prediction errors that guide reinterpretation of previously selected word meanings. Across participants, the amplitude of the ambiguity response showed a marginal positive correlation with comprehension scores, suggesting that sentence comprehension benefits from additional processing around the time of an ambiguous word. Better comprehenders may have increased availability of subordinate meanings, perhaps due to higher quality lexical representations and reflected in a positive correlation between vocabulary size and comprehension success. ■

INTRODUCTION

Most common words are semantically ambiguous (for a review, see Rodd, Gaskell, & Marslen-Wilson, 2002), such that their meaning depends on context. For example, “ace” can refer to a playing card or a tennis serve that an opponent is unable to return. Thus, the ability to make sense of—resolve—ambiguity is a fundamental part of speech comprehension. When listeners (or readers) encounter an ambiguous word (e.g., “ace”), semantic priming studies suggest that they automatically activate the multiple meanings of that word in parallel (irrespective of context) but, within a few hundred milliseconds, settle on a single preferred meaning (Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Swinney, 1979). Initial meaning selection operates on the information available at that time (Cai et al., 2017; Rodd, Cutrin, Kirsch, Millar, & Davis, 2013; Duffy, Morris, & Rayner, 1988; for a review, see Vitello & Rodd, 2015), which will be particularly challenging if disambiguating context is absent or delayed until after the ambiguous word. If a subsequent context supports a subordinate (less frequent, thus more unexpected) meaning, then a later process of reinterpretation is often necessary for accurate comprehension.

Individual differences in comprehension success have been associated with abilities at accessing, selecting, and reinterpreting ambiguous word meanings (Henderson, Snowling, & Clarke, 2013; Szabo Wankoff & Cairns, 2009; Gernsbacher, Varner, & Faust, 1990). Damage to the anterior temporal lobe, a region known to be associated with semantic processing in general (Patterson, Nestor, & Rogers, 2007), has been shown to impair the processing of ambiguous word meanings (Zaidel, Zaidel, Oxbury, & Oxbury, 1995), but it is still unclear how variation in comprehension ability relates to variation in the associated neural processes. The aim of the current study is to understand the neural mechanisms that support two stages of successful ambiguity resolution (initial meaning activation/selection and subsequent reinterpretation) and to explore the relationship between behavioral and neural responses to ambiguity.

The cortical network supporting ambiguity resolution in sentences was first reported in a fMRI study by Rodd, Davis, and Johnsrude (2005). Listeners were presented with high-ambiguity sentences containing multiple ambiguities (e.g., “there were DATES and PEARS on the kitchen table”), and the associated BOLD activation was contrasted with that produced by low-ambiguity control sentences (e.g., “there was beer and cider on the kitchen shelf”). Additional activation during comprehension of high-ambiguity sentences was observed in bilateral inferior frontal gyrus (IFG),

¹University of Cambridge, ²University College London, ³University of Sussex

particularly in pars triangularis and opercularis, and in left posterior temporal regions, including posterior middle temporal gyrus (pMTG), posterior inferior temporal gyrus (pITG), and fusiform. These activations were observed in the absence of explicit awareness of the ambiguities and when listeners were given no explicit task, suggesting involvement of these regions when comprehension occurs automatically as in natural speech comprehension. This basic observation that semantic ambiguity resolution involves frontotemporal regions is now well established, having been replicated using fMRI for spoken (Vitello, Warren, Devlin, & Rodd, 2014; Rodd, Johnsrude, & Davis, 2012; Tahmasebi et al., 2012; Rodd, Longe, Randall, & Tyler, 2010) and written (Mason & Just, 2007; Zempleni, Renken, Hoeks, Hoogduin, & Stowe, 2007) sentences and shown to have a consistent localization across individuals (Vitello et al., 2014). This frontotemporal response to ambiguity has proven useful in translational work, for example, as a neural marker of residual semantic processing of speech at different levels of sedation (Davis et al., 2007) and as evidence for intact speech comprehension, which has prognostic value for patients diagnosed as being in a vegetative state (Coleman et al., 2007, 2009).

However, attempts to attribute specific cognitive operations like initial meaning activation/selection and subsequent reinterpretation to distinct cortical regions have been less successful. One experimental approach has been to compare neural responses to sentences containing ambiguous words with varying meaning frequencies; such sentences are expected to load on different processes in ambiguity resolution. For example, initial meaning selection is assumed to be more difficult for sentences containing ambiguous words with meanings that have similar frequencies (balanced) than for words with a more dominant meaning (biased). Conversely, reinterpretation is assumed to be more difficult or more likely when sentences are disambiguated to a subordinate (less frequent and therefore less expected) meaning. In this way, BOLD responses due to differences in meaning frequency can be related to processes at the time of ambiguity (initial meaning activation/selection) or disambiguation (subsequent reinterpretation). Using this approach, responses to subordinate meanings have been attributed to reinterpretation processes in the left (Vitello et al., 2014) or bilateral (Mason & Just, 2007; Zempleni et al., 2007) IFG, sometimes extending into superior and middle frontal areas (Mason & Just, 2007). However, pMTG/pITG has also been implicated in reinterpretation, with studies observing greater activation for subordinate meanings in left (Vitello et al., 2014) or bilateral (Zempleni et al., 2007) posterior temporal regions, though null results are also reported (Mason & Just, 2007). Initial meaning selection has also been associated with responses in the left IFG (Mason & Just, 2007), but other studies have failed to observe greater activation for balanced compared with biased ambiguous words, and hence, evidence for selection processes is currently lacking (Vitello et al., 2014).

An alternative approach to separating neural responses during initial meaning selection from those involved in subsequent reinterpretation has explored differences used in the timing of frontotemporal responses. Rodd et al. (2012) used a rapid fMRI acquisition sequence to measure the time course of the BOLD response to ambiguous sentences in which the timing of disambiguation was varied. They assumed that additional BOLD responses associated with reinterpretation (relative to unambiguous control sentences) would occur later for ambiguous sentences in which disambiguation occurred after an additional delay. Hence, they contrasted delayed disambiguation sentences, like “The ecologist thought that the PLANT by the river should be *closed down*” with immediate disambiguation sentences, like “The scientist thought that the FILM on the *water* was from the pollution” (AMBIGUOUS and *disambiguation* words highlighted). BOLD responses to immediate and delayed ambiguity resolution showed differences in timing in the left IFG and in posterior temporal areas (fusiform, pITG, and pMTG) consistent with reinterpretation. Furthermore, BOLD responses were also observed in the IFG for sentences in which the disambiguating information occurred before the ambiguous word (“The *hunter* thought that the HARE in the field was actually a rabbit”). Because these sentences should not require reinterpretation Rodd and colleagues concluded that the IFG is also involved in meaning selection.

Taken together, an emerging picture of the differential contribution of inferior frontal and posterior temporal brain regions to semantic ambiguity resolution is that meaning selection may be underpinned by IFG and reinterpretation by IFG and posterior temporal areas together. However, there is a lack of consistent findings in relevant experiments perhaps due to the challenge of associating a slow BOLD response, which has a rise time of around 5 sec (Josephs & Henson, 1999; Boynton, Engel, Glover, & Heeger, 1996), with distinct neurocognitive processes that operate over a shorter time period. This leads to two problems. First, during the comprehension of a single sentence lasting less than 5 sec, the measured BOLD response to different neurocognitive events will inevitably overlap, making it difficult to tease apart initial meaning activation/selection and subsequent reinterpretation. Second, given that meaning selection is thought to occur within a few hundred milliseconds (Seidenberg et al., 1982; Swinney, 1979), the associated neural response may be transient and not detected in the BOLD signal.

Several studies have utilized more temporally sensitive measures of cognition to investigate the processing of ambiguous words. During natural reading, the duration of fixation times have been shown to be longer for ambiguous words in the absence of biasing context compared with unambiguous controls (Frazier & Rayner, 1990; although for evidence that reading times for ambiguous words with biased meanings do not differ from unambiguous controls, see Duffy et al., 1988; Rayner & Duffy, 1986). ERP studies with word-by-word presentation have

shown a sustained frontal negativity for ambiguous words presented in a semantically neutral context compared with unambiguous words (Hagoort & Brown, 1994) and for ambiguous words in a semantically neutral but syntactically constraining context compared with unambiguous controls (Lee & Federmeier, 2006, 2009, 2012; Federmeier, Segal, Lombrozo, & Kutas, 2000). These findings suggest that processing of ambiguous words is more effortful than processing words with single meanings. ERP studies using word-by-word visual presentation have also looked for effects potentially associated with reinterpretation (Gunter, Wagner, & Friederici, 2003; Hagoort & Brown, 1994). In these studies, N400 responses have been observed in response to disambiguating words that resolve an ambiguity to its subordinate meaning. However, these studies did not control for both the presence/absence of ambiguity and the word form itself. Hence, differences in word form and meaning might also be responsible for these neural effects.

In this study, we used combined magnetoencephalography (MEG) and EEG, which provides the temporal resolution required to distinguish neural responses at different time points during sentences and to relate these responses to distinct neurocognitive processes. Our volunteers listened to spoken sentences (see Figure 1A) that manipulated the presence/absence of an ambiguous word (AMBIGUITY) and subsequent disambiguation (*disambiguation*, e.g., “The man thought that one more ACE/SPRINT might be enough to win the *tennis/game*.”). These sets of sentences enable us to specify conditions and time points in which we expect either initial meaning access and selection or reinterpretation to occur.

In the absence of biasing context, an ambiguous word (ACE) should require additional meaning access and selection processes relative to a matched, unambiguous control word (SPRINT). This comparison of sentences with and without an ambiguous word (i.e., the main effect of ambiguity) provides the first experimental contrast in our study. Neural activity during and after the ambiguous word will reflect processes involved in initial meaning activation and selection that are more strongly taxed by ambiguous than control (unambiguous) words. These processes should occur before subsequent context words that drive reinterpretation.

Given that the words that precede the ambiguous word are relatively uninformative, initial meaning access and selection should result in most listeners settling on the dominant (playing card) meaning of the ambiguous word. The subsequent presentation of a sentence-final word (*tennis*) that is incompatible with the dominant meaning of ACE disambiguates the ambiguous word to its subordinate meaning. For listeners to avoid misinterpretation, resource-demanding reinterpretation processes should be triggered by the sentence-final word (*tennis*) but not by an alternative final word (*game*) that is consistent with both meanings (Rodd, Johnsrude, & Davis, 2010; Kambe, Rayner, & Duffy, 2001; Duffy et al., 1988). Because this reinterpretation process will only

occur if the sentence-final word (*tennis*) occurs in a sentence that contains the ambiguous word (ACE), the neural correlates of reinterpretation can be detected using the interaction between ambiguous words and subordinate reinterpretation, time-locked to the sentence-final word.

For both the meaning access/selection (main effect) and reinterpretation (interaction) contrasts, we measured evoked MEG/EEG responses relative to the offset of the critical words. This is a time point at which listeners have heard sufficient phonetic information to recognize the words and are therefore engaged in processing meaning. We used an active comprehension task on noncritical trials (relatedness judgment) during MEG/EEG scanning to ensure attentive listening throughout without contaminating neural measures obtained during critical trials.

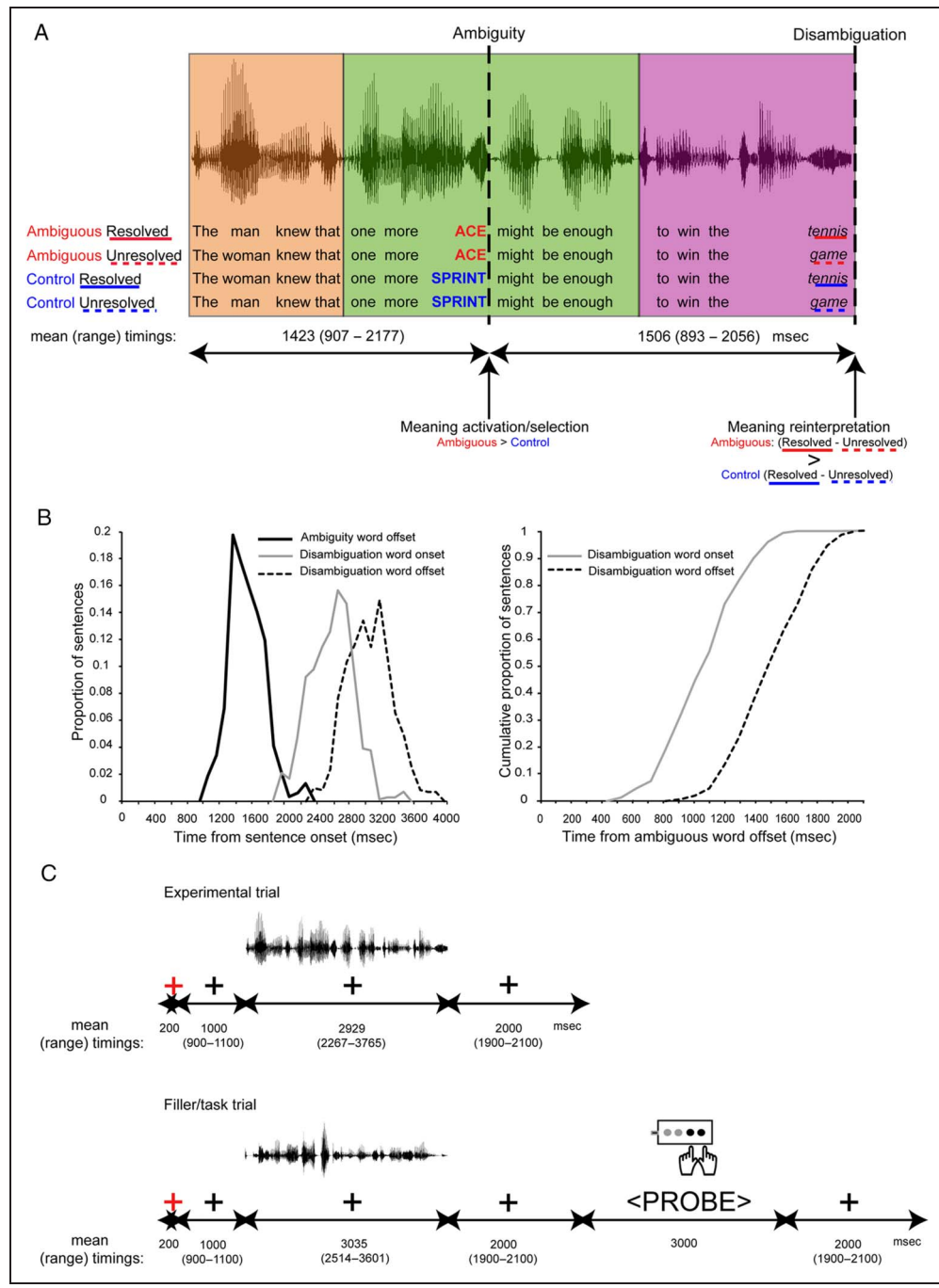
In addition to our analyses of main effects and interactions, we were also interested in relating neural responses to individual differences in sentence comprehension. We therefore administered a postscanning behavioral task to provide a trial-by-trial measure of the comprehension of critical sentences that required reinterpretation of an ambiguous word. We were interested in whether more successful ambiguity resolution would be associated with greater neural engagement or reduced processing effort at the time of ambiguity or reinterpretation. We were also interested in whether there was a relationship between comprehension and verbal and nonverbal abilities (as measured using standard vocabulary and fluid reasoning tests).

METHODS

Stimuli

Sets of 80 spoken sentences were constructed according to a 2×2 factorial design in which we manipulated (1) the presence/absence of an ambiguous word (ambiguity: ambiguous vs. control) and (2) the presence of one of two sentence-final words, which in the ambiguous sentences either disambiguated the ambiguous word so it resolved to a subordinate meaning or left it unresolved (disambiguation: resolved vs. unresolved). Because identical sentence-final words also completed the unambiguous control sentences, we also use the terms resolved/unresolved to refer to the equivalent control conditions (see Figure 1A and Table 1). Ambiguous words occurred midsentence after a neutral context that did not bias interpretation toward either meaning of the ambiguous word (mean word offset of 1423 msec after sentence onset; see Figure 1B) and were followed by additional neutral context words. In the “ambiguous-resolved” sentences, the sentence-final word disambiguated the ambiguous word toward a subordinate meaning (mean word onset and offset were 1068 and 1506 msec after the offset of the ambiguous word; see Figure 1B). In the “ambiguous-unresolved” sentences, the sentence-final word was necessarily more general so that both

Figure 1. Stimulus and experimental examples and timings. (A) Example quartet of the spoken sentence stimuli showing the four experimental conditions, which were designed to investigate neural processes occurring at two critical time points during semantic ambiguity resolution: (1) ambiguity and (2) disambiguation. MEG responses were measured time-locked to the offsets of critical words. At the time of ambiguity, responses to ambiguous words (red) were predicted to be larger than unambiguous control words (blue), reflecting more effortful semantic selection processes. At the time of disambiguation, responses to disambiguating words that resolved the ambiguity to a subordinate meaning (red solid underline) were predicted to be larger than control words that left the ambiguity unresolved (red dotted underline) and control words that completed the unambiguous sentence with each of the words used in the ambiguous sentences (blue solid/dotted underline), reflecting the greater probability of reinterpretation processes. Each sentence was combined from three fragments (highlighted with background colour) from different recordings such that linguistically identical fragments were acoustically identical across conditions, and so that the splice points occurred at least one word before and one word after the ambiguous/control word. (B) Frequency distributions of the time durations (msec) between critical words at ambiguity and at disambiguation, shown as proportions across all 640 sentences (i.e., all conditions). Durations are categorized into 100-msec time bins. The left panel displays the distribution of timings of ambiguity word offsets and of disambiguation word onsets and offsets relative to sentence onsets. The right panel shows the cumulative distribution of timings of the onsets and offsets of the disambiguation words relative to the ambiguity word offsets. The offsets of the disambiguation words occur more than 800 msec after ambiguity word offset for all sentences (i.e., at a time beyond the duration of the analysis window for the ambiguity words), and the onsets of the disambiguation words occur more than 800 msec after ambiguity word offsets for 81% of sentences. (C) Structure and timings (mean and range) of the components of the experimental trials (top) and the filler/task trials (bottom).



meanings of the ambiguous word remained plausible. Identical sentence-final words also completed the control unambiguous sentences. Sentence transcriptions and stimulus properties can be downloaded from <https://osf.io/3jhtb/>.

The critical 80 ambiguous and 80 unambiguous control words were matched on mean frequency of occurrence,

number of syllables, and number of phonemes (Baayen, Piepenbrock, & Gulikers, 1995). Sentence-final words that, in the ambiguous sentences either did or did not resolve the ambiguities, were also matched on the same factors (Table 2).

Analysis of a large database of meaning dominance ratings for single ambiguous words (Gilbert, Betts, Jose, &

Table 1. Examples of Four Stimulus Sets Heard by a Single Participant

<i>Condition</i>	<i>Lead In</i>	<i>Ambiguity Control</i>	<i>Continuation</i>	<i>Sentence-final Word</i>
Ambiguous-resolved	The man knew that one more	ACE	might be enough to win the	tennis
Ambiguous-unresolved	The woman hoped that one more	ACE	might be enough to win the	game
Control-resolved	The woman hoped that one more	SPRINT	might be enough to win the	tennis
Control-unresolved	The man knew that one more	SPRINT	might be enough to win the	game
Ambiguous-resolved	Dan looked over all the	ARTICLES	and found that most of them were	broken
Ambiguous-unresolved	Rob went through all the	ARTICLES	and found that most of them were	useless
Control-resolved	Rob went through all the	HAMMERS	and found that most of them were	broken
Control-unresolved	Dan looked over all the	HAMMERS	and found that most of them were	useless
Ambiguous-resolved	The couple thought that this	JAM	was worse than the one on the	motorway
Ambiguous-unresolved	The man heard that this	JAM	was worse than the one on the	television
Control-resolved	The man heard that this	STORM	was worse than the one on the	motorway
Control-unresolved	The couple thought that this	STORM	was worse than the one on the	television
Ambiguous-resolved	His grandfather joked that this	LEEK	was the biggest he had ever	cooked
Ambiguous-unresolved	His uncle claimed that this	LEEK	was the biggest he had ever	found
Control-resolved	His uncle claimed that this	TROUT	was the biggest he had ever	cooked
Control-unresolved	His grandfather joked that this	TROUT	was the biggest he had ever	found

Rodd, 2017) created using standard word association methods (Twilley, Dixon, Taylor, & Clark, 1994) confirmed that the ambiguous-resolved condition sentences utilized the subordinate meaning of the ambiguous words, with the exception of a small number of sentences (mean dominance = 0.23, *SD* = 0.21, max = 0.76, min = 0). The ambiguous-resolved condition sentences were also tested using the word association method to ensure that disambiguation to the subordinate meaning occurred only at the sentence-final word (not earlier): Participants

who did not take part in the MEG experiment were presented with the ambiguous-resolved condition sentences without the final word, followed by the isolated ambiguous word, and asked to generate a word that was related to the ambiguous word as used in the sentence. Dominance ratings of the ambiguous words in context were comparable to those taken from the database of isolated ambiguous words (mean dominance = 0.25, *SD* = 0.16, max = 0.53, min = 0). Meaning dominance ratings can be downloaded from <https://osf.io/3jhtb/>.

Table 2. Descriptive Statistics, Mean (*SD*), for Key Properties of the Four Keywords

Property	Keyword		Sentence-final Word	
	Ambiguous	Control	Resolved	Unresolved
<i>n</i>	80	80	80	80
Frequency (log-transformed)	1.61 (0.51)	1.3 (0.61)	1.42 (0.64)	1.83 (0.70)
No. syllables	1.28 (0.48)	1.32 (0.47)	1.98 (0.89)	1.98 (0.95)
No. phonemes	3.85 (1.13)	4.05 (1.15)	5.40 (2.08)	5.06 (1.92)

The 80 ambiguous words and their matched unambiguous control words were used to create 80 stimulus sets. Within each set, there were two lead-in contexts (e.g., “The man knew...” and “The woman hoped...”), which were crossed with the ambiguous/control words and the sentence-final ambiguity-resolving/unresolving words, thus resulting in eight stimulus versions. For each set, the eight versions were separated into two lists—List A and List B, each containing one sentence from each of the four conditions such that each ambiguous/control word and sentence-final word occurred twice, but following a different lead-in context. Participants heard stimuli from either List A or List B (320 stimuli in total), which meant that although they heard each ambiguous word twice—in a resolved and an unresolved sentence—each followed a different lead-in context (see Figure 1A and Table 1 for examples of stimulus sets heard by one participant).

The stimuli were spoken by a native speaker of Southern British English (author M. H. D.) and digitally recorded (44.1 kHz sampling rate) in a sound-proofed booth. For each stimulus set, all eight versions of the sentences were recorded, then six segments were extracted from the recordings, corresponding to the lead in portion (two versions), the target word (ambiguous, unambiguous) plus surrounding words, and the sentence-final word plus surrounding words (see shading in Figure 1A). The six segments were then concatenated to make the eight sentence versions, which were carefully checked to ensure no splices were audible. The procedure of splicing and then recombining segments meant that, across conditions, the critical sections of each sentence (e.g., ambiguous word, disambiguation) were acoustically identical. The exact point for splicing was chosen to ensure that the recombined stimuli sounded natural (e.g., by selecting silent periods during plosives). Stimuli were normalized within and between conditions for root-mean-square (RMS) amplitude using Praat software (from www.praat.org).

In addition to the experimental stimuli, 20 sets of filler sentences were constructed with similar lexicosyntactic structures and properties as the experimental stimuli. There were four sentence versions per set in which the ambiguous/control words were crossed with the sentence-final ambiguity-resolving/unresolving words (80 fillers in

total); as with the experimental stimuli, the ambiguous/control words and sentence-final words occurred twice but with a different lead in for each repetition. RMS amplitudes of the fillers were adjusted to match the mean RMS amplitude of the experimental files. Participants heard all filler sentences. For each of the filler sentences, probe words were selected for visual presentation in the relatedness judgment task, which was included to probe for comprehension and to ensure attentive listening. Probe words were either strongly related (50% of probes) or unrelated (50% of probes) to the meaning of the sentence meaning. The probes were never related to the unintended meaning of the ambiguous words.

Cloze Probability Test

Following a suggestion from a reviewer, we ran a sentence completion test on our four experimental sentence types to test whether there were differences in cloze probability across the four conditions. Data were collected from 77 participants (aged 20–39 years, born and residing in the United Kingdom, who had learned English as their first language and had no hearing difficulties) over the Internet using jsPsych (de Leeuw, 2015) and JATOS (Lange, Kuhn, & Filevich, 2015), following recruitment via Prolific (Palan & Schitter, 2018; Peer, Samat, Brandimarte, & Acquisti, 2016). Data from five participants were excluded (see below) and were replaced to meet our a priori goal of analyzing 72 data sets (giving us a cloze probability resolution for each item of 1.4%).

The same sets of 80 sentences from the MEG study were used in this test, except that the final words of each sentence were not presented. Thus, for each of the 80 experimental items, there were four possible sentences created by crossing the two lead-in versions with the two key words (ambiguous or control). To avoid excess stimulus repetition, each participant was tested on only two of four sentence variants (i.e., they heard each lead-in version only once, with one variant presented with an ambiguous word and the other with a control word; 160 experimental item trials in total). We counterbalanced whether ambiguous or control words were presented first for specific items, and which lead-in variant was paired with an ambiguous word

resulting in four experimental versions. Although we aimed at testing 18 participants in each of the four experimental versions, because of accidental overrecruitment in one version, we collected data from 19 participants in one version, 18 in two versions, and 17 in another version.

Participants were told that they would hear sentences in which the ending had been cut off, and their task was to complete the sentence with the word or words that first came to mind. In each trial, a spoken sentence was presented up until the splice point at which the resolved and unresolved sentences diverged acoustically (see Figure 1B; i.e., a silent period between the key word and the sentence-final word). This allowed us to avoid presenting coarticulatory or other cues that could constrain or bias listeners' choice of sentence-final words. However, because the splice point often occurred two or more words before the end of the sentence, we also presented the remaining words before the sentence-final word as written text. For example, for the item "ace," listeners would hear: "The man knew that one more ACE might be enough" (lead-in 1, ambiguous/control key word) and see: "to win the..." followed by a text entry box for a sentence completion response. For splice points occurring in the middle of a word, these words were also presented at the start of the text segment to avoid confusion. Splice points occurred at the same place for all four sentences for each item, and hence, the text presented on the screen was the same for all four versions of each sentence. In addition to the cloze task, as in the MEG/EEG experiment (see below), participants completed the Mill Hill Vocabulary Test (Raven, Raven, & Court, 1998).

Sentence continuations from each participant were scored for whether or not they matched the critical resolved/unresolved sentence. We took only the first word from each response. These first-word responses were checked for spelling errors and corrected when the intended word was obvious (six responses were excluded for being nonwords and therefore uninterpretable). We also checked whether the first-word response was a repetition of the final word(s) in the cutoff sentence and corrected where necessary (e.g., sentence: "The man asked about the nuggets and was told they were...", response: "were chicken."). Data sets from five

participants were excluded (and replaced) because (1) they produced nine or more (5+%) nonresponses or unusable/uninterpretable responses and/or (2) they scored less than 33% correct on the vocabulary test (i.e., 2.5 *SDs* below the sample mean from the main MEG study). From the 11,520 trials (72 participants × 160 sentences), 47 missing and uninterpretable responses were removed, resulting in 11,473 responses for inclusion in the analysis. A response was scored as a match if it was (1) an exact match, (2) an inflected form of the target word (e.g., "tastes" responses matched the target word "taste"), or (3) a longer or contracted form of the target word (e.g., "gymnasium" responses matched the target word "gym"). Responses were combined over participants, lead-in variants and versions.

For each of the 80 experimental items, we calculated the proportions of responses that matched the resolved sentence-final words (e.g., *tennis* and *game*) for sentences containing the ambiguous and control words (e.g., ACE and SPRINT). The resulting cloze probabilities for the critical words in our sentences were low overall (see Table 3; cloze probabilities for all stimuli can be downloaded from <https://osf.io/3jhtb/>), confirming that—as intended—the sentence-final words were only weakly constrained by the preceding context. As the distributions of cloze probabilities for the four conditions were highly skewed, with high frequencies of 0 and near-0 cloze probabilities (i.e., cases where participants never or very rarely responded with the resolved/unresolved sentence-final word), we log-transformed the cloze probabilities to make these distributions more normal. Before this transformation, any probabilities of 0 were changed to a lower bound probability ($\frac{1}{2}$ divided by the total number of responses for that condition) to avoid undefined values that result from taking the natural log of 0.

To quantify the degree of experimental control achieved in our materials, log-transformed cloze probabilities were entered into a Bayesian repeated-measures ANOVA with default priors (JASP Team, 2019; Morey & Rouder, 2015; Rouder, Morey, Speckman, & Province, 2012). This analysis allows us to test for reliable differences in cloze probabilities between conditions as in a conventional ANOVA but importantly to also assess evidence for the null hypothesis (i.e., that our sentence

Table 3. Descriptive Statistics for Cloze Proportions across the Four Sentence Conditions Shown by Key Word (Ambiguous or Control) and Sentence-final Word Response (Matching the Resolved Word or Matching the Unresolved Word)

Keyword	Sentence-final Word Response	Mean Cloze	SD	Median Cloze	Range
Ambiguous (e.g., ACE)	Resolved (e.g., <i>tennis</i>)	0.03	0.06	0.00	0.00–0.34
	Unresolved (e.g., <i>game</i>)	0.09	0.15	0.03	0.00–0.61
Control (e.g., SPRINT)	Resolved	0.06	0.11	0.00	0.00–0.53
	Unresolved	0.08	0.14	0.01	0.00–0.60

materials were well-matched as intended). We included within-item factors for word type (ambiguous or control) and sentence-final word response type (resolved or unresolved word). Model comparisons provide very strong evidence for a difference between resolved and unresolved words ($BF_{10} = 43.217$)—indicating, as expected, that the more specific resolved words (e.g., *tennis*) were less predicted than the more generic unresolved words (e.g., *game*). Model comparisons provide moderate evidence for the null hypothesis that there is no difference between cloze probabilities following ambiguous (e.g., ACE) and control (e.g., SPRINT) words, ($BF_{10} = 0.130$). Most importantly, however, model comparisons also provide moderate evidence for the null hypothesis that the interaction between ambiguity and resolved/unresolved final words is absent ($BF_{10} = 0.258$). Based on standard interpretations of Bayes factors (Lee & Wagenmakers, 2014), this suggests that it is approximately four times more likely that the interaction is absent than present. This therefore makes us confident that any interaction in MEG/EEG response amplitude at the sentence-final word will not be attributable to differences in cloze probabilities.

Participants

Twenty right-handed native British English speakers with normal hearing and no record of neurological diseases took part in the study for financial compensation. Ethical approval was issued by Cambridge Psychology Research Ethics Committee (University of Cambridge), and informed written consent was obtained from all volunteers. No participants had taken part in any of the pretests described or had previously heard the sentences used. Data from four participants were excluded because of high noise in MEG or EEG (greater than 50% of trials were rejected during data processing; see Methods); we report data from 16 participants (10 women), aged 20–39 years (mean = 26.5 years, $SD = 6$ years).

Experimental Procedure

Experimental stimuli from List A or List B were presented auditorily (through in-ear headphones connected via tubing to a pair of Etymotic drivers, www.etymotic.com) in four blocks (80 stimuli in each block; 320 stimuli in total) interspersed with the fillers (20 stimuli in each block; 80 stimuli in total) using E-Prime 2 software (Psychology Software Tools). The four sentences from each stimulus set appeared in separate blocks to avoid repetition of the key words within a block. Across participants, the order of blocks within the list was counterbalanced according to a Latin square design, such that each condition appeared before and after the other conditions for an equal number of times. Each participant heard a different pseudorandomized version for each block. Within a block, there were no more than three sequential presentations of an ambiguous stimulus and no more than two sequential

presentations of stimuli of a particular condition. There were no more than two sequential presentations of fillers/task trials and no more than 10 trials between two fillers/task trials.

Figure 1C, D shows the structure of the experiment. The start of an experimental trial was signaled to the listener by a red fixation cross (200 msec) visually presented on the screen, during which they were encouraged to blink if necessary. The fixation turned black during a silent period (jittered 1000 ± 100 msec) and remained on the screen throughout the duration of the spoken sentence (2267–3765 msec) and for a postsentence silent period (jittered 2000 ± 100 msec). The first part of a filler/task trial followed an identical structure, but spoken sentences were always followed by a relatedness judgment task in which single words were presented visually (3000 msec), followed by a black fixation cross (jittered 2000 ± 100 msec), and participants had to respond whether the word was related or unrelated to the meaning of the sentence they had just heard.

Behavioral Measures

Participants also performed a number of behavioral tasks, allowing us to assess individual differences in comprehension skill, verbal knowledge, and nonverbal ability. Following the MEG/EEG recording, we tested participants' comprehension of the critical sentences in which an ambiguous word was resolved to a subordinate meaning. Participants listened to the 80 ambiguous-resolved sentences they had heard during the MEG/EEG session, each followed by auditory presentation of the ambiguous word from that sentence. They were asked to explain the meaning of that word, as it was used in the preceding sentence, by typing in a synonym or a definition. They were not explicitly told that the words to which they had to respond were ambiguous. These responses were subsequently scored by a native English speaker, naïve to the purpose of the experiment, who indicated whether participants generated the subordinate or dominant meaning of these words.

Participants' vocabulary knowledge was tested using the 34-question multiple-choice Mill Hill Vocabulary Test (Raven et al., 1998). We also measured participants' nonverbal ability with the Cattell 2a Culture Fair Test (Cattell & Cattell, 1960), composed of four multiple choice subtests in which participants (1) complete a sequence of drawings, (2) select the odd one out from a set of drawings, (3) complete a pattern, and (4) identify which drawing fulfils the criteria of an example. Following scoring of the individual behavioral tests, we assess across-participant correlations between test scores using Pearson correlations.

MEG and EEG Data Acquisition and Preprocessing

Magnetic fields were recorded (sampling rate 1000 Hz, bandpass filter 0.03–330 Hz) using a 306-channel Vectorview system (Elekta Neuromag), which contained one

magnetometer and two orthogonal gradiometer sensors at 102 locations within a helmet. Electric potentials were simultaneously recorded from 70 Ag/AgCl electrodes positioned according to the 10–10 system and embedded within an elasticated cap (Easy Cap). Additional electrodes positioned on the nose and one cheek were used as a reference and the ground, respectively. Vertical and horizontal electrooculograms were monitored with electrodes placed above and below the left eye and either side of the eyes, respectively. Electrocardiogram was recorded with electrodes placed at the upper left and lower right area of the torso. Head position relative to the sensor array was recorded (using the Elekta Neuromag cHPI protocol with sampling rate of 200 Hz) by using five head position indicator (HPI) coils that emitted sinusoidal magnetic fields (293–321 Hz). Before the recording, the positions of the HPI coils and 70 EEG electrodes relative to three anatomical fiducials (nasion, left and right preauricular points) were digitally recorded using a 3-D digitizer (Fastrak Polhemus). Approximately 80 additional head points over the scalp were also digitized to allow the offline reconstruction of the head model and coregistration with individual MRI images.

MEG and EEG Data Processing

To minimize the contribution of magnetic sources from outside the head as well as any artifacts closer to the MEG sensor array, the data from the 306 MEG sensors were processed using the signal space separation method (SSS; Taulu & Kajola, 2005) and its temporal extension (tSSS; Taulu & Simola, 2006), as implemented in Maxfilter 2.2 software (Elekta Neuromag): MEG sensors that generated poor quality data were identified and data-interpolated magnetic interference from nonneural sources was suppressed (tSSS buffer of 10 msec and correlation threshold of .98). Within-block movement in head position (as measured by HPI coils with HPI step set to 10 msec) were compensated and data interpolated to adjust for head movement between blocks (interpolation to the first block). Finally, data were downsampled to 250 Hz.

Subsequent preprocessing was performed using MNE Python version 0.14 (Gramfort et al., 2013, 2014). For each participant, continuous data from the four recording blocks were concatenated and visually inspected, and bad EEG channels were identified. To identify components associated with eye blinks and cardiac activity and reduce their contribution to the data, an independent component analysis (ICA; FastICA method) was performed on the raw data (filtered 1–45 Hz, data from bad EEG channels excluded). Before fitting and applying the ICA, the data were whitened (decorrelated and scaled to unit variance—“z-standardized”—also called sphering transformation) by means of a PCA. The number of PCA components entering the ICA decomposition was selected such that a cumulative variance of 0.9 was explained. Bad EEG channels were interpolated after ICA using spherical spline

interpolation (Perrin, Pernier, Bertrand, & Echallier, 1989), continuous data were filtered (fourth order Butterworth, 0.1–40 Hz), and EEG data were rereferenced to the average over all EEG channels suitable for source analysis. Long epochs were created around the offset of the critical words at the two time points of interest (ambiguity: –2800 to 2500 msec; disambiguation: –4400 to 1500 msec), and each data point was baseline-corrected using mean amplitude in the silent period before the sentence onset (ambiguity: –2800 to –2400 msec; disambiguation: –4400 to –4000 msec).

We chose to time-lock MEG and EEG responses to word offset because, at this point, listeners would have sufficient phonological information to recognize the critical words. Because many of our critical words were monosyllabic, word recognition was unlikely to occur before this time point (Marslen-Wilson, 1987). Subsequent processing and analyses were performed on shorter epochs before and after these word offsets (ambiguity: –200 to 800 msec; disambiguation: –500 to 1500 msec). These time windows were chosen in advance based on our expectations regarding the timing of neural responses associated with initial meaning selection and reinterpretation and on the known timing of the critical words in our stimuli (Figure 1B). In all sentences, there was at least 800 msec between the ambiguous word offset and disambiguation word offset (Figure 1B, right, dotted line), and in 81% of sentences, there was at least 800 msec between ambiguous word offset and disambiguation word onset (Figure 1B, right, solid line); thus, we could be confident that effects before 800 msec should be attributable to initial meaning activation and selection triggered by the ambiguity rather than subsequent reinterpretation triggered by the disambiguating word. Epochs were rejected when peak-to-peak amplitudes within the epoch exceeded the following thresholds: 1000 fT/cm in gradiometers, 3500 fT in magnetometers, and 120 μ V in EEG (mean rejection rates: targets 13.3% trials, sentence-final words 21.1% trials), and the remaining epochs were averaged across conditions.

Sensor Space Analysis

Before analysis, between-participant differences in head positions within the helmet were calculated and compensated. To do this, we calculated the mean sensor array across participants and then identified the participant closest to this average (according to both translation and rotation parameters). MEG data from all participants were transformed to this common sensory array using the “-trans” option in MaxFilter 2.2 software (Elekta Neuromag). Data were then analyzed separately for gradiometers, magnetometers, and EEG. Before the gradiometer analysis, for every participant and condition, data from each of the 102 sensor pairs were combined by taking the RMS of the two amplitudes:

$$rms(g) = \sqrt{\frac{g^1 + g^2}{2}}.$$

This is a standard procedure in MEG

analysis, which removes information about the direction of the two orthogonal gradients at each location. The directions of the gradients vary across locations with respect to the brain, and thus are not meaningful for the purposes of our experimental questions. Before EEG analysis, the data were rereferenced to the average of left and right mastoid recordings to allow data to be more comparable to most previous research on language (note that average referencing is required for combined MEG/EEG source analysis).

Between-condition differences were assessed using nonparametric cluster-based permutation tests (Maris & Oostenveld, 2007) to correct for multiple comparisons in time and space. Using this method, conditions were compared, and a t value was calculated for every time point and every sensor. All samples with t values greater than a threshold equivalent to $p < .05$ ($t = 1.753$, one-tailed; $t = 2.131$, two-tailed) were selected and clustered based on temporal and spatial adjacency, and then cluster-level test statistics were calculated by summing all t values in a cluster. To evaluate significance, the maximum cluster-level test statistic was compared against a null distribution generated by permutations: The participant-specific averages were randomly permuted within each participant (5000 times), and the Monte Carlo method was used to create an approximation of the distribution of the test statistics under the null hypothesis. The Monte Carlo p value is the proportion of cluster-level test statistics from the permutation distribution that is larger than the observed cluster-level test statistic. Clusters in which the p value was smaller than the critical alpha level of .05 support the conclusion that the two conditions are significantly different. Across participants, we tested for correlations between the amplitude of neural responses and behavioral scores, using the mean amplitude across the significant sensor time points within the cluster.

Analyses focused on responses at the time of ambiguity and at the time of disambiguation (Figure 1A). To identify neural processes associated with initial meaning activation or selection at the time of ambiguity, we tested for a directional main effect of ambiguity; that is, whether ambiguous words elicit greater neural responses than the unambiguous control words. To identify neural processes associated with reinterpretation at the time of disambiguation, we tested for a directional interaction between ambiguity and disambiguation. The interaction allowed us to avoid confounds due to differences in the informativeness of the sentence-final words within each stimulus set (e.g., *tennis* necessarily has a more specific meaning than *game*). Specifically, disambiguating sentence-final words that resolve the ambiguity to a subordinate meaning should elicit greater activity than sentence-final words that leave the ambiguity unresolved, and this difference in activation should be greater than the difference between responses to the acoustically identical sentence-final words in an unambiguous sentence. For the gradiometer analyses, we performed one-tailed tests because the data had been rectified using RMS transformation and so values

were all positive and monotonically linked to underlying neural activity. We could therefore be confident that ambiguous words would lead to increased signal compared with control words. For magnetometer and EEG analyses, we performed two-tailed tests because we did not have specific predictions regarding the polarity of these effects. Correlation analyses assessing individual differences in comprehension were all two-tailed because, even for comparisons in which we can be confident of observing greater activity for ambiguous than for control items (e.g., ambiguous vs. control items for gradiometers), we could not anticipate whether more successful ambiguity resolution would be associated with greater neural engagement or reduced processing effort (see Taylor, Rastle, & Davis, 2013, 2014, for discussion).

Source Estimation

To estimate the neural sources underpinning the observed sensor data, we used SPM 12 (Wellcome Trust Centre for Neuroimaging). Data from all three neurophysiological measurement modalities (EEG and MEG magnetometers and gradiometers) were integrated using multimodal source inversion, which has been shown to give more precise localization than that obtained by considering each modality in isolation (Henson, Mouchlianitis, & Friston, 2009). With such an approach, sensor types with higher estimated levels of noise contribute less to the resulting source solutions. For each participant, high-resolution structural MRI images (T1-weighted) were obtained using a GRAPPA 3-D MPRAGE sequence (resolution time = 2250 msec, echo time = 2.99 msec, flip angle = 9%, and acceleration factor = 2) on a 3T Tim Trio MR scanner (Siemens) with $1 \times 1 \times 1$ mm isotropic voxels. For each individual, the structural MRI image was normalized to the standard Montreal Neurological Institute (MNI) template brain. The inverse normalization parameters were then used to spatially transform canonical meshes for the cortex (8196 vertices) and scalp and skull (2562 vertices) to the individual space of each participant's MRI. Sensor locations and the scalp meshes were aligned using the three fiducial points measured during digitization with those identified on the MRI scan and with the digitized head shape. Forward models to specify how any given source configuration appears at the sensors were created separately for MEG using a single-shell model and for EEG using a boundary element model (following the recommendations specified in Litvak et al., 2011).

Source inversion was performed using the distributed L2-minimum-norm method (no depth weighting), which attempts to minimize overall source power while assuming all currents are equally likely to be active (Dale et al., 2000). An additional constraint was imposed (SPM "group inversion", as recommended in Litvak et al., 2011), whereby responses for all participants should be explained by the same set of sources, which has been shown to improve group-level statistical power (Litvak & Friston, 2008). In

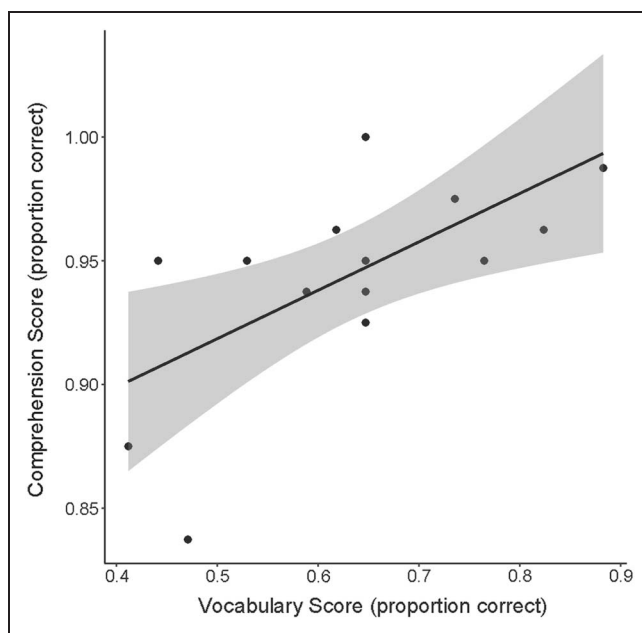


Figure 2. Positive correlation between participants' vocabulary score and their score on the post-MEG comprehension test for the ambiguous-resolved sentences. Shaded areas show the 95% confidence interval of the regression line.

brief, the procedure involves (1) realigning and concatenating sensor-level data across participants, (2) estimating a single source solution for all participants, and (3) using the resulting group solution as a Bayesian prior

on individual participant inversions. Thus, this method exploits the availability of repeated measurements (from different participants) to constrain source reconstruction. Importantly, however, the method does not bias activation differences between conditions to a given source. Source power (equivalent to the sum of squared amplitude) in the 0.1–40 Hz range was calculated from the resulting solutions and converted into 3-D images. Significant effects from sensor space were localized by taking the mean 3-D source power estimates across the relevant time windows and mapping the data onto MNI space brain templates. Between-condition differences were calculated, and statistical significance in each voxel was assessed with a series of one-sample *t* tests at the group level (i.e., mean signal divided by cross-participant variability). Because the aim of the source reconstruction was to localize significant sensor space effects, results are displayed with an uncorrected voxel-wise threshold ($p < .05$, Gross et al., 2013)

RESULTS

Behavioral Results

On the semantic relatedness judgment task, participants scored highly overall (mean proportion correct = 0.93, $SD = 0.05$), indicating they had listened attentively to the sentence stimuli. Overall, participants performed well on the post-MEG/EEG comprehension test, indicating successful disambiguation of the ambiguous-resolved

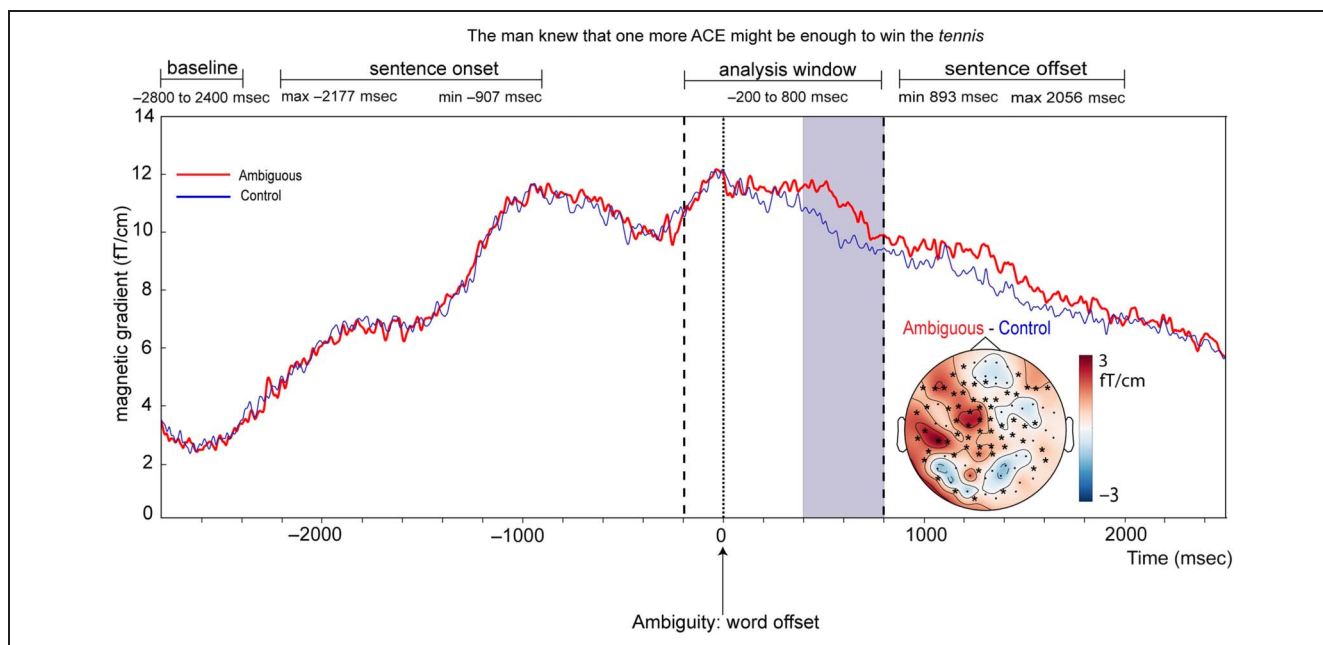


Figure 3. Evoked response at the time of ambiguity for gradiometers. Responses illustrate significantly greater activation for ambiguous (red line) compared with unambiguous control words (blue line) corresponding to a cluster in the data from gradiometer pairs (RMS transformed) beginning approximately 400 msec after word offset, which was prominent over left frontotemporal sensors (analysis time window of -200 to 800 msec relative to word offset). Responses are averaged over all sensors contributing to the significant cluster (highlighted on the topographic plot). Topographic plot shows the distribution over the scalp of the between-condition difference (Ambiguous–Control), averaged over the maximal temporal extent of the cluster (highlighted in purple).

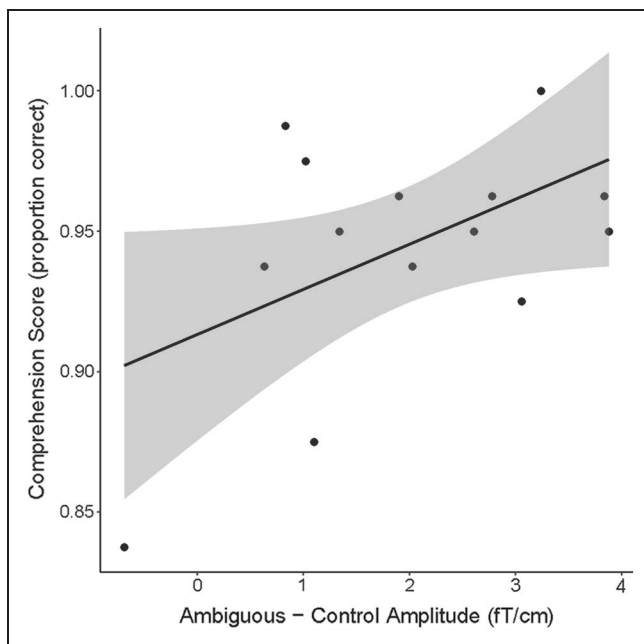


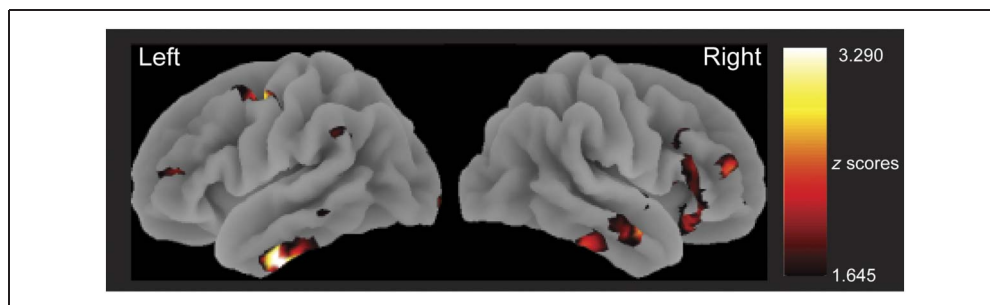
Figure 4. Positive correlation between the amplitude of the MEG effect at the time of ambiguity (Ambiguous–Control) and comprehension scores across participants. Shaded areas show the 95% confidence interval of the regression line.

sentences (mean = 0.94, $SD = 0.04$ proportion correct; scores for one participant were inadvertently not recorded resulting in $n = 15$ for analyses of comprehension scores). Nonverbal IQ scores were above average for the general population (mean = 130.3, $SD = 15.8$ normalized scores). On average, participants knew around two thirds of the words in the vocabulary test (mean = 0.63, $SD = 0.12$ proportion correct). Correlational analysis revealed a positive correlation between sentence comprehension and vocabulary scores, $r(15) = .638$, $p = .0105$ (Figure 2). There were no reliable correlations between any of the other behavioral measures.

MEG/EEG Responses at the Time of Ambiguity

Statistical analysis in sensor space revealed significant effects for gradiometers only (there were no significant clusters for magnetometers or EEG). At the offset of the ambiguous word, there was significantly greater activity

Figure 5. Source localization of the ambiguity-associated response shown in sensor space analysis in Figure 3. Results show activations displayed at $p < .05$ (uncorrected) for clarity.



in response to ambiguous compared with unambiguous control words observed in a single sensor time cluster from approximately 400 to 800 msec after word offset and most pronounced over left frontotemporal sensors (cluster: 392–800 msec, $p = .034$, one-tailed; Figure 3). Across participants, the amplitude of this response (averaged over significant sensor time points) showed a marginally significant positive correlation with comprehension scores, $r(14) = .51$, $p = .052$, two-tailed (Figure 4). These analyses included responses to all trials irrespective of whether the sentence was correctly interpreted. To further explore the relationship between MEG responses and successful comprehension, we reanalyzed the data excluding trials from sentences that were incorrectly understood in the post-MEG comprehension test (one participant was excluded because of a failure in recording the comprehension data). The MEG response at the time of ambiguity remained statistically reliable (cluster: 372–800 msec, $p = .025$, one-tailed) and the cross-participant correlation with comprehension remained marginally significant, $r(14) = 0.45$, $p = .092$, two-tailed. Because, on average, only 6% of sentences were misunderstood, there were insufficient trial numbers to explore comprehension failures in more detail.

To confirm that the ambiguity response occurred before the presentation of any disambiguating information, we carried out a post hoc analysis in which we excluded those trials in which the sentence for at least one condition had less than an 800-msec delay between target word offset and the onset of disambiguating words. That is, we excluded those items for which our analysis window could include a response to the onset of disambiguation words. This resulted in the exclusion of the sentences for 19/80 ambiguous words (76/320 sentences per participant); in the remaining sentences, the onset of the disambiguating word started after the end of the analysis time window (defined a priori as -200 to 800 msec relative to target word offset). Reanalysis of this subset of trials still showed a significant ambiguity effect (cluster: 304–800 msec, $p = .38$, one-tailed) confirming that these effects are due to ambiguous words and not subsequent disambiguation.

Source localization of the significant neural response to ambiguous words showed cortical generators in frontotemporal regions bilaterally (Figure 5; numbered source clusters are reported in Table 4). On the left, increased

Table 4. Peak Voxel Locations (in MNI Space) and Summary Statistics from Source Analysis of the Response to Ambiguity

Cluster	Voxels (<i>n</i>)	Region	Coordinates (mm)			
			<i>x</i>	<i>y</i>	<i>z</i>	<i>Z</i>
1	577	Left inferior temporal gyrus	-52	-14	-36	3.41
2	204	Left precentral gyrus	-36	-4	48	3.12
3	281	Right middle temporal gyrus	58	-6	-22	2.75
4	226	Left calcarine sulcus	-12	-100	-8	2.71
		Left calcarine sulcus	-2	-96	4	1.97
5	218	Right middle frontal gyrus	38	46	12	2.56
6	236	Right inferior temporal gyrus	54	-32	-26	2.49
7	76	Right inferior frontal gyrus (pars triangularis)	36	20	26	2.24
8	463	Right inferior frontal gyrus (pars orbitalis)	48	28	-10	2.23
		Right inferior frontal gyrus (pars orbitalis)	36	26	-18	2.19
		Right inferior frontal gyrus (pars triangularis)	54	26	14	2.07
9	78	Right middle frontal gyrus (orbital)	34	56	-2	2.1
		Right superior frontal gyrus (orbital)	26	48	-4	2.01
		Right superior frontal gyrus (orbital)	24	56	-4	1.71
10	69	Left middle frontal gyrus	-36	44	10	2.05
11	27	Left supramarginal gyrus	-58	-44	32	2.02
12	33	Left inferior frontal gyrus (pars opercularis)	-40	16	22	2.01
13	48	Right inferior frontal gyrus (pars opercularis)	42	18	10	1.93
14	33	Right calcarine sulcus	10	-96	2	1.85
15	44	Left inferior temporal gyrus ^a	-62	-40	-14	1.81
16	47	Right superior temporal gyrus	58	0	-2	1.8
17	34	Left occipital pole	-16	-96	-18	1.75
		Left occipital pole	-32	-92	-18	1.68

Regions are labeled using the AAL atlas (Tzourio-Mazoyer et al., 2002). Activations are thresholded voxel-wise at $p < .5$ (uncorrected) and cluster-wise at $k > 25$ voxels.

^aThis cluster borders the left middle temporal gyrus and in the Harvard–Oxford atlas is labeled as such.

power for ambiguous compared with unambiguous control words was seen in the anterior portion of the ITG extending posteriorly and on the border with MTG (Cluster 1, Cluster 15). On the right, there was an area of activation in homologous regions of the ITG (Cluster 6), which extended into the MTG (Cluster 3), and a small cluster in superior temporal gyrus (Cluster 16). There was also a cluster in supramarginal gyrus (Cluster 11). Frontally, there was a large right-lateralized cluster of activation in the IFG pars triangularis (Cluster 7), extending into IFG pars orbitalis (Cluster 8), and IFG pars opercularis (Cluster 13), and in the middle frontal and superior frontal gyri (Clusters 5 and 9). On the left, similar clusters of activation were seen in IFG pars opercularis (Cluster 12) and middle frontal gyrus (Cluster 10).

MEG/EEG Responses at the Time of Disambiguation

At the sentence-final word, nonparametric cluster-based permutation analysis revealed marginally significant interactions between ambiguity and disambiguation for gradiometers and for EEG. These arise from sensor-time clusters at around the time of word offset (Figure 6). For gradiometers, the interaction corresponded to a cluster in the left and right hemisphere, lasting from approximately 200 msec before to 200 msec after the sentence-final word (cluster: -196 to 156 msec, $p = .078$, one-tailed). For EEG, the interaction corresponded to a cluster for a central cluster of electrodes from over a similar latency range (cluster: -276 to 212 msec, $p = .081$, two-tailed). As predicted, these two marginal effects

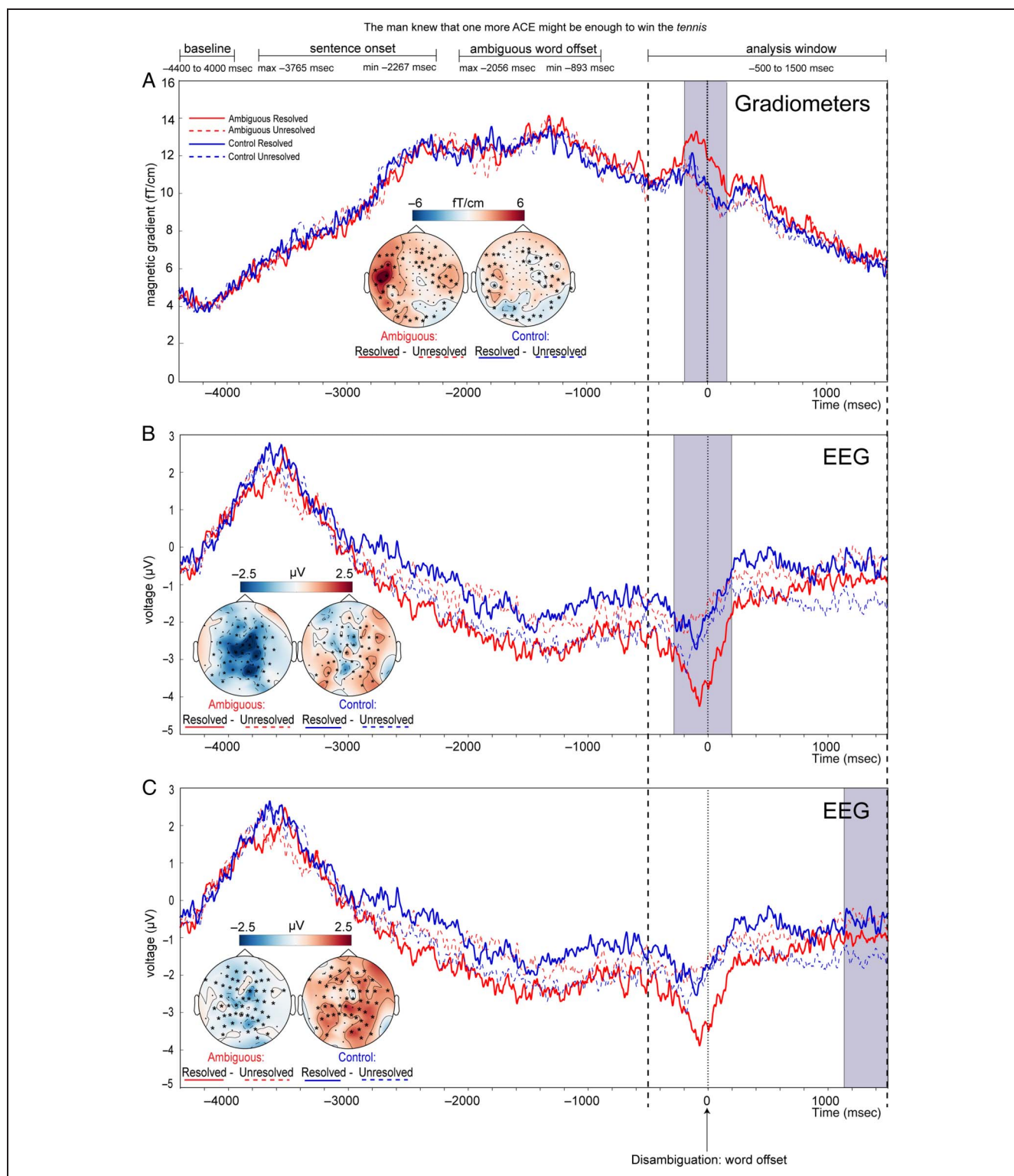
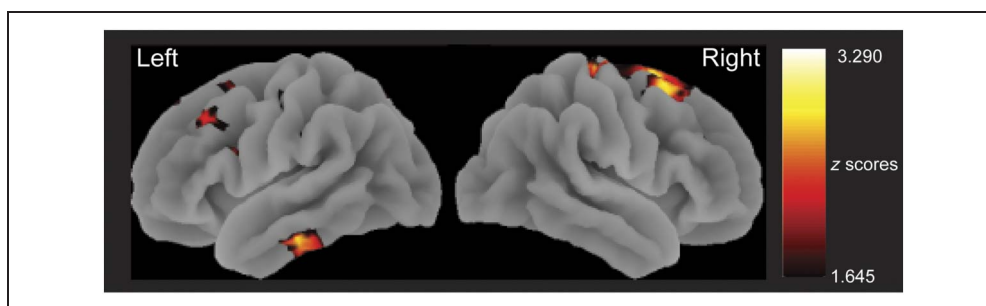


Figure 6. Evoked responses at the time of disambiguation for gradiometers (A) and EEG (B and C). Responses illustrate marginally significant interactions between ambiguity and disambiguation. For the gradiometers (A), responses illustrate significantly greater activation for sentence-final words that resolved the ambiguity (red solid line) minus words that left the ambiguity unresolved (red dotted line) compared with the activation difference between identical sentence-final words (blue solid and blue dotted lines) that completed the unambiguous sentences (analysis time window of -500 to 1500 msec relative to word offset). This effect corresponds to a cluster in the data from gradiometer pairs (RMS transformed), which is prominent around word offset and is visually similar to a cluster in the data from EEG (B). There is a second cluster for EEG data (C), corresponding to a significantly greater difference in activation between the sentence-final words for unambiguous sentences than the difference when these words completed ambiguous sentences. Responses are averaged over all sensors contributing to the significant cluster (highlighted on the topographic plot). Topographic plots show the distribution over the scalp of the between-condition differences (resolved-unresolved), averaged over the maximal temporal extent of the clusters (highlighted in purple), for ambiguous and control conditions separately.

Figure 7. Source localization of the disambiguation-associated response shown in sensor space analysis in Figure 6. Results show activations displayed at $p < .05$ (uncorrected) for clarity.



reflect greater activation for sentence-final words that resolved the ambiguity to a subordinate meaning compared with words that left the ambiguity unresolved; no equivalent difference was observed for resolved/unresolved words that completed the unambiguous sentences. The EEG data also showed a marginally significant interaction for a sensor time cluster in a later time window (cluster: 1144–1500 msec, $p = .083$, two-tailed), but as can be seen in Figure 6C, the effect was driven by a greater difference between sentence-final words in the unambiguous control sentences than the ambiguous sentences. Because the direction of this interaction effect is inconsistent with any

specific functional contribution to reinterpretation, we do not consider it further.

To fully characterize the interaction of interest, we also performed post hoc simple-effect analyses. For the ambiguous sentences, sentence-final words that resolved the ambiguity elicited greater activity than those which left the ambiguity unresolved corresponding to clusters in the gradiometer (cluster: -236 to 336 msec, $p = .002$, one-tailed) and EEG data (cluster: -196 to 236 msec, $p = .047$, two-tailed). There was no significant effect for the unambiguous sentences (i.e., those that contain a control word rather than an ambiguous word). There was also greater

Table 5. Peak Voxel Locations (in MNI Space) and Summary Statistics from Source Analysis of the Response to Disambiguation

Cluster	Voxels (n)	Region	Coordinates (mm)			Z
			x	y	z	
1	1357	Right supplementary motor area	12	6	62	2.95
		Right superior frontal gyrus	24	12	60	2.93
		Right precentral gyrus	12	-22	70	2.75
2	387	Left inferior temporal gyrus	-52	-24	-26	2.82
		Left fusiform gyrus	-42	-32	-18	1.74
3	253	Left middle frontal gyrus	-36	22	40	2.35
4	150	Right middle frontal gyrus	34	44	30	2.16
		Right middle frontal gyrus	34	36	28	1.98
5	68	Left lateral occipital cortex	-8	-74	52	2.13
		Left precuneus	-4	-72	44	1.91
6	101	Left inferior frontal gyrus (pars opercularis)	-50	12	22	2.12
7	51	Left precentral gyrus	-36	-16	48	2.09
8	65	Left superior frontal gyrus	-6	36	46	1.97
9	73	Left superior frontal gyrus	-18	10	52	1.97
10	78	Right supramarginal gyrus	44	-32	42	1.96
		Right supramarginal gyrus	50	-30	48	1.96
11	27	Left supramarginal gyrus	-50	-28	28	1.92
12	25	Left supplementary motor area	-4	-4	60	1.9

Regions are labeled using the AAL atlas (Tzourio-Mazoyer et al., 2002). Activations are thresholded voxel-wise at $p < .05$ (uncorrected) and cluster-wise at $k > 25$ voxels.

activation for words resolving the ambiguity relative to acoustically identical words that completed an unambiguous sentence (gradiometers: cluster -128 to 212 msec, $p = .014$, one-tailed; cluster -436 to -100 msec, $p = .059$, one-tailed; EEG: cluster -172 to 226 msec, $p = .028$, two-tailed) but no difference between the sentence-final words that left the ambiguity unresolved compared with the same words that completed an unambiguous sentence.

To identify the source of the disambiguation effect, we performed source localization on the time window -196 to 156 msec, covering the overlapping time period of effects in the MEG gradiometer and EEG analyses. We looked for regions with increased power for words that resolved the ambiguity than for words that left the ambiguity unresolved, compared with the equivalent difference in power between identical words that completed unambiguous sentences. Results (Figure 7 and Table 5) show generators in left frontotemporal regions, including regions that overlap with those active at the time of ambiguity such as the ITG, extending to fusiform (Cluster 2). There was also a cluster in IFG pars opercularis (Cluster 6) and smaller frontal clusters in superior frontal gyrus (Clusters 8 and 9), middle frontal gyrus (Cluster 3), precentral gyrus (Cluster 7), and SMA (Cluster 13). On the right, there was a large cluster in SMA, extending to superior frontal gyrus and precentral gyrus (Cluster 1) and in the middle frontal gyrus (Cluster 4). We also saw bilateral clusters in the supramarginal gyrus (Clusters 10 and 11).

DISCUSSION

Using MEG/EEG, we investigated the spatiotemporal dynamics of semantic ambiguity resolution by recording neural responses time-locked to the offset of an ambiguous word and to a subsequent disambiguating word that resolved the ambiguity to a subordinate meaning. Building on previous fMRI research, we capitalized on the high temporal resolution of MEG/EEG to distinguish between the neurocognitive processes of initial meaning access/selection versus reinterpretation. These are functionally distinct processes that in our sentences occur just a few hundred milliseconds apart. We feel confident that we have distinguished these neurocognitive effects for two reasons. First, an increased neural response associated with the processing of ambiguous words occurred before the presentation of disambiguating information that triggers reinterpretation. Second, neural manifestations of these processes were assessed with two orthogonal statistical contrasts: Initial ambiguity processing was assessed through a main effect, whereas reinterpretation was assessed with an interaction.

At the time of ambiguity, we observed significantly greater MEG responses for ambiguous words versus unambiguous control words (Figure 3). The effect remained significant when we excluded trials in which the onset of the sentence-final word that triggers reanalysis occurred

within the analysis window. Thus, this neural effect of ambiguity was observed before the presentation or processing of disambiguating information. Furthermore, the amplitude of the MEG response at the time of ambiguity correlated positively with individual differences in comprehension skill, as measured by our post-MEG comprehension test for ambiguous-resolved sentences (Figure 4), although this effect was only marginally significant. Comprehension also correlated positively with vocabulary scores across participants (Figure 2). We discuss the cognitive processes associated with these neural responses in the next section. In a subsequent section, we then turn to neural responses at the time of disambiguation; we observed marginally greater MEG and EEG response amplitudes at the offset of sentence-final words that resolved an ambiguous word to a subordinate meaning (Figure 6).

Source estimation localized ambiguity responses to bilateral frontotemporal regions (Figure 5) and disambiguation responses to bilateral frontal and left temporal regions (Figure 7). Given the overlapping neural localization of the two cognitively distinct processes involved in ambiguity resolution, we will discuss these findings from source localization in a final section of the discussion, drawing on comparisons with the fMRI literature to inform our functional interpretation of these neural responses.

Functional Significance of Neural Responses to Ambiguity

We take the increased neural response after the offset of ambiguous words to reflect more effortful processing of words with more than one meaning compared with matched single-meaning control words. More specifically, we relate the effect to the increased demands of meaning access and selection when multiple possible meanings are known. This neural effect is consistent with fMRI studies, as well as data from eye-tracking and ERP studies on the processing of visually presented ambiguous words in sentences that we reviewed in the introduction.

Although we described the observed response to ambiguity as a neural correlate of initial meaning activation or selection, which we distinguish from subsequent reinterpretation, this still leaves details of its functional contribution unspecified. It is, thus far, unclear whether the ambiguity response reflects processes involved in either (i) accessing and maintaining multiple meanings or (ii) selecting a single meaning of an ambiguous word (e.g., by boosting or suppressing one or other meaning). Both these processes should be more engaged and/or more demanding for words with multiple meanings and hence plausibly observed in our comparison of responses to ambiguous and control words. Critical for distinguishing these two processes is the time course over which listeners select a single meaning of an ambiguous word for sentences in which prior context does not constrain the likely meaning (as in the present experiment). However,

conventional univariate analysis of MEG/EEG data cannot provide information on whether and when both meanings of an ambiguous word are active.

Several sources of experimental evidence have been used to infer the time course of meaning selection of ambiguous words in neutral context sentences. For example, cross-modal priming studies from Seidenberg et al. (1982) and Swinney (1979) are consistent with initial access to multiple meanings followed by selection of a single, dominant meaning. Swinney (1979) provides evidence for selective access by three syllables after word offset, in the time range of 750–1000 msec (p. 657), whereas Seidenberg et al. (1982) suggest it can occur sooner, within 200 msec of word offset (both studies indicate activation of both meanings at word offset but do not test additional time points). Because both sets of data include a speeded response task with latencies between 500 and 1000 msec, of which around 150 msec can be accounted for in motor response planning, the minimum time course over which multiple meanings are maintained before selection occurs is in the region of 550–950 msec. However, it is difficult to infer the specific timing of selection from these studies, in part because meaning activation is only measured indirectly (by lexical decision or naming RTs to targets related to one or other ambiguous word meaning) at discrete points in time and not to the ambiguous word itself. Nonetheless, in the context of this study, these findings would suggest that meaning selection takes place before disambiguating information is presented for the majority of our sentences.

However, successful comprehension of most of our critical sentences ultimately depends on selecting a lower frequency or subordinate meaning. Therefore, initial selection of a single dominant meaning (if that also entails full suppression of alternative meanings) would make reinterpretation even more difficult. Yet, our post-MEG/EEG comprehension test showed that, on average, listeners were able to understand more than 90% of delayed disambiguation sentences, indicating that reinterpretation was for the most part successful. Therefore, even if full suppression occurs, listeners can still semantically reanalyze the sentence when they encounter a disambiguating word that conflicts with the previously selected meaning (perhaps using phonological or working memory). Alternatively, full suppression of alternative meanings may not occur, and multiple meanings of ambiguous words remain accessible and to some degree active, at least up to the point of disambiguation. This proposal is consistent with RT data from a self-paced reading task showing that multiple meanings can be maintained over even longer delays until disambiguation (Miyake, Just, & Carpenter, 1992).

One parsimonious description of longer term maintenance of multiple meanings is through a graded constraint satisfaction process in which listeners make progressively stronger commitments over time as evidence for alternatives increases (MacDonald, Pearlmutter, & Seidenberg,

1994). By this account, neural activity after an ambiguous word reflects the activation of multiple alternative interpretations in a representational space that also provides a mechanism for meaning maintenance such that subsequent context can guide selection. In this account, there is therefore no separation of the neural resources required for initial activation and maintenance in working memory, and meaning selection. At face value, this appears consistent with source localization results that we discuss below.

One hallmark of this constraint satisfaction account is that individual differences in sentence comprehension arise from experience-dependent learning of the probabilities and regularities that underlie language rather than in some external, capacity-limited system (such as working memory; see MacDonald & Christiansen, 2002, for theoretical elaboration along with recurrent neural network implementation). The present data provide tentative findings concerning the relationship between individual differences in comprehension and neural responses to semantically ambiguous words. For a sentence in which preceding context does not provide any specific information to constrain word meaning, the activation and maintenance of multiple semantic alternatives is optimal. Hence, additional activation associated with ambiguous words should be associated with more successful comprehension. In line with this proposal, we observed a positive correlation (albeit, only marginally significant using a two-tailed test) between the amplitude of the ambiguity-related MEG response and comprehension success in individual participants. The positive relationship remained when we excluded sentences containing ambiguous words that specific participants did not interpret correctly in the post-MEG/EEG comprehension test. This association is therefore not explained by reduced responses to sentences for which listeners failed to correctly retrieve the subordinate meaning. Thus, better comprehenders show greater neural processing effort in response to ambiguous words.

We explain this correlation between neural responses and comprehension as indicating that successful comprehension of sentences containing ambiguous words requires additional processes for activation and maintenance of alternative meanings. These result in increased availability of the appropriate meaning, which is required when subsequent context resolves the ambiguity to a subordinate meaning. Interestingly, better comprehenders not only have increased availability of subordinate meanings but also achieved higher vocabulary scores. It might be that higher quality lexical representations are required both for access to low-frequency meanings of unambiguous words (for the more difficult items in the vocabulary test) and for accessing subordinate meanings of ambiguous words (as in our MEG/EEG study). Nonetheless, given the small number of participants and marginally significant results in this study, this correlation between neural activity and successful comprehension requires replication and extension. For

example, we might use more difficult sentences to directly compare neural activity associated with successful and unsuccessful ambiguity resolution or consider other predictors of individual variation to relate ambiguity resolution (specifically) and spoken language comprehension (more generally). This study showed no association between nonverbal IQ and comprehension, but our participants did not show as much variation in cognitive abilities as we might expect in the wider population. More systematic exploration with a larger group of individuals with greater variability in comprehension and measures of other cognitive factors (such as phonological short-term or working memory) would be valuable.

Functional Interpretation of Neural Responses to Disambiguation

In addition to neural activity at the time of the ambiguous word we observed a potential neural marker of reinterpretation during the presentation of sentence-final words that favor the subordinate meaning of a previous ambiguous word. Importantly, reinterpretation effects observed at sentence offset in both MEG and EEG were apparent as an interaction between the presence of an ambiguous word and a sentence-final word that mandated access to an initially nonpreferred, subordinate meaning. This statistical interaction rules out the possibility that these effects are responses simply to the presence of an ambiguous word or a more informative sentence-final word (the potentially disambiguating words necessarily referred to more specific concepts and had a lower cloze probability). Consistent with this conclusion, post hoc simple effects showed that the neural response to a sentence-final word was affected by the presence of an ambiguous word earlier in a sentence only when the sentence-final word disambiguated the ambiguity (and not if the sentence-final word left the ambiguity unresolved). Similarly, response differences between ambiguous and control words were only apparent at sentence offset if the sentence-final word served to resolve the ambiguity (but not if the sentence-final word did not conflict with the dominant meaning of the ambiguous word). Although the neural responses associated with reinterpretation in MEG (gradiometers) and EEG were only marginally significant in analyses correcting for time and sensors, the same pattern of neural difference was observed in both modalities and in overlapping time windows. This similarity gives us greater confidence in the reliability of these observations.

The approximate timing and sensor topography of neural responses to reinterpretation are broadly consistent with interpretation as an N400 effect (Kutas & Hillyard, 1980). Although the N400 has been frequently observed in the EEG and MEG literature on language processing and known to be associated with the processing of meaning, as yet there is no consensus on an underlying functional account or computational mechanisms (for a review, see Kutas & Federmeier, 2011). For example, cognitive accounts suggest it may reflect the ease of

accessing information in semantic memory (Kutas & Federmeier, 2000) or of integrating semantic information into context (Van Berkum, 2009). Computationally, it may be more generally characterized as a semantic prediction error signal (Rabovsky & McClelland, 2014), linked to changes in a probabilistic representation of sentence meaning (Rabovsky, Hansen, & McClelland, 2018).

ERP N400 responses have previously been observed in response to disambiguating words that resolve an ambiguity to its subordinate meaning (Gunter et al., 2003; Hagoort & Brown, 1994), although as discussed in the introduction, there are several differences between these previous studies and ours. First, in previous work, sentences were visually presented word-by-word, whereas our sentences were presented auditorily as connected speech. Second, previous studies did not control for both the presence/absence of ambiguity and the word form itself. We showed a statistical interaction between these two factors for sentence-final words that trigger reinterpretation effects. Unlike previous studies, this interaction cannot be due to simple differences in word form or meaning between the critical words in our sentences.

One possibility raised by a reviewer was that the neural interaction generating this N400-like response to reinterpretation could arise from differences in cloze probability between sentence-final words in our critical conditions. However, a sentence completion test on our materials showed that cloze probabilities were low overall (the median cloze probability was zero in both conditions that contained the resolved word, close to zero for the unresolved words, and did not differ between ambiguous and control words). We did not include highly constrained sentences or semantic anomalies that are typical of N400 studies. More importantly, though, a Bayesian analysis of cloze probability values provided moderately strong evidence that there was not an interaction between ambiguity and reinterpretation (i.e., this analysis provides evidence that the sentences in our critical conditions were matched for cloze probability). Hence, we can conclude that our N400-like effect of reinterpretation is not due to variation in the ease of meaning access due to cloze probability, but rather due to sentence-final words triggering reinterpretation. Nonetheless, future work to determine the functional nature of the neural response to reinterpretation would benefit from comparing this response to the semantic error response evoked by a sentence-final anomalous word. Anomalous words should trigger an N400-like response but would not result in reinterpretation, and hence, differences between anomalous words and words driving reinterpretation may be informative.

The Role of Frontotemporal Regions in Ambiguity Resolution

With regard to the anatomical questions that motivated this study, our source localization provides evidence that frontal and temporal lobe regions are activated both in

response to ambiguous words in a neutral context (before presentation of disambiguating information; Figure 5) and subsequently in response to a disambiguating word that resolves the ambiguity to a subordinate meaning (Figure 7). Previous fMRI evidence has similarly demonstrated the involvement of frontotemporal regions in ambiguity resolution (Musz & Thompson-Schill, 2017; Vitello et al., 2014; Rodd et al., 2005, 2012; Rodd, Longe, et al., 2010; Mason & Just, 2007; Zempleni et al., 2007). However, unlike in fMRI, timing information from MEG/EEG allows us to confidently attribute our ambiguity and disambiguation responses specifically to initial processing of the ambiguous word and also to subsequent reinterpretation of the ambiguous word. Initial meaning activation/selection of an ambiguous word was identified through a statistical main effect, whereas subsequent reinterpretation at a disambiguating word was identified through a statistical interaction. Furthermore, responses associated with initial meaning activation/selection and subsequent reinterpretation could be separated in time; the neural response to ambiguity occurred before the onset of disambiguating words that trigger reinterpretation. Thus, these are two independent effects, and overlap of the neural sources can inform our understanding of the underlying mechanisms.

As we reviewed in the introduction, previous fMRI studies on ambiguity resolution have associated activation in IFG regions on the left (Vitello et al., 2014) or bilaterally (Rodd et al., 2012; Mason & Just, 2007; Zempleni et al., 2007) with reinterpretation, and in one study, activation extended into superior and middle frontal areas (Mason & Just, 2007) in line with the left IFG and superior and middle frontal clusters shown here. Only two previous fMRI studies on ambiguity resolution tentatively associated initial meaning selection with activation in IFG (Rodd et al., 2012; Mason & Just, 2007). Consistent with previous conclusions and our findings that IFG is active both during initial meaning selection and subsequent reinterpretation, one dominant proposal regarding the functional role of the left IFG is its involvement in selecting between competing semantic representation (Jefferies, 2013; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997) or resolving conflict arising from competing stimulus representations of any format (Novick, Trueswell, & Thompson-Schill, 2005; for the suggestion that IFG activation is involved in selection [or conflict resolution] rather than simply reflecting increased competition between semantic representations, see Grindrod, Bilenko, Myers, & Blumstein, 2008).

An alternative account of IFG contributions to language, the unification account (Hagoort, 2005, 2013), proposes a more general role for the IFG in combining individual words into coherent sentence- and discourse-level representations. These are processes that we might also expect to be taxed as the number of meanings increases and multiple meanings are accessed, maintained, or predicted. Although we cannot offer any evidence to adjudicate between these views, we argued above that

meaning selection of the ambiguous words in our study is likely not completed during the time window before disambiguation. This seems to favor a more graded rather than absolute form of selection, perhaps consistent with a constraint satisfaction or unification account.

Previous fMRI studies on ambiguity resolution associated activation in the left MTG and ITG/fusiform to reinterpretation (Vitello et al., 2014; Rodd et al., 2005, 2012; Rodd, Longe, et al., 2010; Zempleni et al., 2007). In line with this, localization of the MEG/EEG response to a disambiguating word indicated a source in the left ITG and fusiform, which we attribute to reinterpretation. Notably, we also observed neural sources of the MEG response to an ambiguous word in MTG and ITG bilaterally, which could be linked to initial meaning activation or selection. Posterior temporal regions have often been proposed to contribute to meaning access for isolated words (see Lau, Phillips, & Poeppel, 2008; Hickok & Poeppel, 2007). These regions would plausibly show greater activation when listeners access multiple meanings of ambiguous words: First, when ambiguity is initially encountered and again at a disambiguating word inconsistent with the previously preferred meaning, which triggers an increase in activation of an alternative. We also note that left posterior MTG activation has previously been observed in response to syntactically ambiguous words, using fMRI (Snijders et al., 2009) and MEG (Tyler, Cheung, Devereux, & Clarke, 2013), although a recent meta-analysis suggests that these posterior temporal regions are recruited more for semantic rather than syntactic processing (Rodd, Vitello, Woollams, & Adank, 2015).

We earlier characterized the MEG/EEG reinterpretation effect as resembling an N400. In line with this proposal, we note there is some overlap between source localization of the reinterpretation response to the left ITG and IFG, and regions proposed to underpin the classic N400 effect, which have been explored using fMRI and MEG/EEG (Lau, Weber, Gramfort, Hamalainen, & Kuperberg, 2016; Lau, Gramfort, Hämäläinen, & Kuperberg, 2013; Maess, Herrmann, Hahne, Nakamura, & Friederici, 2006; Halgren et al., 2002). The N400 is likely to reflect a combination of neural processes originating in multiple cortical region sources but across a number of studies; it has been proposed that the effect may originate in posterior temporal regions before being observed in more anterior portions of the temporal lobe and IFG (for a review, see Lau et al., 2008).

Interestingly, the ITG/fusiform activations we observed at the time of disambiguation and in response to ambiguity extended to more anterior and inferior temporal regions than has been seen in previous fMRI studies of ambiguity resolution. Anterior temporal activations have been less consistently observed in fMRI, perhaps because standard EPI acquisitions give relatively poor signal in these regions (Visser, Jefferies, & Lambon Ralph, 2010; Devlin et al., 2000, although see Musz & Thompson-Schill, 2017, for evidence of anterior inferior temporal

representations of ambiguous words shown by multivariate pattern analysis fMRI). However, damage to the anterior temporal lobe has long been associated with impaired semantic processing in general (Patterson et al., 2007) and of semantically ambiguous words in particular (e.g., measured by patients' ability to produce alternative interpretations of unresolved ambiguous sentences (Zaidel et al., 1995). Thus, the inferior temporal activation we observed when listeners initially encounter an ambiguous word and when disambiguating information is heard is largely consistent with other evidence for semantic contributions of these basal temporal regions.

One other point to consider is that both the frontal and temporal neural sources of responses to ambiguity and disambiguation appear to be somewhat bilateral. Previous fMRI studies have reported significant activation in right frontal regions (Mason & Just, 2007; Zempleni et al., 2007; Rodd et al., 2005), although reports of right temporal lobe responses are more limited, and a meta-analysis of fMRI studies of semantic and syntactic processing demands reveals fewer and less reliable findings of right than left frontotemporal activity (Rodd et al., 2015). However, in the absence of statistical comparison of left- and right-sided activity in fMRI or MEG/EEG, we hesitate to draw strong conclusions from these observations (see Peelle, 2012, for arguments that lateralized effects in thresholded statistical maps provide little or no evidence for functional lateralization). Furthermore, other evidence is consistent with bilateral contributions to ambiguity resolution, for example, from behavioral studies using lateralized word presentations (Faust & Gernsbacher, 1996; Burgess & Simpson, 1988) and neuropsychological studies (Tompkins, Baumgaertner, Lehman, & Fassbinder, 2000; Swaab, Brown, & Hagoort, 1998; Hagoort, 1993). Although functional imaging evidence can potentially play an important role in determining the differential contributions of the left and right hemisphere to ambiguity resolution, published studies, including the present work, have yet to report hemispheric dissociations sufficient to conclude that the left and right hemispheres make distinct functional contributions to initial meaning activation and selection.

Conclusions

Taken together with previous fMRI research, our observations suggest that both temporal and frontal regions play an important role both in initial meaning activation and selection for ambiguous words, as well as later reinterpretation triggered by a disambiguating word. Previous research has tried to fractionate frontal and temporal regions based on the time course of activation during delayed disambiguation sentences (Rodd et al., 2012) or by comparing responses to ambiguous words with balanced and biased meaning frequencies (Vitello et al., 2014; Mason & Just, 2007; Zempleni et al., 2007). However, source localization results from MEG/EEG suggest that

frontal and temporal regions play a coordinated role both in the initial interpretation of ambiguous words presented in neutral sentence contexts and subsequently when interpretations need to be revised. This proposal could be taken to challenge traditional divisions between temporal lobe contributions to semantic representation and frontal contributions to working memory or selection (see Musz & Thompson-Schill, 2017, for a recent statement along these lines).

Rather than the traditional fractionation of temporal and frontal responses, we instead propose a graded, constraint satisfaction account that elides a simple distinction between semantic representations and processing. In this account, neural activity after an ambiguous word reflects the activation of multiple alternative interpretations in a representational space that also supports neural mechanisms for meaning maintenance and eventual selection. During this time period, selection can be construed as stronger, but not exclusive, activation of a particular meaning, which can only be confirmed when disambiguating information is presented. At this point, successful meaning integration and interpretation may require reinterpretation, which can be realized in terms of a reweighting of the activation levels of different meanings. Future work to assess the representational dynamics of these frontal and temporal responses (e.g., using representational similarity or other, similar multivariate methods; Kriegeskorte, Mur, & Bandettini, 2008) might provide additional evidence for this account.

Acknowledgments

This work was supported by the Medical Research Council (SUAG/008 RG91365 to M. H. D. and SUAG058 G101400 to O. H.). The authors thank Jane Warren for helping in creating the stimuli, Maarten van Casteren for helping during data collection, and Isobel Davis helping with scoring and analyzing behavioral data.

Reprint requests should be sent to Lucy J. MacGregor, MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge, CB2 7EF, or via e-mail: lucy.macgregor@mrc-cbu.cam.ac.uk.

REFERENCES

- Baayen, H. R., Piepenbrock, H., & Gulikers, L. (1995). The CELEX lexical database (release 2). Distributed by the Linguistic Data Consortium. University of Pennsylvania, Philadelphia, PA.
- Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience*, *16*, 4207–4221.
- Burgess, C., & Simpson, G. B. (1988). Cerebral hemispheric mechanisms in the retrieval of ambiguous word meanings. *Brain and Language*, *33*, 86–103.
- Cai, Z. G., Gilbert, R. A., Davis, M. H., Gaskell, M. G., Farrar, L., Adler, S., et al. (2017). Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology*, *98*, 73–101.
- Cattell, R. B., & Cattell, A. K. S. (1960). *Handbook for the individual or group culture fair intelligence test*. Champaign, IL: Testing.

- Coleman, M. R., Davis, M. H., Rodd, J. M., Robson, T., Ali, A., Owen, A. M., et al. (2009). Towards the routine use of brain imaging to aid the clinical diagnosis of disorders of consciousness. *Brain*, *132*, 2541–2552.
- Coleman, M. R., Rodd, J. M., Davis, M. H., Johnsrude, I. S., Menon, D. K., Pickard, J. D., et al. (2007). Do vegetative patients retain aspects of language comprehension? Evidence from fMRI. *Brain*, *130*, 2494–2507.
- Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, *26*, 55–67.
- Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F., et al. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. *Proceedings of the National Academy of Sciences, U.S.A.*, *104*, 16032–16037.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*, 1–12.
- Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E., et al. (2000). Susceptibility-induced loss of signal: Comparing PET and fMRI on a semantic task. *Neuroimage*, *11*, 589–600.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*, 429–446.
- Faust, M. E., & Gernsbacher, M. A. (1996). Cerebral mechanisms for suppression of inappropriate information during sentence comprehension. *Brain and Language*, *53*, 234–259.
- Federmeier, K. D., Segal, J. B., Lombrozo, T., & Kutas, M. (2000). Brain responses to nouns, verbs and class-ambiguous words in context. *Brain*, *123*, 2552–2566.
- Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, *29*, 181–200.
- Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning Memory and Cognition*, *16*, 430–445.
- Gilbert, R. A., Betts, H. N., Jose, R., & Rodd, J. M. (2017). *New UK-based dominance norms for ambiguous words*. Paper presented at the Poster session presented at the Experimental Psychology Society Meeting.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*, 267.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage*, *86*, 446–460.
- Grindrod, C. M., Bilenko, N. Y., Myers, E. B., & Blumstein, S. E. (2008). The role of the left inferior frontal gyrus in implicit semantic competition and selection: An event-related fMRI study. *Brain Research*, *1229*, 167–178.
- Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., et al. (2013). Good practice for conducting and reporting MEG research. *Neuroimage*, *65*, 349–363.
- Gunter, T. C., Wagner, S., & Friederici, A. D. (2003). Working memory and lexical ambiguity resolution as revealed by ERPs: A difficult case for activation theories. *Journal of Cognitive Neuroscience*, *15*, 643–657.
- Hagoort, P. (1993). Impairments of lexical semantic processing in aphasia—Evidence from the processing of lexical ambiguities. *Brain and Language*, *45*, 189–232.
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, *9*, 416–423.
- Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology*, *4*, 416.
- Hagoort, P., & Brown, C. (1994). Brain responses to lexical ambiguity resolution and parsing. In C. Clifton, Jr., L. Frazier, & K. Rayner, (Eds.), *Perspectives on sentence processing* (pp. 45–80). Hillsdale, NJ: Lawrence Erlbaum.
- Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., et al. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage*, *17*, 1101–1116.
- Henderson, L., Snowling, M., & Clarke, P. (2013). Accessing, integrating, and inhibiting word meaning in poor comprehenders. *Scientific Studies of Reading*, *17*, 177–198.
- Henson, R. N., Mouchlianitis, E., & Friston, K. J. (2009). MEG and EEG data fusion: Simultaneous localisation of face-evoked responses. *Neuroimage*, *47*, 581–589.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402.
- JASP Team. (2019). JASP (Version 0.10.0.0) [Computer software].
- Jefferies, E. (2013). The neural basis of semantic cognition: Converging evidence from neuropsychology, neuroimaging and TMS. *Cortex*, *49*, 611–625.
- Josephs, O., & Henson, R. N. A. (1999). Event-related functional magnetic resonance imaging: Modelling, inference and optimization. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *354*, 1215–1228.
- Kambe, G., Rayner, K., & Duffy, S. A. (2001). Global context effects on processing lexically ambiguous words: Evidence from eye fixations. *Memory and Cognition*, *29*, 363–372.
- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*, 463–470.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.
- Lange, K., Kuhn, S., & Filevich, E. (2015). “Just another tool for online studies” (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLoS One*, *10*, e0130834.
- Lau, E. F., Gramfort, A., Hämäläinen, M. S., & Kuperberg, G. R. (2013). Automatic semantic facilitation in anterior temporal cortex revealed through multimodal neuroimaging. *Journal of Neuroscience*, *33*, 17174–17181.
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*, 920–933.
- Lau, E. F., Weber, K., Gramfort, A., Hamalainen, M. S., & Kuperberg, G. R. (2016). Spatiotemporal signatures of lexical-semantic prediction. *Cerebral Cortex*, *26*, 1377–1387.
- Lee, C. L., & Federmeier, K. D. (2006). To mind the mind: An event-related potential study of word class and semantic ambiguity. *Brain Research*, *1081*, 191–202.
- Lee, C. L., & Federmeier, K. D. (2009). Wave-ering: An ERP study of syntactic and semantic context effects on ambiguity resolution for noun/verb homographs. *Journal of Memory and Language*, *61*, 538–555.
- Lee, C. L., & Federmeier, K. D. (2012). Ambiguity’s aftermath: How age differences in resolving lexical ambiguity affect subsequent comprehension. *Neuropsychologia*, *50*, 869–879.

- Lee, M., & Wagenmakers, E. (2014). Bayesian model comparison. In *Bayesian cognitive modeling: A practical course* (pp. 101–117). Cambridge: Cambridge University Press.
- Litvak, V., & Friston, K. (2008). Electromagnetic source reconstruction for group studies. *Neuroimage*, *42*, 1490–1498.
- Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., et al. (2011). EEG and MEG data analysis in SPM8. *Computational Intelligence and Neuroscience*, *2011*, 852961.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*, 35–54.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution [corrected]. *Psychological Review*, *101*, 676–703.
- Maess, B., Herrmann, C. S., Hahne, A., Nakamura, A., & Friederici, A. D. (2006). Localizing the distributed language network responsible for the N400 measured by MEG during auditory sentence processing. *Brain Research*, *1096*, 163–172.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition*, *25*, 71–102.
- Mason, R. A., & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain Research*, *1146*, 115–127.
- Miyake, A., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the maintenance of multiple interpretations of lexical ambiguities. *Bulletin of the Psychonomic Society*, *30*, 482.
- Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs (R Package Version 0.9.12-2) [Computer software]. Available at: <https://cran.r-project.org/package=BayesFactor>.
- Musz, E., & Thompson-Schill, S. L. (2017). Tracking competition and cognitive control during language comprehension with multi-voxel pattern analysis. *Brain and Language*, *165*, 21–32.
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective & Behavioral Neuroscience*, *5*, 263–281.
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27.
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976–987.
- Peelle, J. E. (2012). The hemispheric lateralization of speech processing depends on what “speech” is: A hierarchical perspective. *Frontiers in Human Neuroscience*, *6*, 309.
- Peer, E., Samat, S., Brandimarte, L., & Acquisti, A. (2016). Beyond the Turk: An empirical comparison of alternative platforms for online behavioral research. *Journal of Experimental Psychology*, *70*, 153–163.
- Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current-density mapping. *Electroencephalography and Clinical Neurophysiology*, *72*, 184–187.
- Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, *2*, 693–705.
- Rabovsky, M., & McRae, K. (2014). Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition*, *132*, 68–89.
- Raven, J., Raven, J. C., & Court, J. H. (1998). Manual for Raven's progressive matrices and vocabulary scales. Section 5, Mill Hill vocabulary scale (1998 ed). Oxford: Oxford Psychologists Press.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition*, *14*, 191–201.
- Rodd, J. M., Cutrin, B. L., Kirsch, H., Millar, A., & Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language*, *68*, 180–198.
- Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, *15*, 1261–1269.
- Rodd, J. M., Gaskell, G., & Marslen-Wilson, W. D. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*, 245–266.
- Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2010). The role of domain-general frontal systems in language comprehension: Evidence from dual-task interference and semantic ambiguity. *Brain and Language*, *115*, 182–188.
- Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2012). Dissociating frontotemporal contributions to semantic ambiguity resolution in spoken sentences. *Cerebral Cortex*, *22*, 1761–1773.
- Rodd, J. M., Longe, O. A., Randall, B., & Tyler, L. K. (2010). The functional organisation of the frontotemporal language system: Evidence from syntactic and semantic ambiguity. *Neuropsychologia*, *48*, 1324–1335.
- Rodd, J. M., Vitello, S., Woollams, A. M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: An activation likelihood estimation meta-analysis. *Brain and Language*, *141*, 89–102.
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356–374.
- Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, *14*, 489–537.
- Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J., Petersson, K. M., & Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: An fMRI study using word-category ambiguity. *Cerebral Cortex*, *19*, 1493–1503.
- Swaab, T. Y., Brown, C., & Hagoort, P. (1998). Understanding ambiguous words in sentence contexts: Electrophysiological evidence for delayed contextual selection in Broca's aphasia. *Neuropsychologia*, *36*, 737–761.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 645–659.
- Szabo Wankoff, L., & Cairns, H. S. (2009). Why ambiguity detection is a predictor of early reading skill. *Communication Disorders Quarterly*, *30*, 183–192.
- Tahmasebi, A. M., Davis, M. H., Wild, C. J., Rodd, J. M., Hakyemez, H., Abolmaesumi, P., et al. (2012). Is the link between anatomical structure and function equally strong at all cognitive levels of processing? *Cerebral Cortex*, *22*, 1593–1603.
- Taulu, S., & Kajola, M. (2005). Presentation of electromagnetic multichannel data: The signal space separation method. *Journal of Applied Physics*, *97*, 124905.
- Taulu, S., & Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG

- measurements. *Physics in Medicine and Biology*, *51*, 1759–1768.
- Taylor, J. S., Rastle, K., & Davis, M. H. (2013). Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. *Psychological Bulletin*, *139*, 766–791.
- Taylor, J. S., Rastle, K., & Davis, M. H. (2014). Interpreting response time effects in functional imaging studies. *Neuroimage*, *99*, 419–433.
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences, U.S.A.*, *94*, 14792–14797.
- Tompkins, C. A., Baumgaertner, A., Lehman, M. T., & Fassbinder, W. (2000). Mechanisms of discourse comprehension impairment after right hemisphere brain damage: Suppression in lexical ambiguity resolution. *Journal of Speech, Language, and Hearing Research*, *43*, 62–78.
- Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University-of-Alberta norms of relative meaning frequency for 566 homographs. *Memory and Cognition*, *22*, 111–126.
- Tyler, L. K., Cheung, T. P., Devereux, B. J., & Clarke, A. (2013). Syntactic computations in the language network: Characterizing dynamic network properties using representational similarity analysis. *Frontiers in Psychology*, *4*, 271.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*, 273–289.
- Van Berkum, J. J. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In K. Y. U. Sauerland (Ed.), *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Basingstoke, United Kingdom: Palgrave Macmillan.
- Visser, M., Jefferies, E., & Lambon Ralph, M. A. (2010). Semantic processing in the anterior temporal lobes: A meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience*, *22*, 1083–1094.
- Vitello, S., & Rodd, J. M. (2015). Resolving semantic ambiguities in sentences: Cognitive processes and brain mechanisms. *Language & Linguistic Compass*, *9*, 391–405.
- Vitello, S., Warren, J. E., Devlin, J. T., & Rodd, J. M. (2014). Roles of frontal and temporal regions in reinterpreting semantically ambiguous sentences. *Frontiers in Human Neuroscience*, *8*, 530.
- Zaidel, D. W., Zaidel, E., Oxbury, S. M., & Oxbury, J. M. (1995). The interpretation of sentence ambiguity in patients with unilateral focal brain surgery. *Brain and Language*, *51*, 458–468.
- Zempleni, M. Z., Renken, R., Hoeks, J. C. J., Hoogduin, J. M., & Stowe, L. A. (2007). Semantic ambiguity processing in sentence context: Evidence from event-related fMRI. *Neuroimage*, *34*, 1270–1279.