Cogn Affect Behav Neurosci
DOI 10.3758/s13415-016-0492-6

Getting ahead of yourself: Parafoveal word expectancy modulates the N400 during sentence reading

Mallory C. Stites[1] · Brennan R. Payne[1] · Kara D. Federmeier[1,2,3]

Abstract An important question in the reading literature regards the nature of the semantic information readers can extract from the parafovea (i.e., the next word in a sentence). Recent eye-tracking findings have found a semantic parafoveal preview benefit under many circumstances, and findings from event-related brain potentials (ERPs) also suggest that readers can at least detect semantic anomalies parafoveally (Barber, Van der Meij, & Kutas, *Psychophysiology, 50*(1), 48–59, 2013). We use ERPs to ask whether fine-grained aspects of semantic expectancy can affect the N400 elicited by a word appearing in the parafovea. In an RSVP-with-flankers paradigm, sentences were presented word by word, flanked 2° bilaterally by the previous and upcoming words. Stimuli consisted of high constraint sentences that were identical up to the target word, which could be expected, unexpected but plausible, or anomalous, as well as low constraint sentences that were always completed with the most expected ending. Findings revealed an N400 effect to the target word when it appeared in the parafovea, which was graded with respect to the target's expectancy and congruency within the sentence context. Furthermore, when targets appeared at central fixation, this graded congruency effect was mitigated, suggesting that the semantic information gleaned from parafoveal vision functionally changes the semantic processing of those words when foveated.

Keywords ERPs · Lexical access · Semantics · Parafoveal processing

Readers routinely use information about the upcoming parafoveal word—the word to the right of fixation in English—during natural reading. Much work has focused on clarifying the type of information readers get from this word and how it can be used to shape ongoing processing, with a large amount of accumulated evidence showing that readers can effectively extract orthographic and phonological information from the parafovea (for review, see Schotter, Angele, & Rayner, 2012). Recently, there has been a growing interest in readers' ability to extract semantic information from the parafovea, with an increasing number of eye-tracking studies showing evidence of semantic parafoveal processing in English (Rayner & Schotter, 2014; Schotter, 2013; Schotter, Lee, Reiderman, & Rayner, 2015; Schotter & Jia, 2016; cf. Rayner, Schotter, & Drieghe, 2014). Similarly, a recent set of studies using event-related potentials (ERPs) has also shown that readers can detect parafoveal semantic anomalies, as evidenced by effects on a brain response, the N400, that has been closely linked to semantic processing (Barber, Doñamayor, Kutas, & Münte; Barber, Van der Meij, & Kutas, 2013; Li, Niefind, Wang, Sommer, & Dimigen, 2015). At issue is whether such ERP effects are sensitive to more subtle relationships between the parafoveal word and the ongoing sentence context, which would reflect more detailed appreciation of the semantics of the parafoveal word. On the other hand, it could be the case that readers can recognize a parafoveal anomaly because top-down activation of the predicted word's orthography allows a quick assessment of the match between that

✉ Mallory C. Stites
stites2@illinois.edu

[1] Department of Psychology, University of Illinois, Urbana–Champaign, 603 E. Daniel St., Champaign, IL 61820, USA

[2] Program in Neuroscience, University of Illinois, Urbana–Champaign, 603 E. Daniel St., Champaign, IL 61820, USA

[3] Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana–Champaign, 603 E. Daniel St., Champaign, IL 61820, USA

Published online: 18 January 2017

prediction and low-level perceptual information derived from the parafoveal word, without having to access the word's semantics. Here, we test this possibility by examining parafoveal N400 effects across a wider range of expectancy and plausibility.

Parafoveal processing: Evidence from eye-tracking

Much of the evidence regarding parafoveal processing comes from studies in which eye movements are recorded during reading. Some of this evidence comes from skipping rates: readers are more likely to skip words that are predictable from the preceding sentence context (Ehrlich & Rayner, 1981). However, changing even one letter of the parafoveal word reduces skipping rates (Drieghe, Rayner, & Pollatsek, 2005), which has been interpreted as suggesting that there can be relatively complete processing of the parafoveal word. Other evidence regarding parafoveal processing comes from the gaze-contingent boundary change paradigm (Rayner, 1975). In this paradigm, an invisible boundary is placed before a target word, and this target word changes form contingent upon where readers are fixating. The target is manipulated to appear as a different "preview" word before readers reach the boundary; after readers cross the invisible threshold, the target word then changes to its normal appearance in the sentence. In general, reading times on the target are shortest when the preview and the target are identical, longest when they are unrelated, and facilitated to varying degrees when they are related in some way. This reduction in reading times on the target following a related preview relative to an unrelated one is known as the preview benefit.

It is well-replicated that orthographically and phonologically related previews produce a preview benefit (for review, see Schotter, Angele, & Rayner, 2012). However, until recently, evidence for a semantic preview effect was largely negative in both behavioral tasks (Inhoff, 1982; Inhoff & Rayner, 1980) as well as the boundary change paradigm (Balota, Pollatsek, & Rayner, 1985; Rayner, Balota, & Pollatsek,1986; Rayner et al., 2014). For example, Balota et al. (1985) found that in a constraining sentence such as, *Since the wedding was today, the baker rushed the wedding _____ to the reception,* previews that were either identical to the expected target (e.g., *cake*) or visually similar (e.g., *cahc*) reduced gaze durations on the target relative to semantically related (e.g., *pies*) or unrelated (e.g., *bomb*) previews. Because semantically related words are known to prime one another in other contexts (e.g., Balota, 1983; Meyer & Schvaneveldt, 1971; Neely, 1976), the *lack* of a preview benefit in this and similar studies (i.e., Rayner et al, 1986; Rayner et al., 2014) was thus taken as evidence that readers were unable to access the meaning of parafoveal words.

On the other hand, a number of recent studies have provided evidence in favor of parafoveal semantic processing. Semantic preview benefits have been well-documented for reading in Chinese, using both simple characters (Yan, Richter, Shu, & Kliegl, 2009; Yan, Pan, Bélanger, & Shu, 2015), compound characters (Yan, Zhou, Shu, & Kliegl, 2012), and traditional characters (Tsai, Kliegl, & Yan, 2012). These effects have also been replicated to show that they cannot be attributed to foveation of the targets due to fixating very near to the previews (Zhou, Kliegl, & Yan, 2013). Furthermore, *plausibility* effects have also been observed in reading Chinese: Yang (2013) found facilitation for two-character transpositions when the meaning of the transposed word was plausible with the ongoing sentence context, even when it did not overlap in meaning with the target—suggesting that even higher level meaning comprehension can take place for parafoveal words in Chinese. As noted by Hohenstein and Kliegl (2014), there are (at least) two important differences between reading in alphabetic versus nonalphabetic scripts that could make parafoveal semantic processing more likely: (1) the more direct relationships between form and meaning, and (2) the fact that Chinese words are shorter (one or two characters) than words in alphabetic language, bringing parafoveal words closer to the readers' fovea and thereby increasing the acuity with which they can be viewed.

However, parafoveal semantic preview benefits have also been reliably found in German, which is an alphabetic language (e.g., Hohenstein & Kliegl, 2014; Hohenstein, Laubrock, & Kliegl, 2010). Moreover, there is now a growing literature that is beginning to define the circumstances in which preview benefits are observed in English as well. In particular, several studies have found preview benefits for synonyms (i.e., *start–begin*) in English under a variety of conditions (Schotter, 2013; Schotter & Jia, 2016; Schotter et al., 2015). The extent to which benefits are observed for other types of semantically related words, such as semantic associates, seems to be more dependent on factors such as the strength of the context (e.g., Schotter et al., 2015, vs. Schotter, 2013) and the saliency of the word (Rayner & Schotter, 2014). Thus, evidence from the eye-tracking domain is mounting that, at least under some circumstances, readers can extract and use aspects of semantic information from parafoveal words.

ERPs as a tool for studying parafoveal processing

Although the gaze-contingent boundary change paradigm has been considered the gold standard in work on parafoveal processing, it has some drawbacks. In particular, the measurement of preview benefits relies on changing the stimulus presented to readers from one fixation to the next, and then calculating reading times on subsequent targets following the unrelated versus the related preview. Effects in this paradigm thus can reflect both benefits from the valid preview as well as costs from the invalid preview (Kliegl, Hohenstein, Yan, & McDonald, 2012; Tsai et al., 2012; Yan, Risse, Zhou, Kliegl, 2012; Zhou et al., 2013), and these processing effects encompass information extraction from the parafovea, eye movement programming, and integration of parafoveal information with that from the subsequent foveated targets, and can therefore be spread over multiple fixations (e.g., Risse & Kliegl, 2012, 2014). Furthermore, effects observed in the boundary change paradigm have been shown to be modulated by factors such as the pre-target fixation location and duration (i.e., preview time and preview space; Hohenstein & Kliegl, 2014; Kliegl et al., 2012), which cannot be experimentally controlled.

Thus, it is useful to gain converging evidence from other techniques with different strengths and weaknesses. In this study, we used the *RSVP-with-flankers ERP paradigm*, developed by Barber and colleagues (Barber, Ben-Zvi, Bentin, & Kutas, 2011; Barber et al., 2010; Barber et al., 2013). In this paradigm, sentences are presented via rapid serial visual presentation (RSVP), but three words are shown on the screen at once instead of just one: the current word at central fixation, with the upcoming word to the right and the previous word to the left (for reading in English; see Fig. 1). This configuration makes it appear as if the sentence is scrolling across the screen in front of the reader, as the words move from right to left across the center of the screen. Presenting sentences in this way allows the RSVP paradigm to build in some key features of natural reading, in particular by providing readers with a preview of the upcoming word. Critically, with this paradigm, it is possible to measure the brain electrical activity elicited by the parafoveal word itself. Furthermore, because readers maintain central fixation, this paradigm allows control over preview time and space to ensure that readers all receive previews that are as closely matched as possible.

Although the RSVP-with-flankers ERP paradigm differs from natural reading in a number of important ways, including that readers do not have control over the order or timing of their sentence intake, ERPs offer a rich, multidimensional measure with which to study online language comprehension. ERPs provide millisecond-level information about the timing, polarity, and scalp distribution of brain electrical activity elicited by different types of processing, providing functionally well-specified indices of different aspects of processing. One component in particular has come to be widely used in the study of reading comprehension and semantic processing more generally: the N400 (Kutas & Hillyard, 1980). The N400 is part of the normal brain response to potentially meaningful stimuli in all modalities, and its amplitude (but not its latency) has been shown to be modulated by a wide range of factors linked to lexicosemantic processing; for example, N400 amplitudes are reduced when words are repeated, primed by a semantic associate, or are embedded in a congruent sentence context (for review, see Kutas & Federmeier, 2011).

ERP effects, including those on the N400, often align with findings from eye-tracking; for example, at a basic level, word frequency effects result in both shorter reading times (for review, see Rayner, 1998) and smaller N400s (for review, see Kutas & Federmeier, 2000). To the extent that ERPs provide converging evidence with boundary change eye-tracking studies, we can thus be more confident in the robustness and generalizability of those patterns. Indeed, our recent work has shown that the effects of foveal load impact parafoveal processing in similar ways in both the boundary change paradigm and the RSVP-with-flankers ERP paradigm (Payne, Stites, & Federmeier, 2016). In some cases, ERPs can also reveal the

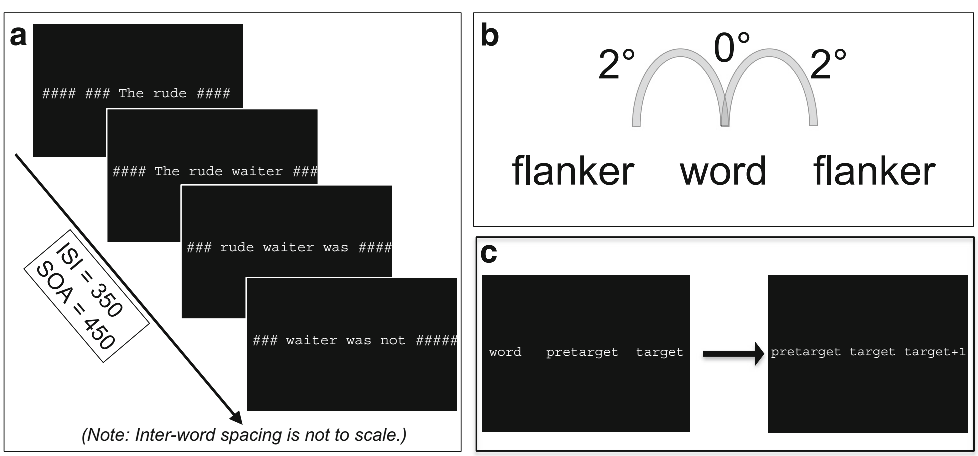

Fig. 1 Diagrams depicting the structure of a single trial. **a**) Sentences were presented serially in triads, with the target word appearing in center of screen (Note: Interword spacing is not drawn to scale). Triads were visible for 100 ms, with a 350 ms interstimulus interval. **b**) The current word of the sentence was presented centrally, and each flanker word started 2° of visual angle from central fixation. **c**) The first set of analyses considers trials on which the target word appears in the parafovea, with the pre-target word at central fixation. On the next triad, which appeared 450 ms after the onset of the previous triad, the target word moved to central fixation, and a neutral post-target word appeared in the parafovea

nature and time course of processing differences between conditions that produce similar eye movement effects; as a simple example, both syntactic and semantic anomalies elicit longer reading times (see Rayner, 1998), but are associated with qualitatively different ERP effects (Kutas & Federmeier, 2007). In this way, ERPs can help uncover what *type* of additional processing readers recruit in the face of processing difficulties that may result in longer reading times. Finally, ERPs sometimes reveal effects that are invisible to behavior, including eye-tracking (e.g., the presence of semantic priming even during the attentional blink; Luck, Vogel, & Shaprio, 1996). Thus, the emerging literature looking at parafoveal processing with ERPs provides a complementary view to the longer-standing literature examining these questions with eye movement measures.[1]

Semantic parafoveal effects in the flanker word ERP paradigm

Using the RSVP-with-flankers ERP paradigm, Barber and colleagues have shown that semantically incongruent parafoveal words embedded in sentences elicit larger N400 amplitudes than do congruent words (Barber et al., 2010)—the classic N400 congruency effect (for review, see Kutas & Federmeier, 2011). Interestingly, Barber et al. (2011) found that parafoveal congruency effects were only observed for flankers appearing in the direction of reading within a given language: readers of English elicited semantic parafoveal effects for words to the *right* of fixation, whereas readers of Hebrew elicited these effects for words to the *left* of fixation. These findings suggest that the parafoveal effects reflect the asymmetrical direction of the perceptual span and are thus influenced by attention rather than visual acuity per se.

By further manipulating sentential constraint, Barber et al. (2013) showed that these effects were larger for semantic anomalies that replaced predictable words in high constraint versus low constraint sentences. This pattern was more pronounced at a slower presentation rate (450 ms SOA); at faster presentation rates (250 ms SOA), the parafoveal N400 congruency effect was only obtained for the high constraint sentences. This suggests that eliciting a parafoveal N400 effect might critically depend on the amount of attention readers have available to allocate to the parafoveal word, which would be reduced under more demanding conditions, like a fast presentation. The parafoveal N400 congruency effect has also been observed in Chinese (Li et al., 2015), with the added benefit of concurrent eye-tracking to closely monitor participant's fixation location on the screen. Their results replicate Barber et al. (2013) in showing that incongruent parafoveal words elicited bigger N400s than congruent ones. Furthermore, their eye-tracking data confirmed that readers are indeed able to maintain central fixation during the RSVP-with-flankers paradigm: only about 6% of trials were rejected because of gaze location falling outside of the central target word.

In sum, studies looking at parafoveal semantic processing using ERPs have replicated sentential congruency effects on the N400 that are well-established in central vision. One drawback to these studies is that most of them have only compared processing of an expected target relative to an anomalous one. Although this is a good starting place for establishing the presence of semantic parafoveal effects, it leaves many questions unanswered. This study is designed to extend these findings by asking whether other, more fine-grained foveal N400 effects—such as facilitation of plausible but unexpected relative to anomalous words (DeLong, Quante, & Kutas, 2014) and graded N400 facilitation with increasing word expectancy (Wlotko & Federmeier, 2012)—can also be observed in the parafovea.

The current study

Although the findings from Barber et al. (2010; 2013) and Li et al. (2015) make it clear that readers can differentiate between an expected and semantically incongruent word in the parafovea, this does not show definitively that they are fully processing the semantics of that word. When sentence context information is sufficiently strong, readers have been shown to predict even the orthography of likely upcoming words (Laszlo & Federmeier, 2009). Thus, one could argue that these findings arose because readers were able to form specific expectations about upcoming words and test those predictions using orthographic information extracted from the parafovea. This argument seems less likely to be able to explain the parafoveal N400 effect in low constraint sentences in Barber et al. (2013), unless one postulates that in low constraint sentences, readers predict multiple words and can then appreciate the mismatch of the unexpected word with all of these likely possibilities from relatively shallow processing. Here, we aim to rule out explanations of these effects based on low-level cues, by moving beyond comparisons of expected versus unexpected words to see whether two unexpected words can be differentiated based on plausibility and whether two expected words can be differentiated based on cloze probability.

[1] In addition to the RSVP-with-flankers paradigm that is the focus of this article, one approach to studying parafoveal processing using ERPs has been to simultaneously record ERPs and eye movements as readers move their eyes over text and measure *fixation-related potentials* (i.e., ERPs time-locked to the onset of a fixation to a target word; Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011). In part because of methodological constraints, most of these studies have looked at effects in word lists, not in connected text (although see Kretzschmar et al., 2009). Results have been mixed, with some studies finding evidence of semantic parafoveal effects (Baccino & Manunta 2005; Kretzschmar et al., 2009), but others failing to find such effects (Dimigen, Kliegl, & Sommer, 2012).

We used words across the full range of predictability—high cloze probability, low cloze probability (the most expected endings of low constraint sentences), and wholly unexpected. This will allow us to ask whether the amplitude of the parafoveal N400 is graded with respect to cloze probability, as has been documented extensively for N400 amplitudes to centrally presented words (e.g., Wlotko & Federmeier, 2012). Low cloze probability words are not unexpected, but (by definition) are unlikely to be strongly predicted, so observing an intermediate level of facilitation for these items—that is, a larger amplitude N400 than for expected words in highly constraining sentences but a reduced N400 relative to unexpected words—would reveal more nuanced semantic processing of parafoveal words than just meeting or violating a context-based expectation. In addition, we presented readers with two different kinds of unexpected words in highly constraining sentences—words that are unexpected but plausible and those that are anomalous. Because both of these words are equally unrelated to the target orthographically, any difference in response to them must be due to their semantic fit with the sentence (e.g., DeLong et al., 2014). Thus, with this design, we can determine whether readers are able to obtain even more nuanced semantic information from words presented in the parafovea.

Method

Participants

Twenty-four participants (19 male, mean age = 20.7 years, range: 18–25 years) from the University of Illinois participated for course credit. All participants were native speakers of English, were right-handed as assessed by Edinburgh handedness inventory (Oldfield, 1971), and had no history of neurological disease or defect.

Materials

Experimental materials consisted of 240 sentence frames, split into four conditions. One hundred eighty of the items were highly constraining sentence frames that could be completed with either their most expected ending (mean cloze = .87, range: .68–1), an unexpected but plausible word (mean cloze = .06, range: 0–.06; cloze norming procedure for these sentences was originally described in Federmeier et al., 2007), or an anomalous word. Anomalous words were drawn from the same grammatical class as the unexpected words and were matched to the unexpected words on frequency measured from Francis and Kucera (1982), the English Lexicon Project (HAL; Balota et al., 2007), and the Corpus of Contemporary American English (CoCA; Davies, 2008). Stimulus characteristics are presented in Table 1, and outcomes of t tests comparing target words across conditions are presented in Table 2. The unexpected plausible and anomalous words were also matched on concreteness and imageability, both of which were measured on a scale from 100 (*very abstract*) to 700 (*very concrete*) and collected from the MRC Psycholinguistic Database (http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm), concreteness: $t(124) = 1.13$, $p = .26$; imageability: $t(126) = 1.34$, $p = .17$. The anomalous words were, however, slightly shorter than the unexpected words, $t(179) = -2.91$, $p < .01$. Within a sentence frame, each of the three targets conditions was preceded by an identical pre-target word (length: mean = 3.5 characters, range: 1–10; frequency: mean = 19,870, range:

Table 1 Mean (and standard deviation) of the lexical characteristics of critical words

Word	Length	Frequency				Concreteness	Imageability
		Kucera & Francis (Raw)	*Kucera & Francis (Log)*	*Log HAL*	*CoCA*		
Pre-target							
High constraint	3.5 (1.8)	18,534 (25,050)				276 (81)	295 (100)
Low constraint	3.0 (1.4)	31,543 (29,832)				246 (53)	259 (75)
Target							
Expected	4.8 (1.2)	116 (163)	4.09 (1.24)	10.07 (1.48)	9.81 (1.46)	523 (104)	546 (86)
Unexpected	5.7 (2.0)	79 (125)	3.43 (1.62)	9.17 (1.79)	9.61 (1.42)	508 (104)	530 (81)
Anomalous	5.3 (1.4)	82 (112)	3.54 (1.44)	9.33 (1.52)	9.59 (1.22)	530 (86)	534 (72)
Low constraint	5.5 (2.0)	163 (405)	4.31 (1.24)	10.07 (1.47)	10.21 (1.32)	502 (102)	533 (79)
Post-target							
High constraint	5.1 (1.4)	4,711 (7,818)				239 (37)	263 (50)
Low constraint	5.4 (1.3)	1,512 (2,278)				282 (102)	324 (103)

Note. Log HAL frequency measures are from the English Lexicon Project; CoCA are from the Corpus of Contemporary American English

Table 2 Summary of *t* tests comparing frequency measures for the target words across conditions

Condition	Kucera & Francis	Log HAL	CoCA
t[HCE-LCE]	-1.19	0	-1.93
t[HCE-HCU]	4.34	5.19	1.26
t[HCE-HCA]	3.88	4.67	1.57
t[LCE-HCU]	4.39	3.87	2.92
t[LCE-HCA]	3.99	3.34	3.2
t[HCU-HCA]	-0.68	-0.91	0.22

Note. HCE = high constraint expected; LCE = low constraint expected; HCU = high constraint unexpected; HCA = high constraint anomalous; Log HAL = English Lexicon Project; CoCA = Corpus of Contemporary American English

12–69,971), which in 121/180 sentences was a short function word or pronoun (i.e., *the*, *a*, *his*).

Additionally, there were 60 low constraint experimental sentence frames that always contained their most expected completion (mean cloze = .23, range: .09–.35) (Federmeier et al., 2007). Including these items allows us to examine the parafoveal N400 to low-to-moderately expected continuations and also ensures that half of the sentences presented to subjects did not violate readers' predictions about upcoming targets and contained no unexpected or anomalous words. Expected words in the low constraint sentences also did not differ from the high constraint expected words in terms of frequency, $t(55) = -.87$, $p = .39$, concreteness, $t(38) = -.24$, $p = .81$, or imageability, $t(39) = -.18$, $p = .85$, although they were longer than the targets in the high constraint sentences, $t(59) = -2.76$, $p < .01$. The targets in the low constraint sentences were preceded by words that were similar to those in the high constraint sentences in terms of length (mean: 3.0 characters, range: 1–8; frequency: mean = 29,832, range: 0–69,971), which in 42/60 sentences was again a short function word or pronoun (i.e., *a, the*, *to*).

Target words appeared sentence medially, on average as the tenth word of the sentence (high constraint contexts mean sentence location = 9.6, range: 4–21; low constraint contexts mean sentence location = 10.4, range: 4–20). To assess ERP effects following the target word's appearance in the parafovea, continuations were written for every sentence (high constraint contexts mean continuation length = 5 words, range: 2–11; low constraint contexts mean = 4.4 words, range: 3–7). For the high constraint sentences, the same continuations were used after expected, unexpected but plausible, and anomalous words; these were judged by the three experimenters to be plausible following both the expected and unexpected endings. Similar continuations were created for the low constraint contexts. The majority (85%) of the immediately post-target words were high frequency function/closed class words (e.g., *after, because, with*), and the rest of the post-target words were distributed across nouns (2%), verbs (2%), adjectives (4%), and adverbs (7%). Including the pre- and post-target regions, experimental sentences were 15-words long on average, ranging from 7 to 27 words total.

In half of the high constraint sentences, the post- target word was replaced by a nonword in parafoveal vision only. This nonword was comprised of an orthographically illegal character string and was matched to the target on length. Foveal data from trials containing this manipulation were removed from the current analysis, as they were the focus of an investigation of different aspects of reading behavior (Payne et al., 2016). Because of this manipulation, there were 60 trials per condition for all parafoveal targets and 30 trials per condition for foveal targets in each of the three high constraint conditions.

To ensure that adding continuations after the target word did not alter the cloze probability of the targets in any of our sentences (particularly when the first word of that continuation is visible simultaneously with the target word), we conducted a large-scale norming study using all 240 experimental stimuli. The entire sentence frame, including the continuation, was presented to participants with a blank space indicating where the target word should go. Participants were instructed to fill in the word that best completed each frame. In total, 103 people participated through Amazon Mechanical Turk (47 male, 56 female; mean age = 38.5 years, range: 19–66 years), and were compensated with cash for their participation. The mean cloze probability for the expected endings in the high constraint sentences from this large-scale norming study (mean cloze = .89, standard deviation = .13) was highly consistent with the previously conducted paper-and-pencil norming study (mean cloze = .87; Federmeier et al., 2007). Only 14 of the 180 sentences had a new cloze that fell below the .67 cutoff typically required for high constraint sentences (mainly caused by participants providing a variety of semantically similar responses that were not perfectly matched to the target word). Moreover, participants only provided the unexpected target word less than 1% of the time (mean = .009, range: 0–.25), indicating that this word was still unexpected even with the sentence completion. Mean cloze probability for target words in the low constraint sentences from the large-scale norming study (mean cloze = .33, standard deviation = .25) was also comparable to the cloze for these words in the sentence frames alone (mean cloze = .23). This numerically higher cloze probability in the new norming study was likely due to later words in the continuation biasing readers to provide the expected target in the offline task. For example, in the sentence, *The shirt was stained with* ***blood*** *from the cut on his arm*, the word *cut* probably made respondents more likely to provide the target *blood* in the off-line task. However, this could not have influenced online processing of the target *blood* in the ERP task, as the biasing word *cut* would not have become visible until after *blood* had already left the screen. The findings from the large-scale norming study thus confirm that adding sentence continuations did not affect the sentential constraint or cloze probability of our stimulus set as a whole. The findings from the norming study also provide validation that conducting a norming study on Amazon Mechanical Turk provides highly convergent findings with conducting a paper-and-pencil test in person.

Items were divided into three lists, so that the high constraint sentences could be rotated through conditions across lists. Thus, participants saw all 60 of the low constraint sentences and each of the 180 high constraint sentence frames once in one of three conditions (expected, unexpected, or anomalous), with 60 items overall in each. Two presentation orders were created for each list: one quasi-random order (with the rule that no more than three sentences of a single condition appear in a row), and the reversal of that order, and each participant saw only one list.

Procedure

Participants were seated 85 cm from a 21-in. CRT computer monitor in a dim, quiet testing room. They were instructed to remain as still as possible with their eyes fixated on the center of the screen and to try to minimize blinks throughout the experiment. Participants were informed that multiple words and symbols would appear to the left and right of the central word, but that their job was to keep their eyes focused on the word presented at the center of the screen and to read the sentence for comprehension, because there would be a memory test at the end.

Sentences were presented serially in triads, with the target word appearing at central fixation, flanked bilaterally by the upcoming word in the sentence to the right, and the preceding word to the left. At the chosen viewing distance, 3.5 letters subtended one degree of visual angle. The beginning letter of the right parafoveal word and the final letter of the left parafoveal word were anchored so as to appear 2° of visual angle from the center of the screen. Because target words were on average between five and six characters long, they typically subtended the 2° to 4° of visual angle to the right of fixation when appearing in the parafovea. The sentence initial word had, as a placeholder, a series of six pound/number signs ("######") appear as its left parafoveal flanker, and similarly, the sentence final word had the identical symbol string appear as its right parafoveal flanker. In all sentences, the parafoveal words were themselves flanked by a series of pound signs ("#####") to maintain a steady outer edge of characters in the participant's peripheral vision, to avoid distracting changes in the overall length of the character string appearing on the screen at any time (see Fig. 1 for a depiction of the screen). The number of additional "#" symbols to appear on each side of the screen varied from five to 11, and was dynamically determined by the length of each flanker word, such that the "#" symbols began approximately one character space away from the flanker word and filled in the screen to the predetermined "edge," which was maintained across all sentences.

Each trial began with a series of fixation crosses ("+++++") that remained on the screen for a duration that was jittered from 500 to 1,500 ms. The sentence then began, in which each word was presented centrally as part of a triad (as described above). Each triad was visible on the screen for 100 ms, with an interstimulus interval of 350 ms. This stimulus duration was chosen because Barber et al. (2013) previously found robust parafoveal N400 effects using similar timing. This short stimulus duration also minimized the possibility that participants could make a saccade to the parafoveal word within the amount of time that it was visible on the screen. At the end of the sentence, a blank screen was presented for 1,950 ms, after which the fixation crosses appeared and the next sentence began.

The experiment began with a practice block of nine sentences to allow participants to become accustomed to the experimental procedure. The practice block could be repeated if participants made excessive amounts of eye movements, and they were continually given feedback about reducing eye movements through the experiment. The 240 sentences were divided into six blocks of 40 sentences each. Each recording block took approximately 7 minutes to complete, and short breaks were offered between each one.

At the end of the experiment, participants completed a memory test. This test consisted of 160 sentences frames with a sentence-medial word left out, 80 of which appeared in the study (20 each from the high constraint expected, unexpected, and anomalous, and the low constraint sentences) and 80 that did not. For each sentence, subjects had to first indicate whether or not they saw the sentence frame in the study; for items they marked as having seen, they then had to fill in the blank with the word that completed the sentence in the study.

EEG recording

The electroencephalogram (EEG) was recorded from 26 evenly spaced silver/silver-chloride electrodes attached to an elastic cap (locations described in Federmeier et al., 2007; MiCe corresponds to Cz in the 10-20 system nomenclature). All scalp electrodes were referenced online to the left mastoid and rereferenced off-line to the average of the left and right mastoids. One electrode was placed on the left infraorbital ridge and referenced to the left mastoid, to monitor for vertical eye movements and blinks. Two electrodes were also placed on the outer canthus of each eye, referenced to each other, to monitor for horizontal eye movements. Electrode impedances for the three electrodes on the face were kept below 10 kΩ, and impedances for scalp electrodes were kept below 5 kΩ. The continuous EEG was amplified through a bandpass filter of 0.02–100 Hz and recorded to hard disk at a sampling rate of 250 Hz.

Epochs of EEG data were taken from 100 ms before stimulus onset to 1,944 ms poststimulus. Those containing artifacts from amplifier blocking, signal drift, eye movements, eye blinks, or muscle activity were rejected off-line before averaging, using threshold selected for each participant through visual inspection of the data. Special care was taken to set very sensitive thresholds for detecting and rejecting trials that contained lateral eye movements, following the procedure described in the following section. Trial loss averaged 18.2%. Artifact-free ERPs were averaged by stimulus type after subtraction of the 100 ms prestimulus baseline. Prior to statistical analyses, ERPs were digitally filtered with a low-pass filter of 30 Hz.

Fixation control

We took steps to ensure that data contained as few eye movements as possible, to ensure that our findings represent circumstances in which readers were actually fixating the central word (and did not, for example, only fixate the right-most word). First, triads were only visible on the screen for 100 ms with a 350 ms ISI. Because it takes approximately 200 ms to program and execute an eye movement, this short stimulus duration would have provided insufficient time for participants to make a saccade from the centrally presented word to the parafoveal word.

Second, stringent artifact rejection procedures were undertaken to ensure the removal of trials that contained eye movements while the target word was present on the screen. Individualized thresholds were separately determined for each participant through visual inspection of the ERPs to identify trials with obvious eye movements in the bipolar HEOG; these thresholds were then applied to every trial for that participant, such that all trials with eye movements were removed prior to performing data analysis (for both the parafoveal and foveal N400 analyses). There is a well-established relationship between the amplitude of saccade-related HEOG deflections and the distance traveled by the eyes: namely, a deflection of 16 μV is approximately equal to a horizontal eye movement of one degree of visual angle (Luck, 2014). Given that parafoveal words started at two degrees of visual angle away from central fixation, this would require a deflection of ~32 μV for the eyes to reach the start of the parafoveal target word. As can be seen in the inset of Fig. 2, after removal of these artifacts, the average HEOG deflection time-locked to the onset of the target in parafoveal vision was very small (<|2|μV), particularly in the first 100 ms after stimulus presentation, thus indicating that participant were not making eye movements to the target word in parafoveal vision.

Finally, we have empirical evidence that readers consistently maintained central fixation. In a separate publication (Payne et al., 2016), we examined foveal-on-parafoveal effects within this same dataset. As mentioned above, when the target words appeared at central fixation, a nonword appeared in the parafovea in half of the high constraint sentences, which then became a real word in foveal vision. We found that readers showed differential effects of a parafoveal nonword versus word based on the processing load induced by the foveal target word. Specifically, results showed a widespread negativity for invalid versus valid previews, which was reduced in size with increasing foveal load (suggesting that readers extracted less parafoveal information when foveal load was higher). Importantly, these results confirm that readers were maintaining fixation on the central word; if they were consistently looking at only the parafoveal word, we would expect to see the same invalid preview effect across all three foveal load conditions (because readers would have actually been fixating the nonwords). Instead, the graded effect of foveal load on this preview validity effect confirms that participants successfully maintained central fixation.

Results

Behavioral results

Participants correctly recognized an average of 42.6 sentences (hit rate = 53%) and false alarmed to an average of 10.2 sentences (false alarm rate = 13%). The sensitivity index, *d'*, was calculated using the Macmillan and Kaplan (1985) correction (because several participants made no false alarms), which yielded an average *d'* of 1.45, 95% CI [1.12, 1.79]. Thus, participants were attending to the experimental materials. Additionally, subjects correctly filled in the target word on an average of 18.5 experimental sentences and incorrectly filled in the target word on an average of 11.5 sentences. The errors subjects made primarily consisted of their guessing high cloze probability words in the strongly constraining sentence frames when they could not recall the unexpected or anomalous target words.

ERPs

ERPs will be reported for two time windows, both of which were measured from the same baseline period: the onset of the target word in the parafovea (following Barber et al., 2013; Li et al., 2015). First, we will describe ERPs, specifically the N400 component, elicited by the target word when it actually appeared in the parafovea. As can be seen in Fig. 2, triads elicited the typical pattern of components elicited by single word presentation: a negative-going N1 component peaking around 100 ms, a positive-going P2 deflection peaking around 200 ms, followed by the N400 component, which was broadly distributed but

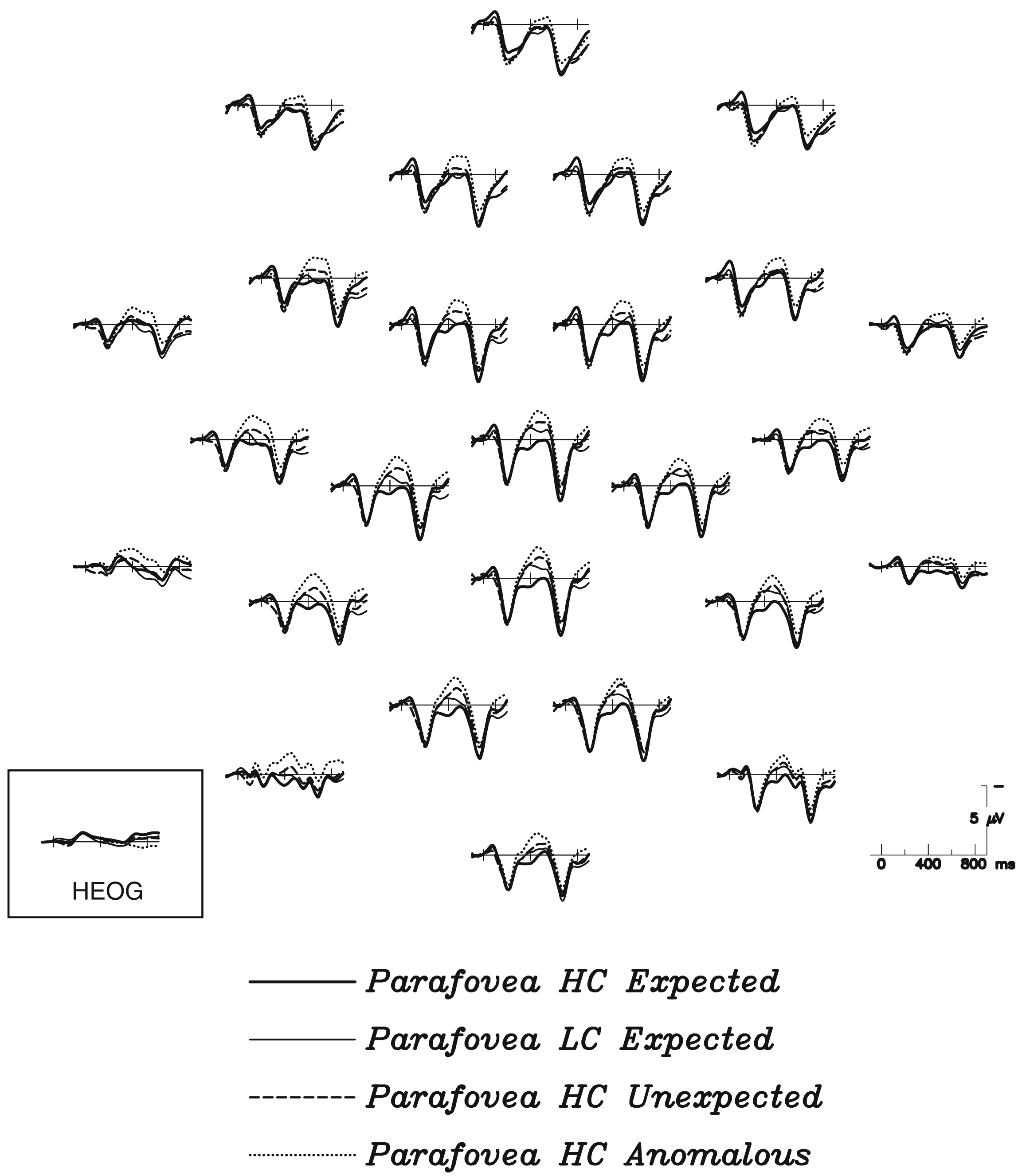

Fig. 2 Grand-average ERPs at all 26 electrode sites for the parafoveal N400 congruency effect. The N400 effect is graded by parafoveal word expectancy, being the most negative for anomalous words and the most facilitated for the expected words with the highest cloze probability. The figure inset depicts the grand average ERP amplitude of the bipolar horizontal eye channel. Deflections of roughly 16 µV are associated with saccades of 1° of visual angle. There was very little activity in this channel, especially during the first 100 ms of processing (the duration that the stimuli were visible on the screen), suggesting that participants consistently maintained central fixation

maximal over centroparietal sites between 300 and 500ms. Second, we will report statistics on the N400 component elicited by the target word's appearance in the fovea, measured from 750 to 950 ms from the start of the epoch, corresponding to 300–500 ms after the target's appearance at central fixation. As can be seen in Fig. 3, these triads still elicited the N1/P2 complex, but the overall N400 amplitude is reduced across all conditions.

Target word in parafovea

The N400 component was measured as the mean amplitude from 300-500 ms after the onset of the triad containing the target in the parafovea, measured over 15 posterior channels (LMCe, RMCe, LDCe, RDCe, MiCe, MiPa, LLTe, RLTe, LDPa, RDPa, LLOc, RLOc, LMOc, RMOc, MiOc), where

N400 effects are typically most prominent. An omnibus repeated-measures ANOVA was conducted with the factors of expectancy (4: high constraint expected, low constraint expected, unexpected but plausible, anomalous) and electrode (15). To correct for violations of sphericity associated with repeated measures, for each ANOVA that follows, we applied the Greenhouse–Geisser correction to the degrees of freedom and report the resulting p values for factors with more than two levels. Main effects of electrode and interactions between electrode and expectancy were not of theoretical interest and will not be discussed. Results revealed a main effect of expectancy, $F(3, 69) = 7.3, p < .001$, indicating an overall difference in the N400 amplitude elicited by these four conditions.

The means of the conditions were graded by contextual fit (HC expected: .42 μV, LC expected: -.31 μV, Unexpected: -.49 μV, Anomalous: -1.12 μV). Follow-up pairwise tests between conditions found that, as predicted, N400 amplitude was facilitated in the high constraint expected condition relative to all three of the other conditions, HC expected vs. LC expected: $F(1, 23) = 6.29$, $p < .05$; HC expected vs. Unexpected: $F(1, 23) = 8.67$, $p < .01$; HC expected vs. Anomalous: $F(1, 23) = 18.21$, $p < .001$. Importantly, the N400 amplitude elicited by the low constraint expected words was also significantly facilitated relative to the anomalous words, $F(1, 23) = 5.00$, $p < .05$, although not significantly different from the unexpected but plausible words, $F(1, 23) = .30$, $p = .59$. The high constraint unexpected words were also marginally facilitated relative to the anomalous words, $F(1, 23) = 3.31$, $p = .08$.

To more directly assess the prediction that N400 amplitude is graded with expectancy of the parafoveal word, single trial ERP amplitudes from eight channels centered around the maximal distribution of the N400 (and previously successfully used to measure N400s in single trial data; Payne, Lee, & Federmeier, 2015: LMCe, RMCe, LDCe, RDCe, MiCe, MiPa, LDPa and RDPa) were submitted to a linear mixed-effects regression (LMER) model (using the lme4 package in R; Bates, Maechler, Bolker, & Walker, 2015) with target word cloze probability as a continuous predictor. Models were fit with random intercepts for subjects, items, and channels, and slope adjustments for the cloze predictor across subjects and channels (excluding correlations between random intercepts and slopes to reduce model complexity; Barr, 2013).[2] There were not random slopes for cloze at the item level because each item had only one cloze value. Additionally, the covariates of word length, frequency, concreteness and imageability were also included in the model to control for effect of these item-level factors, all of which were grand-mean standardized ($M = 0$, $SD = 1$). All trials with artifacts that were excluded from the previous analyses were also excluded from the LMER model, as well as single trials with extreme amplitudes (+/-3 SD), prior to fitting the model. Results showed a significant effect of target word cloze ($b = .52$, $t = 4.41$, 95% CI [.14, .86]), indicating decreasing negative amplitude (i.e., more positive N400s) as target words became more expected (full model output can be found in Table 3). In general, then, the N400 amplitude elicited by parafoveal target words showed the same graded sensitivity to cloze probability as has been documented for foveated words (e.g., Kutas & Hillyard, 1984; Wlotko & Federmeier, 2012).

Target word in fovea

N400 (750–950 ms, or 300–500 ms after target appeared in central fixation)

ERPs were also examined to the immediately following triad, in which the target word appeared at central fixation, while a neutral word N+1 appeared in the parafovea (see Fig. 3). An omnibus ANOVA was conducted in the same manner as described above over the same electrode sites. Results showed no effect of expectancy, $F(3, 69) = 1.82$, $p = .32$. Pairwise comparisons between the two most extreme conditions (high constraint expected vs. anomalous) revealed a marginal difference between their N400 amplitudes, $F(1, 23) = 3.37$, $p = .08$. An LMER model using target word cloze as a continuous predictor, conducted in the same way as described above, replicated the ANOVA findings in showing no effect of cloze probability on ERPs to the target word ($b = .15$. $t = 1.18$, 95% CI [-.10, .40]). This is in striking contrast to the pattern typically seen for these conditions when fixated without prior parafoveal preview, wherein the N400 shows a strong negative relationship with cloze probability. Thus, the results show that the expected graded sensitivity to cloze probability and congruency, seen here when the targets are in the parafovea, does not reemerge when these words move to foveal position. Figure 4b illustrates the difference between the large effect of cloze probability on N400 amplitude for parafoveal targets compared to the nearly absent effect of cloze for foveal targets.

Discussion

This experiment asked whether readers can assess the semantic fit of a parafoveal word with the ongoing sentence context, or whether parafoveal N400 effects instead reflect a more general appreciation of the difference between expected and unexpected words. We tested this by presenting readers with high constraint sentences that contained either the most expected word, an unexpected but plausible word, or a semantic anomaly, as well as low constraint sentences that always continued with the most expected word. We investigated

[2] Because of a recording error, one participant had to be dropped from the item-level analysis, as item-level event codes could not be extracted.

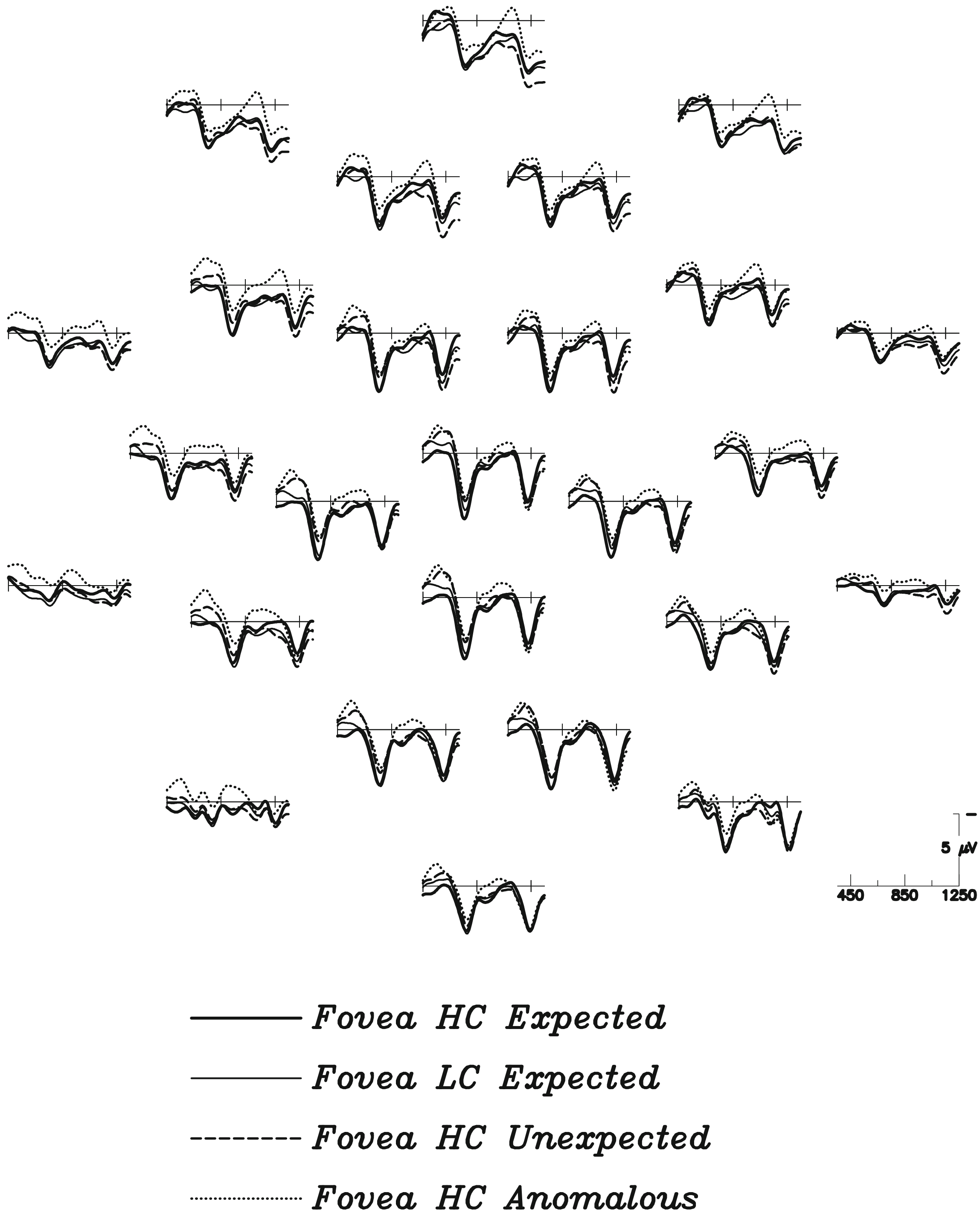

Fig. 3 Grand-average ERPs at all 26 electrode sites for the foveal N400 congruency effect (presented from 100 ms prior to the target word appearing the in fovea). N400 expectancy effect is drastically reduced when target words appear in the parafovea, with only a numeric difference in N400 amplitude across these conditions

processing both when the word appeared in the parafovea as well as when it moved to appear in central fixation.

The findings were clear. First, readers assessed the semantics of the parafoveal word, as indicated by graded N400 cloze probability effects elicited by these words. This finding replicated Federmeier et al. (2007), from which the majority of the stimuli were drawn (with the exception of the anomalous condition), and which found the same pattern of facilitation when these stimuli were presented foveally. Even among items that were all the most expected in their context (comparing expected targets in strongly versus weakly constrained sentences), the expected words in more strongly constrained sentences showed reduced N400 amplitudes when they appeared in the parafovea relative to the weakly constrained sentences. This finding, while not surprising for foveal words, indicates that parafoveal semantic processing can extend beyond simply differentiating expected versus unexpected items to actually allow readers to distinguish between more or less expected items. At the same time, N400 amplitudes were different among words that were not predicted from the context (the unexpected plausible versus anomalous words), as a function of whether they were more or less congruent with the message-level meaning of the sentence. This finding replicates DeLong et al. (2014), who found similar N400 differentiation between unexpected plausible and anomalous targets when presented at central fixation.

These findings show that readers can access the semantics of the parafoveal word and use that information incrementally to determine not only whether an unexpected parafoveal word is anomalous, as was already demonstrated by Barber et al. (2010; 2013) and Li et al. (2015), but also to determine whether an unexpected parafoveal word is *plausible* with the ongoing sentence context. Because it has been shown that readers predict the orthography of high-cloze words (Laszlo & Federmeier, 2009), differences between anomalous and expected words observed in prior work could have been attributed to readers noticing a mismatch in the orthography between the predicted and anomalous word in high constraint sentences without the need to process the parafoveal words beyond this surface level (with the exception of Barber et al., 2013, Experiment 1, who *did* find congruency effects in low constraint sentences). Our results make clear that parafoveal preview affords graded semantic information that goes beyond noticing that a parafoveal word is inconsistent with a contextually formed expectation (or set of expectations), in that unexpected but plausible words were facilitated relative to anomalous words even though both were unexpected and orthographically unrelated to the expected target. Furthermore, the targets in the low constraint sentences were facilitated relative to the two unexpected conditions, but still elicited larger N400s than did the expected words in the high constraint sentences—as would be the case if these words were presented in foveal vision.

Second, support for the idea that readers can extract detailed semantic information from the parafovea comes from the pattern elicited once those same words have moved into foveal vision: the graded N400 effects seen when the targets were in parafoveal vision were *not* elicited again when the same word appeared in the fovea (450 ms later). This outcome is striking, as it reveals that receiving a brief (100 ms) parafoveal preview of an anomalous word drastically reduces even the foveal N400 anomaly effect—one of the most robust N400 effects (Kutas & Federmeier, 2011). These results thus provide additional evidence that readers assessed the semantics of the target word when it was in the parafovea—and, indeed, seemingly largely completed this semantic analysis. When the word moved into fixation in the central location, the classic N400 expectancy effects were not elicited again, indicating that readers integrated their representations of the target as it moved from the parafovea to central fixation and did not need to again access its semantics. In essence, this finding that the N400 congruency effect is eradicated in the fovea parallels the "identity preview" benefit observed in boundary change studies: when readers have an identical preview of the upcoming word in the parafovea, their semantic processing of that word in the fovea as measured by the N400 is greatly facilitated, and, importantly, is statistically *indistinguishable* across the four conditions.

In previous work on the parafoveal N400 effect, only one other study found this eradication of the foveal N400 anomaly effect by a parafoveal preview (Barber et al., 2010). In contrast, both Barber et al. (2013) and Li et al. (2015) observed N400 congruency effects for both parafoveal *and* foveal anomalies, even when the foveal anomalies were immediately preceded by an identical (anomalous) parafoveal preview. One important difference between these two studies and the current design (as well as that of Barber et al., 2010) is that their target words could *change* from being anomalous to congruent (or vice versa) as they moved from the parafovea to the fovea—much like in the boundary change paradigm. The fact that the stimulus changed across fixations could have reduced how much readers' brains treated the parafoveal word as a reliable cue for the upcoming target, making it less likely for information to be integrated across these two representations (i.e., with uncertainty about the validity of the preview, readers may treat every triad as somewhat independent). In the current design, targets remained the same in the parafovea and fovea, which may have encouraged treating the target as the same object moving across the screen, instead of as a new stimulus event—arguably thereby reflecting processing more typical of normal reading.[3] In fact, Li et al. (2015) found what they called "preview validity effects": more negative N1 amplitudes (from 200–300 ms post-target onset) over occipital-temporal electrodes on trials in which the target word changed from the parafovea to the fovea (see also Dimigen et al., 2012). Their findings indicate that readers' brains were sensitive to the fact that parafoveal previews were not always valid, which could explain why their foveal N400 effects were still present whereas ours were not.

In sum, despite the differences between the RSVP-with-flankers ERP paradigm and studies of reading that measure

[3] It is important to note that, as mentioned in the Method section, on half of the trials the target words were followed by an illegal letter string in the parafovea that *always* changed to a real word in the fovea. However, an anomalous word would *never* change to an expected word from parafovea to fovea (or vice versa), so readers could always rely on the meaning of the sentence remaining intact. Readers should refer to the complementary publication on these data (Payne et al., 2016) for a detailed description of the effects of foveal load on the integration of parafoveal and foveal representations.

Table 3 Results from the linear mixed-effects models for parafoveal and foveal N400 effects

	Predictor	Estimate	Standard error	95% CI lower bound	95% CI upper bound	*t* value
Parafoveal N400	Intercept	**-0.72**	**0.17**			**-4.17**
	Cloze	**0.52**	**0.12**	**0.15**	**0.86**	**4.41**
	Length	**0.14**	**0.05**	**-0.32**	**-0.11**	**-2.78**
	Frequency	-0.07	0.04	-0.16	0.02	-1.70
	Concreteness	**-0.28**	**0.10**	**-0.59**	**-0.16**	**-2.77**
	Imageability	**0.17**	**0.10**	**0.11**	**0.53**	**1.76**
Foveal N400	Intercept	-0.16	0.26			-0.64
	Cloze	0.15	0.13	-0.10	0.40	1.18
	Length	**-0.25**	**0.06**	**-0.36**	**-0.14**	**-4.33**
	Frequency	**-0.14**	**0.05**	**-0.23**	**-0.04**	**-2.73**
	Concreteness	**-0.53**	**0.11**	**-0.75**	**-0.31**	**-4.63**
	Imageability	**0.33**	**0.11**	**0.11**	**0.55**	**2.89**

Note: Significant effects are denoted by boldface type

eye movements, the emerging picture from these two fields seems to be converging on the idea that readers can and regularly do extract semantics from the parafovea. For example, Schotter and Jia (2016) found a preview benefit for plausible previews relative to implausible ones, even when these words were orthographically and semantically unrelated to target words—much like our observed parafoveal N400 facilitation for unexpected relative to anomalous words. One question Schotter and Jia (2016) aimed to address was whether preview benefits observed in the boundary change paradigm arise due to integration of the preview and target words or from the contextual fit of the preview with the sentence context. One difficulty with trying to use eye movements to address this question is that "early" and "late" eye-movement effects (i.e., first fixation duration vs. total time) show different patterns of effects, likely because the target and preview

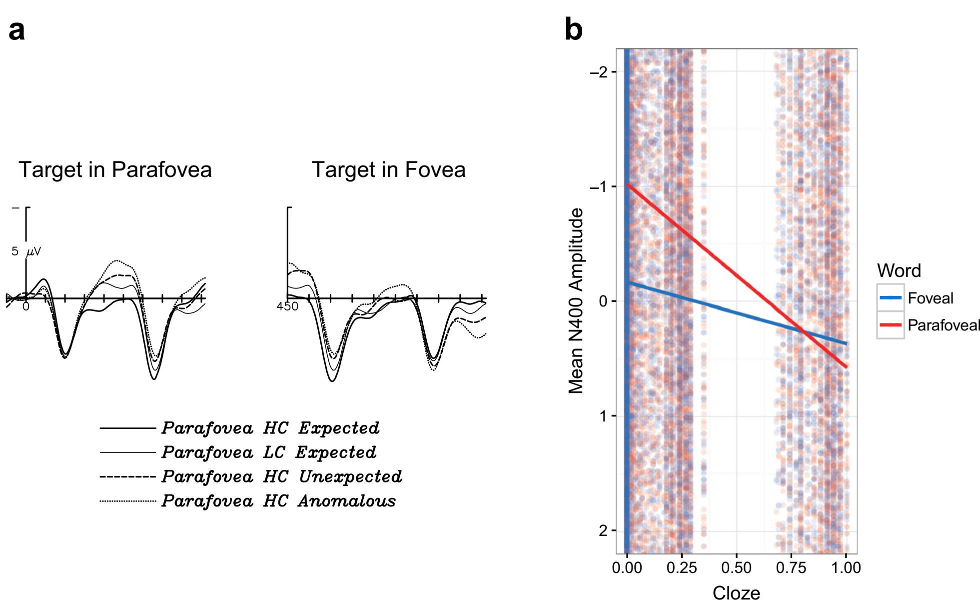

Fig. 4 **a**) Grand-average ERPs at the middle central electrode comparing the parafoveal N400 congruency effect (*left panel*) and foveal N400 congruency effect (*right panel*). Note that negative is plotted up. The same baseline period was used for both measurements: 100 ms prestimulus onset of the target-in-parafovea. As such, the foveal effect is simply plotted starting from the onset of the target word in the fovea, which occurred at 450 ms after the time-locking point (i.e., 450 ms after the target appeared in the parafovea, in the middle of the N400 to the parafoveal word). **b**) Plot of individual item-level N400 amplitude by cloze probability, overlaid with the slope of the cloze effect for the parafoveal N400 (*red*) and foveal N400 effect (*blue*). Note that negative is plotted up, for ease of comparison with the ERP plot in **a**. The steeper slope for the parafoveal effect indicates that parafoveal N400 amplitudes were graded with respect to cloze probability, whereas the flatter slope of the foveal N400 effect shows that this effect was largely eliminated in foveal processing

word are different and so can never be fully "integrated." By providing separable measures of semantic processing in parafoveal vision and the impact of that processing on semantic analysis of the foveated word, our findings show that readers can *both* assess the contextual fit of a parafoveal word with the ongoing context (as indicated by the strong effect of cloze in the parafoveal N400) and also integrate their parafoveal and foveal representations (as indicated by the reduction/elimination of the foveal N400 cloze effect).

Caveats

It is important to point out several caveats to the current findings. First, our pre-target words induced a low foveal load because they were typically short, high frequency words. Past work has shown that a low foveal processing load increases the amount of parafoveal information readers are able to extract (Henderson & Ferreira 1990; Payne & Stine-Morrow, 2012; Payne et al., 2016). This may have increased the likelihood that readers were able to extract parafoveal information. However, in our other recent work, we have shown that increasing foveal load (anomalous foveal words) does reduce the amount of information readers extract from the parafovea, but in a *graded* fashion (Payne et al., 2016), such that even under the highest foveal load some parafoveal information was still transmitted to the language processing system. The implication of this finding for the current results is that the low foveal load induced by our pre-target words may have increased the amount of parafoveal processing readers were able to do—but probably did not fundamentally change *whether* parafoveal processing took place at all.

Secondly, timing—both stimulus duration and ISI—may modulate effect patterns. In terms of ISI, we here chose to employ the slower presentation rate (450 ms SOA) used in Barber et al. (2013), who showed that parafoveal N400 effects were bigger with slower versus faster presentation speeds, particularly for low constraint sentences. This thus leaves open the question of whether readers can still assess the plausibility of parafoveal words when intervals between words are shorter, as they are in typical natural reading patterns. However, it is also important to note that timing is necessarily different between RSVP and normal reading, given the differences in how words enter the system in the two cases. When people are allowed to self-select speeds of RSVP reading (just as they, presumably, self-select speeds when controlling their eyes), they choose rates similar to those employed here (Ditman, Holcomb, & Kuperberg, 2007). Thus, the current stimulus onset asynchrony, while slower than average gaze durations when people are moving their eyes over text, may not be "slow" relative to the different demands imposed by RSVP presentation. At the same time, we used a stimulus duration that is actually *shorter* than that typically selected by readers during natural reading (100 ms stimulus duration vs. 250 ms average gaze durations). Future work is thus needed to explore whether a higher foveal processing load, induced by either more difficult foveal words or faster presentation rates, reduces the parafoveal N400 effects observed in our findings.

Other limitations of the RSVP-with-flankers ERP paradigm are similar to those of typical ERP studies: readers must maintain a central fixation, and they are presented with every word of the sentence for a fixed amount of time, meaning that they cannot control this speed, nor do they have the option to regress back in the sentence. However, the RSVP-with-flankers ERP paradigm does offer an advantage over typical ERP studies in that readers receive a parafoveal preview of the upcoming word, making it a closer approximation of natural reading. Furthermore, it offers an alternative to the boundary change paradigm, in that it allows the stricter manipulation of preview space and time; here, for example, we were able to ensure that our parafoveal words were clearly well outside of foveal vision (although, then, also subject to reduced crowding). Ultimately, we need to seek converging evidence from multiple methods—including, but not limited to, these two paradigms—to truly understand the mechanisms underlying parafoveal processing, rather than simply characterizing the findings obtained from any single method.

Conclusions

The current investigation showed that readers assess the semantics of parafoveal words during online processing, as evidenced by the graded N400 effect elicited by parafoveal words based on their cloze probability. Importantly, the appearance of an unexpected or anomalous word in the parafovea eradicated the N400 effect when that same word appeared at central fixation, suggesting that readers integrated across their foveal and parafoveal representations, such that they did not repeat semantic processing that they had already performed for the same word in the parafovea. At a broader level, this work demonstrates the utility of the flanker-word paradigm for the use of ERPs to study parafoveal processing as a way to bring the eye-tracking and ERP literatures closer to a consensus about the interaction between foveal and parafoveal processing during reading comprehension.

Author note The authors would like to thank Katherine Mimnaugh for assistance with data collection. This work was supported by a James S. McDonnell Foundation Scholar Award and NIH Grant AG026308 to Kara D. Federmeier. Portions of this research were presented at the 2015 Annual Meeting of the Cognitive Neuroscience Society.

References

Baccino, T., & Manunta, Y. (2005). Eye-fixation-related potentials: Insight into parafoveal processing. *Journal of Psychophysiology, 19*(3), 204–215.

Balota, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of Verbal Learning and Verbal Behavior, 22,* 88–104.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17*(3), 364–390.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445–459.

Barber, H. A., Ben-Zvi, S., Bentin, S., & Kutas, M. (2011). Parafoveal perception during sentence reading? An ERP paradigm using rapid serial visual presentation (RSVP) with flankers. *Psychophysiology, 48*(4), 523–531.

Barber, H. A., Doñamayor, N., Kutas, M., & Münte, T. (2010). Parafoveal N400 effect during sentence reading. *Neuroscience Letters, 479*(2), 152–156.

Barber, H. A., Van der Meij, M., & Kutas, M. (2013). An electrophysiological analysis of contextual and temporal constraints on parafoveal word processing. *Psychophysiology, 50*(1), 48–59.

Barr, D. J. (2013). Random effects structure for testing interactions in mixed-effects models. *Frontiers in Psychology 4*(328). doi:10.3389/fpsyg.2013.00328

Bates D., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4* (R package Version 1.1–9). Retrieved from https://CRAN.R-project.org/package=lme4

Davies, M. (2008). *The corpus of contemporary American English: 450 million words, 1990–present.*

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia, 61,* 150–162.

Dimigen, O., Kliegl, R., & Sommer, W. (2012). Trans-saccadic parafoveal preview benefits in fluent reading: A study with fixation-related brain potentials. *NeuroImage, 62*(1), 381–393.

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General, 140*(4), 552–572.

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007). An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology, 44*(6), 927–935.

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

Ehrlich, S., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20,* 641–655.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146,* 75–84.

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage*. Boston: Houghton Mifflin.

Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 417–429.

Hohenstein, S., & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(1), 166–190.

Hohenstein, S., Laubrock, J., & Kliegl, R. (2010). Semantic preview benefit during eye movements in reading: A parafoveal fast-priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1150–1170.

Inhoff, A. W. (1982). Parafoveal word perception : A further case against semantic preprocessing. *Journal of Experimental Psychology: Human Perception and Performance, 8*(1), 137–145.

Inhoff, A. W., & Rayner, K. (1980). Parafoveal word perception: A case against semantic preprocessing. *Perception & Psychophysics, 27,* 457–464.

Kliegl, R., Hohenstein, S., Yan, M., & McDonald, S. A. (2012). How preview space/time translates into preview cost/benefit for fixation durations during reading. *The Quarterly Journal of Experimental Psychology* 1–20. doi:10.1080/17470218.2012.658073

Kretzschmar, F., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). Parafoveal versus foveal N400s dissociate spreading activation from contextual fit. *Neuroreport, 20*(18), 1613–1618.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463–470.

Kutas, M., & Federmeier, K. D. (2007). Event-related brain potential (ERP) studies of sentence processing. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 385–406). Oxford: Oxford University Press.

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.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161–163.

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language, 61*(3), 326–338. doi:10.1016/j.jml.2009.06.004

Li, N., Niefind, F., Wang, S., Sommer, W., & Dimigen, O. (2015). Parafoveal processing in reading Chinese sentences: Evidence from event-related brain potentials. *Psychophysiology, 52*(10), 1361–1374.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). Cambridge: MIT Press.

Luck, S. J., Vogel, E. K., & Shaprio, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature, 383,* 616–618.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98*(1), 185–199.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227–234.

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition, 4*(5), 648–654.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113.

Payne, B. R., Lee, C., & Federmeier, K. D. (2015). Revisiting the incremental effects ofcontext on word processing: Evidence from word-level event-related brain potentials. *Psychophysiology, 52,* 1456–1469. doi:10.1111/psyp.12515

Payne, B. R., & Stine-Morrow, E. A. L. (2012). Aging, parafoveal preview, and semantic integration in sentence processing: Testing the cognitive workload of wrap-up. *Psychology and Aging, 27*(3), 638–469.

Payne, B. R., Stites, M. C., & Federmeier, K. D. (2016). Out of the corner of my eye: Foveal semantic load modulates parafoveal processing in reading. *Journal of Experimental Psychology: Human Perception & Performance, 42*(11), 1839–1857.

Rayner, K. (1975). Parafoveal identification during a fixation in reading. *Acta Psychologica, 39,* 271–282.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Rayner, K., Balota, D. A., & Pollatsek, A. (1986). Against parafoveal semantic preprocessing during eye fixations in reading. *Canadian Journal of Psychology, 40*(4), 473–483. doi:10.1037/h0080111

Rayner, K., & Schotter, E. R. (2014). Semantic preview benefit in reading English: The effect of initial letter capitalization. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1617–1628.

Rayner, K., Schotter, E. R., & Drieghe, D. (2014). Lack of semantic parafoveal preview benefit in reading revisited. *Psychonomic Bulletin & Review, 21,* 1067–1072.

Risse, S., & Kliegl, R. (2012). Evidence for delayed parafoveal-on-foveal effects from word *n+2* in reading. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 1026–1042.

Risse, S., & Kliegl, R. (2014). Dissociating preview validity and preview difficulty in parafoveal processing of word *n* + 1 during reading. *Journal of Experimental Psychology: Human Perception and Performance, 40*(2), 653–668.

Schotter, E. R. (2013). Synonyms provide semantic preview benefit in English. *Journal of Memory and Language 69*(4). doi:10.1016/j.jml.2013.09.002

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception & Psychophysics, 74*(1), 5–35. doi:10.3758/s13414-011-0219-2

Schotter, E. R., & Jia, A. (2016). Semantic and plausibility preview benefit effects in English: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory & Cognition, 42*(12), 1839–1866.

Schotter, E. R., Lee, M., Reiderman, M., & Rayner, K. (2015). The effect of contextual constraint on parafoveal processing in reading. *Journal of Memory and Language, 83,* 118–139.

Tsai, J. L., Kliegl, R., & Yan, M. (2012). Parafoveal semantic information extraction in traditional Chinese reading. *Acta Psychologica, 141*(1), 17–23.

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage, 62,* 356–366.

Yan, M., Pan, J., Bélanger, N. N., & Shu, H. (2015). Chinese deaf readers have early access to parafoveal semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(1), 254–261.

Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2009). Readers of Chinese extract semantic information from parafoveal words. *Psychonomic Bulletin & Review, 16*(3), 561–566.

Yan, M., Risse, S., Zhou, X., & Kliegl, R. (2012). Preview fixation duration modulates identical and semantic preview benefit in Chinese reading. *Reading and Writing, 25*(5), 1093–1111. doi:10.1007/s11145-010-9274-7

Yan, M., Zhou, W., Shu, H., & Kliegl, R. (2012). Lexical and sublexical semantic preview benefits in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(4), 1069–1075.

Yang, J. (2013). Preview effects of plausibility and character order in reading Chinese transposed words: Evidence from eye movements. *Journal of Research in Reading, 36*(Supp. 1), 18–35.

Zhou, W., Kliegl, R., & Yan, M. (2013). A validation of parafoveal semantic information extraction in reading Chinese. *Journal of Research in Reading, 36*(Suppl. 1), 51–64.