


## Comprehending surprising sentences: sensitivity of post-N400 positivities to contextual congruity and semantic relatedness

Katherine A. DeLong & Marta Kutas


To cite this article: Katherine A. DeLong & Marta Kutas (2020) Comprehending surprising sentences: sensitivity of post-N400 positivities to contextual congruity and semantic relatedness, *Language, Cognition and Neuroscience*, 35:8, 1044-1063, DOI: [10.1080/23273798.2019.1708960](https://doi.org/10.1080/23273798.2019.1708960)

To link to this article: <https://doi.org/10.1080/23273798.2019.1708960>

 View supplementary material 

 Published online: 06 Jan 2020.

 Submit your article to this journal 

 Article views: 385

 View related articles 

 View Crossmark data 



## Comprehending surprising sentences: sensitivity of post-N400 positivities to contextual congruity and semantic relatedness

Katherine A. DeLong<sup>a</sup> and Marta Kutas<sup>a,b,c,d</sup>

<sup>a</sup>Department of Cognitive Science, University of California, San Diego, CA, USA; <sup>b</sup>Center for Research in Language, University of California, San Diego, CA, USA; <sup>c</sup>Department of Neurosciences, University of California, San Diego, CA, USA; <sup>d</sup>Kavli Institute for Brain and Mind, University of California, San Diego, CA, USA

### ABSTRACT

Any proposal for predictive language comprehension must address receipt of less expected information. While a relationship between the N400 and sentence predictability is well established, a clear picture is still emerging of the link between post-N400 positivities (PNPs) and processing of semantically unexpected words, as well as any relation to other not-specifically-linguistic and/or syntactic late positivities. The current study employs event-related brain potentials (ERPs) to congruent and anomalous words to assess the impacts of semantic relatedness and contextual plausibility on processing unpredictable sentences. We observe PNPs with different scalp topographies to plausible unexpected words unrelated to predictable continuations (anterior PNP) and to anomalous words, regardless of, but delayed by, relatedness (posterior PNP). We offer functional explanations that reconcile inconsistencies with reported PNP findings and place added constraints on the anterior PNP's proposed link to inhibitory processing. We also suggest a testable general cognitive account for the posterior PNP.

### ARTICLE HISTORY

Received 11 September 2019  
Accepted 7 December 2019

### KEYWORDS

Sentence comprehension; prediction; post-N400 positivity (PNP); ERP; language processing

### Introduction

Comprehending sentences in real time requires individuals to dynamically make use of language input, stored knowledge, and extralinguistic cues. In combination, these sources set the stage for and narrow the scope of information likely to be encountered, with spoken or written sentences creating worlds from whole cloth and confirming or diverting from expectancies at a variety of levels. All of this happens quickly, with comprehenders benefitting from input that continues predictably and confirms expectations. However, not all language is predictable, and if prediction in language is like that in other neural domains, then the parser should be capable of fluidly handling less expected yet informative continuations. Readers and listeners generally count on being able to make sense of even surprising input, but from an experimental standpoint, there is much to be learned from examining brain responses to continuations that are more difficult to interpret. For instance, for a context like “The man wore khaki shorts, a loud Hawaiian shirt, and a camera around his neck. You could definitely tell that he was a ...” understanding the consequences of encountering words as varied as *tourist*, *visitor*, *bachelor*, *resort*, or *hunger* offers a window into the type of information that gets

preactivated when constraint is high, as well as the nature of contextual representations that are formed and perhaps require additional processing upon receipt of less predictable input.

Over the past two decades, online sentence processing studies have demonstrated that the neural processor can anticipate different aspects of likely upcoming continuations, including lexico-semantic features (Federmeier & Kutas, 1999; Kuperberg & Wlotko, 2018; Kwon, Sturt, & Liu, 2017; Otten & Van Berkum, 2008; Szewczyk & Schriefers, 2013), word class (Dikker, Rabagliati, Farmer, & Pyllkkänen, 2010; Luke & Christianson, 2016), morphosyntax (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004), and word forms (DeLong, Urbach, & Kutas, 2005; Laszlo & Federmeier, 2009; Martin et al., 2013), among others. Extensive event related brain potential (ERP) research has shown that when sentence contexts provide sufficient contextual constraint, semantic processing of predictable continuations is eased as reflected in decreased N400 amplitudes to those words. The N400 is a well-studied ERP component, shown to be sensitive not only to the degree of an item's semantic fit within a context, but also to the stored relational/organisational structure of the

information cued by the stimulus in its context. In the related anomaly paradigm, predictable sentence continuations are alternated with words that do not make sense, but are related or not to either the predictable word or the context in some way. Using this paradigm, reduced N400s to unpredictable but related words (compared to unpredictable unrelated words) have been demonstrated for numerous linguistic features, including semantic/categorical information (Federmeier & Kutas, 1999; Kutas, Lindamood, & Hillyard, 1984), event knowledge (Metusalem et al., 2012), word orthography (DeLong, Chan, & Kutas, 2019; Laszlo & Federmeier, 2009), and perceptual attributes (Amsel, DeLong, & Kutas, 2015; Rommers, Meyer, Praamstra, & Huettig, 2013). These N400 reductions have been taken to indicate that aspects of the expected word or the evolving contextual representation have already been preactivated, and have collaterally facilitated the related but anomalous word's processing. On this view, the N400 does not merely index goodness of fit, but rather is sensitive to the neural activation state at the time a physical stimulus is encountered (see Kutas & Federmeier, 2011; Laszlo & Federmeier, 2011).

Since sentences do not always continue as expected, predictions for word information are not always met (at one or more linguistic levels), with information not being activated with sufficient strength and/or rapidly enough to benefit processing. For such input, in addition to modulations of N400 amplitude, later ERP components, in the form of post-N400 positivities (PNPs) have also been observed. Such positivities have recently become a focus of study in the context of linguistic prediction research due to proposals linking them to potential consequences of receiving surprising input, in particular when expectancy for predictable information is high. While the N400 is not considered to reflect a "cost" for mispredicting (its amplitude *decreases* with supportive context, but does not *increase* when predictions are violated, see Van Petten & Luka, 2012; and Luke & Christianson, 2016, for discussions), there are some indications that post-N400 positivities, particularly ones with more anterior scalp distributions, may reflect just such a consequence. However, the functional natures of the late positivities to less predictable continuations are not well understood. Nor is it clear whether there are multiple PNPs with different neural generators, what their sensitivities are, what processing functionalities differentiate them, or what the specific characteristics are of the ERP components themselves (e.g. timing and topography). A clearer delineation of the positivities observed as part of the biphasic N400-PNP complex would thus offer a potential window into fundamental aspects of language comprehension, in particular

into how less predictable linguistic information gets processed.

Theories about post-N400 positivities to semantically unexpected sentence continuations have evolved over the years (refer to Bornkessel-Schlesewsky & Schlesewsky, 2008; DeLong, Quante, & Kutas, 2014; Van Petten & Luka, 2012, for reviews). While early sentence processing ERP studies occasionally noted a late positive complex (LPC) following an N400 to unexpected or improbable sentence continuations (e.g. Juottonen, Revonsuo, & Lang, 1996; Kutas & Hillyard, 1980; McCallum, Farmer, & Pocock, 1984), such responses were not systematically analysed or reported in the literature. Following a period in the 1990s when researchers attributed a parietal late positive brainwave response to processing associated with syntactic violations and ambiguities (with some labelling the effect as a P600 and others as a "syntactic positive shift" or SPS: refer to Gouvea, Phillips, Kazanina, & Poeppel, 2010, for a review), a wave of studies in the mid-2000s found that a morphologically similar posterior late positive ERP response could also be elicited to thematic role violations, animacy violations and so-called semantic illusions (e.g. Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005; Nieuwland & Van Berkum, 2005, respectively; see Kuperberg, 2007, for a review). These positivities occurred sometimes instead of and sometimes in addition to N400s, leading to this class of findings being referred to as "semantic P600s" or "semantic illusion effects". While the experimental manipulations leading to these effects were more semantic than the strictly syntactic ones that had led some researchers to speculate about the "SPS", many of their eliciting conditions involved role reversals, by which a "reordering" approach might have allowed for a licensed interpretation. Other research, however, pointed to similar effects being generated absent these reversal conditions, with the late positivity varying as a function of contextual plausibility (e.g. Van De Meeren-donk, Kolk, Vissers, & Chwilla, 2010) or impossibility (Paczynski & Kuperberg, 2011). Brouwer, Fitz, and Hoeks (2012) proposed that the variety of parietal P600s observed over the years can be unified under a theory whereby the effect broadly reflects the integration of already retrieved word meaning (with retrieval indexed by the N400), with the late positivity's amplitude indexing construction, reorganisation, or updating of a contextual representation.

In an important review of the literature, Van Petten and Luka (2012) cataloged sentence comprehension study post-N400 positivities (offering the theory-neutral label, PNP) to experimental manipulations that were semantic in nature, subdividing comparisons of congruent to incongruent and high to low cloze probability

(but plausible) continuations. Their survey noted that PNP scalp anteriority patterned with the contextual congruity (i.e. plausibility) of the unexpected continuations: less predictable *implausible* continuations elicited more *posterior* PNPs, while less predictable *plausible* words generally elicited more *anterior* PNPs.<sup>1</sup> At the time, aPNP findings to unexpected but plausible sentence continuations were relatively uncommon, likely due to the more limited use of strictly plausible stimuli for contrast with predictable continuations in sentence ERP studies. A few studies employing such conditions are described below.

When Moreno, Federmeier, and Kutas (2002) presented English-Spanish bilinguals with sentences or idioms in which predictable English words were replaced with either their Spanish translations or with within-language synonyms (both interpretable/plausible within the sentence context), an aPNP was observed to both types of substitutions. The researchers speculated at the time that the effect may implicate frontal brain regions involved in switching between stimuli or tasks. Federmeier, Wlotko, De Ochoa-Dewald, and Kutas (2007), too, observed a similar aPNP pattern in a study where two levels of contextual constraint (strong or weak) were crossed with two levels of expectancy (best completions or unexpected/low cloze probability plausible words). Low cloze continuations of high constraint sentences elicited an aPNP (500–900 ms) relative to the other three conditions, which did not differ from each other. Federmeier et al. (2007) proposed this effect related to a possible cost for processing unexpected words in highly predictive sentence contexts. In a somewhat similar study, albeit with different results, Thornhill and Van Petten (2012) tested 2 levels of contextual constraint (high/low) with three types of plausible sentence continuations: best completions (78% mean cloze probability for high and 30% for low constraint contexts) and very low cloze probability words either synonymous or not with best completions. Between 600 and 900 ms, all of the continuations that were not highly predictable (regardless of sentence constraint or relatedness) elicited an aPNP relative to best completions of high constraint sentences, with none of the less predictable conditions differing significantly in mean amplitude. The authors took these results to indicate that the component was sensitive to violations of a specific lexical expectancy (and not just semantic expectancy), since the aPNP effect was elicited independent of a semantic relationship with the expected ending.

DeLong, Urbach, Groppe, and Kutas (2011) also reported an aPNP (measured between 500 and 1200 ms) to plausible low cloze continuations of variably constraining contexts for young adults, with an

amplitude inversely correlated with the cloze probability of the plausible critical words. Over several subsequent studies, DeLong and colleagues have repeatedly encountered PNPs to less predictable continuations of highly constraining sentence contexts, noting differentiation in the scalp anteriority of PNPs to unexpected continuations that – as Van Petten and Luka (2012) suggested – varies as a function of the continuation's contextual plausibility. For instance, DeLong, Groppe, Urbach, and Kutas (2012) observed a sustained aPNP to unexpected plausible sentence continuations beginning as early as the N400 time window (initiating before 400 ms) for young adults, as well as older adults with higher verbal fluency. In a study designed specifically to contrast conditions that may lead to differing anteriority PNP patterns, DeLong et al. (2014) found that the contextual plausibility of prediction violations for highly constraining sentences indeed appeared to modulate the scalp anteriority of the elicited PNPs, with more plausible continuations exhibiting aPNPs and anomalous words pPNPs. Quante, Bölte, and Zwitterlood (2018) replicated these findings in a German language version of DeLong et al. (2014).

Very recently, there have been some attempts to establish the functional roles, in particular, of aPNP patterns. Ness and Meltzer-Asscher (2018) linked an aPNP to inhibitory processes that they suggest may be required when unexpected but congruent sentence continuations appear instead of predictable words. This is akin to a proposal by Kutas (1993) that an aPNP pattern should not be expected to anomalous unexpected words since they cannot be interpreted in context and therefore would not require suppression. (For a contrasting view, see Brothers, Swaab, & Traxler, 2015, who argue that the aPNP does not reflect inhibition). A similar idea has been discussed by Kuperberg and Wlotko (2018), who propose that the aPNP relates to successful updating of a sentence representation when an unexpected (but interpretable) alternative is received, which necessarily involves suppressing the expected word. In contrast, they consider the pPNP to reflect a failure to integrate unpredicted information and possibly the brain's attempt to reanalyse the accrued mental contextual representation.

In sum, the picture emerging regarding anterior and posterior PNP patterns is that: (1) both components are present in a time range following the N400, although the aPNP may start slightly earlier, (2) both occur to stimuli improbable in their contexts, (3) both may be more pronounced for words continuing more constraining contexts, (4) both reflect a different kind of processing than that reflected by preceding N400s, and (5) neither specifically reflects syntactic processing. However,

questions remain about the differing sensitivities and functionalities of the PNPs, with limited attempts, to date, to tease them apart in the context of single studies (but see DeLong et al., 2014; Kuperberg & Wlotko, 2018; Quante et al., 2018). Additionally, although the N400 is sensitive to a sentence continuation's semantic relatedness to a predicted word, the picture is less clear for PNPs. Thornhill and Van Petten's (2012) finding that the aPNP was *insensitive* to degree of relatedness to expected continuations was the basis for their conclusion that the effect may index a violation of an expectation for a specific word. However, the authors outlined a very different set of possible conclusions if the unexpected unrelated continuations had elicited larger aPNPs than the related ones. Specifically, they suggested that such a pattern would have been consistent with proposals linking the effect to "conceptual novelty" triggered by the unrelated word, which would require the comprehender to revise the mental representation they had constructed based on the preceding context.

The current study extends the work of Thornhill and Van Petten (2012), by crossing a related anomaly paradigm with a sentence congruity manipulation to test for differential PNP brain responses to unexpected sentence continuations. Diverging from that study, we limit the current experiment to high constraint sentence contexts (e.g. "The man wore khaki shorts, a loud Hawaiian shirt, and a camera around his neck. You could definitely tell that he was a ...") as well as testing not just congruent unexpected continuations semantically/associatively related (*visitor*) or unrelated (*bachelor*) to best completions (*tourist*), but anomalous ones as well (*resort-related*, *hunger-unrelated*). Our goals are to replicate the general findings of DeLong et al. (2014), in which the anteriority of the PNP patterns split along the dimension of word plausibility, and also to extend the findings of Thornhill and Van Petten (2012), in determining whether the aPNP, in particular, is more sensitive to words semantically unrelated to predictable continuations – words which presumably alter to a greater degree the contextual representation that has been constructed to that point.

Both N400 and PNP analyses will be conducted. Based on the literature, it is anticipated that both congruent and anomalous unexpected words will elicit larger N400s relative to best completions, with additional N400 amplitude reductions for words semantically-associatively related to best completions (a main effect of relatedness). Unclear is whether there will similarly be effects of plausibility that lead to N400 amplitude reductions for more congruent compared to anomalous items. Following the N400, ERPs will be assessed in PNP

time windows over both anterior and posterior electrode sites. If the brain engages in different processing for unexpected continuations based on item plausibility, then we anticipate an increased PNP over posterior sites for anomalous items and anterior sites for congruent items. Furthermore, if congruent items do exhibit an anterior positivity, we will examine whether the effect is modulated by semantic relatedness. If unrelated congruent words elicit larger aPNPs than related ones, this would suggest that the effect does *not* index suppression associated with a violation for a specific lexical prediction, since, in this case, any word that was not the most predictable one should lead to a similar amplitude aPNP (similar to Thornhill & Van Petten, 2012). Instead, such results would suggest that the effect is sensitive to the degree to which the plausible continuation causes the contextual representation to conceptually/semantically shift from the predicted (already constructed) one, potentially aligning with a view in which increased processing (e.g. for overwriting, inhibiting or switching from the most active representation) is required. For the pPNP, since neither related nor unrelated anomalous words are interpretable, we would not necessarily predict a difference in amplitude on the basis of relatedness.

## Materials and methods

### Stimulus materials

Stimulus materials consisted of 200 highly constraining sentence pairs, with contexts and their most likely noun continuations taken from a variety of ERP experiments conducted in our lab (61 items) as well as from studies conducted by Kara Federmeier and colleagues at University of California, San Diego (UCSD) and University of Illinois at Urbana-Champaign (139 items, used with their permission). Sentence pair contexts were continued by 5 possible word types, yielding a total of 1000 unique stimulus items. Critical word conditions were best completions (BC, the highest cloze probability word continuation for a given context), or one of 4 unexpected word types. Unexpected conditions were constructed by crossing factors of congruity of contextual fit (2 levels, congruent or anomalous) and semantic/associative relatedness to the BC (2 levels, related or unrelated), yielding congruent related (CR), congruent unrelated (CU), anomalous related (AR), and anomalous unrelated (AU). See Table 1 for representative stimuli.

### Cloze probability/contextual constraint

Best completion (BC) cloze probabilities were available for all 200 experimental materials, which had been

**Table 1.** Example stimuli.

Sentence Context	Best Completion (BC)	Congruent Related (CR)	Congruent Unrelated (CU)	Anomalous Related (AR)	Anomalous Unrelated (AU)	Sentence Final Context
The tourists had many excursions planned. They visited the pandas at the ...	zoo	sanctuary	website	zebra	volume	... on Tuesday.
The dog stood its ground. The Doberman bared its ...	teeth	fangs	belly	dentist	report	... and growled.
Bill jumped in the lake. He made a big ...	splash	ripple	mistake	mermaid	guide	... with his cannonball.
The homeowners had many valuables. They kept most of them in a ...	safe	box	boat	locksmith	paragraph	... out of sight.
The groom took the bride's hand. He placed the ring on her ...	finger	pinky	pillow	grip	function	... very carefully.

normed in conjunction with previous studies using standard offline cloze probability norming tasks with single sentences truncated prior to the critical words. For the DeLong contexts, 25–30 UCSD student volunteers, compensated with experimental credit or cash, completed each item. Mean contextual constraint across the full set of 200 items was 94.0% (range = 87 to 100%,  $SD = 3.6\%$ ) with similar values for materials from both sources. Contextual constraint was operationalised as the cloze probability (proportion) of the most frequent norming response. For the current study, the single sentences were restructured into sentence pairs to minimise the time required for ERP participants to control their blinking. Critical words were sentence medial, with mean critical word position in the second rapid serial visual presentation (RSVP) sentence being 7.5 ( $SD = 1.2$ ) words, and mean second sentence length being 9.6 ( $SD = 1.2$ ) words. Cloze probabilities for non-BC continuations could not be calculated across the entire stimulus set because raw cloze probability norming data were available only for the DeLong items. However, this seemed unproblematic because such high contextual constraint ensured that non-BC words were unlikely to have been provided during norming, and consequently would have near zero cloze probabilities. Inspection of norming responses for the 61 DeLong items indeed indicated that this was the case.<sup>2</sup>

### Relatedness

For the unexpected words, items in the two related conditions (CR and AR) were chosen on the basis of having a semantic/associative relationship with the BC words. Related words were primarily selected from entries for the BC words, using a variety of word association and word relatedness resources (including the University of South Florida Free Association Norms, Nelson, McEvoy, & Schreiber, 1998; wordassociations.net; MacMillan Online Dictionary and Thesaurus; relatedwords.org; and onelook.com/thesaurus). The semantic/associative

relation of all unexpected words with BC words was further assessed using pairwise latent semantic analysis (LSA, General Reading up to 1st year college, Landauer & Dumais, 1997). Results from pairwise *t*-tests confirmed that on average, the pairwise LSA with BC values were lower for the two unrelated conditions than for the two related conditions (Table 2).

### Congruity/plausibility

For the unexpected conditions, congruent continuations were selected to make sense in context, and anomalous words were chosen to not make sense. One way of assessing item congruity is in terms of plausibility in context. To assess plausibility, we collected offline ratings for all 1000 context + critical word combinations.

Sentence pairs up to and including critical nouns were rated for plausibility in a separate, off-line sentence rating task performed by University of California, San Diego student volunteers, compensated with experimental credit toward coursework. Participants were instructed to read the short sentence passages and rate them on plausibility from 1 to 7, where 7 indicates that the passage describes something *highly plausible* and 1 indicates that the passage describes something that is *highly implausible*. Each of the 1000 sentence pair + critical noun experimental items was rated by 15–18 individuals. See Table 2 for mean condition plausibility ratings.

Pairwise *t*-tests indicated that plausibility ratings for all 5 experimental conditions differed significantly from each other (all comparisons with  $p$ -values  $\leq 0.000001$ ). For both the congruent and the anomalous conditions, relatedness to BC had the unintended consequence of contributing to increases in rated plausibility, although under the experimenters' interpretation of plausibility, CR and CU items were assessed to be equally plausible, and AR and AU conditions similarly implausible. Ultimately, plausibility raters may have found it difficult to

**Table 2.** Condition lexical factor means.

	Cloze probability	Relatedness (Pairwise LSA with BC)	Congruity (Plausibility ratings with 7 = highest, 1 = lowest plausibility)	Word length	Orthographic neighbourhood
BC	0.94	–	6.43	4.77	6.94
CR	Near Zero	0.38	5.77	5.99	4.03
CU	Near Zero	0.13	4.13	6.01	4.11
AR	Near Zero	0.30	2.65	5.93	4.57
AU	Near Zero	0.07	1.95	5.97	3.49

disregard words' relatedness to the BC, and in turn dissociate relatedness to BC from true plausibility. This possibility does not seem unreasonable, given that the same "bump" in plausibility ratings occurred for the anomalous conditions – despite the nonsensicalness of both AR and AU continuations.

### Other lexical factors

Across the four unexpected conditions, we also attempted to minimise differences in lexical factors known to affect the ERP. In particular, word lengths across the four unexpected conditions were similar as indicated by pairwise comparisons (all *t*-tests revealing *p*-values > .05). For orthographic neighbourhood, only the AU and AR conditions differed statistically from each other (*p* < .05): the other unexpected conditions, on average, had similar orthographic neighbourhoods (see Table 2).

### ERP stimulus lists

Each of the 200 sentence pair contexts had 5 possible continuation types, yielding 1000 unique stimuli. The stimuli were divided into 5 lists of 200 items each, with individual ERP participants viewing each context only once. The contexts/conditions were counterbalanced across the 5 lists, with each list containing 40 items from each of the 5 continuation types. Participants were randomly assigned to one of the 5 lists at the beginning of the experiment. Fifty randomly occurring yes/no comprehension questions followed one quarter of the experimental items. Comprehension questions were included to ensure that ERP participants were paying attention. There were no filler items.

### ERP participants

Thirty UCSD volunteers (18 females, 12 males) participated for course credit or cash. Participants were right-handed, native English speakers with normal or corrected-to-normal vision, ranging from 18 to 29 years old (mean = 20.1 years). Ten participants reported a left-handed parent or sibling. Four additional participants were excluded from the analysis due to excessive eye movements or other artefacts.

### Offline tasks and measures

Prior to and following ERP testing, we collected several offline neuropsychological measures from individual participants. Before beginning ERP data collection, we administered the Author and Magazine Recognition Tests (ART and MRT), based on Stanovich and West (1989), to assess general print/reading experience. Following ERP testing, we administered a debriefing questionnaire inquiring about what stood out to the participants during the course of the experiment, as well as what the participants thought was the purpose of the study. Individuals also completed the following tests: verbal fluency, letter and category (Benton & Hamsher, 1978), involving individuals generating as many lexical items of the appropriate type as possible – beginning with a given letter or belonging to a particular category – in a limited time, and a paper version of the word-color Stroop interference task (based on Stroop, 1935) to measure inhibition/cognitive control. The purpose of collecting this offline data was to assess potential variability in the individual ERP results, given past work in our lab, which has suggested that certain neuropsychological factors may correlate with aspects of predictive language processing. However, these results were not part of our main research questions.

### ERP procedure

ERPs were recorded in a single session in a sound-attenuating, electrically shielded chamber. Participants sat one metre in front of a CRT monitor and read sentence pairs for comprehension. Yes/no comprehension questions were responded to with two hand-held buttons, with response hand counterbalanced across participants and lists. Stimuli were presented visually in white type on a black background, in 8 blocks, with short breaks in between. Context sentences were presented in their entirety, with participants advancing to RSVP critical word sentences via button press. RSVP sentences began with an orienting fixation cross (whose duration was jittered between 1000 and 1500 ms), which remained on the screen throughout the RSVP sentence. Individual words were presented centrally, directly above the fixation cross, for a duration of 200 ms and

an interstimulus interval of 300 ms, yielding a stimulus onset asynchrony (SOA) of 500 ms. Following the final word of the RSVP sentence, the fixation cross remained on screen for an additional 1000–1300 ms, after which participants were instructed that they could blink. If a comprehension question did not appear, then the screen automatically advanced to the context sentence of the next experimental item with a 3 s blank screen interval between sentences.

### **Electroencephalographic recording and analysis**

The electroencephalogram (EEG) was recorded from 26 electrodes arranged geodesically in an Electro-Cap, each referenced online to an electrode over the left mastoid. Blinks and eye movements were monitored from electrodes placed on the outer canthi and under each eye, also referenced to the left mastoid process. Electrode impedances were kept below 5 K $\Omega$ . The EEG was amplified with Grass amplifiers with a pass band of 0.01 to 100 Hz and was continuously digitised at a sampling rate of 250 samples/second. Data were re-referenced off-line to the algebraic mean of the left and right mastoids and averaged for each experimental condition, time-locked to the critical word onsets. Single trial epochs spanning 500 ms prestimulus to 1540 ms poststimulus onset were extracted from the continuous EEG. ERPs were computed by time-locking to critical noun onsets and performing baseline correction by subtracting the mean amplitude over the 500 ms precritical word onset. Artefact screening was done via computer algorithm and confirmed by visual inspection. On average, 8% of trials (SD = 6%, range 0 to 23%) were contaminated with eye movements, excessive muscle activity, or amplifier blocking and were excluded from further analysis.

With the experimental goal of testing for N400 and post-N400 positivity (PNP) differences, we conducted ANOVAs on ERP mean amplitudes utilising a canonical N400 time window (300–500 ms) and earlier and later post-N400 positivity (PNP) time windows (600–900 and 900–1200 ms). Splitting the PNP time region into earlier and later portions was motivated by there being little consensus in the literature on precise PNP timing, as well as results from a previous study (DeLong et al., 2014), which suggested a potentially earlier latency for the aPNP than the pPNP in the time window following the N400. We divided the scalp into 3 anteriority regions of interest (ROIs): prefrontal (3 channels: MiPf, LMPf, RMPf), central (3 channels: MiCe, LMCe, RMCe), and occipital (3 channels: MiOc, LMOc, RMOc), roughly suggested by analyses from Kuperberg and Wlotko (2018). Mean amplitude measures for each time

window were analysed by ANOVA over each of the 3 ROIs, with 5 levels of condition (BC, CR, CU, AR, AU). ANOVAs limited to the 4 unexpected conditions were also conducted, testing 2 levels of congruity (congruent, anomalous), 2 levels of relatedness (related, unrelated), and their interactions over each of the 9 time regions/ROIs. All ANOVA *p*-values reported herein are after epsilon correction (Huynh-Feldt) for repeated measures with more than one degree of freedom (*df*) in the numerator, with the original *df*.

### **Behavioural results**

Participants correctly answered an average of 94% (SD = 5%, range 78%–100%) of the comprehension questions. This high performance on the comprehension questions indicates participants were attending to and comprehending the sentences during the experiment.

### **ERP results**

See Figure 1 for ERPs over all 26 scalp electrodes for all 5 conditions, and Tables 3–5 and Figure 2 for reporting from statistical tests. For statistical reporting, not significant (ns) =  $p > .10$ ; marginal (mg) =  $.05 < p < .10$ ;  $.01 < p^* < .05$ ;  $.001 < p^{**} < .01$ ;  $p^{***} < .001$ . Values in Tables 3–5 are *F*-ratios with degrees of freedom as indicated. Mean amplitudes for the three time windows, from three anteriority ROIs, with three medial channels each, were analysed.

#### **N400 time window (300–500 ms)**

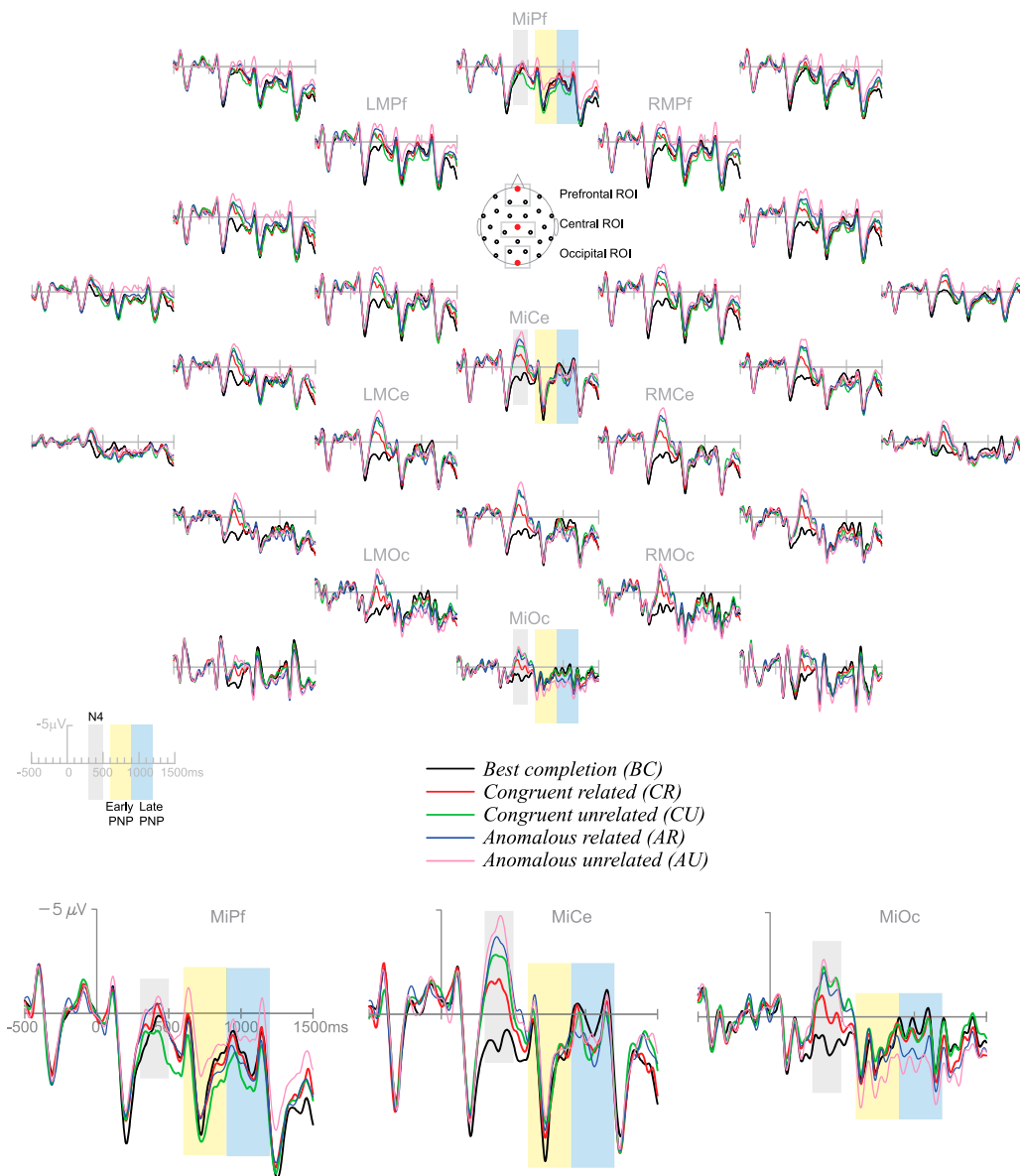
##### **Overview 300–500 ms**

In this time window, canonical N400 mean amplitude reductions were observed for best completions (BC) relative to unexpected conditions at all 3 anteriority ROIs. In addition, there were canonical N400 relatedness effects (related reduced relative to unrelated items) over more posterior sites, although primarily for unexpected congruent items. This pattern differed at prefrontal sites, where congruent unrelated (CU) shifted more positive, hinting at the beginning of an aPNP within the N400 time window. There were also congruity effects (congruent items showing more reduced N400s than anomalous) at all three ROIs.

##### **Prefrontal ROI**

The 5-condition ANOVA revealed a significant main effect during the N400 time window with the following pattern (Most Negative AU > AR > CR > CU > BC Most Positive). Unlike at the more posterior ROIs, there was no significant relatedness effect in the prefrontal N400





**Figure 1** Grand-averaged ( $N = 30$ ) ERPs recorded over 26 scalp channels, negative voltage plotted up. Boxed areas on the schematic scalp diagram indicate the 3 regions of interest (ROIs) used for analyses (prefrontal, central, and occipital), with midline electrodes from each of these ROIs highlighted in red and ERPs for those channels enlarged at the bottom. ERP analysis time windows are highlighted (grey = N400, 300–500 ms; yellow = Early PNP, 600–900 ms; and blue = Late PNP, 900–1200 ms).

**Table 3.** *F*-ratios from ANOVA analyses of 5 levels of condition for 3 time windows at 3 ROIs.

	300–500 ms	600–900 ms	900–1200 ms
PREFRONTAL	7.55***	7.93***	5.23***
CENTRAL	45.08***	1.23 ns	1.36 ns
OCCIPITAL	35.06***	3.48*	8.08***

*F*-values with (4,116) degrees of freedom for each test.

time window, with CU mean amplitude not differing significantly from CR or BC amplitude, suggesting the beginning of a positive shift for the CU condition. The Congruity (2) x Relatedness (2) ANOVA, limited to the unexpected conditions, revealed a significant main

**Table 4.** *F*-ratios from ANOVA analyses limited to 4 unexpected conditions. Main effects of congruity (congruent, anomalous), relatedness (related, unrelated) and interactions for 3 time windows at 3 ROIs.

	300–500 ms	600–900 ms	900–1200 ms
PREFRONTAL			
Congruity (C)	14.15***	14.84***	7.70**
Relatedness (R)	0.01 ns	1.16 ns	1.36 ns
C X R interaction	2.07 ns	16.69***	11.12**
CENTRAL			
Congruity	23.35***	1.78 ns	0.23 ns
Relatedness	13.02**	0.47 ns	0.15 ns
C X R Interaction	2.01 ns	0.40 ns	0.41 ns
OCCIPITAL			
Congruity	8.40**	5.22*	14.77***
Relatedness	14.92***	2.50 ns	1.26 ns
C X R Interaction	9.23**	10.97**	1.44 ns

*F*-values with (1,29) degrees of freedom for each test.

**Table 5.** *F*-ratios for pairwise comparisons of each condition type (5 levels) for 3 time windows at 3 ROIs.

	300–500 ms				600–900 ms				900–1200 ms			
	BC	CR	CU	AR	BC	CR	CU	AR	BC	CR	CU	AR
<b>PREFRONTAL</b>												
CR	7.98**				0.17				0.48			
CU	1.29	1.78			3.93 mg	4.42*			4.66*	2.83		
AR	13.50**	2.19	9.20**		0.07	0.49	5.50*		0.61	0.01	1.66	
AU	20.13***	5.14*	16.47***	0.64	10.68**	13.02**	30.84***	10.67**	5.61*	7.81**	26.77***	8.24**
<b>CENTRAL</b>												
CR	56.26***											
CU	67.29***	16.48***		ns main effect				ns main effect				
AR	84.25***	25.54***	1.5									
AU	99.88***	29.55***	8.14**	3.19 mg								
<b>OCCIPITAL</b>												
CR	33.16***				0.01				1.94			
CU	111.40***	24.62***			0.15	0.66			1.58	0		
AR	61.09***	15.36***	0.00		0.13	0.13	0.82		10.86**	5.49*	4.92*	
AU	72.58***	19.46***	1.19	1.76	4.57*	6.29*	11.06**	7.72**	21.17***	13.00**	13.93***	3.07 mg

*F*-values with (1,29) degrees of freedom for each test.

effect of congruity, with anomalous items more negative than congruent items. Congruity and relatedness did not interact significantly.

### Central ROI

The 5-condition ANOVA at central sites revealed a significant main effect of condition (Most Negative AU > AR > CU > CR > BC Most Positive), with canonical relatedness N400 reduction as well as reductions for congruent compared to anomalous items. The 2 × 2 ANOVA of the unexpected conditions revealed significant main effects of both congruity and relatedness, with anomalous more negative than congruent, and unrelated more negative than related. Congruity and relatedness did not interact significantly.

### Posterior ROI

The 5-condition ANOVA revealed a significant main effect of condition. Pairwise comparisons revealed BC continuations showed significant N400 reduction relative to all other conditions, and CR continuations were reduced relative to the other unexpected continuations. Analyses of the four unexpected conditions revealed significant main effects of both congruity and relatedness, with anomalous more negative than congruent, and unrelated more negative than related. Congruity and relatedness interacted significantly, with a larger relatedness effect (unrelated more negative than related) for congruent than anomalous items.

### Early PNP time window (600–900 ms)

#### Overview 600–900 ms

There were significant main effects of condition over both prefrontal and occipital, but not central, ROIs. At prefrontal sites, CU items were the most positive and

AU items most negative. There were no main effects of relatedness at any of the ROIs between 600 and 900 ms, but significant main effects of congruity at prefrontal and occipital ROIs (but in opposite directions), both of which were mediated by interactions with relatedness.

### Prefrontal ROI

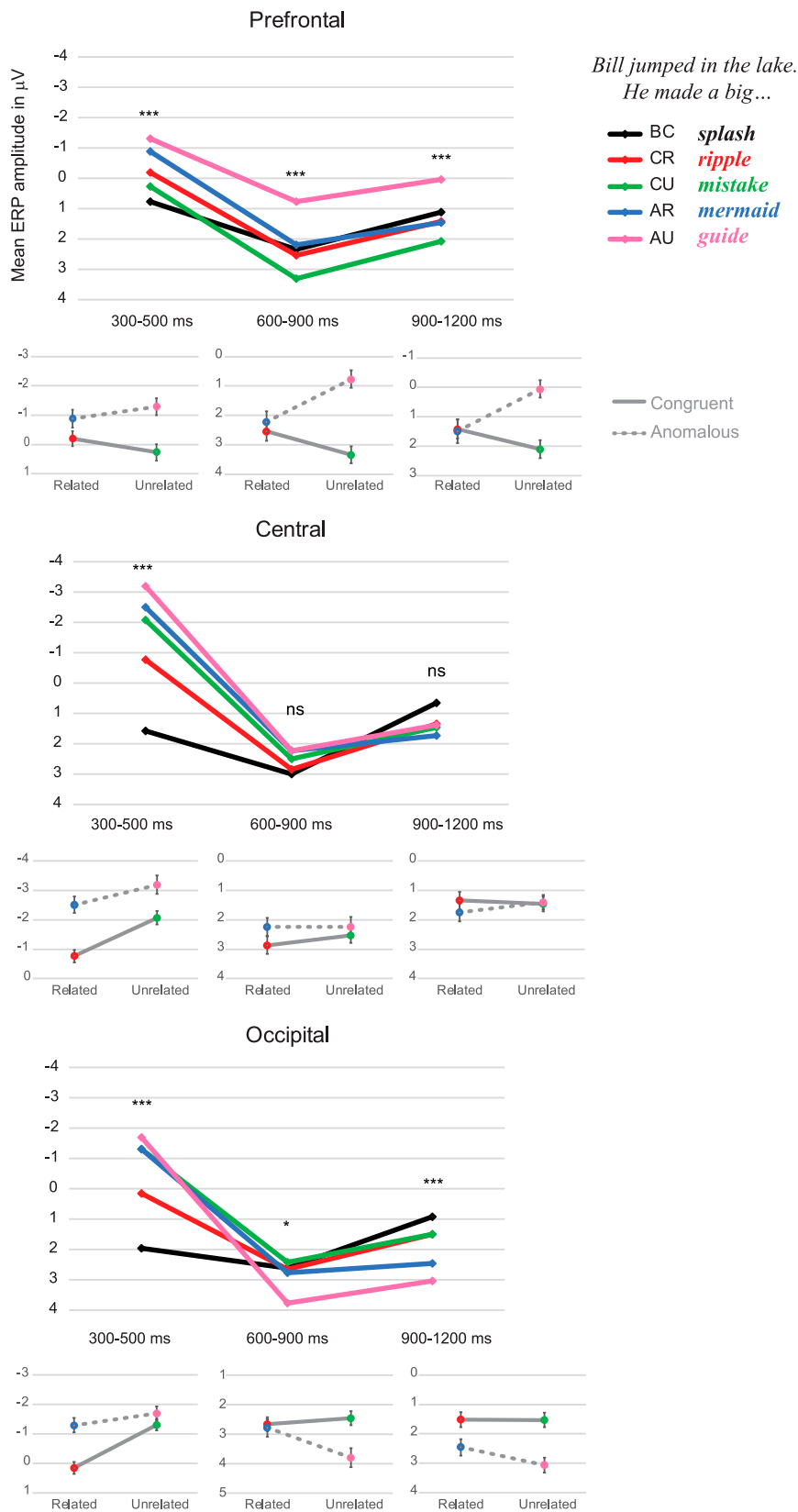
The 5-condition ANOVA over anterior channels revealed a significant main effect of condition. Pairwise comparison indicated that CU was marginally more positive than BC and significantly more positive than the other unexpected conditions, while AU items were significantly more negative than all other conditions. A Congruity (2) × Relatedness (2) ANOVA limited to the four unexpected conditions revealed no main effect of relatedness, but a significant main effect of congruity, with congruent items overall more positive than anomalous. This was mediated by a significant interaction with relatedness, which revealed the congruity effect stemmed from the unrelated items.

### Central ROI

The 5-condition ANOVA indicated no significant effect of condition over central channels, nor were there significant congruity or relatedness effects (or an interaction) revealed by the Congruity (2) × Relatedness (2) ANOVA.

### Posterior ROI

The 5-condition ANOVA over posterior channels revealed a significant main effect of condition, with pairwise testing indicating increased positivity for the AU condition relative to all other conditions. The Congruity (2) × Relatedness (2) ANOVA revealed no significant effect of relatedness, but one for congruity, with anomalous items more positive than congruent. This was,



**Figure 2.** ERP mean amplitudes are plotted in line graph form for the 5 experimental conditions (BC = Best Completion, CR = Congruent Related, CU = Congruent Unrelated, AR = Anomalous Related, AU = Anomalous Unrelated) over 3 time windows (300–500 ms, 600–900 ms and 900–1200 ms) and 3 scalp ROIs (prefrontal, central and occipital scalp sites). Statistical significance is indicated for the 9 corresponding ANOVAs, with full statistical results provided in Table 3. For each ROI, Congruity (2) x Relatedness (2) mean amplitude line graphs are also plotted below for the 3 time windows for analyses limited to the 4 unexpected conditions (CR, CU, AR, AU). ANOVA results for the corresponding main effects and interactions of congruity and relatedness are provided in Table 4. Error bars indicate SEM.

however, mediated by a significant interaction with relatedness, indicating that the congruity effect primarily resulted from differences in unrelated items.

### **Late PNP time window (900–1200 ms)**

#### **Overview 900–1200 ms**

In the late time window, there were very different patterns from the front to back of the head. The anterior pattern was generally a continuation from the 600–900 ms time window, with a prolonged aPNP for the CU condition, and a negativity for the AU condition. In contrast, at occipital sites the anomalous conditions showed the greatest positivity and BC the least, reflecting a reversal in condition mean amplitudes from the earlier 300–500 ms time window over this ROI.

#### **Prefrontal ROI**

The 5-condition ANOVA over anterior channels revealed a significant main effect of condition, with pairwise testing indicating increased positivity for CU and increased negativity for AU relative to BC. The Congruity (2) x Relatedness (2) ANOVA limited to the four unexpected items revealed no main effect of relatedness but one of congruity, showing greater overall positivity for congruent than anomalous items. However, this was mediated by an interaction with relatedness, indicating the effect resulted primarily from the unrelated items.

#### **Central ROI**

The 5-condition ANOVA over central channels revealed no significant main effect of condition. A Congruity (2) x Relatedness (2) ANOVA revealed neither significant main effects of congruity or relatedness, nor an interaction of the two factors.

#### **Posterior ROI**

The 5-condition ANOVA over posterior channels revealed a significant main effect of condition, with pairwise testing indicating that both AU and AR items exhibited increased positivity relative to congruent items and BCs. A Congruity (2) x Relatedness (2) ANOVA revealed a main effect of congruity, with anomalous more positive than congruent. There was no main effect of relatedness, nor did congruity and relatedness interact significantly.

### **Summary of ERP results**

To summarise these findings, canonical cloze-based and relatedness N400 effects were observed between 300 and 500 ms at more posterior sites, with the beginning of a positive shift for CU items at prefrontal sites.

Between 600 and 900 ms, very different ERP patterns were revealed as a function of scalp anteriority, with CU items exhibiting increased prefrontal positivity and AU items exhibiting increased occipital positivity. From 900 to 1200 ms, the prefrontal positivity to CU (and a negativity to AU items) continued from the previous time window, as did the occipital positivity to AU items, which in the late time window was observed for AR items as well.

### **Discussion**

In the current experiment, our first goal was a conceptual replication of the ERP findings observed in DeLong et al. (2014), in which the congruity/plausibility of unexpected word continuations to highly constraining sentences was linked to differential PNP scalp anteriority distributions. The present results are consistent with those patterns as well as distinctive prefrontal and posterior PNP effects noted across the broader literature (see Van Petten & Luka, 2012), by showing that congruent unexpected continuations elicited greater anterior positivities (aPNP), while anomalous words led to greater posterior positivities (pPNP). In addition, our inclusion of a semantic/associative relatedness manipulation allowed us to examine a second line of questioning; namely, whether semantic association with predictable continuations mediates the observed positivities. Our results showed that for unexpected congruent continuations, the unrelated, but not related, words elicited aPNPs, which occurred for an extended duration following the N400 time window. In contrast, both related and unrelated anomalous continuations exhibited posterior PNP, although the response was larger and had an earlier onset for the unrelated items. In the N400 time window, all four unexpected conditions exhibited larger N400s than expected items, displaying canonical N400 amplitude patterns (unexpected > best completions; semantically/associatively unrelated > related), with the data also suggesting that the aPNP to unrelated congruent words may begin its positive shift at frontal sites during this time window. Additionally, there were N400 congruity effects, with anomalous items exhibiting greater N400 amplitude than congruent items. Taken together, these results confirm multiple stages and types of processing for information that continues sentences in unexpected ways, with rapid differentiation based upon how plausible a fit an unexpected word is with the accrued context, as well as the degree to which the unexpected information diverges from or aligns with representations constructed in anticipation of highly predictable, but never received, continuations.

### ***N400 time window: predictability, congruity and relatedness***

The N400 response to continuations of highly constraining sentences followed patterns observed in previous studies. With stimulus materials that included highly predictable best completions and the four unexpected conditions created by crossing factors of congruity with relatedness, results showed that high predictability (indexed by high cloze probability) led to the largest reductions in N400 amplitude, with contributions from semantic/associative relatedness and item congruity. The graded N400 effects were “purest” over central scalp sites where the influence of anterior and posterior PNP amplitudes were presumably minimal, based on the lack of significant PNP results in this scalp region over subsequent time windows. In the prefrontal N400 time window, the ERP to CU words shifted more positive than over more posterior sites, which is notable due to the continued presence of the aPNP in later time windows, suggesting that the aPNP to CU words may initiate fairly early (concurrent with more posterior N400s). This finding is consistent with reports by Thornhill and Van Petten (2012), who also assessed aPNP effects in an early PNP (400–600 ms) time window. Figures 2B and 4 from DeLong et al. (2014), comparing plausible unexpected continuations to best sentence completions, also allude to a similar early onset for the aPNP. The larger implication of an earlier aPNP onset is that it reflects neural processes that can occur in parallel with those reflected by the N400. This is in contrast to the clearly post-N400 timing of the pPNP.

Based on previous findings (e.g. Federmeier & Kutas, 1999; Metusalem et al., 2012; Thornhill & Van Petten, 2012), it is not surprising that both related and unrelated unexpected words exhibited larger N400 amplitudes than best completions, or that unexpected related words showed reduced N400s relative to unrelated ones. On the other hand, prior to testing, it was unclear whether similarly unpredictable congruent and anomalous items would exhibit differential N400s. Results showed that overall, unexpected congruent items elicited greater N400 reductions than anomalous ones. Although the N400’s sensitivity to cloze probability is well established (Kutas & Federmeier, 2011), its relation to plausibility is less clear. Numerous studies have shown that N400 amplitude does not consistently pattern with contextual plausibility (e.g. Federmeier & Kutas, 1999; Fischler et al., 1983; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Urbach & Kutas, 2010), although pragmatic licensing has sometimes been shown to bring N400 amplitudes more in line with plausibility ratings, particularly under conditions of

negation or with quantifier use (e.g. Staab, Urbach, & Kutas, 2008; Urbach, DeLong, & Kutas, 2015). However, in the current study, N400 mean amplitude appeared to closely reflect offline plausibility patterns.<sup>3</sup> Despite this, we maintain that N400 amplitude is not a direct index of item plausibility. Instead, we propose that N400 amplitude is better described as being modulated by the interaction of a variety of factors, of which plausibility is one. Broadly, we propose that plausibility/congruity interacts with both cloze probability/expectancy and contextual/lexical relatedness (semantic or orthographic) – as well as a variety of other lexical factors – to determine an item’s amplitude in the N400 window. In the current study, we explain the high correlation of N400 mean amplitude with plausibility ratings as a byproduct of some inherent biases in the plausibility ratings provided by norming participants. Based on the current study and other anecdotal evidence, we believe that individuals generally rate semantically related items as being more plausible than other types of unexpected continuations, even when they do not fit well with preceding context. This may result either from ratings participants generally not having a clear understanding of what plausibility means, or possibly from the unavoidable tendency of individuals to rate items as being more plausible when they fall within instead of outside of a context’s semantic domain. This may be because in natural language, encountering an unexpected item that is related to the context at hand has more potential to resolve later in the sentence than one that does not. In sum, we suggest that the plausibility ratings for the unexpected words in the current study reflect not just the perceived degree of (in)congruity of the item in context, but also, perhaps unconsciously, the relatedness of the item to either the preceding context or unrepresented best completion.

The N400 effects observed here are consistent with a view that describes the component as reflecting a sort of “snapshot” (Laszlo & Federmeier, 2011) of stimulus driven activity and the neural activation landscape in semantic memory at the time the stimulus is received. In the present study, N400 amplitude reductions manifested as a function of predictability driven by contextual constraint, congruity with the constructed contextual representation, and semantic/associative relatedness to highly expected words. Like N400 results from other related anomaly ERP studies, semantic processing was shown to be eased not only for words that were best fits and congruent words related to best fits, but (with a marginal effect) even for unexpected words that were contextually implausible, but nonetheless shared some semantic/associative relationship with highly expected words, e.g. “Bill jumped in the lake. He made

a big *mermaid* ... " where *splash* is expected. At the same time, unexpected words that were highly plausible but semantically/associatively unrelated also showed reduced N400 amplitude relative to anomalous words, "Bill jumped in the lake. He made a big *mistake/guide* ... ". By the N400 time window, then, there already seems to have been some assessment of an item's contextual plausibility, which is requisite for any interpretation of subsequent scalp divergent PNP patterns that align with this factor.

### Posterior PNP effects to anomalous words

Findings from the current study revealed that anomalous, but not congruent, continuations elicited significant pPNP increases following large N400 amplitudes to those items. Soon after the N400, only anomalous unrelated (AU) words exhibited increased pPNP amplitude, while over a later time window, both anomalous related (AR) and AU items showed this effect. Like in the earlier N400 time window, offline plausibility and pPNP mean amplitude appeared to pattern together.<sup>4</sup> However, also like the N400, we hesitate to interpret this strong correlation as evidence that the pPNP directly indexes item plausibility, because plausibility ratings may reflect an implicit bias to rank semantically related items as more plausible. Perhaps ratings participants cannot ignore an item's relation to a strongly expected word (*mermaid-splash*), or perhaps they can more easily imagine an anomalous related word (*mermaid*) occurring within the same semantic space as the representation constructed from preceding context. Based on the late pPNP patterns in the current study, it appears to take more time for the brain to appreciate the anomalous nature of a word related to the best completion (*mermaid*) than an unrelated continuation (*guide*).

Brouwer and colleagues (Brouwer et al., 2012; Brouwer, Crocker, Venhuizen, & Hoeks, 2017) have proposed that the pPNP (or P600) reflects processing related to integration difficulty/reanalysis of contextual representations that have been constructed to that point. Applying this to the current data would suggest that unexpected congruent continuations (CR and CU conditions, *ripple* and *mistake*, respectively), although unexpected, do not evince the same integration difficulty as anomalous words. The delay in the pPNP to the AR relative to AU condition in the current study also suggests that the brain's parser may hold off on reanalysis/integration attempts when there is some possibility of that word fitting within the context. Given the serial nature of the language input, a word that may seem anomalous upon receipt but which is related to the expected word (*splash*) could potentially make

more (or perfect) sense upon receipt of subsequent context (e.g. "Bill jumped in the lake. He made a big *mermaid* swim away"). That argument seems less likely for AU words (e.g. "Bill jumped in the lake. He made a big *guide* ... ???"). The brain's parser is perhaps sensitive to these probabilities, and for that reason may exhibit variability in the timing of the pPNP response. It is interesting to note that another study similarly utilising a related anomaly paradigm (Metusalem et al., 2012) also describes a pPNP (measured between 500 and 900 ms) with variable onset latencies to event-related versus event-unrelated anomalous sentence continuations. In that study, like ours, the posterior positive shift to related anomalies initiated well after the one to unrelated anomalies.

The eliciting conditions and pPNP responses to semantically anomalous continuations in the current study recall those to some of the garden path conditions that originally propelled theories of the P600 as reflecting sensitivity to syntactic ambiguity (e.g. Osterhout, Holcomb, & Swinney, 1994). The pPNP effect is additionally consistent with a wide variety of ERP sentence studies that have contrasted semantically congruent and incongruent continuations (see Van Petten & Luka, 2012, for a catalogue) and reported more parietal scalp distributions to the violations. The current pPNP also resembles late posterior positivities elicited by "semantic P600" effects to thematic role reversals. In all cases, there is a difficulty in reconciling unexpected input with an already accrued contextual representation, to the extent that the unexpected continuations might be considered (at least temporarily) anomalous – or at least very difficult to salvage in terms of meaning construction for an ongoing sentence. Taken together, all of these results are consistent with proposals linking the pPNP to some kind of reanalysis, with the current study indicating that the parser does not necessarily engage in immediate reanalysis for continuations that are potentially more contextually plausible or interpretable.

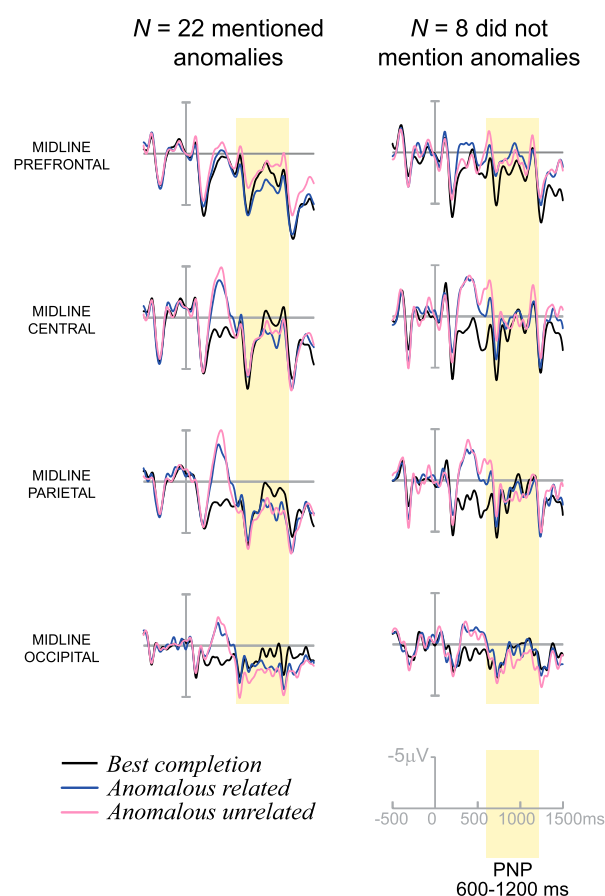
Leckey and Federmeier (2019) have suggested in their examination of late positivities from the language comprehension literature, that there is some renewed support for the proposal that the syntactic P600 and the more domain general P3b are functionally and neurally related, echoing one side of a longstanding debate over the distinctiveness of these two brain responses (see Coulson, King, & Kutas, 1998; versus Osterhout, 1999). They further suggest that the semantic P600 (which, despite its label, is elicited in syntactic manipulations involving semantic anomalies, e.g. to thematic role reversals and/or when there is semantic attraction) may not be a member of this family of effects, with their conclusions based on differential modulations with aging.

Van Petten and Luka (2012), however, suggest that syntactic and semantic P600s, as well as pPNPs to semantically incongruent sentence continuations, may all reflect a similar process; namely, reanalysis or rechecking. These authors propose that in addition to all of these late positivities sharing a surface similarity with the old/new memory ERP effect, they may also be reflecting functionally similar processing. The old/new memory ERP effect manifests as a late positivity over parietal scalp sites when individuals recognise a recently encountered (old) item, compared to falsely or correctly identified new items, as well as unrecognised old items. Similarly, in sentence processing (as in P3b oddball paradigms, Donchin, 1981), encountering an unexpected incongruity may reflect memory related processing associated with context updating. In other words, some reconstruction is involved to sequentially retrieve preceding contextual items, because individual word activation quickly decays in the process of being integrated into some larger meaning/event representation. Incongruent words, then, be they bad fits in terms of meaning, syntax, spelling, would all similarly trigger ordered reactivation of what came before. Following this proposal and based on what is known about the established sensitivities of the P3b, different factors are likely to modulate the speed and/or strength of the retrieval-related response to anomalous sentence continuations, with, for instance, task, attention, stimulus probability, degree of expectancy, or even length or complexity of the preceding context potentially playing a role. All of these are testable proposals.

Indeed, we can specifically question from a functional standpoint whether and how pPNPs like the one observed in the current and similar studies to semantically incongruent information fit into a “family” of posterior positivities; namely, with syntactic anomalies, dispreferred syntactic structures, thematic role reversals, as well as a broader class of late positivities sometimes referred to as late positive components (LPCs), which have been observed to incongruent sentence continuations and also to extra-linguistic stimuli, e.g. to incongruent objects in video clips (Sitnikova, Kuperberg, & Holcomb, 2003); to harmonic violations of musical sequences (Patel, Gibson, Ratner, Besson, & Holcomb, 1998); and even to inaccurate arithmetic solutions (Niedeggen & Rösler, 1999). In determining whether the P3b, P600, LPC and the pPNP indeed reflect some common functional process, one tried method is to explore whether the pPNP shares some of the same known sensitivities of the P3b. For instance, it is well established that the P3b is larger to task relevant stimuli. In the current study, there was no overt task other than reading the sentences for comprehension

and answering occasional content questions. However, it seems possible that participants may have implicitly adopted a strategy about the types of sentences they were encountering, with the most salient items perhaps being those whose continuations were anomalous.

To explore this possibility, we conducted a posthoc analysis, in which the open-ended debriefing questionnaires that participants filled out following the ERP study were consulted. Individuals were sorted based on whether or not they had made note of random, incorrect, unusual, strange, or nonsensical items. Of the 30 participants, 22 commented on the anomalous nature of some stimuli, while 8 did not. Approaching these unbalanced group sizes with caution, we created grand averages to explore the different group responses to the anomalous versus best completions. We were struck by what appears to be a sizeable pPNP effect for the participants who commented on the anomalies versus those who did not (Figure 3). For the anomaly detectors, both AR and AU conditions exhibited increased positivity relative to



**Figure 3.** Midline grand-averaged ERPs for participant groups determined through debriefing questionnaire responses. A collapsed PNP time window (600–1200 ms) is highlighted in yellow, indicating that individuals who made note of anomalies showed increased late positivity to anomalous conditions relative to best completions.

best completions over more posterior scalp locations. In contrast, those individuals not remarking on anomalies showed little pPNP activity for either AR or AU items. Perhaps, then, the pPNPs to anomalous words are more akin to brain responses when words are perceived as “targets”. These patterns, while by no means conclusive, offer some hint that task relevance (even when it is implicit) may play a role in modulating the pPNP, similar to the way it does for the more domain general P3b response.

The above proposal clearly warrants further, more principled investigation. Ways to test whether the pPNP response to the anomalous words has to do with (unconscious) “target detection” might involve manipulating the proportion of anomalous sentences to determine if the amplitude of the pPNP response increases when the proportion of anomalous sentences gets smaller, similar to the behaviour of the P3b. Task relevancy for the anomalous items could also be increased by asking participants to make plausibility judgments on the items, because we know from the literature that P600 amplitude increases when items become more task relevant (Coulson et al., 1998). In conjunction with a plausibility judgment task, it could also be determined whether or not the peak of the pPNP time locks to overt plausibility responses, and additionally whether there is variability in the ERP based on how easily or quickly an item can be classified as anomalous (potentially modulated by, e.g. the relatedness of the anomalous continuation to the expected word or the local or global nature of the incongruity).

### **aPNP**

Results from the current study indicate that only one condition elicited an increased aPNP relative to best completions: congruent unrelated (CU) words. In contrast, neither congruent words related to best completions (CR), nor anomalous words (AR, AU), elicited aPNP responses. Returning to the aPNP checklist presented in the Introduction, we can assess how well our results fit with findings from other ERP sentence comprehension studies (Brothers, Swaab, & Traxler, 2017; DeLong et al., 2014; Federmeier et al., 2007; Kuperberg & Wlotko, 2018; Moreno et al., 2002; Ness & Meltzer-Asscher, 2018; Quante et al., 2018; Thornhill & Van Petten, 2012). First, like previous findings, the current aPNP was elicited by congruent/plausible, low predictability words. In addition, the effect here, like in other studies, was observed in highly constraining contexts. Also, as in the handful of other studies directly contrasting congruent and anomalous unexpected continuations, the aPNP was *not* observed to anomalous

continuations, which instead elicit pPNPs. Finally, the aPNP in the current study initiated quite early but extended well past the N400 (which is consistent with some other reports of the effect beginning during the N400 time window, Thornhill & Van Petten, 2012; DeLong et al., 2014). At the broadest level then, the aPNP to CU words in the current study is consistent with interpretations of additional processing being required for congruent/plausible unexpected words that violate strong sentential expectancies.

On the other hand, there are points of divergence between the current aPNP findings and others. The primary one being that the congruent semantically/associatively related (CR) words here did *not* elicit aPNPs. For instance, Moreno et al. (2002) found that congruent unexpected semantically related words, as well as direct English-Spanish code switches (words with identical meanings but in a different language), elicited aPNPs in both normal sentences and idioms, e.g. “A dog is a man’s best *friend/buddy/amigo*”. Thornhill and Van Petten (2012), too, found similar aPNPs to related and unrelated unexpected plausible sentence endings relative to best completions. Given these results, researchers had concluded that the effect might index the violation of a lexical prediction, reflecting a brain response to *any* lexical form except the highly anticipated one. However, the absence of an aPNP to CR words in the current study would seem to argue against this proposal, as would the lack of an aPNP to anomalous continuations across the broader literature. The current results place yet more constraints on the potential factors that give rise to aPNP responses for some, but not all, congruent unexpected sentence continuations. A closer look at some of the inter-experimental differences may offer some clues about the conditions under which the effect may be observed as well as the functional nature of the component.

Two factors stand out as possibly contributing to different aPNP patterns across studies. The first is that our design combined testing of both unexpected congruent and anomalous sentence continuations, which affected the overall stimulus proportions and probabilities of encountering (different varieties of) congruent sentences, as well as highly predictable ones. Work by Lau and colleagues (e.g. Delaney-Busch, Morgan, Lau, & Kuperberg, 2017; Lau, Holcomb, & Kuperberg, 2013) and others (Brown, Hagoort, & Chwilla, 2000) testing experimental proportions of semantically related word prime pairs, as well as work by Fine, Jaeger, Farmer, and Qian (2013) examining experimental environments with high proportions of dispreferred syntactic parses in garden path sentences, show that broader experimental contexts can impact processing of more and less predictable information, as reflected in N400s in the first



case and reading times in the second. Perhaps even more relevant for the current study, in sentence contexts, Brothers et al. (2017) showed that with increased prediction validity – manipulated by increasing the proportion of highly constraining filler sentences continued by high relative to low cloze endings – reading times for predictive sentence contexts with predictive endings were reduced. In an experiment like ours, then, with relatively low prediction validity (80% of sentences continue unpredictably and of those, half were anomalous) comprehenders may have adapted to prediction ultimately not being a very effective strategy within the context of the experiment.<sup>5</sup> The strength of the predictions for highly probable critical words, and in turn the consequences to not receiving them, may have differed from those in Thornhill and Van Petten (2012) and Moreno et al. (2002), where all conditions were plausible and where higher percentages of items were related to predictable continuations overall.

Another factor that could have contributed to different aPNP patterns across studies is that in the current experiment, critical words were sentence medial, unlike in Thornhill and Van Petten (2012) and Moreno et al. (2002), where critical words were sentence final and where unexpected plausible related (CR) words elicited aPNPs. A potential explanation is that the sentence final status of the critical words in those studies – signalled by a punctuating period – could have forced comprehenders to engage in a similar type of processing as for the sentence medial CU continuations in the current study. In other words, perhaps both the period for the sentence final critical words (signalling that no further information would be received and that integration must be completed) and a sentence medial congruent but unrelated word (“Bill jumped in the lake. He made a big *mistake* ...”) both force an immediate shift from the representation that had been built assuming the predictable continuation. Going a step further, an aPNP to our congruent related continuations (*ripple*) may *not* have been observed because of those words’ sentence medial positions, along with their contextual acceptability and similarity to highly expected continuations. Their relatedness may have caused comprehenders to not fully abandon (suppress) the representation including the expected word (*splash*) because subsequent context could potentially resolve how the related word fits into the existing contextual representation, or better yet, the predictable word could still appear (e.g. *Bill ... made a big ripple ... after his splash*). In contrast, a congruent unrelated continuation like *mistake* requires the comprehender to modify the representation more significantly and immediately, from one that aligns with

a predictable cause–effect event structure of a disturbance to the surface water, to one in which additional information must be recruited. For instance, when *mistake* is encountered, perhaps the reader infers that the water was ice cold. Or that Bill misjudged the water’s depth. Or worse, that he drowned. Note that *ripple* may also require a modification to the contextual representation, albeit more minor: perhaps a minimisation of the force of Bill’s jump, or simply shifting the time focus to a later phase of the same event (a *ripple* being the consequence of the splash). Post-critical word context has the potential to offer this kind of clarification or elaboration to comprehenders. Broadly then, the aPNP may be elicited in situations where it is warranted for the comprehender to abandon or overwrite a strong expectation for the best completion because either 1) no further information will be received (signalled by punctuation) or 2) when the input causes a more dramatic, immediate shift away from the original strongly preactivated contextual representation. This proposal offers a very testable hypothesis, by way of presenting the stimuli used in the current experiment in sentence final positions with punctuating periods, to determine whether congruent related words would elicit aPNPs.

To date, the main proposal for the neural processing indexed by the aPNP has been that the component reflects suppression/inhibition of highly predictable information when an unexpected but plausible alternative continuation is encountered – processing that is not required for anomalous continuations. Ness and Meltzer-Asscher (2018) suggest that a lack of suppression when anomalous words are encountered indicates that “inhibition of the wrongly predicted word is necessary not to facilitate retrieval of the unexpected word, but rather for its integration with the existing representation, because retrieval had to have happened prior to identification of a word as anomalous.” (See Kuperberg & Wlotko, 2018, for similar reasoning.) The current study suggests there may be yet another piece to this positivity puzzle. While the evidence across studies is consistent with the aPNP not simply indexing suppression of just any unexpected word (because otherwise it would be elicited by anomalous continuations, too), our results indicate that the effect also does not seem to be a response to just any surprising but plausible continuation, either. Instead, the aPNP may be mitigated by whether a plausible alternative immediately forces inhibition of a highly predictable continuation so that a plausible alternative can be integrated: this could happen either by virtue of the unexpected word forcing a substantial shift in sentence meaning or potentially by punctuation signalling no further clarifying information.

### **Distinguishing anterior and posterior PNPs**

The current experiment was designed to go beyond examining neural responses to more and less plausible sentence continuations. The relatedness manipulation was included to specifically test for differences in postretrieval processing of information more and less consistent with predictable words. Our analyses revealed that the aPNP is indeed sensitive to the plausibility of unpredictable continuations, with its elicitation potentially signalling points at which the processor is forced to inhibit predicted word information when new information can and must be integrated. In contrast, implausible continuations do not seem to require inhibition of predicted information, but instead may lead to retrieval or reactivation of preceding context, perhaps to assess reparability. In light of some of the pPNP's similarities with other posterior late positivities, this effect may turn out to be sensitive to a variety of factors, including task, stimulus probability, potentially even experimental order effects, if the processor adjusts its probability models throughout the course of the experiment.

The current findings indicate that a word's contextual plausibility is assessed quickly after its appearance. Although the categorical distinction drawn in this experiment between congruent and anomalous continuations was not intended to result in a range of plausibility ratings, it unintentionally did. We argued that this spread in ratings may be a byproduct of some unwitting bias for rating related items as more plausible or perhaps raters not having a clear understanding of what plausibility means. Another possibility is that we take these ratings at face value. If we do that, we see that the crossover point from a more anterior to posterior PNP occurs at the point where congruent items are distinguished from anomalous ones, between the CU and AR conditions (mean plausibility ratings 4.13 and 2.67 out of 7, respectively, on either side of the midway point, 4, on the plausibility scale). It is not, however, clear what it means for an item to be "medium plausibility": whether it means that some individuals rate an item as highly plausible, others as implausible, with an average that is not representative of the extremes; or that raters take those items to be possible but unlikely, which is not the same as being implausible; or that an item does not exactly fit into a context but might if subsequent information were to offer some clarification; or that because the experimental sentences were so constraining, any item that is not the expected one automatically gets rated as less plausible. So although the calculated offline plausibility ratings for the five experimental conditions appear to represent a continuum, the PNP responses tell a different story, with scalp distributions

seeming to reflect two categorically different functional processes in the post-N400 time window.

There are numerous examples from the ERP sentence processing literature where plausibility is a dimension that does not map neatly onto more or less congruent stimuli, e.g. when world and semantic knowledge conflict (e.g. "Dutch trains are *yellow/white*", Hagoort, Hald, Bastiaansen, & Petersson, 2004); when fanciful discourses override semantic knowledge (e.g. when "the peanut was *in love/salted*", Nieuwland & Van Berkum, 2006); or when global and local context constrain for different sentence continuations ("Frank was throwing a birthday party, and he had made the dessert from scratch. After everyone sang, he sliced up some *sweet/healthy cake/veggies*", Boudewyn, Long, & Swaab, 2015). As we attempt to refine our understanding of PNP brainwave patterns and their functional significances, it will be useful to examine contrasts like these, where stored knowledge and representations being constructed on the fly come into conflict, and determine if and which PNP effects manifest. More broadly, the current research should prompt deeper contemplation of extralinguistic factors that potentially shape comprehension of less predictable input, as well as what it means to be plausible, possible or predictable and where these concepts diverge. These considerations should extend not only to how individuals process unexpected linguistic information within specific experimental environments, but also beyond.

### **Notes**

1. From this point forward, for sake of brevity, we will refer to the more posterior positivity pattern as a *pPNP* and the anterior one as an *aPNP*.
2. Only 2 of these 61 contexts had unexpected continuations with non-zero cloze probabilities (.03 and .06 cloze for two CR items). For the remaining 59 contexts, none of the unexpected words ever occurred as norming responses. Cloze probabilities for unexpected continuations across the entire stimulus set were thus presumed to be similarly low.
3. N400 amplitude and plausibility were highly correlated. For N400 condition mean amplitudes over central electrode sites correlated with mean plausibility ratings for the five conditions, Pearson's  $r = .9247$ ,  $p = .0244$ .
4. A calculation of the correlation between offline plausibility and pPNP mean amplitude between 900 and 1200 ms revealed a strong inverse relationship between the two measures (Pearson's  $r = -0.9473$ ,  $p = .0145$  for pPNP condition mean amplitudes over occipital sites correlated with mean condition plausibility ratings for the five conditions).
5. To investigate potential order effects that might be indicative of desensitization to contextual predictability, an exploratory analysis was conducted and is included as supplemental material.

## Acknowledgments

The authors wish to thank Janaina Weissheimer for assistance with ERP and behavioural data collection. This work was supported by the NICHD under Grant R01HD22614 to Marta Kutas.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) under Grant R01HD22614 to Marta Kutas.

## References

- Amsel, B. D., DeLong, K. A., & Kutas, M. (2015). Close, but no garlic: Perceptuomotor and event knowledge activation during language comprehension. *Journal of Memory and Language*, 82, 118–132. doi:10.1016/j.jml.2015.03.009
- Benton, A. L., & Hamsher, K. (1978). *Multilingual aphasia examination manual*. Iowa City: University of Iowa.
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, 59(1), 55–73. doi:10.1016/j.brainresrev.2008.05.003
- Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 15(3), 607–624. doi:10.3758/s13415-015-0340-0
- Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, 136, 135–149. doi:10.1016/j.cognition.2014.10.017
- Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, 93, 203–216. doi:10.1016/j.jml.2016.10.002
- Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, 41, 1318–1352. doi:10.1111/cogs.12461
- Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446, 127–143. doi:10.1016/j.brainres.2012.01.055
- Brown, C. M., Hagoort, P., & Chwilla, D. J. (2000). An event-related brain potential analysis of visual word priming effects. *Brain and Language*, 72, 158–190. doi:10.1006/brln.1999.2284
- Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13(1), 21–58. doi:10.1080/016909698386582
- Delaney-Busch, N., Morgan, E., Lau, E. F., & Kuperberg, G. R. (2017). Comprehenders Rationally Adapt Semantic Predictions to the Statistics of the Local Environment: a Bayesian Model of Trial-by-Trial N400 Amplitudes. In *CogSci*. DeLong, K. A., Chan, W. H., & Kutas, M. (2019). Similar time courses for word form and meaning preactivation during sentence comprehension. *Psychophysiology*, 56(4), e13312.
- DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, 121(3), 226–239. doi:10.1016/j.bandl.2012.02.006
- DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, 61, 150–162. doi:10.1016/j.neuropsychologia.2014.06.016
- DeLong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, 48(9), 1203–1207. doi:10.1111/j.1469-8986.2011.01199.x
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117. doi:10.1038/nn1504
- Dikker, S., Rabagliati, H., Farmer, T. A., & Pykkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science*, 21(5), 629–634. doi:10.1177/0956797610367751
- Donchin, E. (1981). Surprise! ... surprise? *Psychophysiology*, 18(5), 493–513. doi:10.1111/j.1469-8986.1981.tb01815.x
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41(4), 469–495. doi:10.1006/jmla.1999.2660
- 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. doi:10.1016/j.brainres.2006.06.101
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS one*, 8(10), e77661. doi:10.1371/journal.pone.0077661
- Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry Jr, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, 20(4), 400–409. doi:10.1111/j.1469-8986.1983.tb00920.x
- Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, 25(2), 149–188. doi:10.1080/01690960902965951
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *science*, 304(5669), 438–441. doi:10.1126/science.1095455
- Hoeks, J. C., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19(1), 59–73. doi:10.1016/j.cogbrainres.2003.10.022
- Juottonen, K., Revonsuo, A., & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research*, 4(2), 99–107. doi:10.1016/0926-6410(96)00022-5
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52(2), 205–225. doi:10.1016/j.jml.2004.10.002

- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23–49. doi:10.1016/j.brainres.2006.12.063
- Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17(1), 117–129. doi:10.1016/S0926-6410(03)00086-7
- Kuperberg, G., & Wlotko, E. (2018). A tale of two positivities (and the N400): Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *bioRxiv*, 404780. doi:10.1101/404780
- Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, 8(4), 533–572. doi:10.1080/01690969308407587
- 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. doi:10.1146/annurev.psych.093008.131123
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203–205. doi:10.1126/science.7350657
- Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 217–237). Hillsdale, NJ: Lawrence Erlbaum.
- Kwon, N., Sturt, P., & Liu, P. (2017). Predicting semantic features in Chinese: Evidence from ERPs. *Cognition*, 166, 433–446. doi:10.1016/j.cognition.2017.06.010
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211. doi:10.1037/0033-295X.104.2.211
- 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
- Laszlo, S., & Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology*, 48(2), 176–186. doi:10.1111/j.1469-8986.2010.01058.x
- Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, 25(3), 484–502. doi:10.1162/jocn\_a\_00328
- Leckey, M., & Federmeier, K. D. (2019). The P3b and P600 (s): Positive contributions to language comprehension. *Psychophysiology*, e13351. doi:10.1111/psyp.13351
- Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, 88, 22–60. doi:10.1016/j.cogpsych.2016.06.002
- Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, 69(4), 574–588. doi:10.1016/j.jml.2013.08.001
- McCallum, W. C., Farmer, S. F., & Pockock, P. V. (1984). The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 59(6), 477–488. doi:10.1016/0168-5597(84)90006-6
- Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, 66(4), 545–567. doi:10.1016/j.jml.2012.01.001
- Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching palabras (words): An electrophysiological study of code switching. *Brain and Language*, 80(2), 188–207. doi:10.1006/brln.2001.2588
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. <http://www.usf.edu/FreeAssociation/>. doi:10.3758/bf03195588
- Ness, T., & Meltzer-Asscher, A. (2018). Lexical inhibition due to failed prediction: Behavioral evidence and ERP correlates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(8), 1269–1285. doi:10.1037/xlm0000525
- Niedeggen, M., & Rösler, F. (1999). N400 effects reflect activation spread during retrieval of arithmetic facts. *Psychological Science*, 10(3), 271–276. doi:10.1111/1467-9280.00149
- Nieuwland, M. S., & Van Berkum, J. J. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24(3), 691–701. doi:10.1016/j.cogbrainres.2005.04.003
- Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18(7), 1098–1111. doi:10.1162/jocn.2006.18.7.1098
- Osterhout, L. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes*, 14(1), 1–14. doi:10.1080/016909699386356
- Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 786. doi:10.1037/0278-7393.20.4.786
- Otten, M., & Van Berkum, J. J. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, 45(6), 464–496. doi:10.1080/01638530802356463
- Paczynski, M., & Kuperberg, G. R. (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb-argument processing. *Language and Cognitive Processes*, 26(9), 1402–1456. doi:10.1080/01690965.2011.580143
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, 10(6), 717–733. doi:10.1162/089892998563121
- Quante, L., Bölte, J., & Zwitserlood, P. (2018). Dissociating predictability, plausibility and possibility of sentence continuations in reading: Evidence from late-positivity ERPs. *PeerJ*, 6, e5717. doi:10.7717/peerj.5717
- Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, 51(3), 437–447. doi:10.1016/j.neuropsychologia.2012.12.002

- Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology*, 40(1), 160–164. doi:10.1111/1469-8986.00016
- Staab, J., Urbach, T. P., & Kutas, M. (2008). Negation processing in context Is Not (always) delayed. *Center for Research in Language Technical Report*, 20(3), 1–34.
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 402–433. doi:10.2307/747605
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643. doi:10.1037/h0054651
- Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, 68(4), 297–314. doi:10.1016/j.jml.2012.12.002
- Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, 83(3), 382–392. doi:10.1016/j.ijpsycho.2011.12.007
- Urbach, T. P., DeLong, K. A., & Kutas, M. (2015). Quantifiers are incrementally interpreted in context, more than less. *Journal of Memory and Language*, 83, 79–96. doi:10.1016/j.jml.2015.03.010
- Urbach, T. P., & Kutas, M. (2010). Quantifiers more or less quantify on-line: ERP evidence for partial incremental interpretation. *Journal of Memory and Language*, 63(2), 158–179. doi:10.1016/j.jml.2010.03.008
- Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443. doi:10.1037/0278-7393.31.3.443
- Van De Meerendonk, N., Kolk, H. H., Vissers, C. T. W., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, 22(1), 67–82. doi:10.1162/jocn.2008.21170
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2), 176–190. doi:10.1016/j.ijpsycho.2011.09.015
- Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, 16(7), 1272–1288. doi:10.1162/0898929041920487