Neuropsychologia 61 (2014) 150–162

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Predictability, plausibility, and two late ERP positivities during written sentence comprehension

Katherine A. DeLong [a,*], Laura Quante [b], Marta Kutas [a,c,d,e]

[a] University of California, San Diego (UCSD) Department of Cognitive Science, USA
[b] University of Münster, Germany
[c] UCSD Center for Research in Language, USA
[d] UCSD Department of Neurosciences, USA
[e] UCSD Kavli Institute for Brain and Mind, USA

ARTICLE INFO

Article history:
Received 2 May 2013
Received in revised form
11 June 2014
Accepted 12 June 2014
Available online 20 June 2014

Keywords:
Language
Sentence comprehension
Prediction
Plausibility
ERP
P600
N400

ABSTRACT

Van Petten and Luka's (2012, *International Journal of Psychophysiology*, 83(2), 176–190) literature survey of late positive ERP components elicited by more or less predictable words during sentence processing led them to propose two topographically and functionally distinct positivities: a parietal one associated with semantically incongruent words related to semantic reanalysis and a frontal one with unknown significance associated with congruent but lexically unpredicted words. With the goal of testing this hypothesis within a single set of experimental materials and participants, we report results from two ERP studies: Experiment 1, a post-hoc analysis of a dataset that varied on dimensions of both cloze probability (predictability) and plausibility, and Experiment 2, a follow-up study in which these factors were manipulated in a controlled fashion. In both studies, we observed distinct post-N400 positivities: a more anterior one to plausible, but not anomalous, low cloze probability sentence medial words, and a more posterior one to semantically anomalous sentence continuations. Taken together with an observed canonical cloze-modulated N400, these dual positivities indicate a dissociation between brain processes relating to written words' sentential predictability versus plausibility, clearly an important distinction for any viable neural or psycholinguistic model of written sentence processing.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

In everyday language, we often hear or read sentences that continue in semantically unexpected ways. In the laboratory, electrophysiological brain responses can indicate whether such continuations, for instance, constitute novel but sensible continuations on the part of the comprehender (e.g., '*They let the canoe into the water and paddled with Frisbees…*', Chwilla, Kolk, & Vissers, 2007), clash with personal value systems ('*I think euthanasia is… acceptable…*'processed by a strict Christian, van Berkum, Holleman, Nieuwland, Otten, & Murre, 2009), or are interpreted as jokes (e.g., '*I let my accountant do my taxes because it saves time: last spring it saved me 10 years.*', Coulson & Kutas, 2001). Although from a comprehender's perspective these words may not be likely sentence continuations, they are certainly plausible. Sentence studies manipulating semantic expectancy, however, have more frequently utilized incongruent (anomalous) completions than employing only plausible ones (see Van Petten & Luka, 2012). Although two words might share the same near-zero cloze probability rating (a common proxy measure for online predictability), they could fundamentally differ in their contextual plausibility: e.g., '*He pounded the nails with a book/summer.*' Semantically anomalous continuations like *summer* have been a mainstay of psycholinguistic event related brain potential (ERP) research for decades (e.g., Kutas & Hillyard, 1980), with low contextual predictability and anomaly often conflated in studies that have focused on amplitude modulations of the N400—an ERP component related to ease of semantic access. Less frequently, however, studies have made use of more plausible continuations like those mentioned above. And rarely, it seems, have brain responses to plausible and anomalous low cloze probability continuations been directly contrasted within a single study to assess the contributions of these two factors to online sentence comprehension.

Two late ERP positivities may prove useful for addressing this issue. Based on a survey of the ERP sentence processing literature, Van Petten and colleagues (e.g., Thornhill & Van Petten, 2012; Van Petten & Luka, 2012) have hypothesized that there is a late frontal

* Correspondence to: Department of Cognitive Science, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0515, USA. Tel.: +1 858 534 2440; fax: +1 858 534 1128.
E-mail addresses: kadelong@cogsci.ucsd.edu (K.A. DeLong), l.quante@uni-muenster.de (L. Quante), kutas@cogsci.ucsd.edu (M. Kutas).

http://dx.doi.org/10.1016/j.neuropsychologia.2014.06.016
0028-3932/© 2014 Elsevier Ltd. All rights reserved.

positive ERP component that is dissociable from an established posterior/parietal late positive component (LPC, sometimes referred to as the P600 or semantic P600). Namely, they suggest that the parietal post-N400 positivity (PNP) may be linked to reanalysis or repair following impaired interpretation due to syntactic or semantic incongruency. Although for many years the P600 was thought to be an ERP response exclusive to syntactic violations and ambiguities (dating back to Osterhout & Holcomb, 1992), a wave of studies in the past decade has been influential in revising this interpretation, with findings of P600s to anomalies that are more semantic in nature (for instance, to thematic role violations, animacy violations, and so-called semantic illusions, e.g., Hoeks, Stowe & Doedens, 2004; Kim & Osterhout, 2005; and Nieuwland & van Berkum, 2005, respectively; see Kuperberg, 2007 for a review).

In contrast, the more anterior late positivity (sometimes noted beginning between 400 and 600 ms) may relate to violations of lexical predictions involving semantically congruent (plausible) substitutions, this being a consistent factor across a majority of the limited number of studies in which it has been observed. Although many of the reports of this frontal positivity have been quite recent (see below for some representative studies), it has been an incidental finding in the sentence comprehension literature dating back nearly 20 years. Kutas (1993) noted a larger left frontal post-N400 positivity (500–900 ms) to congruent low cloze relative to high cloze probability endings in highly constraining ($\geq$75%) sentence frames, suggesting at the time that the ERP component might index inhibition of predicted words. Coulson and Van Petten (2007), using visual hemifield presentation, also noted a 600–900 ms left hemisphere-biased late frontal positivity elicited by congruent low cloze, relative to high cloze probability, continuations. Moreno, Federmeier, and Kutas (2002) observed in Spanish–English bilinguals that relative to expected sentence completions, both lexical switches (English synonyms of expected endings) and code switches (English-to-Spanish translations of expected endings) elicited late frontal positivities (650–850 ms), especially in highly constraining idioms, but also in constraining non-idiomatic sentences. Federmeier, Wlotko, De Ochoa-Dewald, and Kutas (2007), as well, observed an increased late frontal positivity (500–900 ms) to congruent low cloze continuations of high but not low constraint sentences. In our own work, (DeLong, Urbach, Groppe, & Kutas, 2011; DeLong, Groppe, Urbach, & Kutas, 2012) we have similarly observed frontal positivities temporally overlapping and continuing beyond more succinct posterior N400 effects to congruent low relative to high cloze probability continuations in predictive sentences, both in younger as well as older (age 60+years) adults.

A precise functional correlate of this frontal positivity has yet to be isolated, but a number of the aforementioned studies have proposed that this brain response may reflect a consequence to neurally pre-activating, but not receiving, highly expected (i.e., high cloze probability) continuations. For instance, Federmeier et al. (2007) linked the effect to a cost for the prediction "mismatch", and DeLong et al. (2011, 2012) suggest it may be a "misprediction" response.

However, elicitation of the proposed frontal positivity seems to require *both*: 1) a constraining context, operationalized in terms of either highly convergent cross-participant offline cloze probability norming responses, or possibly through somewhat more divergent responses (e.g., 30% or greater agreement) but supplied with relatively short response times by individual participants in speeded cloze norming tasks[1], *and* 2) a low cloze probability but semantically plausible sentence continuation. A strong test of this hypothesis, and an aim of the current study, is to contrast the ERPs to (1) relatively expected (high cloze) plausible sentence continuations with those to (2) unexpected (low cloze) *plausible* and (3) unexpected (low cloze) *anomalous* continuations in constraining sentence contexts. If, within a single group of readers, there turn out to be ERP differences between the effects of (1) vs. (2) compared to (1) vs. (3), then this would indicate a clear dissociation between the brain's processing of predictability and plausibility, and would prove valuable for assessing the independence of the frontal and posterior positivities and their putative sensitivities to plausible versus implausible unexpected continuations, respectively.

To our knowledge no single study has analyzed late positivities to these specific conditions in predictive contexts. Although Kutas, Lindamood, and Hillyard (1984) included these experimental conditions, they only reported N400 findings. Geyer, Holcomb, Kuperberg, and Pearlmutter (2006) modulated plausibility of sentence continuations, comparing plausible, implausible and anomalous words, but noted only a posterior P600 to anomalous words but no frontal positivities; however, it is unclear the extent to which their contexts were contextually constraining, what the critical word cloze probabilities were, what the proportion of "strange" versus "normal" sentences was, and the role that a plausibility judgment task may have played in eliciting late positivities they observed. Similarly, Van de Meerendonk, Kolk, Vissers, and Chwilla (2010) examined sentences with plausible, mildly implausible and strongly implausible continuations, again observing only a posterior P600 to the strongly implausible words, but controlled for neither sentential constraint nor critical word cloze probability, thus making it difficult to assess whether or not our proposed criteria for eliciting the frontal positivity were met. If the frontal positivity is a response to only semantically *plausible* (i.e., sensical) low cloze probability continuations of constraining contexts, and the posterior PNP is modulated by an item's *implausibility*, this would ultimately limit the candidate functional processes of these ERPs and would indicate that different mechanisms come into play during sentence comprehension for items varying along these dimensions.

In an attempt to dissociate these two effects within a single group of participants and within a single stimulus set, we report results from two separate studies. In Experiment 1 we investigate the sensitivities of the frontal and posterior positivities in an ERP study that afforded a post-hoc analysis of sentence stimuli sorted on their offline plausibilities. Experiment 2 has the same experimental goals, but investigates these questions in an experiment specifically designed to test for a positivity dissociation through manipulations of cloze probability and plausibility.

2. Experiment 1

DeLong, Urbach, and Kutas (2005) argued for sentential-based prediction based on cloze probability-graded modulations of N400 mean amplitude to more and less expected prenominal indefinite articles (*a*/*an*). Post-hoc analyses of those data (DeLong et al., 2011) revealed a prolonged, late frontal positivity to unexpected nouns (e.g., to *lisp* in '*It was difficult to understand the visiting professor because he spoke with a lisp*…' for which '*an accent*' is the most expected continuation). Initially, we hypothesized that this positivity might reflect the violation of a contextual expectancy for a part of speech, namely, adjectives (e.g., *thick*, *heavy*, or *Russian*) that might have been neurally triggered by the unexpected article (e.g., *a*), but which never appeared. Such a strategy could presumably allow the brain's parser to salvage the most contextually expected noun (e.g., *accent*), especially when it is difficult to activate an alternative online. However, once an unexpected noun (e.g., *lisp*) appears instead of an anticipated adjective, additional

[1] See Thornhill and Van Petten (2012) for a discussion of the relevance of "weak" and "divergent" expectations for determining levels of sentence constraint.

processing might be required for integration and resolution of the so-called "syntactic surprise".

To test this idea, we conducted an ERP study (DeLong, Urbach, & Kutas, 2007) that modulated contextual expectancy for adjectives by means of the indefinite articles, asking participants to read sentence contexts that led to expectations for vowel or consonant-initial nouns appropriately preceded by either *a* or *an*. Sentence stems had four possible continuation types: (A) high cloze article+high cloze noun (e.g., '*an accent*'), (B) low cloze article+low cloze noun (e.g., '*a lisp*'), (C) high cloze article+low cloze noun (e.g., '*an apron*'), or (D) low cloze article+congruent adjective+high cloze noun (e.g., '*a thick accent*'), which was included as a filler condition to ensure naturalistic use of adjectives within the experiment. Our logic was that if the frontal positivity was due to an unexpected article-triggered expectation for an adjective instead of a noun, then for nouns like (C) *apron* (preceded by prediction-consistent articles) we would expect to observe only an N400, but not the positivity; however, for article/noun pairs like (B) *a lisp* that warrant a "switch", we would expect to observe both effects. Our results, however, did not support this prediction, with *both* unexpected noun types (B and C) eliciting statistically similar frontal positivities beginning within the N400 time window (in addition to canonical posterior N400s as well as what appeared to be another more distributionally widespread positivity later in the epoch), thus arguing against our syntactic expectancy violation hypothesis.

Upon further scrutiny, we noticed that some of the Unexpected nouns (Conditions B and C) in that study were not just low cloze probability, but anomalous. Given Van Petten and Luka's (2012) recent suggestion that congruent improbable continuations generally elicit more frontal positivities and incongruent ones more posterior positivities, we decided to utilize the variability of our low cloze probability words and sort and analyze the ERP responses to plausible and anomalous unexpected nouns. Our first step was to collect offline plausibility ratings for original Conditions A, B and C and categorize the stimuli according to their contextual plausibility as well as cloze probability. As in our original design, we treated the Condition D adjectives as a filler condition, thus not including them in the current analysis.

2.1. Materials and methods

2.1.1. Stimuli

Participants read generally constraining sentence pairs (mean constraint=85.3%, SD=14.4%)[2] that led to expectations for particular sentence-medial words. In the original DeLong et al. (2007) design, stimuli consisted of 160 contexts continued by the four conditions outlined above (A–D), yielding 640 different stimulus items. For the current study, we utilized these same items, keeping two of the original conditions intact; namely, the 160 high cloze noun items (Condition A) and the 160 adjective items (Condition D), which we continued to treat as a filler condition. The remaining 320 low cloze (Unexpected) nouns (Conditions B and C), however, were re-sorted based on their offline plausibility ratings.

2.1.1.1. Plausibility ratings. Six independent raters scored all 480 context+critical noun experimental items offline for plausibility on a scale of 1 (highly implausible) to 5 (very plausible). Based on the average inter-rater scores (mean plausibility rating across all 3 experimental conditions=3.00, standard deviation=1.64), the high cloze noun condition (A) was judged to have relatively higher plausibility (see Table 1). We classified the remaining 320 Unexpected items as either (B) Somewhat Plausible (USP, plausibility rating > 1.5, yielding 170 items) or (C) Anomalous (ANOM, plausibility rating ≤ 1.5, yielding 150 items). Plausibility ratings for the three experimental conditions differed significantly from each other (plausibility means and standard deviations shown in Table 1).

After resorting, the following stimulus factors turned out to be matched across the new conditions: (1) contextual constraint, (2) written word frequency of critical words (Kucera & Francis, 1967), (3) critical word length, and (4) orthographic neighborhood size for critical words (see Table 1). Because our plausibility rating reclassification of the 320 Unexpected stimuli led to non-identical contexts across conditions, it is conceivable that pre-critical noun ERP differences —in particular at the articles immediately preceding the critical nouns—could potentially propagate into critical noun epochs. An evaluation of the prenominal articles, however, indicated overall similar article cloze probabilities for the USP and ANOM conditions, as well as similar proportions of *a*/*an* article types across all three experimental conditions. Our stimulus reclassification also results in non-identical stimuli in the *post*-critical noun region, which could potentially lead to the commingling of variable ERP responses (e.g., N400s) to the different words immediately following critical nouns, with ERP responses in the critical nouns' late positive time window (which we plan to analyze). However, examination of several relevant lexical factors (word class, word length, written frequency, and orthographic neighborhood) of the post-critical noun words revealed no differences between conditions (see Table 2).

2.1.1.2. Cloze probability norming. Stimulus norming for critical word cloze probability was conducted in a separate, off-line sentence completion task by 30 University of California, San Diego student volunteers, compensated with experimental credit or cash. Contexts were truncated prior to the critical words (following procedures detailed in DeLong et al., 2005). Cloze probability ratings for the three experimental conditions differed significantly from each other (see Table 1 for means and standard deviations).

2.1.1.3. Lists. Table 1 shows representative stimuli. Each ERP participant viewed one of four 160-item lists, with contexts and critical words used once per list. Each list consisted of 40 Expected noun items (Condition A), 80 Unexpected noun items (Conditions B and C), and 40 Adjective+Expected noun filler items. The Unexpected items (Conditions B and C) that any individual subject read were comprised of 39% to 56% ANOM trials (variable across lists) with a complementary percentage of USP trials (44% to 61%). Mean cloze probability and plausibility did not differ significantly between lists.

Yes/no comprehension questions, an example of which is shown in Table 1, followed one quarter of sentences at random intervals.

2.1.2. ERP Participants

Thirty-two UCSD volunteers (23 women, 9 men) participated for course credit or cash. Participants were right-handed, native English speakers with normal or corrected-to-normal vision, ranging from 18 to 23 years old (mean, 19.4 years). Ten participants reported a left-handed parent or sibling. Three additional participants were excluded from the analysis due to excessive eye blink or movement artifacts.

2.1.3. Procedure

ERPs were recorded in a single session in a sound-attenuating, electrically shielded chamber. Participants sat one meter 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 empty fixation frame (jittered between 800 and 1300 ms), with individual words presented centrally in the frame (200 ms duration/500 ms SOA). From 1.5 to 2.5 s of empty frame followed each RSVP sentence, after which a comprehension question appeared if there was one. Either the participant's question response button-press served to advance to the next sentence, or advancement was automatic, with a 2.5-s interval between sentences.

2.1.4. Electroencephalographic recording parameters

The electroencephalogram (EEG) was recorded from 26 electrodes arranged geodesically in an Electro-cap (see depictions in Figs. 1 and 2A), 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 KΩ. The EEG was amplified with Grass amplifiers with a pass band of 0.01 to 100 Hz and was continuously digitized at a sampling rate of 250 samples/second.

2.1.5. Data analysis

Trials contaminated by eye movements, excessive muscle activity, or amplifier blocking were rejected off-line before averaging. These trials (on average, 8%) were excluded from further analysis, with the exception of trials containing eye blinks for four participants: these excessive blink data were corrected using a spatial filter

[2] Contextual constraint was operationalized as the cloze probability of the most frequent response when the sentence contexts truncated prior to the critical articles were normed.

Table 1
Experiment 1 sample stimuli and characteristics.

Condition	Condition label	Example continuation	Number of items	Mean critical noun cloze probability (SD), Range: 0-1	Mean context+noun plausibility rating (SD), Range: 1-5	Mean contextual constraint (SD), Range: 0-1	Mean critical noun KF written frequency (SD)	Mean critical noun word length (SD)	Mean critical noun orthographic neighborhood size (SD)	Yes/no comprehension question
Example stimulus: *It was difficult to understand the visiting professor. Like many foreigners he spoke with a/an…when conversing in English.*										
High cloze/High plausibility	**EXPected**	*…accent…*	160	.88 (.13)	4.9 (.09)	.85 (.14)	55.2 (85.6)	6.7 (2.2)	2.59 (4.16)	*Was the professor hard to understand?*
Low cloze/ Somewhat plausible	**Unexpected Somewhat Plausible (USP)**	*…lisp…*	170	.03 (.08)	2.8 (.86)	.86 (.13)	60.1 (82.5)	6.8 (2.1)	2.49 (3.97)	
Low cloze/ Low plausibility	**ANOMalous**	*…apron…*	150	< .01 (< .01)	1.2 (.18)	.84 (.16)	47.8 (87.9)	6.6 (2.3)	2.69 (4.36)	

algorithm (Dale, 1994). A third order Butterworth band-pass filter set from 0.2 to 15 Hz was used on all data to reduce high frequency noise. 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 noun and adjective onsets. ERPs were computed for epochs extending from 500 ms pre- to 1540 ms post-stimulus onset, using a pre-stimulus baseline of 500 ms.

To detect reliable differences between neural processing of the Expected (EXP), Unexpected Somewhat Plausible (USP), and Anomalous (ANOM) nouns in the current study, ERPs from the three pairwise comparisons of these conditions were submitted to repeated measures *t*-tests at all sampled time points between 0 and 1200 ms (301 total time points) and at all 26 scalp electrodes (i.e., 7826 total comparisons for each conditional comparison: USP minus EXP, ANOM minus EXP, ANOM minus USP). To protect against a large proportion of false discoveries due to the large number of hypothesis tests, the false discovery rate (FDR) control procedure described in Benjamini and Hochberg (1995) was used to determine which *t*-tests were significant using an FDR level of 0.05. This "mass univariate" procedure is only guaranteed to control the FDR when tests are independent or follow positive regression dependency. These assumptions might not hold for ERP data since the individual tests may be negatively correlated. However, it has been shown analytically that for approximately normally distributed data, like ERPs, FDR control behaves as if each univariate test is independent as the number of tests increases (Clarke & Hall, 2009). Moreover, studies of simulated ERPs (Lage-Castellanos, Martínez-Montes, Hernández-Cabrera, & Galán, 2010; Groppe, Urbach, & Kutas, 2011) suggest that this FDR procedure provides accurate control of the FDR rate for ERPs even when a large proportion of tests are negatively correlated. This procedure was used in lieu of more conventional mean amplitude ANOVAs because it provides much better spatial and temporal resolution than conventional ANOVAs while maintaining meaningful suppression of spurious test results. In particular, with the current study examining multiple ERP effects, time windows and scalp regions of interest, we surmised that this method of analysis could offer the best means for examining different brainwave patterns that might overlap temporally and spatially. Based on this procedure, critical *t*-scores of $+/-2.44$, $+/-2.49$, and $+/-2.75$ (each with $df=31$) were derived, respectively, for the three USP-EXP, ANOM-EXP and ANOM-USP comparisons. In other words, any differences in the original data that exceeded these critical *t*-scores were deemed reliable.[3]

As aids to visualizing effects of critical noun cloze probability and plausibility on ERP mean amplitude, we provide the following plots: (1) Fig. 1: grand average ERPs to EXP, USP and ANOM nouns over all 26 scalp electrodes and (2) Fig. 2: raster plots representing the 7826 univariate *t*-tests in two dimensional grids for all three conditional comparisons.

2.2. Behavioral results

Comprehension accuracy was calculated for the yes/no probe questions. Participants correctly answered an average of 96.1% (median=97.4%, range=85–100%) of the questions, indicating they were attending to and comprehending the experimental stimuli during the recording session.

2.3. ERP results

2.3.1. Predictions

If there are indeed two distinct late positivities whose elicitation to low cloze sentence continuations is mediated by contextual (im)plausibility, then we expect to observe increased ERP frontal positivity to the Unexpected Somewhat Plausible (or USP) nouns relative to the Expected (or EXP) nouns, but not to the Anomalous (ANOM) relative to EXP nouns. Conversely, based on the plausibility ratings for the conditions in the current study (Table 1) and previous research indicating that—at least under certain circumstances—the posterior P600 may be sensitive to implausibility (e.g., Geyer et al., 2006; van de Meerendonk et al., 2010), ANOM nouns would be expected to elicit a posterior post-N400 positivity relative to EXP nouns. N400 amplitude of all conditions is

[3] We replicated the results of our 0 to 1200 ms FDR analyses using a more limited set of time points (500–1200 ms). We took this step because strong, broadly distributed effects like the N400 make FDR correction very lenient outside that effect, where findings may be less certain, because the method can allow 5% false positives. Thus, replicating the positivity findings from 500 to 1200 ms ensured that our results were not simply artifacts of including the N400 effect in the longer analysis window.

Table 2
Characteristics of pre- and post-critical noun position words in Experiment 1.

Condition label	Pre-critical noun words (indefinite articles)			Post-critical noun words			
	Article *a* / *an* proportions	Article cloze probability proportions	Mean article cloze probability (SD)	Word class proportions	Mean word length (SD)	Mean KF written frequency (SD)	Mean orthographic neighborhood (SD)
EXP	50% *a* / 50% *an*	100% high cloze	.83 (.16)	89% closed / 11% open class	3.5 (1.7)	9,655 (10, 635)	5.1 (3.0)
USP	54% *a* / 46% *an*	47% high cloze / 53% low cloze	.42 (.41)	89% closed / 11% open class	3.5 (1.8)	10,489 (11,790)	5.0 (3.0)
ANOM	46% *a* / 54% *an*	53% high cloze / 47% low cloze	.47 (.41)	88% closed / 12% open class	3.5 (1.7)	8,712 (9,056)	5.1 (3.0)

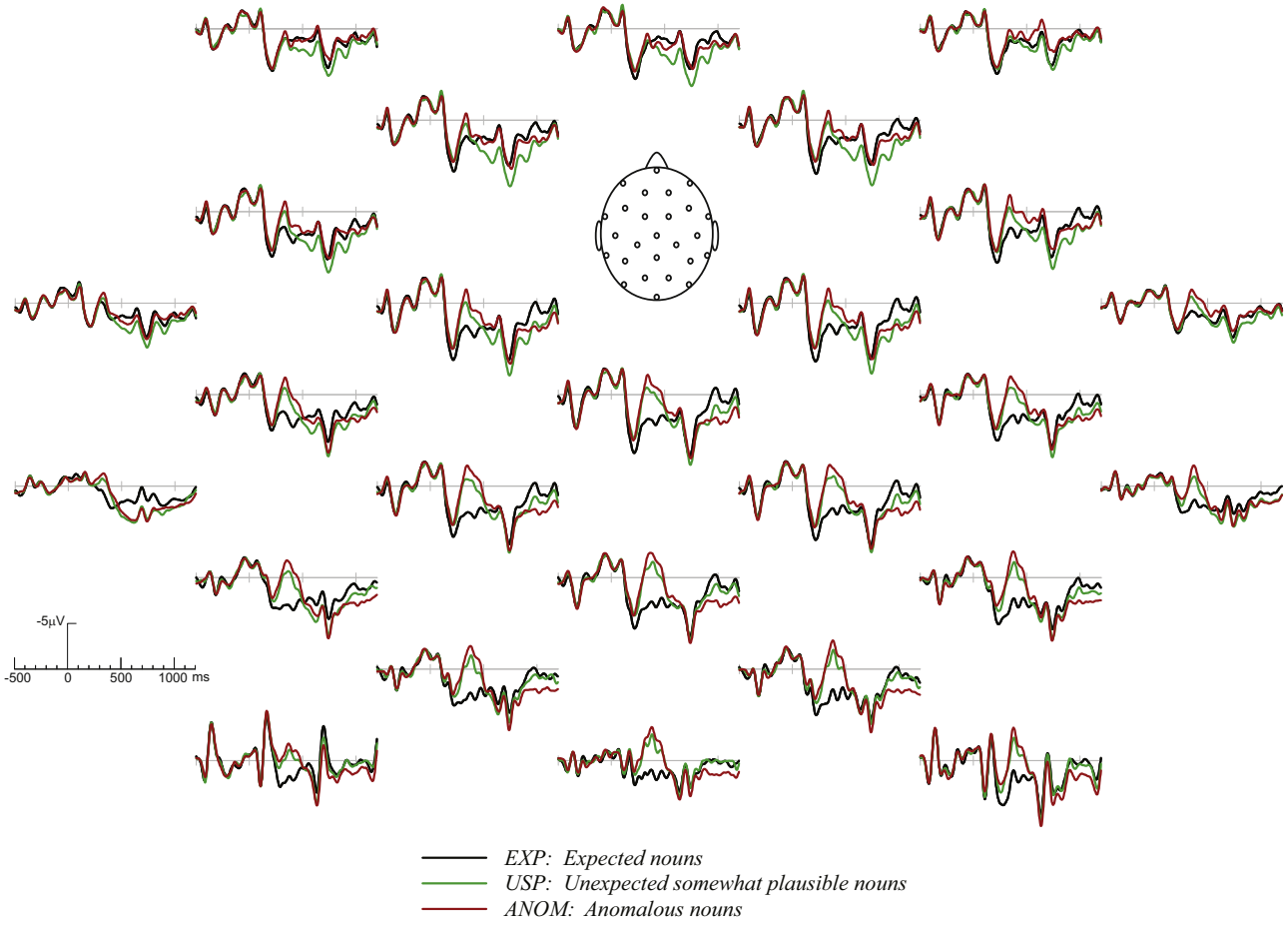

Fig. 1. Experiment 1 grand average (N=32) ERPs recorded over 26 scalp channels.

expected to be modulated by cloze probability in the typical manner, i.e., exhibiting an inverse relationship.

2.3.2. USP minus EXP nouns

We begin by analyzing the ERP differences between USP and EXP nouns. Fig. 1 plots the grand average ERPs to the experimental conditions over all 26 electrodes, with visual inspection hinting that the predicted patterns may obtain. Fig. 2B validates the statistical significance of these findings in demonstrating that by 300 ms a widespread N400 effect (USP more negative than EXP) is well under way and continues until approximately 500 ms, particularly over right posterior scalp sites where N400 effects are known to be largest. Nearly coincident with the N400's offset, a relative positivity of USP to EXP nouns also begins to emerge around 450 ms at frontal scalp locations, as well as at left lateral temporoparietal sites. This positivity continues, albeit less continuously than the N400, until approximately 1050 ms, with a distribution that is best characterized as showing the most consistent effects at medial prefrontal and left temporoparietal channels. Posterior midline and right medial scalp locations show little evidence of the effect. For all time point-by-time point tests (Fig. 2B), significant corrected p-values are between .000001 $\leq p_{adj} \leq$ .05. In sum, the two main effects for this analysis map onto a canonical N400 effect and a later anterior, slightly left-biased positivity of more extended duration.

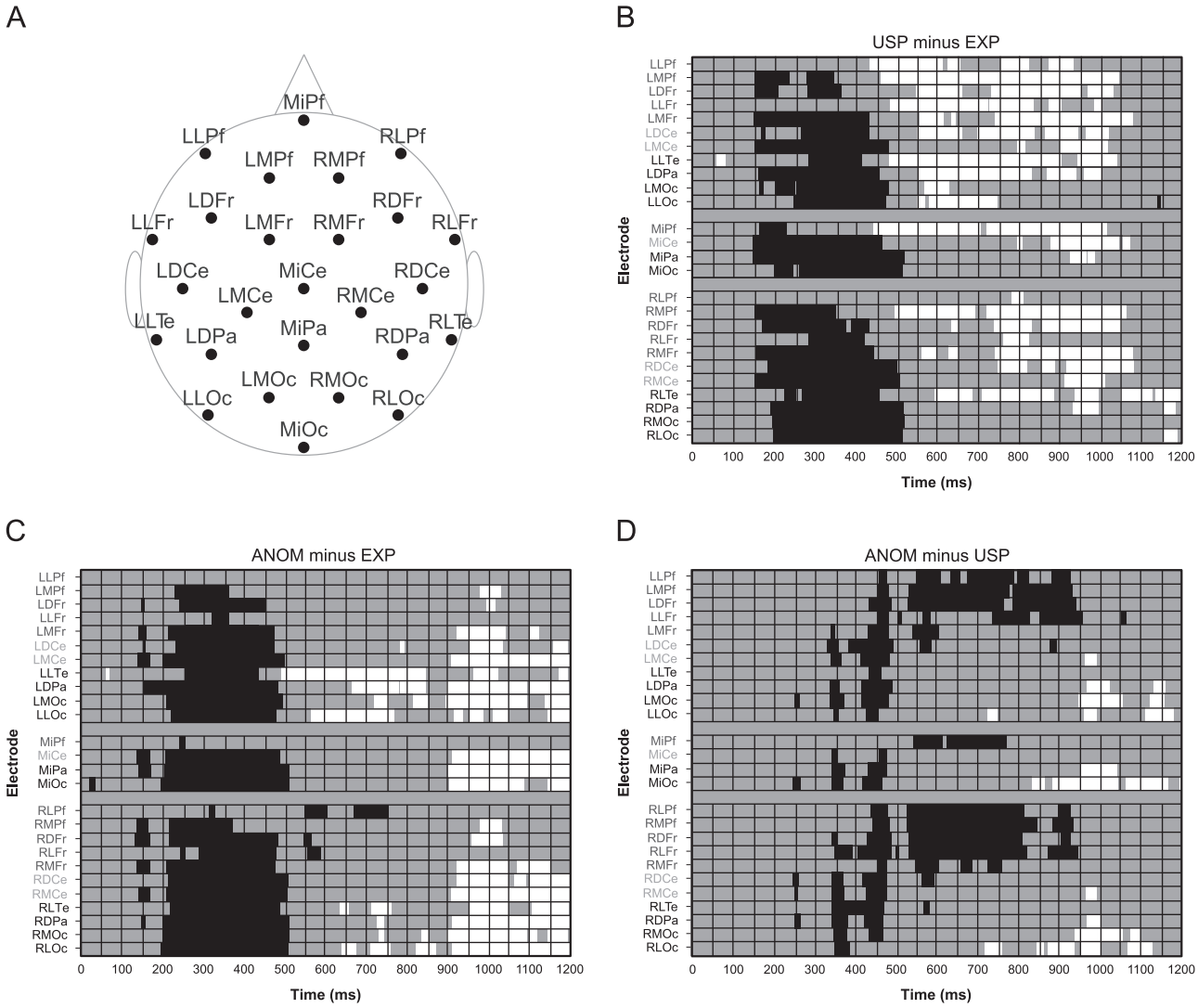

Fig. 2. (A). Schematic diagram showing the array of 26 labeled scalp electrodes from which ERPs were recorded. Panels (B–D): Raster plots of Experiment 1 False Discovery Rate (FDR)-controlled *t*-tests in two dimensional grids from mass univariate analyses of all three conditional comparisons of ERP data. Results are plotted in 4 ms increments for the following: (B) Unexpected Somewhat Plausible (USP) minus expected (EXP) continuations, (C) Anomalous (ANOM) minus EXP continuations, and (D) ANOM minus USP continuations. Left scalp electrodes are depicted uppermost, midline scalp electrodes in the center, and right scalp electrodes in the lower portions of each panel. Significant *t*-tests for negative ERP differences are represented in black and those for positive differences in white.

2.3.3. *ANOM minus EXP nouns*

Our next mass univariate analysis compares the ANOM versus EXP nouns (Fig. 2C). As expected and similar to the USP-EXP differences, we observe a widespread N400 effect (ANOM more negative than EXP nouns) that is significant well before 300 ms, continuing through approximately 500 ms, at all but the most prefrontal electrodes. Unlike the USP-EXP comparison, there is no evidence for significant ANOM-EXP differences at frontal scalp locations coincident with the offset of the N400. However, from 500 through approximately 850 ms a relatively isolated increased positivity of ANOM to EXP nouns over posterior left lateral channels is evident, similar in timing and distribution to the pattern observed for USP-EXP nouns. Following an approximately 50 ms period of no significant differences between conditions, at around 900 ms there is a sustained relative positivity of ANOM to EXP nouns across the posterior half of the scalp, which continues through 1200 ms. This effect aligns with our predicted ERP pattern for this analysis, with a perhaps later onset than has been observed in previous studies. For all time point-by-time point tests (Fig. 2C), significant corrected *p*-values are between .000001 $\leq p_{adj} \leq$ .05. In sum, these plots indicate both a strong, canonical N400 effect, as well as a central-posterior positivity later in the ERP epoch. Unclear is whether or not the left lateral temporoparietal positivity beginning around 500 ms is a distinct ERP effect or the onset of the more widespread posterior positivity observed later in the epoch.

2.3.4. *ANOM minus USP nouns*

Finally, we analyze ERP differences created by subtracting USP nouns from ANOM nouns (Fig. 2D). This comparison, in particular, provides a direct means of investigating how the ERP responses to the two different unexpected continuation types differ from each other. Fig. 2D indicates that within a typical N400 time window (300–500 ms), ERPs to the ANOM words are significantly more negative than those to USP words, though it is possible that this difference may be attributable to the slight difference in these two conditions' cloze probability, with which N400 amplitude is known to negatively correlate (see Table 1). Of greater interest is what happens in the post-N400 time region. Again, in line with our predicted outcomes, Fig. 2D indicates that from approximately 550 ms through 950 ms, there is a clear pattern of USP nouns showing consistently greater positivity than ANOM nouns at frontal

scalp locations over both cerebral hemispheres. Regarding the later posterior positivity, from approximately 900 ms onward there is a less continuous and more focal increased positivity to ANOM relative to USP nouns, limited to occipital and parietal channels. For all time point-by-time point tests (Fig. 2D), significant corrected p-values are between $.001 \leq p_{adj} \leq .05$. In sum, in a typical N400 time window the ANOM nouns generally show slightly greater negativity than the USP nouns, broadly over the scalp. More strikingly, there appear to be two distinctive patterns of late positivities: from after 500 through approximately 950 ms the USP nouns show a larger relative positivity at anterior scalp sites, but beginning after 700 ms continuing through 1200 ms, ANOM nouns show relatively greater positivity at the most posterior channels.

2.4. Discussion

In Experiment 1, we exploited an existing data set to probe a hypothesized dissociation between two late ERP positivities to less expected sentence continuations based on their relative contextual (im)plausibilities. We examined highly constraining sentence pairs with ranges of both critical noun expectancy (indexed by offline cloze probability) and sentential plausibility (determined through offline plausibility judgments). Our analyses suggested that in addition to the typical pattern of larger N400 amplitudes to lower cloze probability words, unexpected critical words indeed appeared to elicit dual late ERP positivities modulated by different linguistic properties, with distinct scalp topographies and somewhat different latencies. For sentence continuations rated anomalous, we observed an ERP positivity that was maximal at posterior sites more focally from around 500 ms and with a wider distribution from around 900 continuing through 1200 ms, relative to more expected words. In contrast, unexpected plausible sentence continuations elicited a more anterior positivity relative to expected continuations, beginning consistently around 500 ms and continuing through approximately 1050 ms. When the two unexpected continuation types were directly contrasted with each other, these distinctions still held, with the relative positivity of plausible unexpected continuations to anomalous ones being earlier, more anterior, longer in duration, more continuous, and more widely distributed than the relatively later positivity of anomalous words to unexpected plausible ones at posterior scalp channels.

In sum, the results from Experiment 1 offer some promise for differentiating brain responses to more and less plausible unexpected sentence continuations. However, the realignment of the original stimuli to create new conditions for our post-hoc analyses, along with the potential for the expected word class of the target items (noun vs. adjective due to the more and less expected indefinite articles) to have contributed to our findings, means that our stimuli were not necessarily optimal to ensure that the observed effects reflect sensitivity to the plausibility of unexpected items and not some other factor.

With this in mind, we conducted a second experiment in an attempt to replicate the abovementioned findings. Experiment 2 was designed explicitly to examine the positivities' sensitivity to the (im)plausibility of unexpected sentence continuations. In Experiment 2, the variability in article expectancy from Experiment 1 was eliminated, as was the Adjective filler condition. We utilized sentence contexts from Experiment 1, but ensured they were consistent across all three experimental conditions by using only high cloze prenominal articles. This led to some of the critical unexpected nouns being replaced with alternatives, but created a balanced design with equal numbers of items per condition. Consistent across both experiments, however, was our examination of conditions differing in noun cloze probability and plausibility, contrasting high cloze noun continuations with low cloze probability ones that could be either contextually plausible or anomalous. The aim of Experiment 2 thus was to replicate, again within a single data set and group of participants, the data patterns observed in Experiment 1, in which plausible unexpected noun continuations elicited a relatively anterior post-N400 positivity, and anomalous continuations elicited a more posterior one.

3. Experiment 2

3.1. Materials and methods

For Experiment 2, we selected 150 of the 160 sentence pair contexts (mean constraint = 87.8%, SD = 10.6%) used in Experiment 1. For each experimental context there were 3 possible sentence medial continuations. All Expected continuations (highest cloze continuations for individual contexts) were the same as those used in Experiment 1. Unexpected Somewhat Plausible (USP) continuations or Anomalous (ANOM) continuations could also continue each sentence pair, for a total of 450 experimental stimuli. Seventy-six of the USP and 74 of the ANOM continuations from Experiment 1 were used, with the remaining continuations (74 USP and 76 ANOM) replaced with alternative nouns for Experiment 2. Experimenter judgment was used to assess the plausibility of the new materials. For an individual context, critical nouns for all three conditions were preceded by the same, congruent indefinite article. Cloze probabilities were determined from the Experiment 1 norming. See Table 3 for means and standard deviations of lexical factors for each condition. Only on the factor of cloze probability did the conditions show statistically significant differences.

3.1.1. Lists

Three stimulus lists with 150 experimental items each (50 items from each of the 3 conditions), were used. Fifty relatively high constraint filler sentence pairs with high cloze continuations (mean constraint/cloze = 73.9%, SD = 14.1%) were also included in each list to hold the proportion of sentences including high cloze nouns constant with that of Experiment 1. This resulted in a total of 200 items per list. All lexical factors reported (refer to Table 3) were matched between lists. Yes/no comprehension questions followed one quarter of all sentences at random intervals.

3.1.2. ERP Participants

Twenty-four UCSD volunteers (12 women, 12 men) participated for course credit or cash. Participants were right-handed, native English speakers with normal

Table 3
Experiment 2 sample stimuli and characteristics.

Experimental Condition	Example continuation	Example stimulus: *For the snowman's eyes the kids used two pieces of coal. For his nose they used a…from the fridge.*				
		Lexical characteristics of critical nouns, condition means and (standard deviations)				Yes/no comprehension question
		Cloze probability	KF written frequency	Word length	Orthographic neighborhood size	
Expected (EXP)	*…carrot…*	.90 (.09)	55.83 (75.65)	6.31 (2.21)	2.91 (4.82)	*Did the children use a tomato for the snowman's nose?*
Unexpected Somewhat Plausible (USP)	*…banana…*	< .01 (.01)	58.06 (80.31)	6.48 (2.10)	2.85 (4.65)	
Anomalous (ANOM)	*…groan…*	< .01 (< .01)	41.68 (71.76)	6.27 (1.97)	2.56 (4.30)	

or corrected-to-normal vision, ranging from 18–29 years old (mean, 20.1 years). Five participants reported a left-handed parent or sibling.

3.1.3. Procedure

ERPs were recorded in a single session in a sound-attenuating, electrically shielded chamber. Participants sat one meter 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. Words in RSVP sentences were presented centrally for a duration of 200 ms (500 ms SOA). If there was a comprehension question, it appeared following the RSVP sentence. Either the participant's question response button-press served to advance to the next sentence, or advancement was automatic, with a 2-s interval between sentences.

3.1.4. Electroencephalographic recording parameters

These were the same as for Experiment 1.

3.1.5. Data analysis

On average, 6.6% of experimental trials contaminated by eye movements, excessive muscle activity, or amplifier blocking were rejected off-line before averaging and excluded from further analysis. 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 noun onsets. ERPs were computed for epochs extending from 500 ms pre- to 1540 ms post-stimulus onset, using a pre-stimulus baseline of 500 ms. Fig. 3 plots ERP waveforms for the 3 experimental conditions over all 26 scalp channels.

For Experiment 1 we utilized mass univariate tests to explore multiple ERP patterns in a post-hoc analysis. Those findings suggested scalp regions and temporal windows where the anterior and posterior post-N400 positivities may be likely to be observed. For Experiment 2 we began by topographically mapping ERP mean amplitude condition differences (USP minus EXP, ANOM minus EXP, and ANOM minus USP) in 100 ms increments over all scalp channels (see Fig. 4) to determine if these patterns approximated those observed in Experiment 1. Visual inspection indicated they did, albeit with slightly different effect latencies for the positivities. Specifically, the following patterns were observable: (1) widespread but posteriorly prominent N400 effects between 300 and 500 ms, (2) an anterior and left-biased post-N400 positivity to USP relative to EXP words which became prominent (>1 μV) around 600 ms—slightly later than in Experiment 1—increasing through approximately 1000 ms, and (3) a clearly posterior post-N400 positivity to ANOM relative to EXP words emerging around the same latency as the anterior positivity (beginning around 600–700 ms), also increasing through around 1000 ms but sustained through 1200 ms.

With the experimental goal of testing for ERP positivity effects under a more controlled experimental design, in Experiment 2 we turned to more traditional mean amplitude analyses, conducting ANOVAs with 3 levels of noun type (EXP, USP, and ANOM). For the N400, we used a canonical time window (300–500 ms) over all 26 scalp channels. For the less established post-N400 anterior and posterior positivity effects, findings from Experiment 1 provided a starting point for determining spatial and temporal regions to analyze. With little previous research available to help determine precise distributions and latencies of post-N400 positivities (particularly the anterior positivity), we analyzed a temporal window apparently common to both patterns (600–1000 ms), but over different scalp areas. For the frontal positivity, we tested for a left-biased frontal effect over 8 anterior, left-midline electrodes (see Fig. 4 for scalp locations), where anterior effects were also prominent in Experiment 1. To investigate the posterior positivity, we analyzed ERPs at 11 posterior electrode sites (non-overlapping with those used for anterior positivity analysis, see Fig. 4), where effects were observed in Experiment 1 as well as more widely in the literature.

All ANOVA p-values reported herein are after epsilon correction (Huynh–Feldt) for repeated measures with more than one degree of freedom. For all pairwise comparisons computed between conditions, statistical significance levels were adjusted according to the Bonferroni procedure.

3.2. Behavioral results

Comprehension accuracy was calculated for the yes/no probe questions. Participants correctly answered an average of 93.9% (median 93.9%, range=87.8–100%) of the questions, indicating they were attending to and comprehending the experimental stimuli during the recording session.

3.3. ERP results

Visual inspection of ERP data in Figs. 3 and 4 reveals patterns similar to those observed in Experiment 1; namely, increased N400s for both unexpected conditions relative to the Expected nouns, an increased post-N400 positivity for the USP condition over frontal scalp channels, and an increased post-N400 positivity for the ANOM condition over posterior channels (see Fig. 5 for enlarged representative frontal and posterior channels from both experiments). As an aid to visualizing the analyzed ERP differences between the three noun types, Fig. 4 shows ERP mean amplitudes in bar graph form of grand average waveforms for each condition for the three analyses for which ANOVAs were conducted.

3.3.1. 300–500 ms

N400 effects appeared to be widespread across the scalp but largest over central posterior channels, as is typical of the N400. An ANOVA with 3 levels of word type and 26 levels of electrode location indeed revealed a main effect of noun type [$F(2,46)=42.05$, $p<.0001$, $\varepsilon_{HF}=0.79$], with ANOM nouns showing the greatest negativity (-1.06 μV), followed by an intermediate amplitude N400 for USP nouns (0.38 μV), with EXP nouns showing the most reduced N400s (2.67 μV). A full set of planned pairwise comparisons indicated that ERPs for the three conditions during the N400 time window all differed significantly from each other (see Fig. 4). These results are roughly consistent with the N400 patterns observed in Experiment 1 and are representative of more general N400 findings.

3.3.2. 600–1000 ms anterior scalp sites

For a time window in which the frontal positivity to USP words was prevalent in Experiment 1, over 8 left-midline anterior scalp electrode locations, we conducted an ANOVA with 3 levels of word type (EXP, USP, ANOM), with our primary interest being in the pairwise comparisons of the 3 conditions. The ANOVA revealed a main effect of word type [$F(2,46)=5.85$, $p=.0054$, $\varepsilon_{HF}=1.00$], with USP nouns (2.57 μV) more positive than both EXP nouns (1.82 μV) and ANOM nouns (1.23 μV). The pairwise comparisons indicated significant mean amplitude differences between USP and both EXP and ANOM words, but not between ANOM and EXP nouns (Fig. 4). This ERP pattern is consistent with the results of Experiment 1 in that the measured ERP positivity was largest to USP nouns relative to the other conditions over frontal left-midline scalp locations.

3.3.3. 600–1000 ms posterior scalp sites

Over 11 posterior electrode sites (non-overlapping with sites used for the anterior 600–1000 ms analysis), an ANOVA with 3 levels of noun type revealed statistically significant mean amplitude differences [$F(2,46)=11.33$, $p=<.001$, $\varepsilon_{HF}=1.00$], with ANOM nouns showing the greatest positivity (3.79 μV), followed by EXP noun (2.66 μV) and then USP nouns (2.19 μV). Planned pairwise comparisons indicated that the ANOM nouns exhibited statistically greater positivity than both EXP and USP nouns in this analysis, with mean amplitudes of USP and EXP nouns not differing significantly from each other (see Fig. 4). Again, this mimics the pattern of ANOM nouns in Experiment 1, in that this same condition showed the largest positivity over posterior scalp locations.

3.4. Discussion

In Experiment 2 we set out to replicate the main findings from a post-hoc analysis of an ERP study we had conducted in which we observed dissociable frontal and posterior late positivities to unexpected plausible and anomalous sentence continuations, respectively,

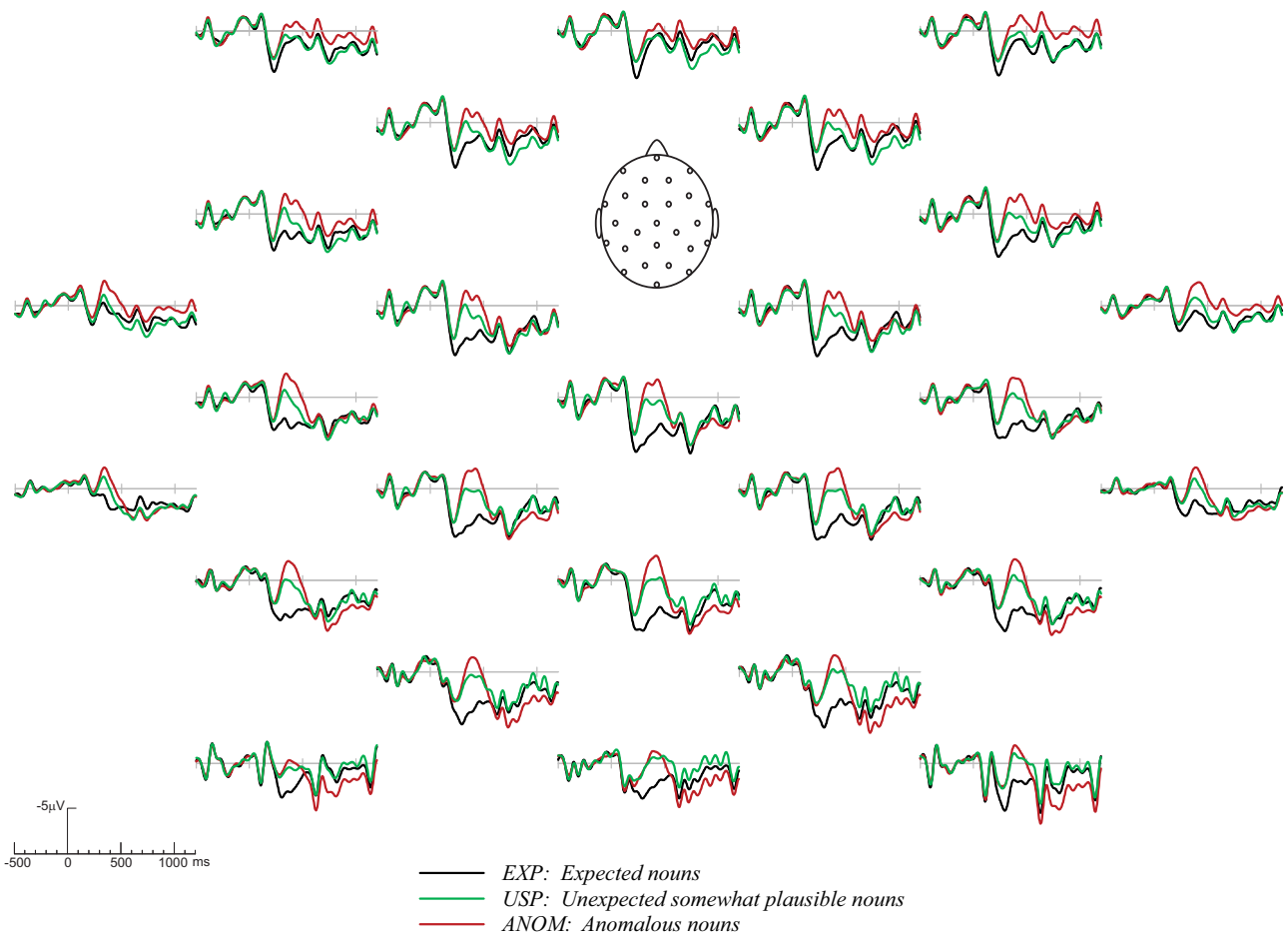

Fig. 3. Experiment 2 grand average (N=24) ERPs recorded over 26 scalp channels.

within a single data set and group of participants. To test the idea that the two positivities were differentially sensitive to the (im)plausibility of unexpected sentence continuations in a more controlled, explicit manner, Experiment 2 was modified from Experiment 1 in several ways: (1) Although the 150 sentence contexts were all taken from Experiment 1, in Experiment 2 the indefinite articles preceding critical nouns were held constant for all three experimental conditions. The consequence to this, in order to maintain article (*a*/*an*)-noun consistency, was that alternative critical nouns were used for a subset of the unexpected continuations. (2) Benefits of using the same sentence contexts for all 3 conditions were that there were equal items per condition and pre- and post-critical word context was matched across conditions. (3) To eliminate the possible influence of expectancy for a part of speech, the Adjective filler condition from Experiment 1 was dropped.

Experiment 2 revealed similar ERP patterns to Experiment 1. In addition to canonical reduced amplitude N400s to more expected nouns compared to both anomalous and plausible unexpected nouns, there was an increased post-400 posterior positivity to anomalous nouns relative to the other two conditions. There was also an increased post-N400 frontal positivity to unexpected but plausible nouns relative to both expected and anomalous nouns. Importantly, these effects dissociated: the anomalous nouns did *not* exhibit the positivity over anterior sites nor did the unexpected plausible nouns exhibit a positivity over posterior scalp locations. In other words, the anterior positivity was specific to plausible unexpected nouns and the posterior positivity to anomalous nouns.

In sum, the Experiment 2 results, in conjunction with those from Experiment 1, offer within subject and within experiment support for Van Petten and colleagues' (Van Petten & Luka, 2012; Thornhill & Van Petten, 2012) proposal that there are two patterns of late (post-N400) ERP positivity that align with the processing of unexpected but plausible versus unexpected but implausible (anomalous) sentence continuations.

4. General discussion

Taken together, the results from Experiments 1 and 2 support a dissociation between neural processing relating to the predictability versus plausibility of sentence continuations. The anterior positivity results from both Experiments 1 and 2 accord with those of Thornhill and Van Petten (2012) and others who have noted a similar ERP pattern to unexpected but plausible sentence continuations in constraining sentence contexts. The additional and important finding from Experiments 1 and 2 is that in a direct comparison of unexpected plausible versus unexpected anomalous continuations, *only* the plausible, and not the anomalous, words exhibit the post-N400 frontal positivity, while *only* the anomalous continuations exhibit the post-N400 occipito-parietal positivity. Thus, the current study contributes to the ERP literature by offering a dissociation—within a single group of participants and a single set of experimental materials—of two late ERP positivities that show different scalp topographies and sensitivities to different stimulus factors.

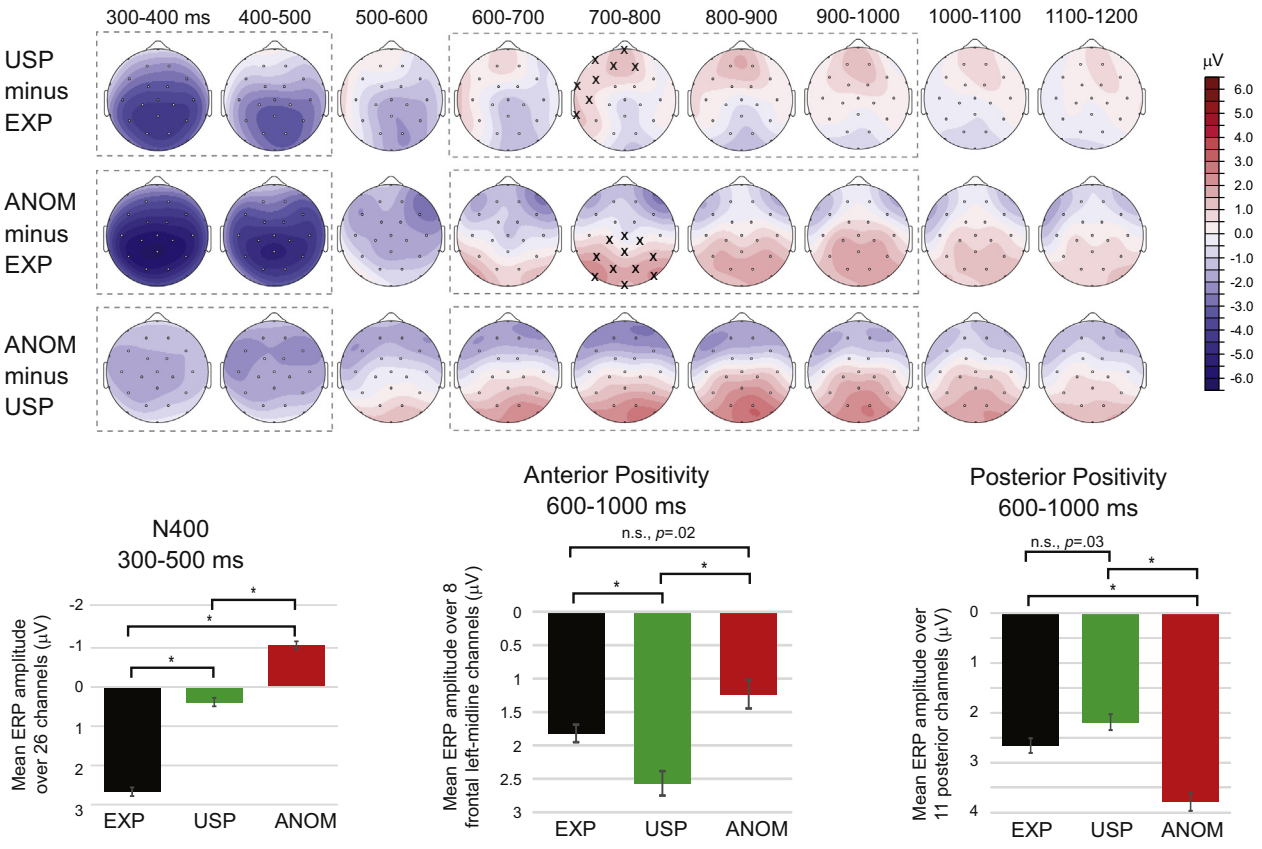

Fig. 4. Experiment 2 topographic scalp maps of ERP mean amplitude voltage differences for all possible conditional comparisons over consecutive 100 ms time windows from 300 to 1200 ms. Experiment 2 analysis time windows over which ERP mean amplitude ANOVAs were performed are indicated in dashed gray line boxes: 300–500 ms for the N400 and 600–1000 ms for the anterior and posterior positivities. Electrode channels highlighted with black "x's" are those for which data was included in the statistical analyses for the frontal and posterior positivities. ERP mean amplitudes are plotted below in bar graph form across channels included in each analysis in the respective time windows. Error bars indicate SEM. All significant pairwise differences (using Bonferroni-adjusted significance levels) for condition contrasts are indicated with an asterisk (*), denoting $p < .0167$: those not reaching statistical significance are indicated with "n.s.".

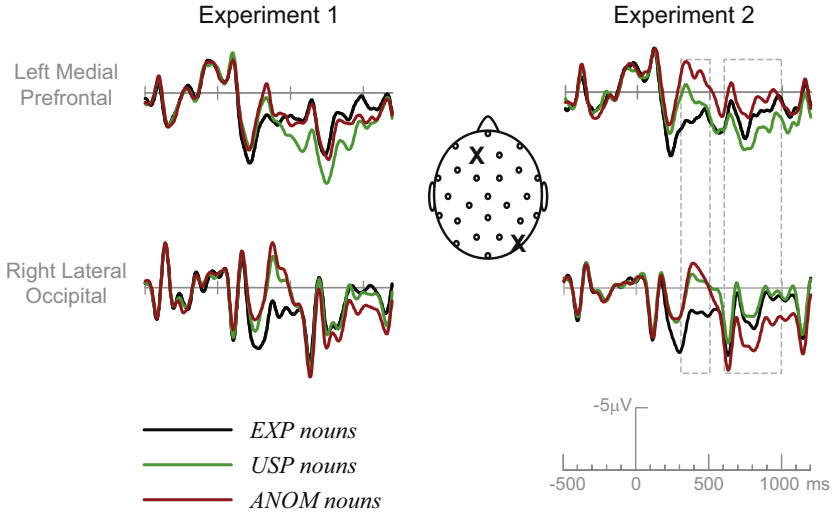

Fig. 5. Representative anterior and posterior electrodes enlarged to show ERP waveforms from Experiments 1 ($N=32$) and 2 ($N=24$). The Experiment 2 noun N400 (300–500 ms) and anterior and posterior positivity (600–1000 ms) mean amplitude analysis time windows are highlighted with dashed-line boxes.

The brain response exhibited in the frontal positivity to unexpected but to some degree plausible continuations, but not to those rated as anomalous, is informative. The fact that this positivity did not occur for both plausibility levels of unexpected words, nor did those levels generate a graded ERP effect when compared with ERPs to expected nouns, suggests that the anterior positivity does not simply reflect some aspect of plausibility evaluation. It also confirms that the frontal positivity is not simply a response to "any word except" the highest cloze completion. In other words, the effect is not just a "mismatch" detector, which is an argument that has similarly been made by Federmeier, Kutas, and Schul (2010) based on frontal positivity findings to atypical

category members (e.g. *ash*) but not to incongruous words (e.g., *tin*) following short phrase prompts (e.g., *a type of tree*), relative to typical category members (e.g., *oak*).

The different ERP responses to the two different unexpected continuation types seem to suggest different downstream processing consequences that might map onto these effects. The plausible and interpretable but contextually unexpected words (see Table 3, e.g., '*For the snowman's eyes the kids used two pieces of coal. For his nose they used a banana...*') may initiate some variety of processing that depends on individuals' world knowledge in order to update their existing contextual representation. These words have the potential to immediately be made sense of in their contexts, without further input (provided, for instance, by the post-critical noun word). In contrast, for the anomalous sentences, critical words are illogical continuations and contextually uninterpretable based on the meaning of the preceding context (e.g., '*…For his nose they used a groan...*', see Table 3). These are conditions under which receipt of post-target words may sometimes offer the potential of disambiguation, to resolve or make sense of the (apparent) anomaly.

One point worth noting is that in Experiment 2, the two positivities, although dissociable in their stimulus sensitivities and scalp distributions, exhibited similar timing. However, across the limited number of studies which have reported frontal positivities, and indeed even in our own more exploratory Experiment 1, this is not always the case. Processing indexed by the frontal positivity sometimes appears to initiate fairly early (within the N400 time region, as suggested by Experiment 1; see also Thornhill & Van Petten, 2012, as well as DeLong et al., 2012, for N400-concurrent frontal positivities). In Experiment 2 of this study we observed the frontal positivity becoming prominent slightly later (by around 600 ms), and other studies have reported mean amplitude findings of frontal positivities over extended post-N400 mean amplitude time windows (e.g., between 500 and 900 ms, Federmeier et al., 2010; Federmeier et al., 2007). Some of this noted latency variability might well be due to experimental differences that could hinge on both the conditions being contrasted, as well as potential overlap in components (e.g., if concurrent with large and widespread N400s, a clear onset to the positivities might be difficult to discern). So while the frontal positivity sometimes appears to precede the posterior one, their at-least-partial-temporal overlap and possible (for instance, task-related) temporal variability in the two effects' timing make it difficult to use "timing" as a criterion for dissociating the two effects.

Another framework for exploring the positivities elicited by the more and less plausible unexpected sentence continuations is in terms of the *impossibility* of the word in context once the critical noun is encountered. In the eye-tracking literature, for instance, some researchers (e.g., Rayner, Warren, Juhasz, & Liversedge, 2004; Warren & McConnell, 2007) have proposed that selection restriction violations (anomalous verb arguments) have immediate effects on reading times; in contrast, they argue that implausible (unlikely) sentence continuations are not detectable as quickly (but see Matsuki et al., 2011, who *did* observe early effects to less plausible continuations, countering that selection restrictions may be one-in-the-same with event knowledge). In the ERP literature, Paczynski and Kuperberg (2012), among others (e.g., Bornkessel-Schlesewsky et al., 2011; van de Meerendonk, Kolk, Chwilla, & Vissers, 2009) have written about this, suggesting that the (posterior) P600 effect may be evoked by selection restriction violations from items that are both implausible *and* impossible with preceding context. For example, they note that contextually impossible animacy violations (i.e., inanimate nouns encountered when animate agents are expected) elicited P600s, in contrast to possible but unlikely world knowledge violations, which elicited N400s but no P600s. Informal inspection of our stimuli from Experiments 1 and 2 indicates that although the majority of items categorized as anomalous were *not* animacy violations, approximately half to two thirds of the items might be considered contextually "impossible" based on other factors. We raise this point because *impossibility*, whether defined as such based on criteria of selection restriction violations or violation of some other factor (e.g., concreteness, suitable affordances, spatial or temporal constraints, or event knowledge), may have contributed to the posterior positivity in our studies; however, since the current studies did not control for *impossibility*, its contribution to the elicitation of the P600-like effect is unclear. We highlight this for future investigation, noting that a more fine-grained analysis of the factors contributing to an item being classified as impossible or anomalous in context might be fruitful.

Although in the past nearly decade there has been much postulated about the functional nature of posterior "semantic" P600s in the sentence processing literature, considerably fewer ideas have been put forth about the nature of the anterior positivity. One tentative explanation for the type of processing indexed by the frontal positivity may have to do with inhibition. For instance, Levy and Anderson (2002) in a review of inhibitory processes and semantic memory retrieval, argue that executive control mechanisms (in prefrontal cortex) may be required to override prepotent responses when a memory must be selectively retrieved in the face of other competing memories. With respect to the present experiments (using the example stimulus from Table 3), when a likely continuation (*carrot*) is strongly pre-activated by a constraining sentence context (as we among others have proposed as a possible sentence comprehension mechanism, e.g., DeLong et al., 2005), but that sentence continues with a less likely but plausible alternative (*banana*) instead, the representation of the pre-activated continuation *carrot*, or the contextual representation that has been built up until that point, may need to be suppressed in order to minimize interference with ongoing meaning construction. Similar suppression may not be required for anomalous continuations such as *groan* because (as we discuss above) the resultant sentence here is more evidently impossible (a *groan* does not have touch affordances, is not an object, is not associated with the act of building snowmen, etc.), and *groan* represents a concept orthogonal to the highly pre-activated one. Thus, anomalous words may demand less (or different) allocation of resources potentially associated with processes of evaluating expectancy-based outcomes, monitoring conflict, negotiating selection competition, or redirecting attentional focus, to suggest a few possibilities. At a very minimum, the rough correspondence (taken at face value) of the gross prefrontal brain regions implicated in inhibition networks with the scalp regions where we observe the frontal positivity is tempting.

Our frontal and posterior positivities also bring to mind a more extensively studied ERP component dissociation between members of the P300 family. Namely, we note that the P3a and P3b ERP components exhibit similar frontal versus parietal scalp distribution patterns (respectively), with our positivities' latencies falling roughly within similar temporal boundaries as these varieties of P300s. The more frontal P3a component is known to reflect orienting of attention to unexpected (novel) events, and is elicited by non-target distracter stimuli in oddball paradigms (Polich, 2003). In similar paradigms, the more posterior task-sensitive P3b is elicited to rare targets (oddballs) and is considered to index the updating of working memory (Donchin & Coles, 1988) and/or categorization and decision making (Kok, 2001). Polich (2007) has suggested that these two P300 subprocesses may reflect neural inhibition of ongoing brain activity in order to facilitate transmission of stimulus or task information from more frontal to parietal sites, based on the need to devote attention to an incoming stimulus relative to the representations that have been constructed in working memory. In the context of the present experiments—for which the tasks were simply to read for comprehension and answer occasional content-related questions—plausible unexpected words could have led to greater inhibition of highly contextually

pre-activated words than the anomalous ones, because conceptual representations can be constructed from these sentences, but perhaps not from the anomalous ones. These comparisons draw on the assumption that the late positivities and P300 effects may belong to the same family of components, which is speculative and indeed runs parallel to the longstanding debate centering on the relation between P3b and P600 components (see Osterhout & Hagoort, 1999, and Coulson, King, & Kutas, 1998, for contrasting views). These issues are beyond the scope of the current paper but suggest avenues for continued investigation.

A final intriguing proposal for a functional correlate of the anterior ERP positivity we observed to plausible but unexpected sentence continuations in the present studies was raised in Kutas, Federmeier and Urbach (in press). They speculate that the frontal component may be linked to activity in the basal forebrain cholinergic system relating to the degree of environmental "expected uncertainty", for which top-down and bottom-up processing influences must be balanced while maintaining attentional sets (Avery, Nitz, Chiba, & Krichmar, 2012; Yu & Dayan, 2005). This more domain general proposal suggests that future explorations of the anterior positivity effect look beyond the sentence processing literature to investigate whether (or the conditions under which) similar scalp recorded ERP patterns obtain.

5. Conclusions

Taken together, the three ERP effects described in this paper point to a neural parser that is both contextually predisposed and immediately sensitive to multidimensional relations of words in written sentences. We offer strong evidence that the brain's sensitivities to variations in cloze probability (evident in the graded sensitivity of the N400 already by 300 ms) and contextual implausibility (visible most clearly in the posterior positivity to anomalous words in the post-N400 time region) are dissociable. We suggest here that the frontal positivity may reflect converging sensitivities to both factors; namely, the pattern seems to be elicited by lexically unexpected words, however, *only* if they are contextually plausible continuations. We speculate that this frontal positivity may relate to a necessary suppression of mental representations arising from pre-activation of highly probable but not presented sentence continuations, when a plausible alternative is encountered. We would thus predict that if used as post-critical word probes, expected (high cloze) continuations might still show some residual activation immediately following presentation of anomalous, but not plausible, unexpected continuations. Continued research into the nature of these proposed interactions will undoubtedly benefit from further systematic manipulations of cloze probability, constraint and plausibility. In the meantime, our finding of both frontal and posterior post-N400 positivities within a single dataset offers a clear demonstration that anomalous and merely unlikely written sentence continuations are processed in qualitatively different ways. The brain seems to engage in differential processing of variably plausible continuations not pre-activated from preceding context, with apparent contributions from different brain regions but with partially overlapping temporal dynamics. Such data offer welcome constraints to psycholinguistic and neurobiological theories of written sentence processing.

Acknowledgments

This research was supported by NICHD Grant HD22614 and NIA Grant AG08313 to M.K., and by UCSD Center for Research in Language and Institute for Neural Computation training predoctoral fellowships to K.A.D. Special thanks to Alex Kuo for his assistance in data collection, and to Tom Urbach and David Groppe for their helpful comments on earlier versions of this manuscript. Our gratitude to David, as well, for his development of the Matlab tools used for statistical analyses in Experiment 1.

References

Avery, M. C., Nitz, D. A., Chiba, A. A., & Krichmar, J. L. (2012). Simulation of cholinergic and noradrenergic modulation of behavior in uncertain environments. *Frontiers in Computational Neuroscience*, *6* (5).

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., et al. (2011). Think globally: cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language*, *117*(3), 133–152.

Chwilla, D. J., Kolk, H. H., & Vissers, C. T. (2007). Immediate integration of novel meanings: N400 support for an embodied view of language comprehension. *Brain Research*, *1183*, 109–123.

Clarke, S., & Hall, P. (2009). Robustness of multiple testing procedures against dependence. *The Annals of Statistics*, *37*(1), 332–358.

Coulson, S., King, J., & Kutas, M. (1998). Expect the unexpected: event-related brain responses to morphosyntactic violations. *Language and Cognitive Processes*, *13*, 21–58.

Coulson, S., & Kutas, M. (2001). Getting it: human event-related brain response to jokes in good and poor comprehenders. *Neuroscience Letters*, *316*, 71–74.

Coulson, S., & Van Petten, C. (2007). A special role for the right hemisphere in metaphor comprehension? ERP evidence from hemifield presentation. *Brain Research*, *1146*, 128–145.

Dale, A. M. (1994). *Source localization and spatial discriminant analysis of event-related potentials: linear approaches*. San Diego, La Jolla, CA: Unpublished Dissertation, University of California.

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*, 226–239.

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), 1117–1121.

DeLong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, *48*, 1203–1207.

DeLong, K. A., Urbach, T. P. & Kutas, M. (2007, March). A cost to mispredicting: effects of sentential constraint violations. In Proceedings of the poster session presented at 20th annual CUNY conference on human sentence processing, La Jolla, CA.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, 357–374.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*(3), 149–161.

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.

Geyer, A., Holcomb, P., Kuperberg, G., & Pearlmutter, N. (2006). Plausibility and sentence comprehension. An ERP Study (Abstract). *Journal of Cognitive Neuroscience, Supplement*.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields II: simulation studies. *Psychophysiology*, *48*(12), 1726–1737.

Hoeks, J. C. J., 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.

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.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*, 557–577.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, Rhode Island: Brown University Press.

Kutas, M. (1993). In the company of other words: electrophysiological evidence for single word versus sentence context effects. *Language and Cognitive Processes*, *8* (4), 533–572.

Kutas, M., Federmeier, K.D., & Urbach, T.P. The "negatives" and "positives" of prediction in language. In: M. S. Gazzaniga and G. R. Mangun (Eds.). *The Cognitive Neurosciences.* MIT Press, Cambridge, (5th Edition), in press.

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.

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.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Research*, *1146*(1), 23–49.

Lage-Castellanos, A., Martínez-Montes, E., Hernández-Cabrera, J. A., & Galán, L. (2010). False discovery rate and permutation test: an evaluation in ERP data analysis. *Statistics in Medicine*, *29*(1), 63–74.

Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences*, *6*(7), 299–305.

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 913.

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching palabras (words): an electrophysiological study of code switching. *Brain and Language*, *80*, 188–207.

Nieuwland, M. S., & van Berkum, J. J. A. (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.

Osterhout, L., & Hagoort, P. (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*, 1–14.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806.

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of memory and language*, *67*(4), 426–448.

Polich, J. (2003). Overview of P3a and P3b. In: J. Polich (Ed.), *Detection of change: event-related potential and fMRI findings* (pp. 83–98). Boston: Kluwer Academic Press.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128–2148.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1290–1301.

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*, 382–392.

van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong? the brain's fast response to morally objectionable statements. *Psychological Science*, *20*(9), 1092–1099.

van de Meerendonk, N., Kolk, H. H., Chwilla, D. J., & Vissers, C. T. W. (2009). Monitoring in language perception. *Language and Linguistics Compass*, *3*(5), 1211–1224.

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, *22*(1), 67–82.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190.

Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin and Review*, *14*(4), 770–775.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692.