



The Impact of Linguistic Prediction Violations on Downstream Recognition Memory and Sentence Recall

Ryan J. Hubbard^{ID} and Kara D. Federmeier^{ID}

Abstract

■ Predicting upcoming words during language comprehension not only affects processing in the moment but also has consequences for memory, although the source of these memory effects (e.g., whether driven by lingering pre-activations, re-analysis following prediction violations, or other mechanisms) remains underspecified. Here, we investigated downstream impacts of prediction on memory in two experiments. First, we recorded EEG as participants read strongly and weakly constraining sentences with expected, unexpected but plausible, or semantically anomalous endings (“He made a holster for his gun / father / train”) and were tested on their recognition memory for the sentence endings. Participants showed similar rates of false alarms for predicted but never presented sentence endings whether the prediction violation was plausible or anomalous, suggesting that these arise from pre-activation of the expected words during reading. During sentence reading, especially in strongly constraining

sentences, plausible prediction violations elicited an anterior positivity; anomalous endings instead elicited a posterior positivity, whose amplitude was predictive of later memory for those anomalous words. ERP patterns at the time of recognition differentiated plausible and anomalous sentence endings: Words that had been plausible prediction violations elicited enhanced late positive complex amplitudes, suggesting greater episodic recollection, whereas anomalous sentence endings elicited greater N1 amplitudes, suggesting attentional tagging. In a follow-up behavioral study, a separate group of participants read the same sentence stimuli and were tested for sentence-level recall. We found that recall of full sentences was impaired when sentences ended with a prediction violation. Taken together, the results suggest that prediction violations draw attention and affect encoding of the violating word, in a manner that depends on plausibility, and that this, in turn, may impair future memory of the gist of the sentence. ■

INTRODUCTION

Numerous psychological and neuroscientific studies have identified that the brain generates predictions of features of upcoming stimuli and environmental states to more efficiently process the rapid input of information we constantly receive (De Lange, Heilbron, & Kok, 2018; Bar, 2009; Friston, 2005). Prediction appears to play an important role in several areas of perceptual and cognitive processing (Den Ouden, Kok, & De Lange, 2012). For instance, it is now widely accepted that language comprehension involves a combination of integrative processing to build a message-level meaning, as well as engagement of predictive mechanisms to facilitate rapid comprehension (Federmeier, 2007, 2022; Pickering & Gambi, 2018; Kuperberg & Jaeger, 2016; Kutas, DeLong, & Smith, 2011; Altmann & Mirković, 2009). The benefits of contextual information and predictive processing during language comprehension have been observed in behavioral studies examining RTs to lexical decisions and word naming times (Simpson, Peterson, Casteel, & Burgess, 1989; Schwanenflugel & LaCount, 1988; Schubert, Spoehr, & Lane, 1981; Fischler & Bloom, 1979), as well as in eye-tracking studies during passive reading (Staub, 2015; Rayner,

Slattery, Drieghe, & Liversedge, 2011; Frisson, Rayner, & Pickering, 2005; Ehrlich & Rayner, 1981) and visual world paradigms (Kamide, 2008; Altmann & Kamide, 1999). In addition, cognitive neuroscientific studies have identified evidence of facilitation through prediction by examining how neural signals related to stimulus processing vary based on predictability. For instance, the N400 component of the ERP, a centroparietal negativity reflecting access of semantic long-term memory (Federmeier, 2022; Kutas & Federmeier, 2000, 2011), is reduced in a graded manner following words that are more predictable (Szewczyk & Federmeier, 2022; Wlotko & Federmeier, 2012; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Federmeier & Kutas, 1999). In addition, recent work using multivariate analysis approaches have demonstrated that features of an upcoming word are neurally pre-activated before the word’s occurrence (Hubbard & Federmeier, 2021a; Wang, Kuperberg, & Jensen, 2018). Thus, there is ample evidence that individuals can engage predictive processing mechanisms during comprehension of language and that the engagement of these mechanisms can facilitate processing.

Facilitation in terms of language comprehension often refers to speed of processing; language information accrues rapidly, and prediction allows comprehenders to

better keep up with that rapid input. However, pre-activation of upcoming information could have unintended negative consequences as well. Comprehenders may process stimuli that were predicted in a “top–down verification mode” (Van Berkum, 2010), such that, if the encountered stimulus matches what was predicted, the bottom–up stimulus input is then processed less deeply or fully. For instance, in one study, predictable sentence-ending words that were then repeated later in a different context elicited reduced ERP repetition effects compared with (similarly repeated) unpredictable words, suggesting that less information had been taken in and retained about those predictable words (Rommers & Federmeier, 2018a). This effect was not observed for unexpected words in the same sentence contexts (Lai, Rommers, & Federmeier, 2021), indicating that prediction played a crucial role in creating the reduced processing.

On the other hand, when comprehenders are predicting and then encounter inputs that do not match what they predicted, the (erroneously) pre-activated features of the stimulus that was not encountered can linger. This lingering pre-activation can result in repetition-like ERP effects for predicted words that were never actually observed (Rommers & Federmeier, 2018b). In addition, these lingering activations can affect later memory judgments: When individuals are tested on their memory for previously read words, they are more likely to falsely remember seeing predictable words that they did not actually read compared with unpredictable words they did not read (Haeuser & Kray, 2022; Hölting & Mecklinger, 2022; Hubbard, Rommers, Jacobs, & Federmeier, 2019; Smith, Hasinski, & Sederberg, 2013), and they are slower to correctly reject predictable words when tested immediately after reading a sentence (Rich & Harris, 2021). Thus, prediction may essentially generate a representation of the upcoming expected stimulus, which can in some ways facilitate processing of the expected stimulus in the moment, but can also have downstream consequences for how that stimulus is then processed in future encounters.

Prediction can also serve a second purpose—namely, prediction errors, or instances in which the encountered stimulus does not match the predicted stimulus, can be potentially useful signals for learning and updating of internal models (Dell & Chang, 2014; Chang, Dell, & Bock, 2006) and may critically impact memory (Sinclair & Barense, 2019). Prediction errors can arise when the context provided by a sentence leads to the prediction of a particular word (e.g., “The rude waiter was not given a *tip*”), but a different word than what was predicted is encountered (e.g., “The rude waiter was not given a *tray*”). Although eye-tracking studies of natural reading behavior have reported little evidence of a processing penalty in reading or fixation times following unexpected words (Frisson, Harvey, & Staub, 2017; Luke & Christianson, 2016), studies employing electrophysiological measurements have identified unique neural responses elicited by words that violate predictions (Van Petten & Luka,

2012). These effects have three important characteristics: First, they are observed most robustly when a particular word is highly expected in the context (i.e., the sentence context is strongly constraining [SC] toward that word) and are less likely to be observed when there is no highly predictable word (i.e., the sentence context is weakly constraining [WC]), suggesting elicitation requires there to be a prediction violation (Federmeier et al., 2007). Second, they are observed at a processing stage later in time with respect to the N400, suggesting they reflect engagement of mechanisms that arise after initial semantic processing of the incoming word. Last, the pattern of ERP responses to prediction-violating words depends on plausibility, suggesting these post-N400 signals may reflect some sort of updating or revision of the previously built message-level representation.

More specifically, unexpected words that violate predictions but are still plausible given the preceding sentence context (e.g., “The rude waiter was not given a *tray*”) elicit an extended positivity, observed roughly 600–1000 msec poststimulus, with an anterior maximum over the scalp. This contrasts with the response to semantically anomalous words (e.g., “The rude waiter was not given a *cabin*”), which often (although not always) elicit a positivity with a posterior scalp distribution in a similar time-frame to the anterior positivity (Brothers, Wlotko, Warnke, & Kuperberg, 2020; DeLong & Kutas, 2020; Kuperberg, Brothers, & Wlotko, 2020; Brothers, Swaab, & Traxler, 2015; DeLong, Quante, & Kutas, 2014; Paczynski & Kuperberg, 2012; Thornhill & Van Petten, 2012; DeLong, Urbach, Groppe, & Kutas, 2011; Van De Meerendonk, Kolk, Vissers, & Chwilla, 2010; Otten & Van Berkum, 2008; Federmeier et al., 2007). Importantly, recent MEG work found that separable neural sources produced these different responses to the different types of unexpected words, suggesting they reflect the engagement of different processes (Wang et al., 2023), although the exact mechanisms engaged is currently a debated topic.

One recent account (Kuperberg et al., 2020) posits a hierarchical generative framework in which the anterior positivity following plausible violations reflects successful updating of the individual’s internally constructed situation model to fit with the unexpected information, whereas the posterior positivity reflects an initial (and potentially continued) failure to update the situation model because of the semantically anomalous nature of the information. Others have posited that the anterior positivity reflects engagement of the frontal cortex to inhibit the predicted word, because of the appearance of a plausible alternative (Ness & Meltzer-Asscher, 2018; Kutas, 1993). The posterior positivity has often been compared with the P600, an ERP component initially associated with the processing of syntactic anomalies (Hagoort, 2003; Gunter, Stowe, & Mulder, 1997; Osterhout & Holcomb, 1992, 1993) and in recent accounts linked to difficulty in semantic integration (Brouwer, Crocker, Venhuizen, & Hoeks, 2017; Brouwer, Fitz, & Hoeks, 2012; Kuperberg,

2007). Thus, the posterior positivity could index integration success, whereas the anterior positivity reflects suppression of the predicted word.

These differing accounts raise questions regarding the locus of the previously discussed effects of prediction on future memory. Individuals may falsely remember seeing predictable words they never actually read because the pre-activated representations linger in memory; however, it may also be the case that prediction violations lead to re-analysis of the originally constructed situation model, causing some level of encoding of the predictable target word. For instance, reading “The rude waiter was not given a *tray*” may cause the reader to re-evaluate the scenario involving the waiter—potentially that the waiter is not given as many tables to work—but still may lead to the inference that the waiter is not given tips for their work or will receive less tips because of not being given as many tables. Critically, then, the impacts on false memory for predictable words may differ based on the type of prediction violation that is read instead of the predictable word. Returning to the previous example, reading “the rude waiter was not given a *cabin*” may lead the reader to consider a completely different scenario, or to simply not engage the same re-evaluation of the scenario, leading to a lack of encoding of the predictable word “tip” into memory at all. In this case, false recognition of predictable words would only occur following unexpected but plausible prediction violations, as only these violations lead to some level of encoding of the predictable word.

The mechanisms engaged to deal with different violation types may also lead to differences in memory for the prediction violations themselves. In our previous work (Hubbard et al., 2019), we found that unexpected but plausible sentence endings elicited larger N1 and late positive complex (LPC) amplitudes during a recognition test compared with expected sentence endings. The increased N1 amplitude may have been caused by participants performing some internal target discrimination of predictable and unpredictable words (Curran, Tanaka, & Weiskopf, 2002; Hopf, Vogel, Woodman, Heinze, & Luck, 2002; Vogel & Luck, 2000), whereas the increased LPC amplitude likely reflected deeper encoding of the unexpected words, leading to greater episodic detail associated with these words during recognition (Yu & Rugg, 2010; Woodruff, Hayama, & Rugg, 2006; Rugg et al., 1998). Unpredictable sentence endings might draw more attention than expected endings, leading to greater depth of encoding (Röer, Bell, Körner, & Buchner, 2019; Craik & Tulving, 1975; Craik & Lockhart, 1972; Wallace, 1965; Von Restorff, 1933). However, semantically anomalous sentence endings were not included in this study, and it is unclear if these words would draw even greater attention than plausible unexpected words, leading to deeper encoding.

Recent work has compared false recognition for predicted words, as well as recognition of prediction violations, following plausible versus anomalous violations (Haeuser & Kray, 2022, 2023). Participants self-paced their

reading of sentences and then were tested on their memory for predictable sentence endings and prediction violations, predictable but unseen words, and unpredictable unseen (new) words. Increased luring was observed for predictable but unseen words, and this luring did not differ based on the type of prediction violation that was read. In addition, a slight benefit to recognition memory was found, specifically for semantically anomalous violations compared with plausible violations or expected endings. However, there were important caveats to this study. First, although self-paced reading paradigms are likely more ecologically valid and closer to actual reading than the rapid-serial visual presentation (RSVP) paradigms used in ERP studies (Ditman, Holcomb, & Kuperberg, 2007), they allow for individual variability in reading rates of critical items, which could influence recognition results. Indeed, participants in this study devoted longer reading times to prediction violations, which could have influenced their recognition memory for these items (Tullis & Benjamin, 2011; Son & Metcalfe, 2000; Mazzoni & Cornoldi, 1993). Second, EEG was not recorded in this experiment, making it difficult to directly examine how processes engaged during reading are related to the observed pattern of recognition memory results. Last, only highly predictable (SC) sentences were read in this study, and thus the effect of constraint could not be tested. Including WC sentences serves as an important baseline comparison, because post-violation ERPs are more robustly observed in SC contexts, and thus the impacts of prediction violations on memory may differ when constraint is low.

In the current study, we expanded upon the previous results of Haeuser and Kray (2022) and Hubbard and colleagues (2019) in two experiments examining the impact of prediction violations on downstream memory. The design of the first experiment largely replicated the design of Hubbard and colleagues (2019), but we now added semantically anomalous endings, in which the real-word plausibility of the scenario described by the sentence was violated, as well as the unexpected but plausible sentence endings. With this design, we could evaluate the impact of different types of prediction violations, as well as how these effects interact with sentential constraint, on downstream luring effects and recognition memory. By recording EEG while sentences were read, we could more directly examine how engagement of processes following prediction violations during reading might influence downstream memory, by relating the magnitude of post-N400 ERP responses to later recognition. This design also allowed us to examine the processes engaged during recognition of the sentence final words themselves, by examining LPC and N1 ERP responses elicited by sentence final words during the recognition test to better understand how prediction errors impact the formation of memories.

In the second experiment, we attempted to provide a fuller understanding of how predictive processing influences downstream memory by testing individuals’ free recall of sentences they had read. Prediction violations

may lead to message-level revision to incorporate the unexpected information, which in turn may impact memory for the sentence as a whole. Few studies have investigated the impact of predictability on sentence recall; although classic psychological research does suggest that more predictable sentences are more readily recalled from memory (Holmes & Murray, 1974), “predictability” in that work was defined based on the real-life plausibility of the events of the sentence, not quantitatively based on the cloze probability of the words within the sentence. Recent work from our laboratory examining the impact of value-driven strategies on sentence recall found that different strategies influenced sentence recall and that plausible final-word prediction violations tended to reduce sentence recall (Chung & Federmeier, 2023). Other work has shown that sentences containing more highly associated words tend to be recalled more easily (Vanevery & Rosenberg, 1970; Rosenberg, 1968, 1969). Thus, it stands to reason that unpredictable or semantically anomalous words may disrupt the semantic associations between words of the sentence, leading to an impairment in sentence recall. Alternatively, given that predictability was manipulated only on the single, sentence-ending word, whether that final word was predictable could end up having little impact on the overall gist-level representation of the sentence in memory, which may be more important than the specific lexical units (Potter & Lombardi, 1990; Graesser, 1978).

Thus, we were able to test a set of hypotheses across the two experiments. We expected that predictions for sentence ending words would linger, leading to more false alarms for lure items compared with new items; of interest, then, was whether the preceding sentence context or the type of prediction violation would have any effect on the false alarm rate. We also hypothesized that predictable sentence endings would not be recognized as easily as unpredictable endings, but that sentences with expected endings would be recalled more easily from memory. We expected that anomalous sentence endings would draw the most attention and thus likely elicit larger N1s than plausible unexpected endings, but that this would also create a greater detriment on recall of the sentences containing the anomalies. Finally, we expected that the magnitude of the post-N400 positivities at the time of sentence reading would be related to downstream memory for the prediction violations, but that the magnitude of this relationship would differ based on the type of prediction violation.

EXPERIMENT 1

Materials and Methods

Participants

Forty-five individuals from the Champaign–Urbana area participated in the experiment in exchange for cash. The sample size was chosen based on the previously conducted study (Hubbard et al., 2019); we increased the

sample size for this study, as an additional condition was included in the current study. All participants were right-handed, reported normal or corrected-to-normal vision, were native English speakers, and had no history of any neurological or psychiatric disorder. Following data collection, three participants were removed because of excessive noise or artifacts in the EEG data, leaving 42 participants in the reported data. Mean age was 20.2 years (range: 18–30 years), and 25 of the participants were women. The study was approved by the institutional review board at University of Illinois, Urbana-Champaign (UIUC), and all participants provided written informed consent and were debriefed following participation. Typical methods of conducting power analyses are not appropriate for mixed-effects model analyses (Faul, Erdfelder, Lang, & Buchner, 2007). Therefore, we used a modern simulation-based approach to estimate power (Kumle, Vö, & Draschkow, 2021). We used the ERP data from Hubbard and colleagues (2019) to estimate the power of detecting an anterior positivity ERP effect, as this effect is typically smaller than other ERP effects and likely more difficult to detect. A linear mixed-effects model was constructed predicting anterior positivity amplitude, with a fixed effect of condition, random intercepts for participants and items, and a random slope of condition for participants. We then used the mixedpower package in R to estimate the power of detecting this effect with an N of 42. This analysis suggested the power to detect this effect was 0.89; thus, the study was adequately powered to detect the effects of interest.

Materials

The stimuli were composed of 240 English sentences, a subset of the stimuli originally used in Federmeier and colleagues (2007). Half of the sentences (120) were SC toward a particular ending word (final word cloze > 0.68, mean final word cloze = 0.83), whereas the other half were WC (final word cloze < 0.42, mean final word cloze = 0.28). A third of the sentences (80) ended with the highest cloze probability sentence ending (i.e., “expected”), a third of the sentences ended with an unexpected but plausible sentence ending (cloze probability approximately 0), and the final third of the sentences ended with a semantically anomalous ending word. Anomalous sentence endings were never produced during the cloze norming and thus had a cloze probability of 0. These words were not normed for plausibility, but were highly implausible given the scenario of the preceding sentence and thus determined to be semantically incongruous by experimenter judgment. Note that although an anomalous ending is also unexpected by the participant, we use the term “unexpected” here to refer specifically to the unexpected but plausible sentence endings and not the semantically anomalous endings. Thus, participants read 40 SC sentences with expected endings (strongly constrained expected [SCE]), 40 with unexpected but plausible endings (SCU), and 40 SC sentences with anomalous endings

(SCA); this was also the case for the WC sentences (40 weakly constrained expected [WCE], 40 WCU, 40 WCA). These stimuli were evenly split into 10 blocks (four of each condition in each block). The lexical properties (word frequency, concreteness, imageability, familiarity) of sentence ending words were controlled such that there were no significant differences across these variables between the experimental conditions.

Participants were tested on their memory for sentence ending words in the recognition memory blocks of the experiment. In each block, participants were presented with sentence ending words, new words, and sentence medial fillers to ensure participants attended to and read the entire sentence. Sentence ending words were either matches, in which the word presented at test matched what the participant read during encoding, or lures, in which the word presented at test was the expected ending to a sentence with an unexpected but plausible ending, or an anomalous ending, at encoding. As an example of a lure item, a participant may read the sentence “Shuffle the cards before you *forget*” (an unexpected but plausible ending) during the encoding period, and be tested on the word “*deal*” (the expected ending) during the memory test. New words were stimuli that were never presented during the course of the encoding period, and were selected to match the stimulus characteristics of the lure items. Over the course of the experiment, participants were tested on 20 items in each condition, along with 40 new items to ensure an equal number of old and new items were tested. Thus, in each of the 10 recognition blocks, participants were tested on two items from each condition, as well as four new items and two sentence medial words, resulting in 26 items per test block. Table 1 provides an overview and examples of the different types of test items.

Lexical properties (word frequency, concreteness, imageability, familiarity) were mostly controlled across test items; however, unexpected but plausible test items did significantly differ from expected test items in word frequency ($E = 107, U = 85, t = 1.98, p = .05$), obtained from Kučera and Francis (1967). Anomalous test items did not significantly differ from expected or unexpected but plausible stimuli in word frequency. Table 2 summarizes the lexical properties of the stimuli used in the recognition memory test. To account for the potential differences in word frequency between conditions, frequency was included as a covariate in statistical analyses.

Similar to Hubbard and colleagues (2019), the stimuli were presented in a pseudorandomized order to prevent issues of stimulus repetition during the memory test. Namely, a sentence ending word used as a test item (“Shuffle the cards before you *deal*”) could appear in the middle of another sentence (“He learned to *deal* with it”), and this repetition could influence recognition memory. To avoid this issue, presentation order was set up to avoid participants reading both sentences before being tested; namely, any sentence containing a critical test item in the middle of it was presented only after the item had already been tested. All participants read the same list of stimuli; although the order of presentation of each stimulus within blocks was randomized, the order of presentation of the blocks was not.

Procedure

Participants were seated approximately 100 cm from a CRT computer monitor in a sound-attenuated and electrically shielded recording booth. Participants were given an explanation of the experimental procedure, as well as a short practice session to familiarize them with the task.

Table 1. Examples of Stimuli Presented during Encoding and Test Blocks during the Experiment

Constraint	Encoding			Test	
	Sentence	Ending Type	Ending	Test Item Type	Test Item
SC	Tim threw a rock and broke the	Expected (E)	window	Match	window
		Unexpected (U)	camera	Match	camera
			Lure	window	
		Anomalous (A)	novel	Match	novel
			Lure	window	
		WC	His ring fell into a hole in the	Expected (E)	sink
Unexpected (U)	couch			Match	couch
	Lure			sink	
Anomalous (A)	banana			Match	banana
	Lure			sink	

Table 2. Lexical Properties of Test Stimuli

<i>Condition</i>	<i>Frequency</i>	<i>Concreteness</i>	<i>Imageability</i>	<i>Familiarity</i>	<i>Word Length</i>
SCE match	4.14	519	567.56	533.94	5.05
SCU match	3.34	455.83	552.31	491.77	6.15
SCA match	4.10	560	547.65	558.25	5.5
SCU lure	4.05	494.59	564.42	515.26	5.45
SCA lure	3.73	554.06	572.38	567.44	4.65
WCE match	4.13	492.77	582.33	533.85	5.35
WCU match	3.10	463.2	553.6	498.27	5.85
WCA match	3.68	561.55	556.9	556.9	5.3
WCU lure	3.34	511.21	578.4	535.27	5.3
WCA lure	2.93	516.7	566.91	512.36	5.45
New	3.80	493.68	563.13	518.53	5.65

Values represent means across items. Frequency values are log transformed and obtained from Kučera and Francis (1967). Concreteness, imageability, and familiarity values are obtained from the MRC psycholinguistic database.

The experiment was divided into 10 encoding-test blocks. During each encoding phase, participants read sentences word by word in a RSVP format and were instructed to try to remember what they read, as their memory would be tested later. Each word appeared in the center of the screen for 200 msec, followed by a 300-msec interstimulus interval (a blank screen). After the last word of the sentence was presented, a blank screen was presented for 500 msec, followed by a fixation cross for 1000 msec. Participants were instructed to try not to blink when they were reading the sentence and to blink and rest their eyes once the fixation cross appeared. Following each encoding phase, participants were given math problems to complete for 30 sec as a distractor between the encoding and test phases.

Following each encoding phase, participants were tested on their memory in the test phase. Each trial of the test phase began with a fixation cross in the center of the screen for 1000 msec, which was then replaced by a test item (a sentence ending word, a new word, or a sentence medial filler). After 1000 msec, a confidence scale appeared underneath the test word, at which point participants could make their response. Upon making a response, the trial would end and the next trial would begin. The confidence scale consisted of 4 points—“Sure New,” “Maybe New,” “Maybe Old,” and “Sure Old.” Participants were instructed to try not to blink during the initial presentation of the word, but once the confidence scale appeared and they could make their response, they could blink. The test phase was self-paced, in that participants could take as long as they needed to respond.

EEG Recording and Preprocessing

EEG data were recorded from 26 Ag/AgCl electrodes embedded into a flexible elastic cap and distributed over

the scalp in an equidistant arrangement. Additional facial electrodes were attached for monitoring of EOG artifacts, including one adjacent to the outer canthus of each eye and one below the lower eyelid of the left eye. Electrode impedances were kept below 5 k Ω . Signals were amplified by a BrainVision amplifier with a 16-bit A/D converter, an input impedance of 10 M Ω , an online bandpass filter of 0.016–250 Hz, and a sampling rate of 1 kHz. The left mastoid electrode was used as a reference for online recording; offline, the average of the left and right mastoid electrodes was used as a reference.

Preprocessing of the EEG data was completed using functions from the EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) toolboxes in the MATLAB programming environment. Following data collection, each raw EEG time series was passed through a 0.1- to 30-Hz bandpass Butterworth filter with a 12-dB/oct roll-off. The signal then was segmented into epochs from –200 msec prestimulus to 1000 msec poststimulus, relative to the onset of each sentence ending word during encoding and each test item during the test phase. The 200 msec prestimulus was used as a baseline period and was averaged and subtracted from the poststimulus data. Ocular artifacts were corrected using the same procedure as in Hubbard and colleagues (2019). For participants' data containing a large number of blinks, the data were decomposed into independent components with Adaptive Mixture Independent Components Analysis (Hsu et al., 2018; Palmer, Kreutz-Delgado, & Makeig, 2012), and the correlation between each independent component timecourse and the bipolar VEOG channel (lower eye channel - left prefrontal channel) was calculated to find the component(s) containing blinks. Components with a high correlation were removed from trials marked as containing blinks. This ocular artifact correction process was performed for 36 of the 42 participants' data, and,

on average, two components were removed. The remaining components were then recombined to reconstruct the EEG data, which were then scanned with an additional sliding window amplitude threshold (300-msec sliding time window, 50-msec step size, 90- μ V threshold), and finally manually checked by the experimenter for any additional artifacts. In total, an average of 6% of trials were removed, with a range of 2%–16% across participants. Following these preprocessing steps, epochs were averaged together to create an ERP for each participant, and grand average ERPs were created by averaging participant ERPs. The grand average plotted ERPs were filtered with a 10-Hz lowpass filter for clarity of visualization and were only calculated for visualizing effects, not statistical measurements.

Statistical Analyses

Statistical analyses were conducted at the trial level to account for variance because of individual items. Both the behavioral and electrophysiological data were statistically analyzed using linear mixed-effects models, using the *lme4* package in R (Bates, Mächler, Bolker, & Walker, 2015). Recognition memory performance was analyzed with generalized mixed-effects logistic regression models (Jaeger, 2008), predicting whether participants made a correct or incorrect recognition response (0 = incorrect, 1 = correct) on trial-level behavioral data. Fixed effects predicting accuracy for each analysis are detailed in the Results section, and the numerical coding of these fixed effects (i.e., dummy coding, contrast coding) are explicitly stated (Brehm & Alday, 2022).

Our analytic approach was to conduct statistical tests of specific condition contrasts when replicating effects from prior studies where the effects were known (e.g., when testing sentence-final ERP effects), and to use an omnibus ANOVA-like approach for conducting tests when the outcomes were less studied and our goal was to test main effects and interactions before specific condition contrasts (e.g., when examining the recognition memory results). In cases where a variable with three levels (i.e., expectancy) was tested, the procedure outlined by Levy (2014) was used, in which two numerical coding variables were entered into the model (e.g., X1, Level 1 = 0, Level 2 = 0.5, Level 3 = -0.5; X2, Level 1 = 0.5, Level 2 = -0.5, Level 3 = 0). To then test for significance of main effects and interactions, likelihood-ratio tests between mixed-effects models, differing only in the presence or absence of a fixed main effect or interaction, were conducted. Random effects were the same between models, and included intercepts for items and slopes and intercepts for participants, with slopes for the fixed effects of the model. When comparing different conditions of interest for testing simple effects, the *relevel* function in R was used to change the reference condition level in the mixed-effects model, and the model was recomputed. Note that no other aspects of the model were changed in these cases, and this process

only changes the reference level to conduct specific tests (Linck, 2016). Word frequency of the target word was included as a fixed effect, because of the potential differences in frequency between conditions that could explain effects of interest (Sassenhagen & Alday, 2016), and because of the long-standing literature suggesting that word frequency can have significant effects on recognition memory (Glanzer & Adams, 1985, 1990). Word frequency values were log-transformed, scaled, and centered. When testing for simple effects (e.g., differences between specific conditions), Wald's z scores were computed for the coefficients of interest. Recognition performance data were plotted as bar plots, with error bars representing 95% within-subject confidence intervals calculated using the Cousineau–Morey method (Morey, 2008; Cousineau, 2005).

For the ERP analyses, statistical analyses were performed on trial-level measurements extracted from averaged activity in specified time windows and channel clusters, chosen to correspond with the previous study (Hubbard et al., 2019). For the N400, measurements were extracted from a central-posterior cluster of six channels (shown in Figure 2), from 250 to 500 msec. For the anterior positivity, a frontal cluster of six channels and a time window of 700–1000 msec was used (shown in Figure 2). For the posterior positivity, an occipital cluster of three channels and a time window of 700–1000 msec was used (shown in Figure 2; DeLong et al., 2014). For the LPC, a posterior cluster of seven channels and a time window of 500–800 msec was used (shown in Figure 4), and for the N1, a posterior cluster of five channels and a time window of 50–150 msec was used, based on the results of the cluster analysis in the previous study (shown in Figure 4). Trial-level amplitudes were predicted with linear mixed-effects models, with random effects including intercepts for items and slopes and intercepts for participants. Significance testing of main effects and interactions was conducted with likelihood-ratio tests between mixed-effects models, and testing of simple effects was conducted with t tests using the Satterthwaite approximation method in the *lmerTest* package in R (Kuznetsova, Brockhoff, & Christensen, 2017).

Brain–Behavior Correlations

Engagement of the post-N400 processes following prediction violations (the anterior and posterior positivities) may impact downstream recognition memory for the prediction violations, or the predicted lures that the prediction violations replaced. To directly test this hypothesis, we conducted brain–behavior correlation analyses to relate the magnitude of the ERP responses during encoding to later recognition memory. Isolating the processes indexed by ERPs generally involves measuring a difference in amplitude between conditions, which renders relating single trial ERP measurements to behavioral outcomes

difficult (Klawohn, Meyer, Weinberg, & Hajcak, 2020; Meyer, Lerner, De Los Reyes, Laird, & Hajcak, 2017). In addition, correlations of ERP measurements with performance within a single condition of the recognition memory test may be contaminated by participants' overall response bias during the test (Rotello & Macmillan, 2007). Therefore, correlations were made at the participant level, rather than the trial level, and partial correlations were used to control for overall memory performance. In addition, Spearman partial correlations were used to relate variables, as this method is more robust to outliers than Pearson correlations (Pernet, Wilcox, & Rousselet, 2013; Rousselet & Pernet, 2012).

For each participant, anterior positivity amplitudes were calculated as the difference between the average anterior positivity in the SCU condition and the average anterior positivity in the WCE condition (where no engagement of the anterior positivity, or other frontal effects that have sometimes been observed in this paradigm, was expected; see Hubbard & Federmeier, 2021b), and posterior positivity amplitudes were calculated as the difference between the average posterior positivity in the SCA condition and the average posterior positivity in the WCE condition. Partial correlation analyses entailed correlating subject-level anterior positivity amplitudes with average "old" responses for SCU matches, controlling for average old responses across all match conditions (thus controlling for response bias). Anterior positivity amplitudes were also correlated with average old responses for SCU lures, controlling for average old responses across all lure conditions. The same process was conducted for posterior positivities, SCA matches, and SCA lures. Partial correlation analyses were conducted with the *ppcor* package in R

(Kim, 2015), which also provided *t* and *p* values for testing significance of the correlations.

Results

Recognition Memory Performance

Overall recognition performance across all of the test item types, plotted as proportion old response, is presented in Figure 1A. Participants were successfully able to discriminate old items (matches) from new items (lures and new words). There appeared to be an increase in false alarms for lure items compared with new items, as was observed previously (Hubbard et al., 2019). To differentiate conditions, we labeled expected ending lures that were replaced by unexpected but plausible words as unexpected lure (UL), whereas expected ending lures that were replaced by semantically anomalous words as anomalous lure (AL).

To assess the differences between conditions statistically, mixed-effects logistic regression models predicting accuracy were used to compare conditions, with word frequency included to account for this variance. First, we examined the false alarms to lure items by comparing the four lure conditions (SC-UL, WC-UL, SC-AL, and WC-AL) to the new item condition. The first analysis predicted recognition rates with a fixed effect of condition consisting of the four lure conditions combined, compared with new items (contrast coded, lure = 0.5, new = -0.5), as well as a fixed effect of confidence (contrast coded, sure = 0.5, maybe = -0.5). This analysis resulted in significant main effects of Condition, $\chi^2(1) = 22.69, p < .01$, and Confidence, $\chi^2(1) = 16.41, p < .01$, but no significant interaction, $\chi^2(1) = 1.41, p = .24$. To assess if each of the lure

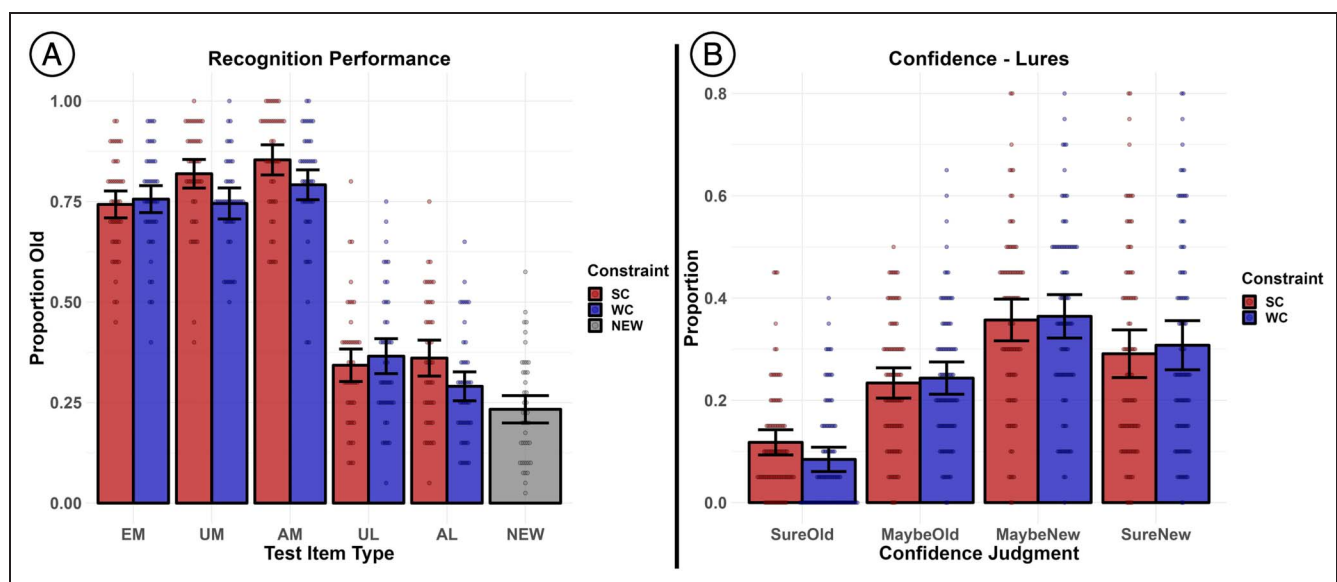


Figure 1. Experiment 1 recognition memory accuracy and confidence. (A) Overall recognition performance across all conditions and levels of constraint. Proportion "Old" responses are plotted on the y axis. Individual dots reflect participant accuracies. (B) Proportion of confidence judgments for lure items. Individual dots reflect participant confidence judgments. SC = strong constraint; WC = weak constraint; EM = expected match; UM = unexpected match; AM = anomalous match.

conditions differed in false alarms, an analysis was conducted with a fixed effect of condition that was dummy coded, such that each of the four lure conditions were compared with the new item condition (i.e., simple effects). Every lure condition significantly differed in accuracy to new items (SC-UL, $\beta = 0.66, z = 2.91, p < .01$; WC-UL, $\beta = 0.94, z = 4.05, p < .01$; SC-AL, $\beta = 0.88, z = 3.84, p < .01$; WC-AL, $\beta = 0.75, z = 3.09, p < .01$), demonstrating that false alarms to lures were observed for all conditions. Word frequency also predicted false alarms in the expected direction, with higher frequency words leading to greater false alarms ($\beta = 0.36, z = 6.65, p < .01$). Thus, participants made more false alarms and were more confident in their judgments to lure items compared with new items.

We next tested whether there were differences in false alarms between the lure conditions. Fixed effects of Constraint (two levels, SC and WC; contrasted coded, SC = 0.5, WC = -0.5), Confidence (two levels, sure and maybe; contrast coded, sure = 0.5, maybe = -0.5), and Expectancy (two levels, UL and AL; contrasted coded, AL = 0.5, UL = -0.5), as well as the interaction between these effects, were included in a mixed-effects logistic regression model predicting accuracy. There were no significant main effects of Constraint or Expectancy, Constraint, $\chi^2(1) = 0.06, p = .80$; Expectancy, $\chi^2(1) = 0.001, p = .97$, and the interaction between these variables was not significant, $\chi^2(1) = 1.40, p = .24$. However, there was a significant main effect of Confidence, $\chi^2(1) = 11.80, p < .01$, as well as a significant interaction of Confidence and Constraint, $\chi^2(1) = 6.42, p = .01$. This is observable in Figure 1B; namely, participants were more confident in their false alarms when the lure was from a SC sentence. Thus, increased false alarms were observed for each of the four lure conditions, but the rate of false alarms did not differ between conditions, although confidence judgments were higher for SC lures.

The next analyses focused on the match items. We tested whether accuracy differed between match conditions with a mixed-effects logistic regression model including fixed effects of Constraint (two levels, SC and WC; contrasted coded, SC = 0.5, WC = -0.5), Expectancy (three levels, E, U, and A; contrasted coded with two variables), Confidence (contrast coded, sure = 0.5, maybe = -0.5), the interaction of these factors, and word frequency. The model reported significant main effects of Constraint, $\chi^2(1) = 9.15, p < .01$, and Expectancy, $\chi^2(2) = 12.91, p < .01$, as well as a significant interaction of these factors, $\chi^2(2) = 7.42, p = .02$. There was a significant main effect of Confidence, $\chi^2(1) = 63.75, p < .01$, but Confidence did not significantly interact with any other factor. In addition, word frequency significantly impacted recognition in the expected direction, with lower frequency words leading to more hits ($\beta = -0.10, z = -2.02, p = .04$).

Follow-up simple effects tests were conducted to unpack the interaction between Constraint and

Expectancy. Recognition performance did not significantly differ between SC-E match items and WC-E match items ($\beta = 0.03, z = 0.15, p = .88$), and accuracy for WC-E match items did not significantly differ from WC-U match items ($\beta = 0.21, z = 1.02, p = .31$). However, accuracy was significantly higher for SC-U match items compared with SC-E match items ($\beta = 0.48, z = 2.22, p = .03$), as well as compared with WC-U items ($\beta = 0.66, z = 3.06, p < .01$). Finally, performance for WC-A match items was greater than WC-U match items ($\beta = 0.43, z = 2.07, p = .04$), and accuracy for SC-A match items showed a similar (but nonsignificant) pattern compared with SC-U match items ($\beta = 0.46, z = 1.92, p = .06$). Thus, there was an overall pattern for better recognition memory for prediction violation test items, and this pattern was larger for violations of SC sentences compared with WC sentences.

To summarize the behavioral results, participants were more likely to false alarm to lure test items compared with new test items, and this effect did not differ based on sentence constraint or sentence ending type (unexpected or anomalous), although individuals were more confident in their memory judgments when the false alarms were from SC sentences. On the other hand, sentence constraint and ending type did have an impact when examining recognition memory for match test items. Prediction violations were better recognized than expected endings, and this effect was larger when constraint was high.

Sentence Final ERPs

Grand average ERPs to sentence final words during the encoding phase are plotted in Figure 2. ERPs were statistically analyzed to determine if the prior effects on ERPs of interest seen with these materials (e.g., Federmeier et al., 2007) were replicated.

First, N400 amplitudes were analyzed to determine if the effects of sentence constraint and expectancy on the N400 were replicated. N400 amplitudes were compared between WCE endings and SCE endings, as well as between WCE and unexpected (U) endings (collapsed across constraint) and anomalous (A) endings (collapsed across constraint), because previous research has demonstrated that N400s to prediction violations largely do not differ based on constraint. There were significant differences in N400 amplitude between WCE and U endings ($\beta = -1.64, t = -4.66, p < .01$), as well as between WCE and A endings ($\beta = -2.87, t = -6.97, p < .01$). Surprisingly, the difference in N400 amplitudes between SCE and WCE endings was not statistically significant ($\beta = 0.54, t = 1.37, p = .17$), possibly because of post-N400 differences in the measurement window.¹ Thus, the graded N400 effect was largely replicated in this experiment.

Next, we examined anterior positivity amplitudes to determine if SCU sentence endings elicited greater anterior positivity ERPs. Previous work has operationalized the anterior positivity effect as a significant difference between SCU and WCU endings (Federmeier et al.,

2007), or a difference between expected (E, collapsed across constraint) endings and SCU endings (DeLong et al., 2014). However, frontally distributed negativity responses have also been observed following expected endings to more highly constraining sentences (Wlotko & Federmeier, 2012). Therefore, the WCE condition, where neither a positivity nor a negativity is expected to be elicited, was used as baseline comparison condition. The difference between SCU endings and anomalous (A) endings was tested as well. The difference in anterior positivity amplitudes between SCU and WCU endings was not significant ($\beta = -0.75, t = -1.55, p = .13$). In contrast, the difference in anterior positivity amplitudes elicited by SCU and WCE endings was significant ($\beta = -1.24, t = -2.59, p = .01$), as was the difference between SCU and A endings ($\beta = -1.43, t = -3.11, p < .01$). To determine the specificity of the anterior positivity effect, an additional analysis compared anterior positivity amplitudes elicited by SCA endings to WCE endings; this comparison was not significant ($\beta = 0.08, t = 0.14, p = .89$). Thus, unexpected but plausible endings to SC sentences elicited an anterior positivity response.

Last, we examined posterior positivity amplitudes to determine if SCA sentence endings elicited greater posterior positivity ERPs. We tested the difference between SCA and WCA endings, as well as the difference between SCA endings and WCE endings, and SCA and unexpected (U) endings. All three comparisons were significant; specifically, posterior positivity amplitudes elicited by SCA endings were more positive compared with WCA endings

($\beta = -1.07, t = -2.70, p = .01$), WCE endings ($\beta = -1.22, t = -2.93, p < .01$), and U endings ($\beta = -1.24, t = -3.43, p < .01$). To determine the specificity of the posterior positivity effect, an additional analysis compared posterior positivity amplitudes elicited by SCU endings to WCE endings, which we used as a baseline condition for consistency with the anterior positivity analysis. This analysis produced a nonsignificant result ($\beta = 0.58, t = 1.62, p = .11$). Thus, semantically anomalous endings to SC sentences elicited a posterior positivity response.

To summarize the sentence-final ERP results, we replicated effects that were observed in prior studies. Namely, N400 amplitudes varied based on sentential constraint and expectancy, anterior positivities were elicited by unexpected endings of SC sentences, and posterior positivities were elicited by anomalous endings of SC sentences. Participants elicited neural responses that were indicative of engagement of anticipatory processing during reading, and likely did not engage in radically different processing during sentence reading, allowing us to link responses during sentence reading to future behavior on the recognition test.

Brain–Behavior Correlations

Partial correlation analyses were conducted to relate the magnitude of post-N400 ERPs (the anterior and posterior positivities) to downstream memory for matches and lures. These analyses related ERP amplitudes to

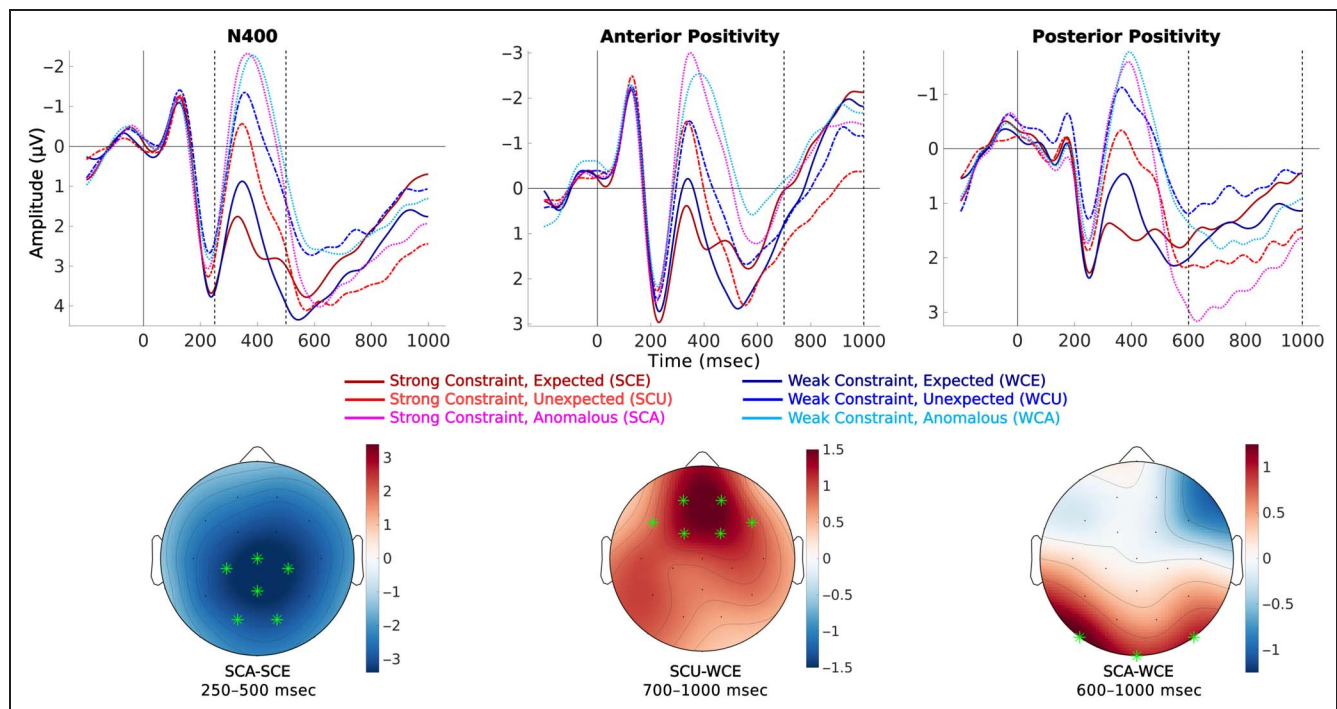


Figure 2. ERPs time-locked to sentence final words during the encoding phase. Top: Time-course of ERPs at three different channel ROIs, highlighting different ERP components (the N400, anterior positivity, and posterior positivity). Dotted lines show the averaged time window for the topography plots. Bottom: Topography plots of the ERP effects in the top plot. The condition difference and time window are shown below each topography plot. The bolded channels depict the averaged channel ROIs for the top ERP plots.

recognition performance at the participant level, while controlling for overall memory performance. The results of these partial correlations are plotted in Figure 3.

At the participant level, anterior positivity amplitudes elicited by SCU endings at the time of encoding were not correlated with successful recognition memory for SCU match items ($r = .01, t = 0.04, p = .97$), or for false recognition of predicted lures that were replaced by unexpected but plausible words ($r = -.01, t = -0.06, p = .96$). In contrast, although posterior positivity amplitudes elicited by SCA endings at the time of encoding were not correlated with false recognition of predicted lures that were replaced by anomalous words ($r = .09, t = 0.53, p = .60$), posterior positivity amplitudes were significantly correlated with recognition of SCA match items ($r = .35, t = 2.33, p = .02$). Thus, the magnitude of elicitation of posterior positivities was related to future memory for

semantically anomalous endings. In summary, elicitation of the anterior positivity to unexpected endings during sentence reading was unrelated to recognition of matches or lures during the recognition test, whereas elicitation of the posterior positivity to anomalous endings during sentence reading was related to downstream recognition of these endings, although not to predictable lures, during the memory test.

Recognition ERPs

Grand average ERPs to match items during the recognition memory test are plotted in Figure 4. Only correct responses were included. ERP components were statistically analyzed to assess differences in recognition memory processes between the different types of items.

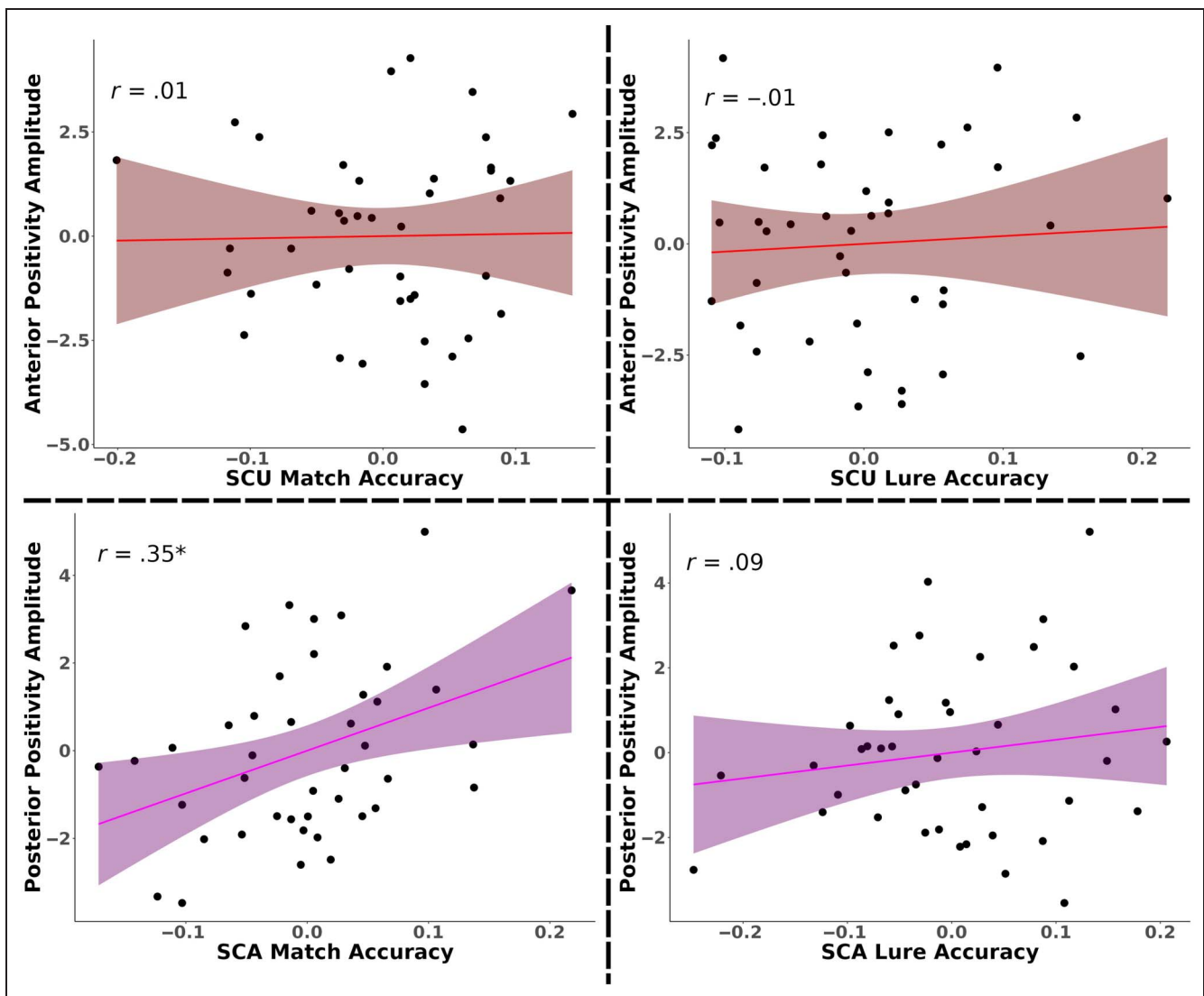


Figure 3. Partial correlations between post-N400 ERP magnitudes (the anterior and posterior positivities) and downstream memory for matches and lures. ERP amplitudes were calculated as average measurements of the difference of the condition of interest (SCU for anterior positivities, SCA for posterior positivities) and the WCE condition. Accuracy reflects average “Old” response for the specific condition of interest, controlling for overall memory performance.

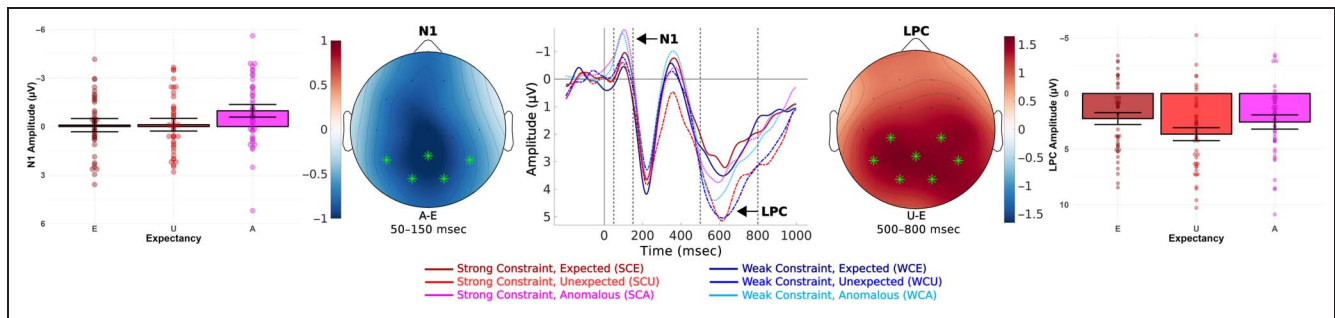


Figure 4. ERPs time-locked to match items during the recognition phase. Middle: Time-course of ERPs at the posterior channel cluster, with different ERP components (the N1 and LPC) labeled. Dotted lines show the averaged time window for the topography plots. Left: Topography plot of the N1 ERP effect, and bar plots of condition differences in amplitude. Right: Topography plot of the LPC ERP effect, and bar plots of condition differences in amplitude. Error bars depict 95% within-subject confidence intervals calculated using the Cousineau–Morey method.

The first analysis examined differences in N400 amplitudes between conditions. N400 amplitudes elicited by match items were predicted in a mixed-effects linear regression model including fixed effects of Constraint (two levels, SC and WC; contrasted coded, SC = 0.5, WC = -0.5), Expectancy (three levels, E, U, and A; contrasted coded with two variables), and the interaction of these factors. The model reported no significant main effect of Constraint, $\chi^2(1) = 1.95, p = .16$, or Expectancy, $\chi^2(2) = 4.57, p = .10$, as well as no significant interaction of these factors, $\chi^2(2) = 3.65, p = .16$. Thus, there were no statistical differences in N400 amplitudes between match item conditions during the recognition test.

The next analysis examined differences in LPC amplitudes between conditions. A mixed-effects linear regression model analysis with the same fixed effects as previously described was conducted. The model reported no significant main effect of Constraint, $\chi^2(1) = 0.15, p = .70$; however, the effect of Expectancy was significant, $\chi^2(2) = 9.04, p = .01$. The interaction of these factors was not statistically significant, $\chi^2(2) = 0.12, p = .94$. Follow-up simple effects tests were conducted to unpack the significant effect of Expectancy. Amplitudes of LPCs elicited by unexpected (U) items were significantly more positive than for expected (E) items ($\beta = -1.49, t = -2.99, p < .01$), as well as for anomalous (A) items ($\beta = -1.17, t = -2.15, p = .03$). There was no significant difference in LPC amplitudes between E and A items ($\beta = -0.32, t = -0.62, p = .54$). In summary, match items that were previously unexpected but plausible sentence endings elicited larger LPCs than previously expected or semantically anomalous endings.

The final analysis examined differences in N1 amplitudes between conditions. A mixed-effects linear regression model analysis with the same fixed effects as previously described was conducted. The model reported no significant main effect of Constraint, $\chi^2(1) = 0.46, p = .50$, whereas the effect of Expectancy was significant, $\chi^2(2) = 12.80, p < .01$. The interaction of these factors was not statistically significant, $\chi^2(2) = 1.47, p = .48$. Follow-up simple effects tests were conducted to unpack the significant effect of Expectancy. Amplitudes of N1s elicited by anomalous (A) items were significantly more

negative than for expected (E) items ($\beta = 0.86, t = 2.99, p < .01$), as well as for unexpected (U) items ($\beta = 0.88, t = 3.10, p < .01$). There was no significant difference in N1 amplitudes between E and U items ($\beta = 0.02, t = 0.06, p = .95$). In summary, match items that were previously semantically anomalous sentence endings elicited larger N1s than previously expected or unexpected but plausible endings.

To summarize the recognition test ERP results, we found that unexpected test items elicited larger LPC amplitudes than both expected and anomalous test items. In addition, we found that anomalous test items elicited greater amplitude N1 responses compared with expected and unexpected test items.

EXPERIMENT 2

Materials and Methods

Participants

Forty-six participants were recruited to participate in the experiment through Prolific, an online data collection platform (www.prolific.co) that uses bot detection, prescreening, participant requirements, and ethical payment practices to ensure higher data quality than alternative online platforms (Palan & Schitter, 2018; Peer, Brandimarte, Samat, & Acquisti, 2017). The experiment took approximately 60 min, and participants were paid \$12 for their participation. Participants were required to be native English speakers located in the United States, with a minimum age of 19 years and a maximum age of 45 years. The mean participant age of the remaining sample was 31 years. The study was approved by the institutional review board of UIUC, and all participants provided written informed consent and were debriefed following participation.

Materials

The stimuli were composed of the 120 SC sentences from Experiment 1. Because the effects of prediction violations on recognition memory were largely only observed for SC

sentences, and to make the experiment shorter to include memory recall as well, the WC sentences were omitted for Experiment 2. A third of the sentences (40) ended with the expected or highest cloze probability sentence ending, a third of the sentences ended with an unexpected but plausible sentence ending (cloze probability approximately 0), and the final third of the sentences ended with a semantically anomalous ending word. Thus, participants read 40 SC sentences with expected endings (SCE), 40 with unexpected endings (SCU), and 40 SC sentences with anomalous endings (SCA).

Participants were also tested on their recognition memory for sentence ending words in the recognition memory blocks of the experiment. In each of the two recognition blocks, participants were presented with 60 test items. These included 30 sentence ending words that were read in the encoding period (matches: 10 expected, 10 unexpected, and 10 anomalous), 20 expected endings to sentences that ended with prediction violations (lures: 10 previously unexpected, and 10 previously anomalous), and 10 words that were not read during the encoding period (new words), leading to 120 items across the two recognition blocks.

Procedure

Participants first provided informed consent to participate and then answered demographic questions. Following this, participants were told they would read a series of sentences and that they should try to remember the entire sentences as best as they could, as their memory for the sentences would be tested. Importantly, because the experiment was online and participants were not monitored, participants were told not to use any external aids or take notes during the study to remember the sentences and that our goal was to study what people are able to naturally remember after reading sentences.

The procedure was similar to that of Chung and Federmeier (2023). Participants completed 10 blocks of sentence reading. In each block, 12 sentences (4 from each condition) were presented in a random order and were presented one word at a time. The timing of the presentation of the words was identical to the EEG experiment in an attempt to match memory performance as closely as possible; each word appeared for 200 msec, followed by a 300-msec interstimulus interval, and a 500-msec blank screen followed the final word, followed by a fixation cross for 1000 msec. Following the sentence reading, participants were given math problems to complete for 30 sec as a distractor between the encoding and test phases, identical to Experiment 1.

Following the math problems, participants were tested for their free recall of the sentences they had read. Participants were provided a text entry box and were told to write out as many sentences that they could remember in any order in the provided text box. In addition, they were told to write out full sentences whenever possible,

but that if they could only remember single words or phrases, then to provide those in the text box as well. Participants were given 3 min to enter their responses. Participants could not end the recall period early, and thus the 3 min were equivalent across blocks and participants. A free recall test was given at the end of each block, leading to 10 free recall tests.

Although the primary goal of Experiment 2 was to test sentence recall, the design also allowed us to replicate the previously observed recognition memory results of Experiment 1. Thus, participants were also given two recognition memory tests (one after the fifth block, and one at the end of the experiment) on sentence endings that they had read. Each recognition test contained 60 words, presented in randomized order. Similar to Experiment 1, on each trial, a fixation cross appeared for 1000 msec, followed by a test item. After 1000 msec, the words “OLD” and “NEW” appeared on the screen, at which point participants could make their response. Participants were instructed that if they remembered seeing the word when they were reading the sentences, they should respond OLD, whereas if they did not remember seeing the word, they should respond NEW. Confidence judgments were not tested in Experiment 2, as only SC sentences were included in this experiment.

Statistical Analyses

Statistical analysis of the recognition memory results was carried out similarly to Experiment 1. Because only SC sentences were presented in Experiment 2, constraint was not included in the mixed-effects logistic regression models predicting trial accuracy, and thus simple effects tests were conducted to analyze differences in recognition accuracy between conditions of interest.

To conduct statistical analysis of the recall data, the sentences were coded into three separate categories, to conduct quantitative analyses. The sentences were coded by the experimenter and an additional independent coder. Chung and Federmeier (2023) had previously used six categories, but collapsed across “Verbatim,” “Almost Verbatim,” and “Gist” categories to assess memory of the primary message of the sentence; therefore, a similar strategy was used here. “Full” category responses referred to recall responses that perfectly matched the originally presented sentence, had slight changes to the surface structure, or conveyed the main message of the sentence with less detail or with a few words missing. “Fragment” category responses referred to recall responses that did capture a part of a studied sentence, but did not capture the whole sentence and missed the main message of the sentence. Last, “Single” category responses referred to recall responses in which a single, identifiable content word from a studied sentence was recalled, but no other words or details could be recalled. Participants’ recall responses across the 10 blocks of free recall were compared with the studied sentences and coded into these three categories.

We were interested in differences in recall rates between conditions (expected, unexpected, and anomalous), as well as between different recall categories (full, fragment, or single), and how these factors might interact. However, this made fitting a mixed logistic regression model at the trial level difficult, because, if recall was successful, each trial could only fall in one of the three recall categories, which would lead to fitting three different regression models for the different recall categories. We instead opted to treat the participants' recall responses as count data and fit a mixed-effect, zero-inflated negative binomial regression model to predict the recall count data (Brooks et al., 2017; Moghimbeigi, Eshraghian, Mohammad, & Mcardle, 2008). In this way, a model predicting recall count could include both factors of interest and their interaction. Fixed effects of the model included recall type (three levels, full, fragment, or single; contrast coded with two variables) and expectancy (three levels, E, U, and A; contrast coded with two variables), and the interaction. A random intercept was included for participants; because of model convergence issues, only a random slope of recall type could be included in the analysis.

second experiment was expected, as the recognition test contained a longer list of items compared with the first experiment.

We first examined if the increased false alarms to predicted lures was observed in Experiment 2. A mixed-effects logistic regression model predicting accuracy, with word frequency included, reported that both lure conditions significantly differed in false alarms compared with new items (SC-UL, $\beta = 0.73, z = 5.90, p < .01$; SC-AL, $\beta = 0.60, z = 4.50, p < .01$); however, false alarm rates did not differ between lure conditions ($\beta = 0.13, z = 1.22, p = .22$). Word frequency also predicted false alarms in the expected direction, with higher frequency words leading to greater false alarms ($\beta = 0.21, z = 4.53, p < .01$). Thus, the increased false alarms to predicted lures compared with new words was replicated in Experiment 2. In contrast, when examining recognition memory for match items, there were no significant differences in recognition across conditions (E vs. U, $\beta = 0.11, z = 0.98, p = .33$; E vs. A, $\beta = 0.14, z = 1.27, p = .21$; U vs. A, $\beta = 0.03, z = 0.25, p = .80$); only the effect of word frequency was significant ($\beta = -0.11, z = -2.10, p = .04$).

Results

Recognition Memory Performance

Overall recognition performance across all of the test item types, plotted as proportion old response, is presented in Figure 5A. Although recognition performance in Experiment 2 was lower overall compared with performance in Experiment 1, participants in Experiment 2 were successfully able to discriminate old items (matches) from new items (lures and new words). Lower performance in the

Sentence Recall Performance

Overall sentence recall performance across levels of expectancy and recall category, plotted as the number of items recalled, is presented in Figure 5B. On average, participants were able to recall at least some information from the sentences they had read for 39% of the sentences (47.35 items, $SD = 17.56$). Full recall of the sentences was more difficult, and, on average, participants fully recalled only 22% of the sentences (26.67 items, $SD = 14.35$).

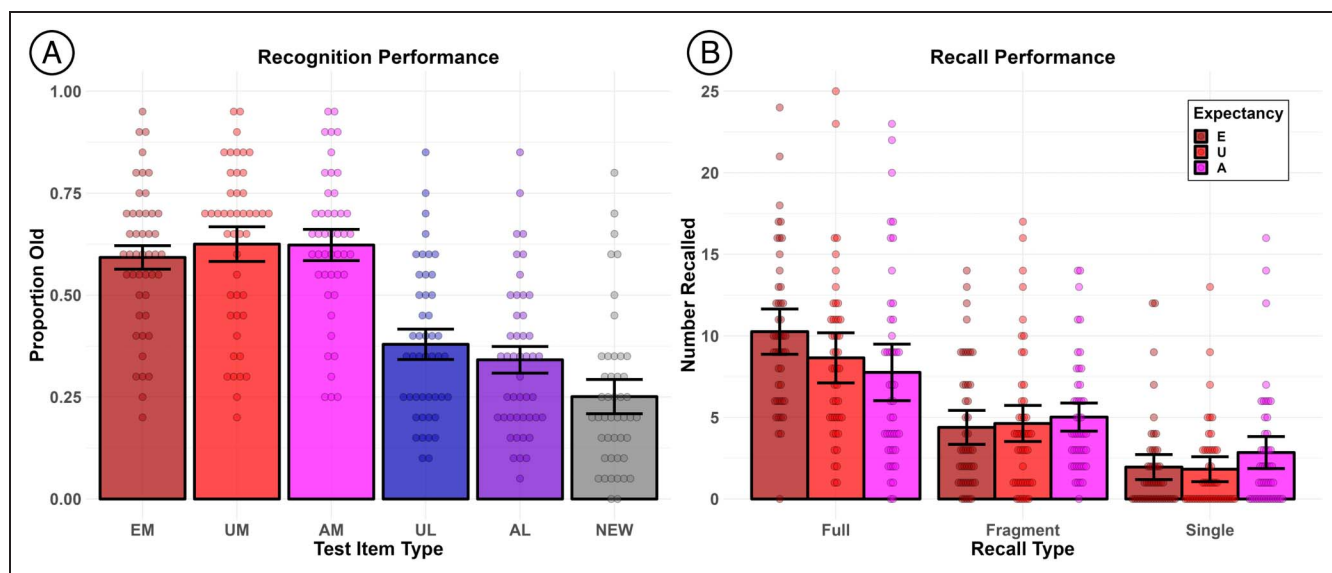


Figure 5. Experiment 2 recognition and recall accuracy. (A) Recognition results. Proportion “Old” responses are plotted on the y axis. EM = expected match; UM = unexpected match; AM = anomalous match. (B) Sentence recall performance. Number of items recalled is plotted on the y axis. E = expected; U = unexpected; A = anomalous. Error bars depict 95% within-subject confidence intervals calculated using the Cousineau–Morey method.

A mixed-effect, zero-inflated negative binomial regression model predicting the recall count data reported a significant main effect of Recall Type, $\chi^2(2) = 38.54, p < .01$, demonstrating significant differences in recall count for the different coding categories. Although the main effect of Expectancy was not significant, $\chi^2(2) = 5.23, p = .07$, the interaction between Expectancy and Recall Type was statistically significant, $\chi^2(4) = 27.47, p < .01$. To unpack this interaction, simple effects tests comparing levels of expectancy were conducted separately for each recall type. When examining recall of full sentences, recall counts were significantly higher for sentences with expected endings (E vs. U, $\beta = 0.17, z = 2.51, p = .01$; E vs. A, $\beta = 0.28, z = 3.98, p < .01$), whereas recall counts did not differ for sentences with prediction violations (U vs. A, $\beta = 0.11, z = 1.49, p = .14$). Recall counts did not statistically significantly differ between levels of expectancy for recall of sentence fragments (E vs. U, $\beta = 0.05, z = 0.54, p = .59$; E vs. A, $\beta = 0.13, z = 1.39, p = .16$; U vs. A, $\beta = 0.08, z = 0.85, p = .39$). Last, when examining recall of single words, recall counts were significantly higher for sentences with anomalous endings (E vs. A, $\beta = 0.38, z = 2.74, p = .01$; U vs. A, $\beta = 0.44, z = 3.18, p < .01$), whereas recall counts did not differ between E and U conditions (E vs. U, $\beta = 0.07, z = 0.45, p = .65$). In summary, full recall of sentences was greater for sentences with expected endings, whereas recall of single words was higher for sentences with anomalous endings. We note that, qualitatively, the single word that was recalled was often the semantically anomalous ending itself.

To summarize the results of Experiment 2, we found that participants were more likely to falsely recognize lure test items than new items during the recognition test, but prediction violations were not recognized more than expected endings. When testing sentence recall, we found that sentences were recalled with greater detail more often when they ended with an expected ending than with a prediction violation. In contrast, participants were more likely to recall single words from sentences that ended with anomalous sentence endings.

DISCUSSION

In two experiments, we investigated the impact of different types of prediction violations on recognition memory of predicted and unpredicted words, as well as on sentence recall. We additionally examined how electrophysiological responses elicited by reading prediction violations in the moment were related to future memory. Our results indicate that the engagement of post-N400 processes following prediction violations is unlikely to explain the heightened false alarms to predicted lures, as violation type had no impact on the tendency to lure. In particular, the result pattern is inconsistent with the notion that the anterior positivity reflects inhibition of the predictable

word, as this would have been expected to correlate with more false alarms to predictable lures following semantically anomalous prediction violations compared with unexpected but plausible violations (Ness & Meltzer-Asscher, 2018). In addition, the results did not indicate that a revision process following an unexpected but plausible word led to some level of encoding of the expected word, as the rates of false alarms were equivalent following plausible and implausible sentence endings. Instead, our results suggest that the neural pre-activation of predictable words leads to the encoding of a representation of the predicted words into memory.

The nature of the mnemonic representations of predicted lures requires further investigation than just the scope of this study; however, two interesting observations can be made from the currently available data. First, the recognition memory results of Experiment 2, as well as work from Haeuser and Kray (2022), demonstrate that false alarms to predicted lures occur even when recognition testing is not immediate (i.e., shortly following reading of the sentences). Work from Hölting and Mecklinger (2022) has even demonstrated that participants will false alarm to predicted lures when tested an entire day after reading, suggesting that the mnemonic representations of the predicted lures are not simply maintained in working memory over a delay and that, once the reading of a particular sentence has been completed, the information that was pre-activated from the sentence context may be stored in long-term memory (Ericsson & Kintsch, 1995). Second, the rate of false alarms to predicted lures did not differ based on the constraint of the sentence, but individuals were more confident in their false alarms when predicted lures were from SC sentences, potentially suggesting more vivid remembering or engagement of recollective processes when remembering these items (Yonelinas, 2001). Greater false alarms to lures from SC sentences were observed when recognition testing occurred a day later (Hölting & Mecklinger, 2022) and when prediction was strategically encouraged (Chung & Federmeier, 2023). A greater rate of false alarms to lures from WC sentences compared with new items suggests that some degree of predictive pre-activation may occur even for words in WC contexts; indeed, previous work examining neural pre-activation using representational similarity analysis indicated some level of pre-activation even for lower cloze probability words (Hubbard & Federmeier, 2021a). However, the magnitude or fidelity of neural pre-activation may differ based on contextual constraint, as well as the degree of strategic engagement of anticipatory processing by the individual, and this greater pre-activation may essentially lead to greater “depth of encoding” for the predicted word (Craik & Tulving, 1975). Further work manipulating the engagement of anticipatory processes and examining memory for predicted lures in populations that engage a lesser degree of prediction, such as older adults (Wlotko, Federmeier, & Kutas, 2012) and less skilled readers (Ng, Payne, Steen, Stine-Morrow, & Federmeier, 2017), will

be necessary to fully understand the link between pre-activation and downstream memory.

The representation of these lure words may not be the same as other types of false memories, such as autobiographical false memories (Pezdek & Lam, 2007; Pezdek, Finger, & Hodge, 1997; Loftus & Pickrell, 1995). The mechanisms that give rise to false recognition of predicted lures may be similar to those that lead to false memory in the Deese–Roediger–McDermott (DRM) task, in which participants generally exhibit higher false alarms to lures that are semantically related to a studied list of stimuli (Gallo, 2010; Roediger & McDermott, 1995; Deese, 1959). Although increased false alarms to lures in the DRM task can be observed up to 60 days following study (Seamon et al., 2002; Thapar & McDermott, 2001; McDermott, 1996), there is little to no correlation between rates of false alarms in the DRM task and the magnitude of false memory judgments following misinformation (Calvillo & Parong, 2016; Ost et al., 2013; Zhu, Chen, Loftus, Lin, & Dong, 2013). The false alarms in the DRM task potentially arise due in some part to spreading activation in semantic memory networks to the lure item during encoding of the semantically related words (Robinson & Roediger, 1997; Underwood, 1965; although see Meade, Watson, Balota, & Roediger, 2007; Zeelenberg, Boot, & Pecher, 2005), as well as more top–down strategic processes that lead to greater associative encoding of the items. For instance, giving participants in a DRM task encoding instructions that focus attention on item-specific details reduces false alarms to lure items (Thomas & Sommers, 2005; McCabe, Presmanes, Robertson, & Smith, 2004), whereas false alarms to lures are increased when encoding is prioritized with high-value cues (Bui, Friedman, McDonough, & Castel, 2013). A similar case may occur here with predictable sentence ending lure words: When reading a particular sentence, words in the sentence that are semantically related to the predictable lure are activated, and the subsequent spreading activation may cause the lure to be activated to some degree, leading to a later false recognition. On the other hand, the rate of false alarms in the DRM task is highly dependent on the number of semantic associates presented during study, and generally many semantic associates are presented during encoding (Jou, Arredondo, Li, Escamilla, & Zuniga, 2017; Robinson & Roediger, 1997). In the current set of experiments, the studied sentences were not semantically related to each other, and many of the sentences had little semantic association between words (e.g., “She dropped the glass and woke up the baby”). It is possible that the mechanisms giving rise to the observed false alarms to predictable lures differs from other processes that give rise to the formation of false memories, such as in the DRM task. On the other hand, this may provide evidence to support the claim that activation states in semantic memory are noncompetitive, and multiple semantic associates of different categories can remain in an activate state concomitantly without interference (Federmeier, 2022).

Brain–behavior correlation analyses provided novel results that the amplitudes of posterior positivities elicited by anomalous endings during sentence reading were related to successful downstream recognition of those words. These results have important implications for psycholinguistic interpretations of the post-N400 positivities, such as the hierarchical generative framework proposed by Kuperberg and colleagues (2020). Although anterior positivities may reflect updating of the constructed internal situation model to incorporate the unexpected information, the results here suggest that, although anomalous sentence endings are clearly not successfully integrated with their sentence contexts, the posterior positivity may not directly reflect this integration failure, but may reflect episodic encoding of the anomalous word itself (Rugg et al., 1998; Wilding, Doyle, & Rugg, 1995; Paller & Kutas, 1992). The sentence recall results of Experiment 2 are in line with this notion, as sentences completed by anomalous endings were less likely to be recalled compared with sentences completed with expected endings, clearly demonstrating a failure to integrate the anomalous word with the sentence, but the anomalous ending words were also more likely to be recalled on their own compared with other ending types, demonstrating greater encoding of the words. From these results, we posit that, during reading, anomalous ending words receive greater attention and encoding into memory, at the cost of disrupting the encoding of the sentence-level message. The mechanisms giving rise to this effect may be similar to other phenomena in the memory literature; for instance, emotional or unusual words or events are often remembered better and receive more attention than neutral ones, but at the cost of memory for the surrounding words or peripheral details (Waring & Kensinger, 2009, 2011; Hope & Wright, 2007; Kensinger, Garoff-Eaton, & Schacter, 2007; Loftus, Loftus, & Messo, 1987).

Our results suggest that linguistic prediction violations are generally more likely to be remembered than expected or predictable words, and this result has been corroborated by other studies reporting better memory for prediction violations (Chung & Federmeier, 2023; Haeuser & Kray, 2022; Corley, MacGregor, & Donaldson, 2007; Federmeier et al., 2007). However, this pattern of results was not replicated in the recognition memory results of Experiment 2. One possibility is that the intervening free recall tests influenced the participants’ later recognition of the sentence endings, as successful recall can facilitate later recognition (Wenger, Thompson, & Bartling, 1980). In addition, other related work has demonstrated that information or stimuli that are congruous with events or mental schema are better remembered than incongruous stimuli (Höltje, Lubahn, & Mecklinger, 2019; Van Kesteren et al., 2013; DeWitt, Knight, Hicks, & Ball, 2012; Staresina, Gray, & Davachi, 2009; Neville, Kutas, Chesney, & Schmidt, 1986); indeed, the recall results of Experiment 2 in some ways corroborated these results, as sentences were recalled more often

when they ended with the expected, congruent ending. This suggests that retrieval of message-level or gist-based information from memory benefits from schema-consistent information, whereas distinctiveness or unexpectedness may benefit retrieval of specific stimuli from memory.

Turning to the ERP results during the recognition test, we found that larger N1 amplitudes were elicited by semantically anomalous endings, and the magnitude of N1 amplitudes between unexpected and expected endings did not differ. This result is in line with an “attentional tagging” account or the interpretation that task-based categorization of stimuli can influence N1 amplitudes elicited by those stimuli later (Curran et al., 2002; Hopf et al., 2002; Vogel & Luck, 2000). It is possible that when reading, participants may mentally categorize read words into predictable versus unpredictable stimuli, which influences the elicited N1 ERP amplitudes when they are encountered later. However, the plausibility or the “magnitude” of the prediction error may also be important for this categorization process, as larger N1 ERPs were elicited by unexpected but plausible words in Hubbard and colleagues (2019) but, strikingly, were elicited only by anomalous sentence endings in the current study. Thus, the amplitude of the N1 at retrieval may indicate the stimuli marked as most “deviant” during encoding, but in a categorical rather than graded manner. Interestingly, other work has demonstrated that expectancy violations in experiments with more simplistic visual stimuli may modulate N1 ERP amplitudes, with violations eliciting larger N1 amplitudes (Baker, Pegna, Yamamoto, & Johnston, 2021; Robinson, Breakspear, Young, & Johnston, 2020; Johnston et al., 2017; Roussel, Hughes, & Waszak, 2014). This work suggests there may be a more fundamental relationship between prediction errors and early visual processing, although we note that N1 ERP modulations by semantically anomalous words are not generally observed at the time of reading (e.g., Kutas, Neville, & Holcomb, 1987), and, even in the current experiment, were only observed during the recognition test. This effect may be more readily observed in simplistic visual experiments with more “obvious” prediction violations, whereas more complex stimuli (words in sentences) require an initial attentional tagging or categorization for the N1 effect to emerge.

Successful recognition of unexpected but plausible sentence endings was associated with larger amplitude LPCs compared with recognition of expected endings. Strikingly, LPC amplitudes elicited by semantically anomalous endings were also smaller than those elicited by unexpected endings and did not statistically differ from the LPCs elicited by expected endings. This result may be in line with other work, in which participants read target words that were either congruous or incongruous with a previously presented category statement (e.g., “A type of bird: *robin / hammer*”) and were tested for their recognition memory of the studied target words (Neville et al.,

1986). Interestingly, LPC amplitudes were larger for correctly recognized study words compared with new words, but amplitudes did not differ between congruous and incongruous study words. It is possible that semantic incongruity on its own, which can potentially draw attention and influence recognition rates, does not lead to downstream modulation of LPC amplitudes. Larger LPCs during recognition memory tests are thought to reflect greater episodic recollection of details and higher memory confidence associated with the study stimuli (Rugg & Curran, 2007; Curran, 2000; Rugg et al., 1998), although task demands and decision-related factors can also influence LPC amplitudes (Yang et al., 2019; Guillaume & Tiberghien, 2013; Finnigan, Humphreys, Dennis, & Geffen, 2002). Unexpected but plausible endings may elicit larger LPCs during recognition memory testing because of the retrieval of episodic details that are associated with these words during study, which may be because of the message-level revision processes necessary to incorporate the unexpected information. Thus, one possibility is that unexpected but plausible words may be better linked to some aspect of the sentence representation than implausible violations, and that sentence representation is then retrieved in more detail when these words are encountered. Yet, why then would expected endings, which are highly congruent with the sentence representation, be recognized less often? We posit that individuals do not thoroughly process predictable endings, as prediction leads to engagement of a “top-down verification mode” in which predicted information is simply confirmed and not deeply encoded (Van Berkum, 2010). This is also in line with a predictive coding account of memory; incoming sensory information that is consistent or expected is essentially “explained away” by top-down predictive signals, potentially reducing encoding, but during memory recall, predictive signals can essentially reinstate the information that was presented (Barron, Aukstulewicz, & Friston, 2020). This demonstrates a potential trade-off of predicting information: Individual words are not encoded as deeply, but message-level representations are preserved.

In summary, our results provide novel insights into the impacts of prediction confirmations and violations on downstream memory. Taken together, the results suggest that the engagement of prediction during reading leads to pre-activation of upcoming words, which, when encountered, are then processed to a lesser extent compared with unexpected words. Although this top-down verification can lead to reduced encoding of individual components of a sentence, it is potentially to the benefit of the message-level representation, as sentences with expected endings are more easily recalled from memory. Unexpected words draw attention and, when plausible, deeper encoding; however, this attention to unexpected information can come at the cost of disruption of the message-level representation of the sentence in memory.

Corresponding author: Ryan J. Hubbard, 405 N Mathews Ave, Urbana, IL 61801, or via e-mail: rjhubba2@illinois.edu.

Data Availability Statement

The conditions of our ethics approval do not permit public archiving of the EEG data in this study because of risk of identification of individuals through biological signals. Readers seeking access to the data should contact the corresponding author.

Author Contributions

R. J. H. and K. D. F. designed the experiment; R. J. H. collected data, and performed the analyses. R. J. H. and K. D. F. wrote the paper together.

Funding Information

This work was supported by a National Institute of Aging grant (<http://dx.doi.org/10.13039/100000049>), grant number: R01AG026308, awarded to K. D. F.

Diversity in Citation Practices

A retrospective analysis of the citations in every article published in this journal from 2010 to 2020 has revealed a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience (JoCN)* during this period were $M(\text{an})/M = .408$, $W(\text{oman})/M = .335$, $M/W = .108$, and $W/W = .149$, the comparable proportions for the articles that these authorship teams cited were $M/M = .579$, $W/M = .243$, $M/W = .102$, and $W/W = .076$ (Fulvio et al., *JoCN*, 33:1, pp. 3–7). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

Note

1. Visually, there appeared to be component overlap influencing the N400 measurements in the SCE condition. A post hoc analysis comparing N400 amplitudes of SCE and WCE endings with a shorter time window (300–400 msec) replicated the well-attested N400 cloze probability effect between these conditions ($\beta = 0.95$, $t = 2.03$, $p = 0.04$).

REFERENCES

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264. [https://doi.org/10.1016/S0010-0277\(99\)00059-1](https://doi.org/10.1016/S0010-0277(99)00059-1), PubMed: 10585516

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33, 583–609. <https://doi.org/10.1111/j.1551-6709.2009.01022.x>, PubMed: 20396405

Baker, K. S., Pegna, A. J., Yamamoto, N., & Johnston, P. (2021). Attention and prediction modulations in expected and unexpected visuospatial trajectories. *PLoS One*, 16, e0242753. <https://doi.org/10.1371/journal.pone.0242753>, PubMed: 34624029

Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 1235–1243. <https://doi.org/10.1098/rstb.2008.0310>, PubMed: 19528004

Barron, H. C., Auksztulewicz, R., & Friston, K. (2020). Prediction and memory: A predictive coding account. *Progress in Neurobiology*, 192, 101821. <https://doi.org/10.1016/j.pneurobio.2020.101821>, PubMed: 32446883

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>

Brehm, L., & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, 125, 104334. <https://doi.org/10.1016/j.jml.2022.104334>

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., et al. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R Journal*, 9, 378–400. <https://doi.org/10.32614/RJ-2017-066>

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. <https://doi.org/10.1016/j.cognition.2014.10.017>, PubMed: 25497522

Brothers, T., Wlotko, E. W., Warnke, L., & Kuperberg, G. R. (2020). Going the extra mile: Effects of discourse context on two late positivities during language comprehension. *Neurobiology of Language*, 1, 135–160. https://doi.org/10.1162/nol_a_00006, PubMed: 32582884

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, 41(Suppl. 6), 1318–1352. <https://doi.org/10.1111/cogs.12461>, PubMed: 28000963

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. <https://doi.org/10.1016/j.brainres.2012.01.055>, PubMed: 22361114

Bui, D. C., Friedman, M. C., McDonough, I. M., & Castel, A. D. (2013). False memory and importance: Can we prioritize encoding without consequence? *Memory & Cognition*, 41, 1012–1020. <https://doi.org/10.3758/s13421-013-0317-6>, PubMed: 23576217

Calvillo, D. P., & Parong, J. A. (2016). The misinformation effect is unrelated to the DRM effect with and without a DRM warning. *Memory*, 24, 324–333. <https://doi.org/10.1080/09658211.2015.1005633>, PubMed: 25664935

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113, 234–272. <https://doi.org/10.1037/0033-295X.113.2.234>, PubMed: 16637761

Chung, Y. M. W., & Federmeier, K. D. (2023). Read carefully, because this is important! How value-driven strategies impact sentence memory. *Memory & Cognition*, 51, 1511–1526. <https://doi.org/10.3758/s13421-023-01409-3>, PubMed: 37458967

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, 105, 658–668. <https://doi.org/10.1016/j.cognition.2006.10.010>, PubMed: 17173887

- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*, 42–45. <https://doi.org/10.20982/tqmp.01.1.p042>
- Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268. <https://doi.org/10.1037/0096-3445.104.3.268>
- Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition, 28*, 923–938. <https://doi.org/10.3758/BF03209340>, PubMed: 11105518
- Curran, T., Tanaka, J. W., & Weiskopf, D. M. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cognitive, Affective, & Behavioral Neuroscience, 2*, 1–18. <https://doi.org/10.3758/CABN.2.1.1>, PubMed: 12452581
- de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences, 22*, 764–779. <https://doi.org/10.1016/j.tics.2018.06.002>, PubMed: 30122170
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17–22. <https://doi.org/10.1037/h0046671>, PubMed: 13664879
- Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*, 20120394. <https://doi.org/10.1098/rstb.2012.0394>, PubMed: 24324238
- DeLong, K. A., & Kutas, M. (2020). Comprehending surprising sentences: Sensitivity of post-N400 positivities to contextual congruity and semantic relatedness. *Language, Cognition and Neuroscience, 35*, 1044–1063. <https://doi.org/10.1080/23273798.2019.1708960>, PubMed: 36176318
- DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia, 61*, 150–162. <https://doi.org/10.1016/j.neuropsychologia.2014.06.016>, PubMed: 24953958
- 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. <https://doi.org/10.1111/j.1469-8986.2011.01199.x>, PubMed: 21457275
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*, 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>, PubMed: 15102499
- Den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology, 3*, 548. <https://doi.org/10.3389/fpsyg.2012.00548>, PubMed: 23248610
- DeWitt, M. R., Knight, J. B., Hicks, J. L., & Ball, B. H. (2012). The effects of prior knowledge on the encoding of episodic contextual details. *Psychonomic Bulletin & Review, 19*, 251–257. <https://doi.org/10.3758/s13423-011-0196-4>, PubMed: 22231723
- Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007). An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology, 44*, 927–935. <https://doi.org/10.1111/j.1469-8986.2007.00593.x>, PubMed: 17850242
- Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*, 641–655. [https://doi.org/10.1016/S0022-5371\(81\)90220-6](https://doi.org/10.1016/S0022-5371(81)90220-6)
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211–245. <https://doi.org/10.1037/0033-295X.102.2.211>, PubMed: 7740089
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. <https://doi.org/10.3758/BF03193146>, PubMed: 17695343
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*, 491–505. <https://doi.org/10.1111/j.1469-8986.2007.00531.x>, PubMed: 17521377
- Federmeier, K. D. (2022). Connecting and considering: Electrophysiology provides insights into comprehension. *Psychophysiology, 59*, e13940. <https://doi.org/10.1111/psyp.13940>, PubMed: 34520568
- 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*, 469–495. <https://doi.org/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. <https://doi.org/10.1016/j.brainres.2006.06.101>, PubMed: 16901469
- Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP 'old/new' effects: Memory strength and decisional factor(s). *Neuropsychologia, 40*, 2288–2304. [https://doi.org/10.1016/S0028-3932\(02\)00113-6](https://doi.org/10.1016/S0028-3932(02)00113-6), PubMed: 12417459
- Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior, 18*, 1–20. [https://doi.org/10.1016/S0022-5371\(79\)90534-6](https://doi.org/10.1016/S0022-5371(79)90534-6)
- Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language, 95*, 200–214. <https://doi.org/10.1016/j.jml.2017.04.007>
- Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 862–877. <https://doi.org/10.1037/0278-7393.31.5.862>, PubMed: 16248738
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*, 815–836. <https://doi.org/10.1098/rstb.2005.1622>, PubMed: 15937014
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition, 38*, 833–848. <https://doi.org/10.3758/MC.38.7.833>, PubMed: 20921097
- Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13*, 8–20. <https://doi.org/10.3758/BF03198438>, PubMed: 4010518
- Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 5–16. <https://doi.org/10.1037/0278-7393.16.1.5>, PubMed: 2136752
- Graesser, A. C. (1978). Tests of a holistic chunking model of sentence memory through analyses of noun intrusions. *Memory & Cognition, 6*, 527–536. <https://doi.org/10.3758/BF03198241>, PubMed: 24203386
- Guillaume, F., & Tiberghien, G. (2013). Impact of intention on the ERP correlates of face recognition. *Brain and Cognition, 81*, 73–81. <https://doi.org/10.1016/j.bandc.2012.10.007>, PubMed: 23174431

- Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, *34*, 660–676. <https://doi.org/10.1111/j.1469-8986.1997.tb02142.x>, PubMed: 9401421
- Haeuser, K. I., & Kray, J. (2022). How odd: Diverging effects of predictability and plausibility violations on sentence reading and word memory. *Applied Psycholinguistics*, *43*, 1193–1220. <https://doi.org/10.1017/S0142716422000364>
- Haeuser, K. I., & Kray, J. (2023). The dark side of prediction: Pervasive false memories for nouns predicted but not seen. Talk held at *The Human Sentence Processing Conference*, Pittsburgh, PA. March 9–11.
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, *15*, 883–899. <https://doi.org/10.1162/089892903322370807>, PubMed: 14511541
- Holmes, P. J., & Murray, D. J. (1974). Free recall of sentences as a function of imagery and predictability. *Journal of Experimental Psychology*, *102*, 748. <https://doi.org/10.1037/h0036111>
- Höltje, G., Lubahn, B., & Mecklinger, A. (2019). The congruent, the incongruent, and the unexpected: Event-related potentials unveil the processes involved in schematic encoding. *Neuropsychologia*, *131*, 285–293. <https://doi.org/10.1016/j.neuropsychologia.2019.05.013>, PubMed: 31112723
- Höltje, G., & Mecklinger, A. (2022). Benefits and costs of predictive processing: How sentential constraint and word expectedness affect memory formation. *Brain Research*, *1788*, 147942. <https://doi.org/10.1016/j.brainres.2022.147942>, PubMed: 35562077
- Hope, L., & Wright, D. (2007). Beyond unusual? Examining the role of attention in the weapon focus effect. *Applied Cognitive Psychology*, *21*, 951–961. <https://doi.org/10.1002/acp.1307>
- Hopf, J.-M., Vogel, E., Woodman, G., Heinze, H.-J., & Luck, S. J. (2002). Localizing visual discrimination processes in time and space. *Journal of Neurophysiology*, *88*, 2088–2095. <https://doi.org/10.1152/jn.2002.88.4.2088>, PubMed: 12364530
- Hsu, S.-H., Pion-Tonachini, L., Palmer, J., Miyakoshi, M., Makeig, S., & Jung, T.-P. (2018). Modeling brain dynamic state changes with adaptive mixture independent component analysis. *Neuroimage*, *183*, 47–61. <https://doi.org/10.1016/j.neuroimage.2018.08.001>, PubMed: 30086409
- Hubbard, R. J., & Federmeier, K. D. (2021a). Dividing attention influences contextual facilitation and revision during language comprehension. *Brain Research*, *1764*, 147466. <https://doi.org/10.1016/j.brainres.2021.147466>, PubMed: 33861998
- Hubbard, R. J., & Federmeier, K. D. (2021b). Representational pattern similarity of electrical brain activity reveals rapid and specific prediction during language comprehension. *Cerebral Cortex*, *31*, 4300–4313. <https://doi.org/10.1093/cercor/bhab087>, PubMed: 33895819
- Hubbard, R. J., Rommers, J., Jacobs, C. L., & Federmeier, K. D. (2019). Downstream behavioral and electrophysiological consequences of word prediction on recognition memory. *Frontiers in Human Neuroscience*, *13*, 291. <https://doi.org/10.3389/fnhum.2019.00291>, PubMed: 31555111
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446. <https://doi.org/10.1016/j.jml.2007.11.007>, PubMed: 19884961
- Johnston, P., Robinson, J., Kokkinakis, A., Ridgeway, S., Simpson, M., Johnson, S., et al. (2017). Temporal and spatial localization of prediction-error signals in the visual brain. *Biological Psychology*, *125*, 45–57. <https://doi.org/10.1016/j.biopsycho.2017.02.004>, PubMed: 28257807
- Jou, J., Arredondo, M. L., Li, C., Escamilla, E. E., & Zuniga, R. (2017). The effects of increasing semantic-associate list length on the Deese–Roediger–McDermott false recognition memory: Dual false-memory process in retrieval from sub- and supraspan lists. *Quarterly Journal of Experimental Psychology*, *70*, 2076–2093. <https://doi.org/10.1080/17470218.2016.1222446>, PubMed: 27606720
- Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, *2*, 647–670. <https://doi.org/10.1111/j.1749-818X.2008.00072.x>
- Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007). Effects of emotion on memory specificity: Memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language*, *56*, 575–591. <https://doi.org/10.1016/j.jml.2006.05.004>
- Kim, S. (2015). ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Communications for Statistical Applications and Methods*, *22*, 665–674. <https://doi.org/10.5351/CSAM.2015.22.6.665>, PubMed: 26688802
- Klawohn, J., Meyer, A., Weinberg, A., & Hajcak, G. (2020). Methodological choices in event-related potential (ERP) research and their impact on internal consistency reliability and individual differences: An examination of the error-related negativity (ERN) and anxiety. *Journal of Abnormal Psychology*, *129*, 29–37. <https://doi.org/10.1037/abn0000458>, PubMed: 31868385
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Kumle, L., Vö, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, *53*, 2528–2543. <https://doi.org/10.3758/s13428-021-01546-0>, PubMed: 33954914
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49. <https://doi.org/10.1016/j.brainres.2006.12.063>, PubMed: 17400197
- Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, *32*, 12–35. https://doi.org/10.1162/jocn_a_01465, PubMed: 31479347
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*, 32–59. <https://doi.org/10.1080/23273798.2015.1102299>, PubMed: 27135040
- Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, *8*, 533–572. <https://doi.org/10.1080/01690969308407587>
- Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). New York, NY: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195395518.003.0065>
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*, 463–470. [https://doi.org/10.1016/S1364-6613\(00\)01560-6](https://doi.org/10.1016/S1364-6613(00)01560-6), PubMed: 11115760
- 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. <https://doi.org/10.1146/annurev.psych.093008.131123>, PubMed: 20809790

- Kutas, M., Neville, H. J., & Holcomb, P. J. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening and signing. *Electroencephalography and Clinical Neurophysiology Supplement*, *39*, 325–330. PubMed: 3477442
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lai, M. K., Rommers, J., & Federmeier, K. D. (2021). The fate of the unexpected: Consequences of misprediction assessed using ERP repetition effects. *Brain Research*, *1757*, 147290. <https://doi.org/10.1016/j.brainres.2021.147290>, PubMed: 33516812
- Levy, R. (2014). Using R formulae to test for main effects in the presence of higher-order interactions. *ArXiv* <https://doi.org/10.48550/arXiv.1405.2094>
- Linck, J. A. (2016). Analyzing individual differences in second language research. *Cognitive Individual Differences in Second Language Processing and Acquisition*, *3*, 105–128. <https://doi.org/10.1075/bpa.3.06lin>
- Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about “weapon focus”. *Law and Human Behavior*, *11*, 55–62. <https://doi.org/10.1007/BF01044839>
- Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, *25*, 720–725. <https://doi.org/10.3928/0048-5713-19951201-07>
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213. <https://doi.org/10.3389/fnhum.2014.00213>, PubMed: 24782741
- Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88*, 22–60. <https://doi.org/10.1016/j.cogpsych.2016.06.002>, PubMed: 27376659
- Mazzoni, G., & Cornoldi, C. (1993). Strategies in study time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General*, *122*, 47. <https://doi.org/10.1037/0096-3445.122.1.47>
- Mccabe, D. P., Presmanes, A. G., Robertson, C. L., & Smith, A. D. (2004). Item-specific processing reduces false memories. *Psychonomic Bulletin & Review*, *11*, 1074–1079. <https://doi.org/10.3758/BF03196739>, PubMed: 15875978
- McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language*, *35*, 212–230. <https://doi.org/10.1006/jmla.1996.0012>
- Meade, M. L., Watson, J. M., Balota, D. A., & Roediger, H. L., III. (2007). The roles of spreading activation and retrieval mode in producing false recognition in the DRM paradigm. *Journal of Memory and Language*, *56*, 305–320. <https://doi.org/10.1016/j.jml.2006.07.007>
- Meyer, A., Lerner, M. D., De Los Reyes, A., Laird, R. D., & Hajcak, G. (2017). Considering ERP difference scores as individual difference measures: Issues with subtraction and alternative approaches. *Psychophysiology*, *54*, 114–122. <https://doi.org/10.1111/psyp.12664>, PubMed: 28000251
- Moghimbeigi, A., Eshraghian, M. R., Mohammad, K., & Mcardle, B. (2008). Multilevel zero-inflated negative binomial regression modeling for over-dispersed count data with extra zeros. *Journal of Applied Statistics*, *35*, 1193–1202. <https://doi.org/10.1080/02664760802273203>
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*, 61–64. <https://doi.org/10.20982/tqmp.04.2.p061>
- 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*, 1269–1285. <https://doi.org/10.1037/xlm0000525>, PubMed: 29283606
- Neville, H. J., Kutas, M., Chesney, G., & Schmidt, A. L. (1986). Event-related brain potentials during initial encoding and recognition memory of congruous and incongruous words. *Journal of Memory and Language*, *25*, 75–92. [https://doi.org/10.1016/0749-596X\(86\)90022-7](https://doi.org/10.1016/0749-596X(86)90022-7)
- Ng, S., Payne, B. R., Steen, A. A., Stine-Morrow, E. A. L., & Federmeier, K. D. (2017). Use of contextual information and prediction by struggling adult readers: Evidence from reading times and event-related potentials. *Scientific Studies of Reading*, *21*, 359–375. <https://doi.org/10.1080/10888438.2017.1310213>
- Ost, J., Blank, H., Davies, J., Jones, G., Lambert, K., & Salmon, K. (2013). False memory ≠ false memory: DRM errors are unrelated to the misinformation effect. *PLoS One*, *8*, e57939. <https://doi.org/10.1371/journal.pone.0057939>, PubMed: 23573186
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806. [https://doi.org/10.1016/0749-596X\(92\)90039-Z](https://doi.org/10.1016/0749-596X(92)90039-Z)
- Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes*, *8*, 413–437. <https://doi.org/10.1080/01690969308407584>
- Otten, M., & Van Berkum, J. J. A. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, *45*, 464–496. <https://doi.org/10.1080/01638530802356463>
- 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*, 426–448. <https://doi.org/10.1016/j.jml.2012.07.003>, PubMed: 23284226
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>
- Paller, K. A., & Kutas, M. (1992). Brain potentials during memory retrieval provide neurophysiological support for the distinction between conscious recollection and priming. *Journal of Cognitive Neuroscience*, *4*, 375–392. <https://doi.org/10.1162/jocn.1992.4.4.375>, PubMed: 23968130
- Palmer, J. A., Kreutz-Delgado, K., & Makeig, S. (2012). *AMICA: An adaptive mixture of independent component analyzers with shared components*. University of California San Diego, Tech. Rep.: Swartz Center for Computational Neuroscience.
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. <https://doi.org/10.1016/j.jesp.2017.01.006>
- Pernet, C. R., Wilcox, R., & Rousselet, G. A. (2013). Robust correlation analyses: False positive and power validation using a new open source MATLAB toolbox. *Frontiers in Psychology*, *3*, 606. <https://doi.org/10.3389/fpsyg.2012.00606>, PubMed: 23335907
- Pezdek, K., Finger, K., & Hodge, D. (1997). Planting false childhood memories: The role of event plausibility. *Psychological Science*, *8*, 437–441. <https://doi.org/10.1111/j.1467-9280.1997.tb00457.x>
- Pezdek, K., & Lam, S. (2007). What research paradigms have cognitive psychologists used to study “false memory,” and what are the implications of these choices? *Consciousness*

- and Cognition, 16, 2–17. <https://doi.org/10.1016/j.concog.2005.06.006>, PubMed: 16157490
- Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*, 1002–1044. <https://doi.org/10.1037/bul0000158>, PubMed: 29952584
- Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language, 29*, 633–654. [https://doi.org/10.1016/0749-596X\(90\)90042-X](https://doi.org/10.1016/0749-596X(90)90042-X)
- Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 514–528. <https://doi.org/10.1037/a0020990>, PubMed: 21463086
- Rich, S., & Harris, J. (2021). Unexpected guests: When disconfirmed predictions linger. *Proceedings of the Annual Meeting of the Cognitive Science Society, 43*, 2246–2252.
- Robinson, J. E., Breakspear, M., Young, A. W., & Johnston, P. J. (2020). Dose-dependent modulation of the visually evoked N1/N170 by perceptual surprise: A clear demonstration of prediction-error signalling. *European Journal of Neuroscience, 52*, 4442–4452. <https://doi.org/10.1111/ejn.13920>, PubMed: 29602233
- Robinson, K. J., & Roediger, H. L., III. (1997). Associative processes in false recall and false recognition. *Psychological Science, 8*, 231–237. <https://doi.org/10.1111/j.1467-9280.1997.tb00417.x>
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814. <https://doi.org/10.1037/0278-7393.21.4.803>
- Röer, J. P., Bell, R., Körner, U., & Buchner, A. (2019). A semantic mismatch effect on serial recall: Evidence for interlexical processing of irrelevant speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*, 515–525. <https://doi.org/10.1037/xlm0000596>, PubMed: 29985037
- Rommers, J., & Federmeier, K. D. (2018a). Predictability's aftermath: Downstream consequences of word predictability as revealed by repetition effects. *Cortex, 101*, 16–30. <https://doi.org/10.1016/j.cortex.2017.12.018>, PubMed: 29414458
- Rommers, J., & Federmeier, K. D. (2018b). Lingering expectations: A pseudo-repetition effect for words previously expected but not presented. *NeuroImage, 183*, 263–272. <https://doi.org/10.1016/j.neuroimage.2018.08.023>, PubMed: 30107258
- Rosenberg, S. (1968). Association and phrase structure in sentence recall. *Journal of Verbal Learning and Verbal Behavior, 7*, 1077–1081. [https://doi.org/10.1016/S0022-5371\(68\)80071-4](https://doi.org/10.1016/S0022-5371(68)80071-4)
- Rosenberg, S. (1969). The recall of verbal material accompanying semantically well-integrated and semantically poorly-integrated sentences. *Journal of Verbal Learning and Verbal Behavior, 8*, 732–736. [https://doi.org/10.1016/S0022-5371\(69\)80037-X](https://doi.org/10.1016/S0022-5371(69)80037-X)
- Rotello, C. M., & Macmillan, N. A. (2007). Response bias in recognition memory. *Psychology of Learning and Motivation, 48*, 61–94. [https://doi.org/10.1016/S0079-7421\(07\)48002-1](https://doi.org/10.1016/S0079-7421(07)48002-1)
- Roussel, C., Hughes, G., & Waszak, F. (2014). Action prediction modulates both neurophysiological and psychophysical indices of sensory attenuation. *Frontiers in Human Neuroscience, 8*, 115. <https://doi.org/10.3389/fnhum.2014.00115>, PubMed: 24616691
- Rousselet, G. A., & Pernet, C. R. (2012). Improving standards in brain-behavior correlation analyses. *Frontiers in Human Neuroscience, 6*, 119. <https://doi.org/10.3389/fnhum.2012.00119>, PubMed: 22563313
- Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences, 11*, 251–257. <https://doi.org/10.1016/j.tics.2007.04.004>, PubMed: 17481940
- Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature, 392*, 595–598. <https://doi.org/10.1038/33396>, PubMed: 9560154
- Sassenhagen, J., & Alday, P. M. (2016). A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. *Brain and Language, 162*, 42–45. <https://doi.org/10.1016/j.bandl.2016.08.001>, PubMed: 27543688
- Schuberth, R. E., Spoehr, K. T., & Lane, D. M. (1981). Effects of stimulus and contextual information on the lexical decision process. *Memory & Cognition, 9*, 68–77. <https://doi.org/10.3758/BF03196952>, PubMed: 7231171
- Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 344–354. <https://doi.org/10.1037/0278-7393.14.2.344>
- Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., et al. (2002). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Memory & Cognition, 30*, 1054–1064. <https://doi.org/10.3758/BF03194323>, PubMed: 12507370
- Simpson, G. B., Peterson, R. R., Casteel, M. A., & Burgess, C. (1989). Lexical and sentence context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 88–97. <https://doi.org/10.1037/0278-7393.15.1.88>
- Sinclair, A. H., & Barense, M. D. (2019). Prediction error and memory reactivation: How incomplete reminders drive reconsolidation. *Trends in Neurosciences, 42*, 727–739. <https://doi.org/10.1016/j.tins.2019.08.007>, PubMed: 31506189
- Smith, T. A., Hasinski, A. E., & Sederberg, P. B. (2013). The context repetition effect: Predicted events are remembered better, even when they don't happen. *Journal of Experimental Psychology: General, 142*, 1298–1308. <https://doi.org/10.1037/a0034067>, PubMed: 23957285
- Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 204–221. <https://doi.org/10.1037/0278-7393.26.1.204>, PubMed: 10682298
- Staresina, B. P., Gray, J. C., & Davachi, L. (2009). Event congruency enhances episodic memory encoding through semantic elaboration and relational binding. *Cerebral Cortex, 19*, 1198–1207. <https://doi.org/10.1093/cercor/bhn165>, PubMed: 18820289
- Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass, 9*, 311–327. <https://doi.org/10.1111/lnc3.12151>
- Szewczyk, J. M., & Federmeier, K. D. (2022). Context-based facilitation of semantic access follows both logarithmic and linear functions of stimulus probability. *Journal of Memory and Language, 123*, 104311. <https://doi.org/10.1016/j.jml.2021.104311>, PubMed: 36337731
- Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition, 29*, 424–432. <https://doi.org/10.3758/BF03196393>, PubMed: 11407419

- Thomas, A. K., & Sommers, M. S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language*, *52*, 71–86. <https://doi.org/10.1016/j.jml.2004.08.001>
- 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. <https://doi.org/10.1016/j.ijpsycho.2011.12.007>, PubMed: 22226800
- Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language*, *64*, 109–118. <https://doi.org/10.1016/j.jml.2010.11.002>, PubMed: 21516194
- Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*, 122–129. <https://doi.org/10.1037/h0022014>, PubMed: 14315122
- Van Berkum, J. J. A. (2010). The brain is a prediction machine that cares about good and bad - any implications for neuropragmatics? *Italian Journal of Linguistics*, *22*, 181–208.
- 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*, 67–82. <https://doi.org/10.1162/jocn.2008.21170>, PubMed: 19199401
- Van Kesteren, M. T. R., Beul, S. F., Takashima, A., Henson, R. N., Ruiter, D. J., & Fernández, G. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: from congruent to incongruent. *Neuropsychologia*, *51*, 2352–2359. <https://doi.org/10.1016/j.neuropsychologia.2013.05.027>, PubMed: 23770537
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*, 176–190. <https://doi.org/10.1016/j.ijpsycho.2011.09.015>, PubMed: 22019481
- Vanevery, H., & Rosenberg, S. (1970). Semantics, phrase structure, and age as variables in sentence recall. *Child Development*, *853*–859. <https://doi.org/10.2307/1127231>
- Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*, 190–203. <https://doi.org/10.1111/1469-8986.3720190>, PubMed: 10731769
- Von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld. *Psychologische Forschung*, *18*, 299–342. <https://doi.org/10.1007/BF02409636>
- Wallace, W. P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin*, *63*, 410–424. <https://doi.org/10.1037/h0022001>
- Wang, L., Kuperberg, G., & Jensen, O. (2018). Specific lexico-semantic predictions are associated with unique spatial and temporal patterns of neural activity. *eLife*, *7*, e39061. <https://doi.org/10.7554/eLife.39061>, PubMed: 30575521
- Wang, L., Schoot, L., Brothers, T., Alexander, E., Warnke, L., Kim, M., et al. (2023). Predictive coding across the left fronto-temporal hierarchy during language comprehension. *Cerebral Cortex*, *33*, 4478–4497. <https://doi.org/10.1093/cercor/bhac356>, PubMed: 36130089
- Waring, J. D., & Kensinger, E. A. (2009). Effects of emotional valence and arousal upon memory trade-offs with aging. *Psychology and Aging*, *24*, 412–422. <https://doi.org/10.1037/a0015526>, PubMed: 19485658
- Waring, J. D., & Kensinger, E. A. (2011). How emotion leads to selective memory: Neuroimaging evidence. *Neuropsychologia*, *49*, 1831–1842. <https://doi.org/10.1016/j.neuropsychologia.2011.03.007>, PubMed: 21414333
- Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 135–144. <https://doi.org/10.1037/0278-7393.6.2.135>
- Wilding, E. L., Doyle, M. C., & Rugg, M. D. (1995). Recognition memory with and without retrieval of context: An event-related potential study. *Neuropsychologia*, *33*, 743–767. [https://doi.org/10.1016/0028-3932\(95\)00017-W](https://doi.org/10.1016/0028-3932(95)00017-W), PubMed: 7675165
- Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *Neuroimage*, *62*, 356–366. <https://doi.org/10.1016/j.neuroimage.2012.04.054>, PubMed: 22565202
- Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: Age-related differences in the use of sentential context. *Psychology and Aging*, *27*, 975–988. <https://doi.org/10.1037/a0029206>, PubMed: 22775363
- Woodruff, C. C., Hayama, H. R., & Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Research*, *1100*, 125–135. <https://doi.org/10.1016/j.brainres.2006.05.019>, PubMed: 16774746
- Yang, H., Laforge, G., Stojanoski, B., Nichols, E. S., McRae, K., & Köhler, S. (2019). Late positive complex in event-related potentials tracks memory signals when they are decision relevant. *Scientific Reports*, *9*, 9469. <https://doi.org/10.1038/s41598-019-45880-y>, PubMed: 31263156
- Yonelinas, A. P. (2001). Components of episodic memory: The contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*, 1363–1374. <https://doi.org/10.1098/rstb.2001.0939>, PubMed: 11571028
- Yu, S. S., & Rugg, M. D. (2010). Dissociation of the electrophysiological correlates of familiarity strength and item repetition. *Brain Research*, *1320*, 74–84. <https://doi.org/10.1016/j.brainres.2009.12.071>, PubMed: 20051232
- Zeelenberg, R., Boot, I., & Pecher, D. (2005). Activating the critical lure during study is unnecessary for false recognition. *Consciousness and Cognition*, *14*, 316–326. <https://doi.org/10.1016/j.concog.2004.08.004>, PubMed: 15950885
- Zhu, B., Chen, C., Loftus, E. F., Lin, C., & Dong, Q. (2013). The relationship between DRM and misinformation false memories. *Memory & Cognition*, *41*, 832–838. <https://doi.org/10.3758/s13421-013-0300-2>, PubMed: 23397226