Downstream Behavioral and Electrophysiological Consequences of Word Prediction on Recognition Memory

Ryan J. Hubbard[1,2]*, Joost Rommers[3], Cassandra L. Jacobs[4] and Kara D. Federmeier[1,2,5]

[1]Department of Psychology, University of Illinois, Urbana-Champaign, IL, United States, [2]Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana-Champaign, IL, United States, [3]Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands, [4]Department of Psychology, Center for Mind and Brain, University of California, Davis, Davis, CA, United States, [5]Program in Neuroscience, University of Illinois, Urbana-Champaign, IL, United States

When people process language, they can use context to predict upcoming information, influencing processing and comprehension as seen in both behavioral and neural measures. Although numerous studies have shown immediate facilitative effects of confirmed predictions, the downstream consequences of prediction have been less explored. In the current study, we examined those consequences by probing participants' recognition memory for words after they read sets of sentences. Participants read strongly and weakly constraining sentences with expected or unexpected endings ("I added my name to the list/basket"), and later were tested on their memory for the sentence endings while EEG was recorded. Critically, the memory test contained words that were predictable ("list") but were never read (participants saw "basket"). Behaviorally, participants showed successful discrimination between old and new items, but false alarmed to the expected-item lures more often than to new items, showing that predicted words or concepts can linger, even when predictions are disconfirmed. Although false alarm rates did not differ by constraint, event-related potentials (ERPs) differed between false alarms to strongly and weakly predictable words. Additionally, previously unexpected (compared to previously expected) endings that appeared on the memory test elicited larger N1 and LPC amplitudes, suggesting greater attention and episodic recollection. In contrast, highly predictable sentence endings that had been read elicited reduced LPC amplitudes during the memory test. Thus, prediction can facilitate processing in the moment, but can also lead to false memory and reduced recollection for predictable information.

Keywords: language comprehension, prediction, false memory, recognition, ERP

INTRODUCTION

The process of prediction has been suggested to play a role in many areas of cognition and behavior, with some arguing that one of the core functions of the brain is to use previously learned associations and top-down control to predict future events (Bar, 2007, 2009; Bubic et al., 2010; Clark, 2013). This function of predicting upcoming information may play a particularly important role in language comprehension (Federmeier, 2007; Kuperberg and Jaeger, 2016), as incoming linguistic information must be processed rapidly. Essentially, by using the bottom-up sensory information provided by written and spoken words, combined with previously learned world knowledge, semantic, and syntactic information, the brain can quickly create and continuously update a representation of likely upcoming linguistic information, which facilitates processing when this information is encountered.

As evidence of the impact of predictability on language comprehension, behavioral work has shown that words that are highly predictable and fit into the ongoing sentence context are processed more rapidly than less predictable words (West and Stanovich, 1978; Fischler and Bloom, 1979; Schuberth et al., 1981; Schwanenflugel and LaCount, 1988; Duffy et al., 1989; Simpson et al., 1989; Hess et al., 1995). Similarly, eye-tracking studies have demonstrated that predictable words are anticipated and read more quickly than unpredictable words (Ehrlich and Rayner, 1981; Altmann and Kamide, 1999; Frisson et al., 2005; Kamide, 2008). Research using event-related potentials (ERPs) has identified that the predictability of words affects the amplitude of the N400, a centroparietal negativity peaking around 400 ms that is thought to index access of semantic memory (Kutas and Hillyard, 1984; Federmeier and Kutas, 1999; Wlotko and Federmeier, 2007, 2012; Kutas and Federmeier, 2011; DeLong et al., 2014). Additionally, unexpected but plausible words that disconfirm a prediction elicit a late, frontally-distributed positivity, which has been hypothesized to index a revision process of some kind (Federmeier et al., 2007; Otten and Van Berkum, 2008; DeLong et al., 2011, 2014; Thornhill and Van Petten, 2012).

There is thus substantial evidence that predictability can lead to facilitated processing of expected information when it is encountered. There are also consequences of processing inputs that violate predictions, as indexed by the late frontal positivity. Do these consequences that are evident in ERPs have corresponding behavioral costs? In early work using lexical decision tasks, identification of predictable words was consistently faster than unpredictable words, but prediction violations did not always lead to response slowing when compared to "baseline" conditions, which varied across the literature (Schuberth and Eimas, 1977; Fischler and Bloom, 1979; Schwanenflugel and Shoben, 1985). Other recent work, in which subjects read sentences at their own pace while eye movements were tracked, reported no evidence of slowing or an increase in re-reading for unexpected words (Luke and Christianson, 2016; Frisson et al., 2017). Therefore, across multiple behavioral paradigms of language processing, convincing evidence of

behavioral costs associated with prediction violations has been lacking.

In addition, behavioral and electrophysiological effects of prediction have predominantly been measured at the time of encountering the predicted or unexpected stimulus. Although this has been useful for identifying the immediate effects of prediction, it leaves open what downstream effects confirmed or disconfirmed predictions might have on later cognition. In order to investigate these potential downstream effects, the present study tested participants' episodic memory for sentence final words of sentences that varied in contextual constraint. The memory test contained words that had been highly predictable, weakly predictable, or unexpected. This allowed for a comparison of the downstream effects of having predictions confirmed or disconfirmed. Critically, the test also included words that were likely to have been predicted but were never actually observed during reading (because the sentence instead had ended with an unexpected word); we will refer to these items as lures.

In addition to behavioral memory measures, the present study recorded EEG to further probe how predictability influences memory processes. Examining ERPs during the memory test allowed us to draw inferences about the neurocognitive processes involved in successfully recognizing, or false alarming to, predictable and unexpected words. Previous studies have identified two major components associated with recognition memory (Rugg and Curran, 2007)—the N400, which has been linked to conceptual fluency or familiarity (Paller and Kutas, 1992; Curran, 2000, 2004; Voss and Federmeier, 2011), with greater familiarity leading to smaller N400s, and the LPC, a left-lateralized posterior component temporally extending from 500 to 800 ms, which is related to recollection or retrieval of more detailed episodic information (Düzel et al., 1997; Rugg et al., 1998; Woodruff et al., 2006; Yu and Rugg, 2010), with greater recollection eliciting more positive LPCs. The amplitudes of these ERP components during the memory test may differ based on the prior predictability of the words or the constraint of the sentences they were presented in, which would provide information about the state of the representations of these items.

Two main issues were of interest: first, we compared memory for predictable words and unexpected words. Here, context-driven prediction could influence the encoding of information into long-term memory by modulating the level of attention given to the predictable or unpredictable information that is being encoded (Craik et al., 1996). Paying more attention to certain stimuli could modulate the depth or level of processing, leading to a more stable and persistent memory representation (Craik and Lockhart, 1972; Craik and Tulving, 1975). In eye-tracking experiments with natural reading, individuals spend less time looking at and exhibit fewer regressions to predictable words (Ehrlich and Rayner, 1981), suggesting they may, in fact, pay less attention to them. Rommers and Federmeier (2018a), investigating ERP repetition effects for previously predictable and unpredictable words, also found that previously predictable words showed reduced downstream repetition effects, suggesting that prediction can lead to an impoverished initial representation. In the case of unpredictable words, some evidence points toward attentional enhancement of encoding: an

item in a list of words that is physically or semantically distinct from the others will be more likely to be recalled (Von Restorff, 1933), unexpected sentence endings draw more attention away from and lead to disruption of serial recall (Röer et al., 2019), and unexpected or error-related events modulate early attention-related ERPs (Wills et al., 2007), suggesting that distinctive, unpredictable events might be more attended to and then more easily remembered. Indeed, studies have reported better recognition memory performance for sentence endings that had been unpredictable (compared with predictable endings), supporting the idea that such words are encoded more strongly (Corley et al., 2007; Federmeier et al., 2007). We further probed the memory processes underlying the recognition of previously predictable and previously unexpected words. In particular, if previously encountered sentence endings increase conceptual priming at test, they should show a reduced N400, whereas if they increase recollection processes, they should elicit an enhanced LPC.

We were also interested in the responses to the lures. If prediction during sentence comprehension leads to pre-activation of information associated with an upcoming word, then participants may show greater false alarms to lures as compared to completely new items. This would constitute a cost of prediction, in that lingering representations can cause false recognition. Alternatively, if the prediction disconfirmation leads to strong revision processes that suppress previously expected information, participants may show fewer false alarms to lures as compared to completely new items. Previous studies have employed an implicit memory paradigm in which participants predict a high cloze ending, are given an unexpected ending, and then must complete a mid-cloze sentence that could potentially be completed by a previous high-cloze or unexpected ending (Hartman and Hasher, 1991; Lorsbach et al., 1996; Hasher et al., 1997). These studies have focused mainly on inhibition and control processes; however, they have demonstrated that individuals tend to retain the expected but disconfirmed endings in some form. In terms of explicit memory, classic studies using the Deese-Roediger-McDermott (DRM) paradigm have shown that individuals will recall an unstudied semantic associate (e.g., "sleep") following study of a list of related words ("dream," "bed," "night," et cetera), suggesting that the representation of the lure was activated and erroneously selected during retrieval (Deese, 1959; Roediger and McDermott, 1995; Steffens and Mecklenbräuker, 2007). In these studies, false alarming is largely driven by semantic similarity of items, and generally occurs immediately following the study. In the current experiment, participants read sentences that were not semantically similar, and were tested after reading several items; thus, a finding of increased false alarming to lures would be a powerful demonstration of prediction's lasting effects on recognition memory.

In addition to behavioral effects, we were also interested in the processes involved in false recognition, as revealed by electrophysiological responses; however, previous results of when and how false recognition manifests in the ERP have been mixed (Curran et al., 2001; Wolk et al., 2006; Geng et al., 2007; Beato et al., 2012; Chen et al., 2012). A recent ERP study showed that words that were previously expected, but not presented, elicited a "pseudo-repetition" effect (Rommers and Federmeier, 2018b); namely, these items showed ERP effects similar to repeated words, suggesting they were not fully suppressed. We hypothesized that, if similar processes also influence end-state recognition responses, these predicted but unobserved lures would show higher false alarms than new items. Furthermore, we used the N400 and LPC to help clarify the neurocognitive mechanisms involved in prediction-based false alarms vs. correct rejections of lures, focusing on whether these responses were associated with priming and/or recollection during the recognition test.

MATERIALS AND METHODS

Participants

Thirty-three right-handed, native speakers of English with normal or corrected-to-normal vision from the University of Illinois, Urbana-Champaign participated in the experiment and were paid $10 an hour or received course credit for their participation. No participant had a history of neuropsychological or psychiatric disorders. Procedures were approved by the IRB of the University of Illinois, and all participants signed consent forms prior to participation. Based on previous work using these same materials to examine ERPs during sentence comprehension (Federmeier et al., 2007), the *a priori* number of subjects was set to 32; mid-way through data collection, a participant's recorded data was noisy, and thus an extra subject was run. Data analysis led to the removal of another subject's data due to high trial loss, leading to a sample size of 31 participants in the final analyses.

Materials

The stimuli were comprised of 192 English sentences, a subset of the sentences used in Federmeier et al. (2007). The cloze probabilities of the endings of the sentences were previously determined in a norming study in which the subjects filled in the final word of the sentence frame with the word they "would generally expect to find completing the sentence fragment." In the current experiment, half of the stimuli (96 sentences) were strongly constraining, while the other half were weakly constraining. A sentence was considered strongly constraining if the cloze probability of the most commonly completed word was 0.68 or higher, and was considered weakly constraining if the cloze probability was 0.42 or lower. Additionally, half of the strongly constraining sentences (48 sentences) ended with the expected word, while the other half ended with an unexpected word; this was also true of the weakly constraining sentences. Unexpected words all had a cloze probability close to 0 (max = 0.088). Thus, participants read 48 strongly constraining sentences with expected endings (SCE; mean cloze = 0.83), 48 with unexpected endings (SCU), 48 weakly constraining sentences with expected endings (WCE; mean cloze = 0.28), and 48 with unexpected endings (WCU). These stimuli were evenly split into eight blocks (six of each condition in each block). **Table 1** provides the lexical properties (word frequency, concreteness, imageability, familiarity, word length) of the sentence ending words. Target words averaged 5–6 letters

TABLE 1 | Lexical properties of sentence ending words.

Condition	Frequency	Concreteness	Imageability	Familiarity	Word length
SCE	4.16	506.10	527.31	574.00	4.98
SCU	3.06	502.45	520.82	554.67	6.29
WCE	3.96	501.10	518.38	575.94	5.52
WCU	3.22	485.83	520.19	554.027	5.85

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

TABLE 2 | Examples of experimental materials.

SC	Tim threw a rock and broke the	E	window	(Match)	window
		U	camera	(Match)	camera
				(Lure)	window
WC	His ring fell into a hole in the	E	sink	(Match)	sink
		U	couch	(Match)	couch
				(Lure)	sink

SC, strong constraint; WC, weak constraint; E, expected; U, unexpected. Match and Lure refer to the items that appear during the memory test.

TABLE 3 | Lexical properties of test words.

Condition	Frequency	Concreteness	Imageability	Familiarity	Word length
SCE Match	4.21	526.58	539.95	573.63	5.04
SCU Match	2.94	477.00	503.84	546.33	6.21
SC Lure	4.13	507.45	527.86	571.77	5.29
WCE Match	4.04	503.87	525.00	582.44	5.38
WCU Match	2.99	476.22	509.67	552.94	6.08
WC Lure	3.43	509.81	527.18	580.94	5.29
New	3.70	497.77	525.31	560.15	5.77

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

in length and were fairly concrete, imageable, and familiar; unexpected endings tended to be of lower frequency on average than expected endings but were similar across constraint.

After each block of sentence reading, participants took a memory test. For the recognition memory test, participants were presented with single words, the majority of which were words that had ended the previously read sentences. "Matches" were words that had previously been seen as sentence endings (either expected or unexpected). "Lures" were words that might have been expected (they were the most likely completion of a sentence from the prior block) but were never actually presented (because the sentence had ended with an unexpected word instead). As an example of a "Lure" item, during the encoding phase a participant might read the sentence "I added my name to the basket," where basket is an unexpected ending, and in the test phase read the word "list," the expected ending of the sentence. "New" words had never been presented in the block. The test also contained some sentence-medial words, to ensure that participants would be motivated to pay attention to and encode the entire sentence. Half of the test items that were previously sentence ending words were from strongly constraining sentences while the other half were from weakly constraining sentences. **Table 2** provides examples of the different types of test items. There was an equal number of items presented in each of the conditions during each test block (totaling 24 in each condition, along with 48 new items, over the course of the experiment), as

well as an equal number of "old" and "new" items each test, so as not to bias responses. As with the sentence endings, lexical properties of the test items (see **Table 3** for details) were similar across conditions, with some variation in frequency; we aimed to test the impact of the frequency variability in our statistical models.

The memory test constrained the stimuli used and the order of presentation, in that each test item had to be unique, as well as not repeat. For example, participants might read the sentence "he played with the dog," see the word "dog" during the memory test, and later read the sentence "the dog ate the food." To avoid participants reading both sentences before being tested, the stimulus list was pseudo-randomized, such that 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 in an electrically shielded EEG recording booth approximately 100 cm from a CRT computer monitor. Prior to starting the experiment, we verified that all participants could easily read the presented information from this distance. Additionally, participants were given an explanation of the experimental procedure, as well as a short practice session to familiarize them with the task. Words that appeared in the

practice sentences and test items did not appear as critical test words in the actual experiment.

The experiment was divided into eight study-test blocks, in which participants first studied a set of sentences, and then were tested on their memory for critical words. During the encoding phase of each study-test block, participants were instructed to read the sentences silently and to try to remember what they read, as their memory would be tested. Sentences were presented word by word on the screen, with each word appearing in the center of the screen for 200 ms, followed by a 300 ms interstimulus interval. After the last word of the sentence was presented, a blank screen was presented for 500 ms, followed by a fixation cross for 1,000 ms. 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 the encoding phase, participants were given math problems to complete for 30 s. The math problems were simply given as a distractor between the study and test phases—thus, performance on the math section was not analyzed.

After the math section, participants started the test phase of the block. Each trial began with a fixation cross in the center of the screen for 1,000 ms, which was then replaced by a test word. After 1,000 ms, 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 four points—"Sure New," "Maybe New," "Maybe Old," and "Sure Old." Participants were instructed to respond with "Old" if they thought the test word was a word they had seen during the encoding phase and otherwise to respond "New." Additionally, they were told to try to use the whole scale of confidence and to use the "Maybe" option if they felt like they were guessing or unsure. Finally, participants were instructed to try not to blink during the initial presentation of the word, but told that once the confidence scale appeared and they could make their response, as well as during the fixation cross, they could blink. The test phase was self-paced, in that participants could take as long as they needed to respond.

EEG Recording and Processing

EEG data were recorded from 26 Ag/AgCl electrodes embedded into a flexible elastic cap and distributed over the scalp in an equidistant arrangement; see icon in **Figure 2**. Five additional electrodes were attached, including one on each mastoid bone behind the ear, one adjacent to the outer canthus of each eye, used for monitoring of the horizontal electro-oculogram (EOG), and one below the lower eyelid of the left eye, used for monitoring of blinks. Electrode impedances were kept below 5 kΩ. Signals were amplified by a BrainVision amplifier with a 16-bit A/D converter, an input impedance of 10 MΩ, a bandpass filter of 0.016–100 Hz, and a sampling rate of 1 kHz. The left mastoid electrode was used as a reference for on-line recording; offline, the average of the left and right mastoid electrodes was used as a reference.

Following data collection, each raw EEG time series was passed through a 0.1–30 Hz Butterworth filter with a 12 dB/oct roll-off. The signal was segmented into epochs from −200 to 1,000 ms relative to the onset of each sentence ending word

during encoding and each test item during the test phase. Following subtraction of the 200 ms prestimulus baseline, and artifact correction (described below), epochs within each bin were averaged together to create an ERP for each subject and bin. Prior to calculating statistics, individual subject ERPs were passed through an additional 20 Hz lowpass filter.

To correct for ocular artifacts, a bipolar VEOG channel was created by subtracting data in the lower eye channel from the most frontocentral channel (MiPf), and that channel was then scanned with a sliding window step function to detect blinks. For subjects who had a large number of blinks, the data were run through AMICA (Palmer et al., 2011), an ICA decomposition algorithm that generalizes Infomax and multiple mixtures approaches adaptively. Following decomposition, the correlation between the timecourse of each component and the VEOG channel was calculated in order to find the component(s) containing blinks. Components with a high correlation were removed from trials marked as containing blinks. The remaining components were then recombined to reconstruct the EEG data, which were then scanned with an additional sliding window amplitude threshold (300 ms sliding time window, 50 ms step size, 90 μV threshold), and finally manually checked by the experimenter for any additional artifacts. In total, an average of 8% of trials were removed, with a range of 2% to 11% across participants. Artifacts were spread fairly evenly across conditions, resulting in an average of 22 trials in each condition of the memory test.

For the ERP analyses, statistical analyses were performed on channel clusters as opposed to single channels to improve the signal to noise ratio. Component-based analyses were done using the signal-averaged across channel clusters and time windows based on prior work: N400 at a central cluster (shown in **Figure 2**; Federmeier et al., 2007), 300–500 ms; frontal positivity at a frontal cluster, 700–1,000 ms (shown in **Figure 2**; DeLong et al., 2014); LPC at a left parietal cluster (shown in **Figure 3**; Woroch and Gonsalves, 2010; Addante et al., 2012), 500–800 ms. For other effects, as described in the results, cluster-based permutations with restricted time windows were used in order to explore the data while retaining statistical power and maintaining Type I error rate (Fields and Kuperberg, 2018). Plotted ERPs were filtered with a 10 Hz lowpass filter for clarity of visualization.

RESULTS

Behavior

Proportion "Old" responses is plotted in **Figure 1**. For Matches, "Old" was a correct response, whereas for New items and Lures, "Old" was an incorrect response. Analyses revealed no differences in confidence across experimental conditions and generally low trial numbers for "Maybe" responses; thus, "Maybe" responses were combined with "Sure" responses for behavioral and ERP analyses. Overall, participants successfully discriminated Matches from New items. Collapsing across Match conditions and comparing to New items, the average d' was
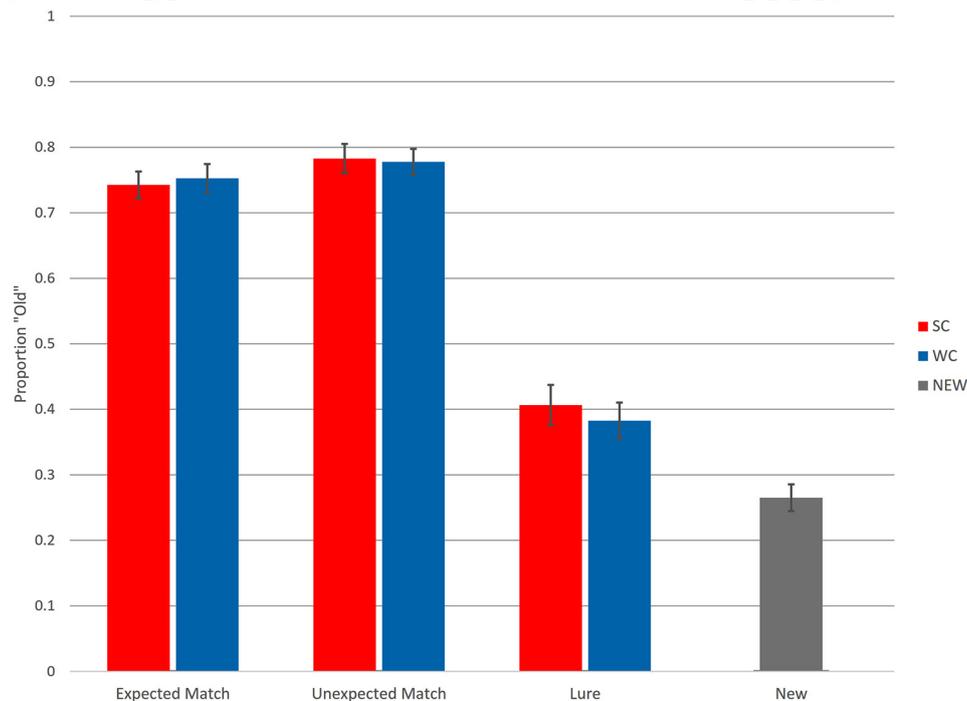
FIGURE 1 | Recognition memory accuracy. Proportion "Old" responses are plotted on the Y axis. SC, strong constraint; W, weak constraint. Error bars reflect standard error around the mean.

1.41, with a range of 0.72–2.76. Recognition accuracy between Expected and Unexpected Matches appeared similar, whereas participants false alarmed more to Lures compared to New items.

To assess the pattern statistically, behavioral responses (Old or New) on each trial were submitted to a mixed-effects logistic regression model fit by maximum likelihood using the lme4 package in R (Jaeger, 2008). Random factors included intercepts for items and slopes and intercepts for participants for each fixed effect. Correlations between random factors were not calculated to ease convergence of the models. Wald's z-scores were computed for each coefficient to test for significance.

The first model compared responses to Lures with responses to New items by modeling responses to those items with Condition (Lures, New) as a fixed factor. Recognition accuracy differed between Lures and New items ($\beta = 0.75$, $z = 3.18$, $p < 0.01$), but accuracy did not differ between Strong Constraint Lures and Weak Constraint Lures ($\beta = 0.11$, $z = 0.42$, $p = 0.68$). Thus, participants showed greater false alarms to Lures compared to New items.

Although we attempted to control the lexical properties of stimuli, it could be the case that a subset of the Lures were more frequent than other Lures or New items, and this could have contributed to the false alarm effect. To assess this, a second model was fit with Condition and log-transformed Word Frequency as fixed effects. Frequency had a significant effect on responses ($\beta = 0.30$, $z = 4.72$, $p < 0.01$), with higher Frequency leading to a greater number of "Old" responses, but recognition

accuracy still differed between Lures and New items ($\beta = 0.64$, $z = 2.91$, $p < 0.01$). Thus, word frequency did not completely explain the false alarm effect that we observed.

The next model assessed responses for Matches by modeling responses with Constraint, Expectedness, and the interaction (Constraint * Expectedness) as fixed factors. None of the coefficients, Constraint ($\beta < 0.01$, $z = 0.06$, $p = 0.95$), Expectedness ($\beta = 0.21$, $z = 1.26$, $p = 0.21$), or the interaction ($\beta = 0.13$, $z = 0.27$, $p = 0.79$), returned significant z-scores. Including word frequency in the model (C*E*F) did not change previous results, although word frequency seemed to have a tendency to reduce "Old" responses ($\beta = 0.10$, $z = -1.94$, $p = 0.05$). Thus, behavioral accuracy for Match items did not differ based on constraint or expectedness.

Sentence Final Word ERPs

ERPs to sentence final words were analyzed to determine if prior effects seen with these materials (e.g., Federmeier et al., 2007) were replicated. Grand average ERPs at the sentence final word at the frontal and central cluster are plotted in **Figure 2**. To assess effects statistically, linear mixed-effects models were used (Baayen et al., 2008), using the lme4 and lmerTest packages in R. Random factors included intercepts for items and slopes and intercepts for participants. As with the behavioral analyses, correlations between random factors were not calculated to ease convergence of the models. The reported t-tests used the Satterthwaite approximations to calculate degrees of freedom (Satterthwaite, 1946).
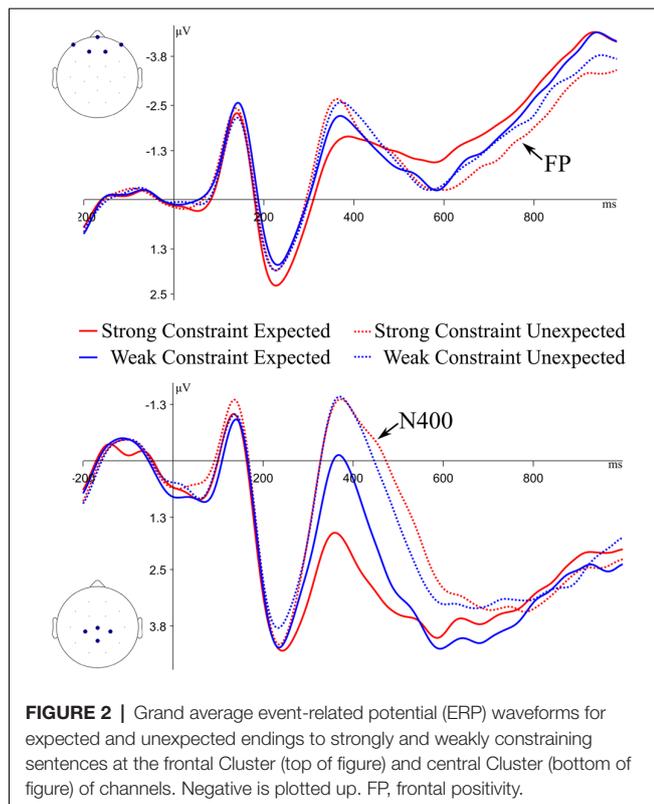
FIGURE 2 | Grand average event-related potential (ERP) waveforms for expected and unexpected endings to strongly and weakly constraining sentences at the frontal Cluster (top of figure) and central Cluster (bottom of figure) of channels. Negative is plotted up. FP, frontal positivity.

FIGURE 3 | Grand average ERP waveforms to Match items during the memory test. ERPs plotted at the Central Cluster (top of figure) and the Left Parietal Cluster (bottom of figure). SC, strong constraint ("Tim threw a rock and broke the. . ."); WC, weak constraint ("His ring fell into a hole in the. . ."); E, expected; U, unexpected. In quotations are example stimuli, based on examples from **Table 1**.

N400 amplitudes were compared between weakly constrained expected (WCE) endings and strongly constrained expected (SCE) endings, as well as between WCE and unexpected (U) endings (collapsed across constraint, as this has repeatedly been shown not to affect N400 responses). There were significant differences in N400 amplitude between WCE and SCE endings ($\beta = 1.32$, $t = 2.87$, $p < 0.01$), as well as between WCE and U endings ($\beta = 1.70$, $t = 4.42$, $p < 0.01$). Thus, the graded N400 effect was replicated in this experiment.

ERPs to sentence final words were also analyzed to determine if the frontal positivity to Strong Constraint Unexpected endings was replicated. The frontal positivity has been operationalized as a difference between Strong Constraint Unexpected (SCU) and Weak Constraint Unexpected (WCU) endings (Federmeier et al., 2007), or a difference between expected (E) endings and SCU endings (DeLong et al., 2014), so both of these differences were tested. There were no significant differences in frontal positivity amplitudes between the SCU and WCU conditions ($\beta = 0.41$, $t = 0.95$, $p = 0.35$); however, SCU endings elicited larger positivities than E endings ($\beta = 0.84$, $t = 2.01$, $p = 0.05$). A follow-up comparison of WCU and E conditions showed no significant differences ($\beta = 0.42$, $t = -1.04$, $p = 0.31$). Thus, the frontal positivity from SCU endings was more positive than other conditions, replicating prior work, but did not differ significantly from the WCU condition.

Recognition Memory ERPs: Matches

ERPs to correctly recognized test items were analyzed to assess recognition memory processes. The grand average ERPs at the central cluster to expected and unexpected Matches from strongly and weakly constraining sentences are plotted in **Figure 3**. ERPs are time-locked to the onset of the test item, and only correct responses are included.

LPC mean amplitudes from 500 to 800 ms at the Left Parietal cluster were submitted to a linear mixed effect model with fixed effects of Expectancy (E vs. U) and Constraint (SC vs. WC). The fixed effect of Expectancy was significant ($\beta = 1.14$, $t = 2.53$, $p = 0.01$), whereas Constraint ($\beta = 0.51$, $t = -0.97$, $p = 0.34$) and the interaction ($\beta = 0.80$, $t = 0.94$, $p = 0.35$) were not. A follow-up comparison of LPCs from SCE Matches and WCE Matches trended toward significance ($\beta = 0.94$, $t = -1.93$, $p = 0.06$). Unexpected Matches generated more positive LPC amplitudes compared to Expected Matches, and SCE Matches generated the smallest LPC amplitudes.

Visual inspection suggested an additional effect on the N1, a component that is part of the visual evoked potential and is sensitive to attention (Mangun and Hillyard, 1991). To assess this effect, we performed a *post hoc* exploratory analysis, using a cluster-based permutation test with a restricted time window based on previous literature to increase statistical power (Maris and Oostenveld, 2007; Groppe et al., 2011; Fields and Kuperberg, 2018). In this test, *t*-tests were calculated at each time-point and channel, and significant *t*-values that were adjacent in space and time were clustered together. Clusters were

characterized by taking the sum of *t*-values within the adjacent points. These observed clusters were compared to a permutation distribution, generated by shuffling the condition labels of the data, finding clusters, and summing the *t*-values of the clusters 2,500 times. Distributions of the most extreme cluster sums were created for comparison to the observed cluster sums. Reported *p*-values represent the percentile ranking of the observed clusters compared to the permutation distribution. Here, *t*-tests tested differences between Expected Matches and Unexpected Matches at each channel and time-point within the 50–175 ms window, and a family-wise alpha of 0.05 was used.

The results of this analysis are displayed in **Figure 4**. A significant difference between Expected and Unexpected Matches was found (cluster-wise $p < 0.05$). This difference had a temporal extent from 81 to 153 ms and a central-posterior topography, similar to previously reported posterior visual N1 effects, though somewhat earlier in time (Di Russo et al., 2001; Hopf et al., 2002). Thus, Unexpected Matches elicited more negative N1 potentials compared to Expected Matches[1]. To test for the possibility of pre-stimulus activity leading to the appearance of an N1 effect, an addition permutation test was run on the same contrast in the 0–80 ms time window. No significant clusters were found ($p = 0.29$).

Recognition Memory ERPs: Lures

Of particular interest for the analysis of ERPs to Lures was if ERPs differed between false alarms and correct rejections, and whether this ERP difference was affected by constraint. However, few studies have investigated ERP differences to false alarms and correct rejections, particularly for previously predicted information. Thus, while we were interested in early vs. late differences, there were not *a priori* predictions about particular ERP components to target in the post-N400 time window. We thus used time-constrained permutation tests, as described for the N1 analyses (Fields and Kuperberg, 2018). ERPs to SC and WC Lures were separated into Correct and Incorrect bins based on the response given (pooled across "Maybe" and "Sure"), and the difference between these ERPs was calculated. These difference waves were submitted to cluster-based permutation tests to test time-points for significant differences from 0, using a family-wise alpha value of 0.05. Separate permutation tests were run for Strong Constraint and Weak Constraint lures, and to increase statistical power and focus on times of interest, separate permutation tests were run for time windows of 300–500 ms (N400) and 500–1,000 ms.

Results of the permutation tests and ERPs are plotted in **Figure 5**. For the Strong Constraint Lure comparison, a significant difference (cluster-wise $p = 0.04$) between false alarms and correct rejections was found in the 300–500 ms time window, while no significant differences were found in the late

window. This difference began from the onset of the analysis window and continued to 488 ms, with a central-posterior topography. For the Weak Constraint Lure comparison, a significant difference (cluster-wise $p < 0.01$) between false alarms and correct rejections was found in the late time window, while no significant differences were found in the earlier window. This cluster showed a broad right-lateralized topography, with a right frontal maxima, and a temporal extent of 594–1,000 ms. These results suggest that mechanisms with different timecourses led to false alarming based on the constraint of the item[2].

The behavioral effect of interest was the comparison of false alarm rates of Lure items compared to false alarm rates of New items; therefore, we were also interested in how the electrophysiological differences associated with false alarming to Lures compared to those associated with false alarming to New items. **Figure 6** plots correct rejection and false alarm ERPs for Weak Constraint Lures as well as New items; although the ERPs at the same channel as before are plotted, the ERP patterns between these conditions were fairly similar across other channels as well. Permutation tests testing for differences between correct rejections and false alarm ERPs to New items in both the 300–500 ms and 500–1,000 ms windows were not significant (early, $p = 0.09$; late, $p = 0.11$), but numerically, false alarming to Weak Constraint Lures seemed to have engaged similar neurocognitive processes as false alarming to New items.

DISCUSSION

In this study, participants read strong and weak constraint sentences that ended with either an expected or unexpected-but-plausible word and then were tested on their memory for sentence ending words, new words, and predictable endings that had never been seen (lures). ERP responses during sentence reading replicated previously shown effects. We observed a graded N400 pattern (Federmeier et al., 2007), such that N400s were smallest to expected items in strong constraint sentences, intermediate to expected items in weak constraint sentences, and largest to unexpected items. We also found a post-N400 frontal positivity, larger for unexpected than expected words and numerically largest for unexpected words in strongly constraining sentences (where predictions can be correspondingly stronger). Different from the pattern in Federmeier et al. (2007), we did not observe a significant difference between unexpected items in strongly and weakly constraining contexts, seemingly because there was also some level of frontal positivity for the unexpected items in the weakly constraining sentences. It is possible that the memory task induced different reading strategies than the passive comprehension task in Federmeier et al. (2007). For example, Brothers et al. (2017) reported a larger frontal positivity to

---

[1]A mixed effect analysis was also run on single trial N1 amplitudes derived from significant cluster timepoints and channels, with fixed effects of expectancy and word frequency, to control for lexical confounds. The effect of expectancy was significant ($\beta = 0.21$, $t = 3.41$, $p < 0.01$), while frequency was not ($\beta = 0.03$, $t = 1.53$, $p = 0.13$). However, since estimates were derived based on cluster analyses, this mixed effect analysis could be considered double-dipping, and further replication of this effect will be necessary.

[2]Mixed effect analyses were also run on single trial SC and WC lure amplitudes derived from significant clusters, with fixed effects of correct/incorrect and word frequency. For both analyses, the fixed effect of correctness was significant (SC: $\beta = 0.36$, $t = 2.64$, $p = 0.01$; WC: $\beta = 0.87$, $t = 3.61$, $p < 0.01$), while word frequency was not significant (SC: $\beta = 0.09$, $t = 1.45$, $p = 0.16$; WC: $\beta = 0.05$, $t = 0.72$, $p = 0.48$). As with the N1 effect, these results could be considered double dipping and replication will be necessary.

FIGURE 4 | Permutation test results and ERP plots for analyses of N1 recognition memory effect. The raster plot show channels and time-points which make up the significant cluster found in the permutation tests. Colors represent the *t*-value at the time-point. The ERP topography plot shows the mean amplitude in the time window of the significant cluster, with significant channels highlighted in white. The ERP plot shows the Expected and Unexpected Match ERPs at the channel with the largest *t*-value within the cluster (MiOc). The black dashed lines indicate the time range of the permutation test.

unexpected words when participants were instructed to predict upcoming information compared to when they simply read for comprehension. Anticipating an imminent memory test may have encouraged participants to read more attentively and devote more resources to prediction.

The central question for this study concerned participants' later memory for sentence-ending words they had predicted and/or read. Behaviorally, hit rates were numerically higher for unexpected than for expected matches, though no reliable effect was found. A similar pattern had previously been seen for word recognition at the end of the experiment using these stimuli; higher hit rates were also found for expected words that had completed weakly vs. strongly constraining sentences (Federmeier et al., 2007; see also Corley et al., 2007). The ERPs during the memory test in the present study, however, revealed that LPC responses elicited by unexpected Matches were more positive than those to expected Matches, suggesting greater recollection for unexpected words. Additionally, LPC amplitudes differed between strongly and weakly constrained expected matches, with more positive LPCs for weak constraint matches. This LPC pattern mirrors the behavioral memory performance pattern observed in Federmeier et al. (2007). This pattern may arise because prediction trades off with depth of encoding, such that participants process—and hence encode—predicted words

less. In other words, the information needed to verify that an expectation is met may require less attention and less stimulus-driven processing than that needed to encode a stimulus that readers could not predict. A recent ERP repetition study supports this account (Rommers and Federmeier, 2018a). Words that had first been encountered as expected sentence endings of strongly constraining sentences showed reduced ERP repetition effects (when seen again in a weakly constraining sentence) compared to those that had first been seen in weakly constraining sentences. Thus, predictability may have downstream costs: when information is pre-activated, comprehension may take place in a top-down "verification mode" (Van Berkum, 2010), in which readers need only confirm that the stimulus matches with the expectation. This process achieves speedier processing in the moment by sacrificing thorough processing of the bottom-up input, ultimately leading to impoverished representations. Future studies investigating memory for predicted information could assess this further by examining ERPs for misses or incorrect responses, as trial numbers were too low to assess misses here.

Surprisingly, unexpected matches also elicited larger (more negative) N1 amplitudes than did expected matches. N1 amplitude modulations are not routinely reported in electrophysiological studies of recognition memory. Although unexpected sentence endings may have received greater depth of

FIGURE 5 | Results and ERP plots for analyses of Lures. The top half **(A)** focuses on Strong Constraint Lures, with a time window of 300–500 ms, whereas the bottom half **(B)** focuses on Weak Constraint Lures, with a time window of 500–1,000 ms. The raster plots show channels and time-points which make up the significant cluster found in the permutation tests. Colors represent the *t*-value at the time-point. The ERP topography plots show the mean amplitude in the time window of the significant cluster, with significant channels highlighted in white. The ERP plots show the SC and WC Lures at the maximal channel within the observed cluster. The black dashed lines indicate the time range of the permutation tests.

FIGURE 6 | Grand average ERP waveforms for correct rejections and false alarms for Weak Constraint Lures and New items at the previously described maximal channel from the WC Lure cluster analysis. The pattern of responses for New items appears similar to the WC Lure items.

processing during encoding, ERP studies examining retrieval of words that were deeply or shallowly encoded have not reported modulations of the N1 (Rugg et al., 1998, 2000; Allan et al., 2000). However, N1 modulations have been observed in the context of visual attention and categorization. The N1 is sensitive to the allocation of attentional resources (Mangun and Hillyard, 1991; Hillyard and Anllo-Vento, 1998) and may reflect an early, attention-dependent visual discrimination process that is sensitive to category membership (Vogel and Luck, 2000; Hopf et al., 2002). In one study (Curran et al., 2002), participants were trained in separating abstract blob images into two separate categories (similar or dissimilar to a prototype) and were later given a recognition memory test on the images. The N1 during the recognition test was sensitive to category membership, but not to old/new differences, similar to the current reported
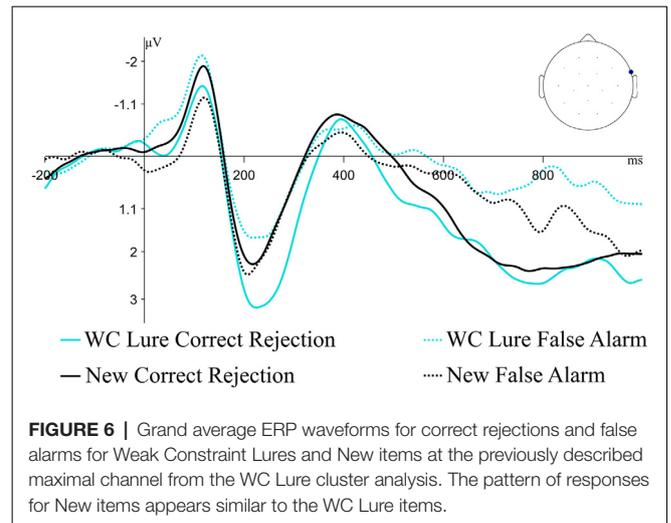
results. Differences in predictability during sentence reading may have led to separable categories during recognition testing; however, given the *post hoc* nature of the analysis of the N1 in the current study, it will be important to replicate the effect in future work, as well as to confirm that the results cannot be explained by other factors (such as lexical variables).

A critical manipulation in the current study was the inclusion of lures—items that were likely to have been predicted during sentence reading but that were never actually presented (because an unexpected word appeared instead). Behaviorally, individuals were significantly more likely to false alarm to Lures than to New items that had not been studied, suggesting increased accessibility or fluency for these items. This pattern is consistent with claims that words are predicted and pre-activated as a sentence unfolds (Federmeier, 2007; Kutas et al., 2011) and further reveals that such predictive pre-activation can have long-lasting effects. Here, several sentences were presented in each block, and each block was followed by interfering math problems, and yet participants still showed increased false alarming to these lures. This finding mirrors previously reported effects from studies on false memory using the DRM paradigm, in which subjects falsely recall—and are more likely to falsely recognize (Gallo, 2010)—critical lures that are semantically similar to studied items. However, a number of differences between the paradigms make the current findings particularly striking. First, in DRM experiments, the lure items are usually closely related to an entire list of words. Here, instead, each predicted sentence ending used as a Lure test item was related to only one sentence in a block, and the sentences were not semantically related to each other. Moreover, different from the DRM paradigm, in the present study predictions were explicitly disconfirmed, *via* the presentation of an unexpected word (which was always semantically unrelated to the predicted ending). Thus, these findings suggest that expected representations are not fully suppressed when a prediction is disconfirmed and that false memories can arise for such disconfirmed information. This presents another cost of prediction during language comprehension: individuals may falsely remember reading or

hearing words that were not actually experienced, simply because they were predicted in the moment, and those predictions linger.

An alternative explanation of the luring effect is that participants could have tried to use the word presented during the test as a cue to perform a retrospective search through memory for a sentence that might have included it. By this account, when a Lure was presented, subjects were able to retrieve a likely sentence frame for that word, and thus more false alarming occurred. Similar to Neely and Keefe's (1989) hybrid prospective-retrospective processing theory, this retroactive search could be performed regardless of any pre-activation of the test item. However, in the case of the Lures in the present study, the associated sentence was completed by an unexpected word. For a retroactive search strategy to work, the unexpected word that originally completed the sentence and its effect on the sentence-level meaning that was extracted would need to be ignored, thus rendering the Lures as ineffective search cues.

Behaviorally, participants did not show a greater rate of false alarms to lures from strongly constraining sentences compared to lures from weakly constraining sentences. However, electrophysiological analyses revealed that different underlying patterns of brain activity were associated with false alarming across constraint. False alarming to strong constraint lures correlated with an earlier, N400-like effect, whereas false alarming to weak constraint lures was associated with a later, right-lateralized effect that was fairly broadly distributed. The N400-like pattern to the lures from the strong constraint sentences is consistent with the idea that false alarms to these items were driven by an increase in conceptual fluency or familiarity (Voss and Federmeier, 2011; Wang et al., 2015). A plausible account of this effect is that when words or concepts are strongly predicted, they linger, such that when the word is encountered again, it is processed more fluently or is more familiar, which behaviorally is associated with a tendency to mark these words as "old" and electrophysiologically is associated with a reduced N400 response. The later right-lateralized effect observed following false alarms to weak constraint lures may be comparable to the right frontal old/new effect in the recognition memory literature, which is thought to index decision making, evaluation, and post-retrieval monitoring processes (Hayama et al., 2008; Cruse and Wilding, 2009; Hayama and Rugg, 2009) and has been related to lure discrimination (Morcom, 2015). Thus, despite a lack of behavioral differences in false alarming based on constraint, it appears different processes may have led to false alarms depending on the prior constraint of the item: a more rapid semantic matching based process for strong constraint lures and a slower, more top-down decision process for weak constraint lures. Future studies could use experimental manipulations to dissociate these effects; for instance, employing speeded recognition decisions would likely increase false alarm

rates for weak constraint lures, but might not affect strong constraint lures.

Overall, these results demonstrate that prediction during language comprehension has important downstream effects on recognition memory. Participants were more likely to false alarm to predictable, but never observed words compared to unexpected and unstudied words, suggesting unobserved predictions are not fully suppressed and remain accessible in memory. Individuals also had enhanced memory for unexpected information, as evidenced by larger LPC amplitudes during recognition testing, along with a larger N1 response. Finally, ERPs revealed sentential constraint-based differences in the neurocognitive mechanisms involved in false alarming to lures, with earlier semantic matching processes contributing to false alarms to strongly predicted information, but later decision-making processes contributing to false alarms to weakly predicted information. Ultimately, prediction during language comprehension does have costs: namely, predicting upcoming words in sentences can produce more rapid processing in the moment, but can lead to impoverished memory of predictable information and false remembering of unobserved predictions.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

Human Subject Research
The studies involving human participants were reviewed and approved by University of Illinois Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RH, JR, CJ, and KF contributed to the conception, design of the study and wrote the manuscript. RH and JR collected data. CJ created code for generating stimulus lists with non-repeating stimuli. RH performed the statistical analysis. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

REFERENCES

Addante, R. J., Ranganath, C., and Yonelinas, A. P. (2012). Examining ERP correlates of recognition memory: evidence of accurate source recognition without recollection. *Neuroimage* 62, 439–450. doi: 10.1016/j.neuroimage.2012. 04.031

Allan, K., Robb, W. G., and Rugg, M. D. (2000). The effect of encoding manipulations on neural correlates of episodic retrieval. *Neuropsychologia* 38, 1188–1205. doi: 10.1016/s0028-3932(00)00013-0

Altmann, G. T., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. doi: 10.1016/s0010-0277(99)00059-1

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* 11, 280–289. doi: 10.1016/j.tics.2007.05.005

Bar, M. (2009). The proactive brain: memory for predictions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1235–1243. doi: 10.1098/rstb.2008.0310

Beato, M. S., Boldini, A., and Cadavid, S. (2012). False memory and level of processing effect: an event-related potential study. *Neuroreport* 23, 804–808. doi: 10.1097/WNR.0b013e32835734de

Brothers, T., Swaab, T. Y., and Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *J. Mem. Lang.* 93, 203–216. doi: 10.1016/j.jml.2016.10.002

Bubic, A., von Cramon, D. Y., and Schubotz, R. I. (2010). Prediction, cognition and the brain. *Front. Hum. Neurosci.* 4:25. doi: 10.3389/fnhum.2010.00025

Chen, H., Voss, J. L., and Guo, C. (2012). Event-related brain potentials that distinguish false memory for events that occurred only seconds in the past. *Behav. Brain Funct.* 8:36. doi: 10.1186/1744-9081-8-36

Clark, A. (2013). Whatever next? Predictive brains, situated agents and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

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

Craik, F. I., Govoni, R., Naveh-Benjamin, M., and Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *J. Exp. Psychol. Gen.* 125, 159–180. doi: 10.1037/0096-3445.125.2.159

Craik, F. I., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verbal Learn. Verbal Behav.* 11, 671–684. doi: 10.1016/S0022-5371(72)80001-X

Craik, F. I., and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *J. Exp. Psychol. Gen.* 104, 268–294. doi: 10.1037/0096-3445.104.3.268

Cruse, D., and Wilding, E. L. (2009). Prefrontal cortex contributions to episodic retrieval monitoring and evaluation. *Neuropsychologia* 47, 2779–2789. doi: 10.1016/j.neuropsychologia.2009.06.003

Curran, T. (2000). Brain potentials of recollection and familiarity. *Mem. Cognit.* 28, 923–938. doi: 10.3758/bf03209340

Curran, T. (2004). Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia* 42, 1088–1106. doi: 10.1016/j.neuropsychologia.2003.12.011

Curran, T., Schacter, D. L., Johnson, M. K., and Spinks, R. (2001). Brain potentials reflect behavioral differences in true and false recognition. *J. Cogn. Neurosci.* 13, 201–216. doi: 10.1162/089892901564261

Curran, T., Tanaka, J. W., and Weiskopf, D. M. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cogn. Affect. Behav. Neurosci.* 2, 1–18. doi: 10.3758/cabn.2.1.1

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. Exp. Psychol.* 58, 17–22. doi: 10.1037/h0046671

DeLong, K. A., Quante, L., and Kutas, M. (2014). Predictability, plausibility and two late ERP positivities during written sentence comprehension. *Neuropsychologia* 61, 150–162. doi: 10.1016/j.neuropsychologia.2014.06.016

DeLong, K. A., Urbach, T. P., Groppe, D. M., and Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology* 48, 1203–1207. doi: 10.1111/j.1469-8986.2011.01199.x

Di Russo, F., Martínez, A., Sereno, M. I., Pitzalis, S., and Hillyard, S. A. (2001). Cortical sources of the early components of the visual evoked potential. *Hum. Brain Mapp.* 15, 95–111. doi: 10.1002/hbm.10010

Duffy, S. A., Henderson, J. M., and Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 791–801. doi: 10.1037/0278-7393.15.5.791

Düzel, E., Yonelinas, A. P., Mangun, G. R., Heinze, H. J., and Tulving, E. (1997). Event-related brain potential correlates of two states of conscious awareness in memory. *Proc. Natl. Acad. Sci. U S A* 94, 5973–5978. doi: 10.1073/pnas.94.11.5973

Ehrlich, S. F., and Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *J. Verbal Learn. Verbal Behav.* 20, 641–655. doi: 10.1016/s0022-5371(81)90220-6

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* 44, 491–505. doi: 10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., and Kutas, M. (1999). A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang.* 41, 469–495. doi: 10.1006/jmla.1999.2660

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., and Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Res.* 1146, 75–84. doi: 10.1016/j.brainres.2006.06.101

Fields, E. C., and Kuperberg, G. R. (2018). Having your cake and eating it too: flexibility and power with mass univariate statistics for ERP data. *PsyArXiv* [Preprint]. doi: 10.31234/osf.io/qfkgc

Fischler, I., and Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *J. Verbal Learn. Verbal Behav.* 18, 1–20. doi: 10.1016/s0022-5371(79)90534-6

Frisson, S., Harvey, D. R., and Staub, A. (2017). No prediction error cost in reading: evidence from eye movements. *J. Mem. Lang.* 95, 200–214. doi: 10.1016/j.jml.2017.04.007

Frisson, S., Rayner, K., and Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 862–877. doi: 10.1037/0278-7393.31.5.862

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Mem. Cognit.* 38, 833–848. doi: 10.3758/MC.38.7.833

Geng, H., Qi, Y., Li, Y., Fan, S., Wu, Y., and Zhu, Y. (2007). Neurophysiological correlates of memory illusion in both encoding and retrieval phases. *Brain Res.* 1136, 154–168. doi: 10.1016/j.brainres.2006.12.027

Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48, 1711–1725. doi: 10.1111/j.1469-8986.2011.01273.x

Hartman, M., and Hasher, L. (1991). Aging and suppression: memory for previously relevant information. *Psychol. Aging* 6, 587–594. doi: 10.1037/0882-7974.6.4.587

Hasher, L., Quig, M. B., and May, C. P. (1997). Inhibitory control over no-longer-relevant information: adult age differences. *Mem. Cognit.* 25, 286–295. doi: 10.3758/bf03211284

Hayama, H. R., Johnson, J. D., and Rugg, M. D. (2008). The relationship between the right frontal old/new ERP effect and post-retrieval monitoring: specific or non-specific? *Neuropsychologia* 46, 1211–1223. doi: 10.1016/j.neuropsychologia.2007.11.021

Hayama, H. R., and Rugg, M. D. (2009). Right dorsolateral prefrontal cortex is engaged during post-retrieval processing of both episodic and semantic information. *Neuropsychologia* 47, 2409–2416. doi: 10.1016/j.neuropsychologia.2009.04.010

Hess, D. J., Foss, D. J., and Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *J. Exp. Psychol. Gen.* 124, 62–82. doi: 10.1037/0096-3445.124.1.62

Hillyard, S. A., and Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proc. Natl. Acad. Sci. U S A* 95, 781–787. doi: 10.1073/pnas.95.3.781

Hopf, J. M., Vogel, E., Woodman, G., Heinze, H. J., and Luck, S. J. (2002). Localizing visual discrimination processes in time and space. *J. Neurophysiol.* 88, 2088–2095. doi: 10.1152/jn.2002.88.4.2088

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Lang. Linguist Compass* 2, 647–670. doi: 10.1111/j.1749-818x.2008.00072.x

Kučera, H., and Francis, W. N. (1967). *Computational Analysis of Present-Day American English.* Providence, RI: Brown University Press.

Kuperberg, G. R., and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* 31, 32–59. doi: 10.1080/23273798.2015.1102299

Kutas, M., DeLong, K. A., and Smith, N. J. (2011). "A look around at what lies ahead: prediction and predictability in language processing," in *Predictions in the Brain: Using Our Past to Generate a Future*, ed. M. Bar (Oxford: Oxford University Press), 190–207.

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163. doi: 10.1038/307161a0

Lorsbach, T. C., Wilson, S., and Reimer, J. F. (1996). Memory for relevant and irrelevant information: evidence for deficient inhibitory processes in language/learning disabled children. *Contemp. Educ. Psychol.* 21, 447–466. doi: 10.1006/ceps.1996.0030

Luke, S. G., and Christianson, K. (2016). Limits on lexical prediction during reading. *Cogn. Psychol.* 88, 22–60. doi: 10.1016/j.cogpsych.2016.06.002

Mangun, G. R., and Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 1057–1074. doi: 10.1037/0096-1523.17.4.1057

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Morcom, A. M. (2015). Resisting false recognition: an ERP study of lure discrimination. *Brain Res.* 1624, 336–348. doi: 10.1016/j.brainres.2015.07.049

Neely, J. H., and Keefe, D. E. (1989). Semantic context effects on visual word processing: a hybrid prospective-retrospective processing theory. *Psychol. Learn. Motiv.* 24, 207–248. doi: 10.1016/s0079-7421(08)60538-1

Otten, M., and Van Berkum, J. J. (2008). Discourse-based word anticipation during language processing: prediction or priming? *Discourse Processes* 45, 464–496. doi: 10.1080/01638530802356463

Paller, K. A., and Kutas, M. (1992). Brain potentials during memory retrieval provide neurophysiological support for the distinction between conscious recollection and priming. *J. Cogn. Neurosci.* 4, 375–392. doi: 10.1162/jocn.1992.4.4.375

Palmer, J. A., Kreutz-Delgado, K., and Makeig, S. (2011). *AMICA: An Adaptive Mixture of Independent Component Analyzers with Shared Components.* San Diego, CA: Technical Report, Swartz Center for Computational Neuroscience.

Roediger, H. L., and McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 803–814. doi: 10.1037/0278-7393.21.4.803

Röer, J. P., Bell, R., Körner, U., and Buchner, A. (2019). A semantic mismatch effect on serial recall: evidence for interlexical processing of irrelevant speech. *J. Exp. Psychol. Learn. Mem. Cogn.* 45, 515–525. doi: 10.1037/xlm0000596

Rommers, J., and Federmeier, K. D. (2018a). Predictability's aftermath: downstream consequences of word predictability as revealed by repetition effects. *Cortex* 101, 16–30. doi: 10.1016/j.cortex.2017.12.018

Rommers, J., and Federmeier, K. D. (2018b). Lingering expectations: a pseudo-repetition effect for words previously expected but not presented. *Neuroimage* 183, 263–272. doi: 10.1016/j.neuroimage.2018.08.023

Rugg, M. D., Allan, K., and Birch, C. S. (2000). Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. *J. Cogn. Neurosci.* 12, 664–678. doi: 10.1162/089892900562291

Rugg, M. D., and Curran, T. (2007). Event-related potentials and recognition memory. *Trends Cogn. Sci.* 11, 251–257. doi: 10.1016/j.tics.2007.04.004

Rugg, M., Mark, R., Walla, P., Schloerscheidt, A., Birch, C., and Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature* 392, 595–598. doi: 10.1038/33396

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics* 2, 110–114. doi: 10.2307/3002019

Schuberth, R. E., and Eimas, P. D. (1977). Effects of context on the classification of words and nonwords. *J. Exp. Psychol. Hum. Percept. Perform.* 3, 27–36. doi: 10.1037/0096-1523.3.1.27

Schuberth, R. E., Spoehr, K. T., and Lane, D. M. (1981). Effects of stimulus and contextual information on the lexical decision process. *Mem. Cognit.* 9, 68–77. doi: 10.3758/bf03196952

Schwanenflugel, P. J., and LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 344–354. doi: 10.1037/0278-7393.14.2.344

Schwanenflugel, P. J., and Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *J. Mem. Lang.* 24, 232–252. doi: 10.1016/0749-596x(85)90026-9

Simpson, G. B., Peterson, R. R., Casteel, M. A., and Burgess, C. (1989). Lexical and sentence context effects in word recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 88–97. doi: 10.1037/0278-7393.15.1.88

Steffens, M. C., and Mecklenbräuker, S. (2007). False memories: phenomena, theories and implications. *Z. Psychol. J. Psychol.* 215, 12–24. doi: 10.1027/0044-3409.215.1.12

Thornhill, D. E., and Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: frontal positivity and N400 ERP components. *Int. J. Psychophysiol.* 83, 382–392. doi: 10.1016/j.ijpsycho.2011.12.007

Van Berkum, J. J. (2010). The brain is a prediction machine that cares about good and bad-any implications for neuropragmatics? *Ital. J. Linguist.* 22, 181–208. Available online at: http://linguistica.sns.it/RdL/2010.html

Vogel, E. K., and Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology* 37, 190–203. doi: 10.1017/s0048577200981265

Von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld. *Psychol. Forsch.* 18, 299–342. doi: 10.1007/bf02409636

Voss, J. L., and Federmeier, K. D. (2011). FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing. *Psychophysiology* 48, 532–546. doi: 10.1111/j.1469-8986.2010.01085.x

Wang, W., Li, B., Gao, C., Xiao, X., and Guo, C. (2015). Electrophysiological correlates associated with contributions of perceptual and conceptual fluency to familiarity. *Front. Hum. Neurosci.* 9:321. doi: 10.3389/fnhum.2015.00321

West, R. F., and Stanovich, K. E. (1978). Automatic contextual facilitation in readers of three ages. *Child Dev.* 49, 717–727. doi: 10.1111/j.1467-8624.1978.tb02373.x

Wills, A. J., Lavric, A., Croft, G. S., and Hodgson, T. L. (2007). Predictive learning, prediction errors, and attention: evidence from event-related potentials and eye tracking. *J. Cogn. Neurosci.* 19, 843–854. doi: 10.1162/jocn.2007.19.5.843

Wlotko, E. W., and Federmeier, K. D. (2007). Finding the right word: hemispheric asymmetries in the use of sentence context information. *Neuropsychologia* 45, 3001–3014. doi: 10.1016/j.neuropsychologia.2007.05.013

Wlotko, E. W., and Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology* 49, 770–785. doi: 10.1111/j.1469-8986.2012.01366.x

Wolk, D. A., Schacter, D. L., Lygizos, M., Sen, N. M., Holcomb, P. J., Daffner, K. R., et al. (2006). ERP correlates of recognition memory: effects of retention interval and false alarms. *Brain Res.* 1096, 148–162. doi: 10.1016/j.brainres.2006.04.050

Woodruff, C. C., Hayama, H. R., and Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Res.* 1100, 125–135. doi: 10.1016/j.brainres.2006.05.019

Woroch, B., and Gonsalves, B. D. (2010). Event-related potential correlates of item and source memory strength. *Brain Res.* 1317, 180–191. doi: 10.1016/j.brainres.2009.12.074

Yu, S. S., and Rugg, M. D. (2010). Dissociation of the electrophysiological correlates of familiarity strength and item repetition. *Brain Res.* 1320, 74–84. doi: 10.1016/j.brainres.2009.12.071