PRECURSORS AND DOWNSTREAM CONSEQUENCES OF PREDICTION IN LANGUAGE COMPREHENSION

BY

RYAN J. HUBBARD

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Kara D. Federmeier, Chair
Professor Neal J. Cohen
Professor Gary S. Dell
Associate Professor Lili Sahakyan
Assistant Professor Darren Tanner

ABSTRACT

During language comprehension, the brain rapidly integrates incoming linguistic stimuli to not only incrementally build a contextual representation, but also predict upcoming information. This predictive mechanism leads to behavioral facilitation of processing of expected words, as well as a reduction in amplitude of the N400, a neural response reflecting access of semantic memory. However, little research has identified a behavioral or neurophysiological cost of errors in prediction. Additionally, only recent work has begun to investigate neural activity related to prediction prior to encountering a predicted stimulus. Most work has focused on what happens immediately after a predicted or unpredicted stimulus is encountered. Here, I explore new avenues of research by examining downstream consequences of prediction during language comprehension on future recognition memory. Additionally, I test whether these consequences occur following any violation of predictions, or whether the semantic fit of the violation to the established context plays a role. Finally, I adapt a classic paradigm, word stem completion, to investigate electrophysiological activity following a cue that is modulated by how predictive the outcome is. With this work, I not only have discovered costs of failed and successful predictions and identified neural signals potentially related to generation of predictions, but also have researched prediction in novel ways that can continue to expand and further elucidate how this mechanism affects cognition and changes across populations.

ACKNOWLEDGEMENTS

My time as a graduate student led to great professional growth and allowed me to develop a clear goal of a career path as a scientist. It's been rewarding, certainly challenging at times, but clearly worth any hardship. I'd like to thank my advisor, Kara Federmeier, my committee, Neal Cohen, Gary Dell, Lili Sahakyan, and Darren Tanner, the faculty members who worked with me and instructed me, including Aaron Benjamin, Duane Watson, Sarah Brown-Schmidt, Diane Beck, Mark Hasegawa-Johnson, Paris Smaragdis, and others, as well as the numerous postdocs and graduate students I collaborated with and took courses with for teaching me things and sharpening my scientific acumen. It's been a real privilege to work with such great people, and I'm certain I would not know what I know and be able to do what I do without their input, advice, and collaboration.

But, more than that, graduate school was a time of great personal growth and change for me, which not only affected my work and made me a better scientist, but also made me a better person. For whatever reason, for much of my life I've viewed myself as some sort of nihilistic curmudgeon, holding the belief that relationships and interactions with others were enjoyable, but ultimately temporary and meaningless, and that I'd essentially make my way through life on my own. I thought I'd get to this point and acknowledge a few select people for their scientific contributions, but as I look back and think of the numerous individuals that took the time to work with me, to help me, to advise me, to be my friend, and to support me, I realize how meaningful the connections I have made have been, and how much they have impacted not just my career, but my life.

Of course I must thank my parents, Sandra and Geoff Hubbard, and my brother, Adam Hubbard. I'm so deeply appreciative of the support they've given me through undergraduate and graduate school. They never pushed me into anything, and allowed me to explore and find

what truly interested me. They've also kept me grounded in life outside of academia – it's easy to get stuck up in the ivory tower, so I appreciate them pulling me down every once in a while.

I'm thankful to the entirety of the Cognition and Brain Lab, for doing great work and allowing me to be a part of it, and for creating the most fun and enjoyable lab experience I've had. I thank Joost Rommers, who collaborated on the memory projects with me and took extra time to teach me and work with me on proper mixed effect modeling practices and time frequency analysis methods, and shared hikes and beers and silly phrases. I'm grateful to Brennan Payne, who imparted a magnificent amount of statistical knowledge and a magnificent amount of friendship (quad bass forever!). I had numerous discussions about experimental design, statistical practices, career paths, relationships, music, and all sorts of other things with Dan Kleinman, and I'm thankful for every one of them. I had a blast working with Jamie Norton and Cybelle Smith on machine learning techniques, not just because the work was fun, but those meetings felt like I was hanging out with two good friends of mine. To Mallory, Danielle, Michelle, Heather, Manoj, Melinh, Shukhan, Katie, Melissa – thank you all!

I'm enormously grateful and indebted to the numerous graduate students and postdocs who showed me friendship and kindness along the way. Ben, who I met my very first day as a graduate student, has been a best friend ever since, and will be a best friend forever onward. Cassie, who I can't thank enough for numerous scientific discussions in which I probably got too defensive, fantastic hikes and camping trips, snobby dissections of barrel aged stouts, tremendous amounts of curry and chili, and for emotional support during some of the most stressful times of my life. Pat, who was like a second advisor to me, an incredible friend to me, and the best roommate I've ever had. Natalie, with whom I had some of the most important and maturing discussions of life and relationships (mixed in with a good dose of silliness, of course). Payel, who I thank for countless philosophical discussions, and for giving me a deeper appreciation of art and friendship. Steph and Dan, who cooked me tons of dinners and went on

great trips with me – I only wish I'd known how awesome they were sooner (oh, and thanks to Princess and June too!). Carry, who I thank for video games, science talks, metal shows, political discussions, and general good friendship. Aki, Brion, Kailen, Nyssa, Erika, Amogh, Ryan, Leyshon, Mike, Metzger, Emma, Evan & Lexi, Sara, Rachel – thank you all for your invaluable friendship.

I'll conclude by thanking Kara again –this time for the generosity and kindness she showed me, and the wisdom and perspective she imparted to me. Kara has shown me that one can pursue a career in science and do well thought-out rigorous work, while still having balance and pursuing other life interests. I consider her a model scientist, and I will strive to bring the same joy and excitement, as well as compassion for others, that she has throughout my own career. I am also so appreciative of the friendship I have with Kara - it seems a rare thing for one's graduate advisor to become one's very good friend, and I consider myself quite lucky that this has been the case.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: BEHAVIORAL AND ELECTROPHYSIOLOGICAL CONSEQUENCES OF PREDICTION ON RECOGNITION MEMORY 25

CHAPTER 3: INVESTIGATING CONSEQUENCES OF PREDICTION ON RECOGNITION MEMORY FOR ANOMALOUS STIMULI 62

CHAPTER 4: ELECTROPHYSIOLOGICAL PRECURSORS OF PREDICTION 80

CHAPTER 5: CONCLUSION 114

REFERENCES 115

CHAPTER 1: INTRODUCTION

The process of prediction has been suggested to play a role in many areas of cognition and behavior, including visual processing (Cheung & Bar, 2012), auditory perception (Chennu et al., 2013), motion and motor control (Wolpert & Flanagan, 2001), and processing of emotions (Gilbert & Wilson, 2009). This generality across multiple domains has led many to posit that one of the core functions, if not the most important function, of the brain is to utilize previously learned associations and top-down control to predict future events (Bar, 2007, 2009; Bubic, von Cramon, & Schubotz, 2010; Clark, 2013). This function of predicting upcoming information may play a particularly important role in language comprehension (Federmeier, 2007; Kuperberg & Jaeger, 2016). While all language processing may not reflect prediction, and prediction may not be *necessary* for language processing to occur (Huettig & Mani, 2016), a substantial body of evidence suggests that prediction often influences language comprehension. Essentially, by using the bottom-up sensory information provided by written and spoken words, as well as previously learned contextual, semantic, and syntactic information, the brain quickly creates and updates a representation of upcoming linguistic information, which facilitates processing when this information is encountered. Here, I review previous seminal work on prediction in language processing, and provide new research examining mechanisms of predictive processing in the brain, and how this processing can impact downstream cognition.

Evidence for Prediction in Language Comprehension

Historically, the idea that prediction plays a role in language comprehension has not always been accepted, with some arguing that language input is far too noisy and unpredictable for prediction to be a worthwhile strategy (Forster, 1981). It is at least generally accepted that

language processing is incremental, in that meaning is accrued gradually, and that semantic associations and predictive sentence contexts can facilitate this processing (Altmann & Steedman, 1988; Marslen-Wilson & Tyler, 1980). Theories differ in explaining the mechanism by which this facilitation occurs. One explanation is *prediction*: a predictive sentence context allows comprehenders to generate or pre-activate upcoming information (Federmeier, 2007). The other is *integration*: expected words or semantically related words are more easily integrated into the currently built representation, but this information is not necessarily activated beforehand (Moss & Marslen-Wilson, 1993; Schwanenflugel & Stowe, 1989). One argument is that the state of the language system is affected by context even before bottom-up input, so at least some minimal prediction or pre-activation is involved (Kuperberg & Jaeger, 2016); however, the extent to which prediction is engaged remains an open debate. Here, I summarize multiple lines of research on contextual facilitation and integration, and argue that the overall body of work argues for a predictive account.

A substantial body of behavioral research has provided empirical data that suggests language comprehension can be facilitated by contextual information. For instance, reading aloud and lexical decisions of words is faster when these words are preceded by a congruous sentence context (Duffy, Henderson, & Morris, 1989; Hess, Foss, & Carroll, 1995; Kleiman, 1980; Masson, 1986; Schuberth, Spoehr, & Lane, 1981; Simpson, Peterson, Casteel, & Burgess, 1989; Tulving & Gold, 1963; West & Stanovich, 1978) or a series of semantically congruent words (Schvaneveldt, Meyer, & Becker, 1976). The magnitude of these effects are larger when sentences are highly constraining than when they are less constraining (Fischler & Bloom, 1979; Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985), but facilitation can be found for words semantically related to upcoming targets of low-constraining sentences. Predictive contexts may even influence early-stage perception of expected information, as expected words are more easily discriminable from lures that differ by one letter (Jordan & Thomas, 2002), and also do not show transposed-letter priming effects (Luke & Christianson,

2012). Sentential contexts also enhance identification of mispronunciations of expected words (Cole & Perfetti, 1980), as well as identification of expected words in noisy listening conditions (Fallon, Trehub, & Schneider, 2002), demonstrating that contextual facilitation occurs in speech listening as well.

Some early work in the area of word recognition and semantic priming did essentially postulate a predictive mechanism. Several studies have shown that lexical decisions are made faster when a target word (*dog*) is preceded by a semantically associated word (*cat*), an effect known as semantic priming (Meyer & Schvaneveldt, 1971; Neely, 1976). According to Posner and Snyder's (1975) theory of priming, these effects are caused by two mechanisms: a rapid automatic spreading activation mechanism, in which semantic nodes near to an activated node within an individual's semantic network passively become active, and a slower expectancy mechanism, in which top-down expectations about upcoming stimuli lead to activation of nodes in the network. Evidence for this comes from manipulating the *relatedness proportion*, or proportion of semantically related pairs that were presented in a list of word pairs. Priming occurs even if there is only one semantically related pair on the list (Fischler, 1977); however, increasing the relatedness proportion increases the magnitude of the priming effect (den Heyer, 1985; Tweedy, Lapinski, & Schvaneveldt, 1977). These results argue that individuals track contextual information, and expectancy mechanisms can be engaged in certain processing environments that are relatively predictive.

Eyetracking research has also demonstrated facilitative effects of predictive contexts on language processing. Subjects spend less time fixating on upcoming words that are predictable, and are even more likely to skip them entirely (Ehrlich & Rayner, 1981; Frisson, Rayner, & Pickering, 2005; McDonald & Shillcock, 2003a, 2003b; Morris, 1994; Staub & Clifton, 2006). Eye movements and fixation durations of a word are affected when a parafoveal preview, or view of upcoming words, is given (Rayner, 1975), and importantly, this benefit is increased when the

parafoveal information is highly predictable (Balota, Pollatsek, & Rayner, 1985; McClelland & O'Regan, 1981; though see Brothers, Hoversten, & Traxler, 2017). Further evidence shows that the magnitude of parafoveal benefits is related to the expectancy of information in the fovea (Payne, Stites, & Federmeier, 2016). Expected words led to the largest parafoveal benefits, unexpected but plausible words reduced these benefits, and unexpected semantically anomalous words eliminated parafoveal benefits. Thus, multiple levels of contextual constraint influence language processing, and when information is consistent with the context, attention can be allocated to upcoming information, though it remains unclear if this attentional allocation essentially acts as a form of active prediction, or if more easily integrated information simply frees up resources.

Numerous studies have also investigated prediction using the visual world paradigm, in which participants hear a sentence while simultaneously viewing a collection of images, some of which are related to the sentence information (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In these studies, individuals will move their eyes toward the image of an upcoming word prior to actually hearing the word if the word is highly predictable from the context (Altmann & Kamide, 1999; Kamide, 2008; Kamide, Altmann, & Haywood, 2003; Knoeferle, Crocker, Scheepers, & Pickering, 2005; Sussman & Sedivy, 2003). This data provides a stronger argument for a predictive account, as individuals demonstrate knowledge of the upcoming word prior to receiving the bottom-up input of the word. Individuals will also fixate on competitor objects that are similar in shape to the target (Rommers, Meyer, Praamstra, & Huettig, 2013), indicating that prediction can potentially lead to pre-activation of multiple features of a word.

One problem with behavioral work investigating prediction is that behavioral responses are required, which may alter the way individuals process linguistic input, perhaps encouraging prediction. Additionally, inferences are made about pre-activation of information prior to

receiving input based on data from responses after the input has been received. As previously stated, this makes teasing apart theories of prediction and integration quite difficult. The overlap of semantic features between the context and the target word could allow the expected word to be processed and integrated into the discourse representation more easily, leading to faster responses (Moss & Marslen-Wilson, 1993; Traxler & Foss, 2000; Tyler & Wessels, 1983), but these features are not necessarily pre-activated. Thus, an account of prediction need not be taken to explain the data.

While the semantic priming literature manipulating relatedness proportions seems to necessitate a predictive account, it is unclear if the results from this work can be generalized to other areas of language processing. First, the mechanism of expectancy is postulated to be quite slow, and even when relatedness proportion is high the effect is diminished at shorter SOAs (den Heyer, Briand, & Dannenbring, 1983). When information is more complex, as in a sentence, this mechanism might be too slow to be engaged in time. Additionally, it is unclear if relatedness proportion effects are apparent with more complex stimuli. Finally, Neely & Keefe (1989) proposed that retrospective strategies, in which participants mentally compare the target to the previously presented prime to make a decision, may play an important role in priming effects as well. Though they argued for a hybrid prospective-retrospective theory, it is possible to imagine that using retrospective strategies could sidestep the need for prediction. Interestingly, retrospective processing is not often considered in the sentence processing literature.

Visual world eyetracking studies have received criticism as well. For instance, it is unclear exactly how and in what timescale the perceived visual information interacts with the auditory language input. Additionally, the presence of a specific set of pictorial options could lead participants to narrow processing or create predictions for those images (Henderson & Ferreira, 2004; Huettig, Rommers, & Meyer, 2011). When reading or listening to language,

likely more often than not the information discussed is not immediately visible to the individuals. Thus, the effects observed in visual world paradigms may not generalize to reading or everyday language comprehension, and the predictive effects observed here may be highly task-specific.

Based on these criticisms, an ideal method for investigating integration and prediction in language comprehension would not require an explicit response or task demand, and also allow for measuring processing before, during, and after an expected or unexpected stimulus. An optimal method that fits these requirements is electroencephalography (EEG), in which sensors placed on the scalp measure brain activity primarily generated by post-synaptic potentials from cortical pyramidal cells (Nunez & Srinivasan, 2006). Using EEG, researchers can constantly monitor a subject's neural activity on a millisecond timescale, which allows for tracking of the onsets and timecourses of the cognitive processes occurring during language comprehension. Importantly, these neurocognitive events will occur whether an explicit response is made or not; thus, activity can be assessed during passive sentence reading with no additional task. Most headway in this area has been made by analyzing event related potentials (ERPs), or the resulting waveforms that are produced after averaging many trials of EEG data, usually time-locked to the onset of some stimulus or event (Luck, 2014). By comparing the ERP waveforms of different conditions, researchers can identify components, or segments of the time-series thought to be related to a cognitive process of interest.

Arguably the most studied ERP component in investigations of language processing is the N400 (Kutas & Federmeier, 2011). The N400 is a negative-going wave with a central-posterior topography that onsets approximately 250 ms and peaks around 400 ms after the presentation of a meaningful stimulus, such as a word. Originally observed as a large amplitude response to a semantically anomalous word in a sentence context (Kutas & Hillyard, 1980a), the N400 has been further characterized over several years of research, and is not simply a response

to linguistic anomalies (Fischler et al., 1983; Hillyard & Kutas, 1983; Kutas & Hillyard, 1980b). Rather, the N400 is now thought to index the demands on semantic processing of any meaningful stimulus, including gestures and actions, or perhaps retrieval of features related to that stimulus from semantic memory, and may reflect the integration of this semantic information with perceptual input (Freunberger & Roehm, 2016; Ganis, Kutas, & Sereno, 1996; Hagoort & van Berkum, 2007; Lau, Phillips, & Poeppel, 2008; Kutas & Federmeier, 2000; Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008; Willems & Hagoort, 2007). This semantic processing occurs implicitly and automatically, as neither attention (Ibáñez, López, & Cornejo, 2006; Luck, Vogel, & Shapiro, 1998; Stenberg et al., 2000) nor intact episodic memory (Olichney et al., 2000) are necessary for N400 modulations to semantic manipulations to appear.

An important point that must be stressed is that the N400 is not an "ERP component of prediction". The N400 does not index some neurocognitive process of prediction, and a change in amplitude in the N400 does not necessarily mean that prediction of information has occurred. For example, words with many lexical neighbors elicit larger N400s compared to words with fewer neighbors (Holcomb, Grainger, & O'Rourke, 2002), regardless of any predictive context. Additionally, in sentences containing a false negation in which the words are semantically related (e.g. "a robin is not a bird"), the amplitude of the N400 at the final word will still be reduced (Fischler et al., 1983). It seems unlikely that participants are actively predicting these words, as they are highly implausible given the context, suggesting that the N400 reductions are due to semantic association. However, in cases where upcoming information could be predicted, the pre-activation of semantic features prior to encountering the expected stimulus would reduce the demands on semantic processing of that stimulus, potentially even if that stimulus isn't very semantically related. Thus, the N400 likely reflects multiple aspects of processing, including contextual constraint and semantic association;

however, in certain situations, it can potentially be used as a proxy to measure the degree to which information has been pre-activated due to prediction.

There is clear evidence that N400 amplitudes are modulated by predictive contexts (Kutas & Hillyard, 1984). A good example of the N400's relationship to predictability of information comes from Wlotko & Federmeier (2007), in which participants read sentences word by word while their EEG was recorded. Importantly, these sentences had been previously normed on a cloze probability task (Taylor, 1953), in which subjects are given a sentence with the final word missing and are instructed to complete the sentence with the word that best completes it. Using this method, one can quantify how predictable a particular sentence ending is from its sentence context based on the number of individuals that produced the word. Sentences ending with high cloze probability words are considered *high constraint*, as they essentially constrain the number of expected outcomes, whereas sentences with low cloze probability endings are *low constraint*, as several potential words could adequately complete the sentence. In the EEG experiment, sentences were varied across the entire range of constraint (0–1 cloze probability, and ended either with an expected ending (the highest cloze probability word) or an unexpected, but still plausible, ending. The plausibility of the unexpected words were an important control, as the words were not semantically anomalous, just unpredictable from the context. At the sentence final word, the amplitude of the N400 was inversely related to cloze probability, such that words with higher cloze probability produced lower amplitude responses. Unexpected but still plausible words of produced the largest amplitude N400s. This pattern of the N400's relationship to predictability has been replicated in other studies (DeLong, Quante, & Kutas, 2014; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Thornhill & Van Petten, 2012; Wlotko, Federmeier, & Kutas, 2012). Thus, predictive contexts can modulate the amplitude of the N400, suggesting that semantic access is eased or facilitated by predictability of information.

Of course, one could still make the argument that the N400 amplitude modulations observed due to predictive sentence contexts reflect integration. A sentence ending word might have a greater cloze probability because it is more predictable, or simply because it is more semantically associated to the preceding context. Additionally, even if a word is more "predictable", individuals may not actively predict or pre-activate information. To contrast accounts of prediction and integration, Federmeier and Kutas (1999) presented participants with high and low constraint sentences that ended with an expected ending, an unexpected ending from the same category as the expected ending, or an unexpected ending from a different category. N400 amplitudes were smallest for expected words, followed by unexpected within-category words, followed by unexpected between-category words. Additionally, within-category violations elicited smaller N400s when constraint was higher, even though plausibility ratings for these completions were lower. These results argue against an integration account, as both unexpected endings would be difficult to process, and strong constraint sentences would lead to equal facilitation of both types of unexpected endings. While these findings also argue against an account of predicting a specific word, they support the notion that predictive contexts allow for pre-activation of certain semantic features.

Another debate has been the mechanism through which prediction occurs, and the extent to which it is obligatory. Essentially, rather than being actively predicted through a top-down mechanism, high cloze probability words might become activated through passive spreading activation due to the message-level semantics of the context (Gerrig & McKoon, 1998; Myers & O'Brien, 1998). Information might be pre-activated, but this might a passive as opposed to an active process. To investigate prediction vs. spreading activation, Lau, Holcomb, & Kuperberg (2013) presented participants with prime-target pairs that were either semantically associated (*"dog-cat"*) or unassociated (*"dog-table"*). Critically, the relatedness proportion differed between blocks. Essentially, if pre-activation of information is simply due to passive spreading activation, then changing the relatedness proportion should not affect N400

amplitudes. In fact, N400 amplitudes to target words were most positive for associated words in high-relatedness blocks, followed by associated words in low-relatedness blocks, and finally by unrelated words in either block. Here, semantic relatedness was carefully controlled, but the predictive context of the block led to changes in N400 amplitude; thus, top-down prediction can have separable effects from spreading activation on the N400.

Reading prime-target pairs may not be completely comparable to natural sentence reading, and perhaps N400 facilitation effects from sentence contexts are due to passive spreading activation. Results from Wlotko & Federmeier (2015) oppose this notion. Here, participants read the sentences from Federmeier & Kutas (1999); in one block, the SOA was regular (500 ms), while in the other it was speeded (250 ms). When the speeded block came first, the N400 facilitation for within-category violations was found at 500 ms SOA, but was diminished at 250 ms SOA. However, when the block order was reversed, the facilitation was found at both SOAs. It could be the case that, at 250 ms SOA, information is presented too rapidly for passive spreading activation to occur; however, the facilitation effect should then be abolished regardless of block order, which was not observed. These results better support the notion that prediction mechanisms are top-down and are engaged based on demands on the system and the processing environment. Indeed, recent work has shown that N400 effects are enhanced when prediction is emphasized in the task, and predictability effects on reading times are essentially abolished when overall predictive validity of the stimuli is low, suggesting top-down volitional control of prediction (Brothers, Swaab, & Traxler, 2017).

The previously discussed results have clearly demonstrated facilitative effects of prediction on N400 amplitudes, but there has been little evidence of a cost to semantic access from violating predictions. In theory, prediction errors, or encountering an unpredicted stimulus, should lead to a cost as well; if a specific mental representation is built ahead of time, but then must be revised or inhibited, then this additional step should lead to some increased

processing effort. However, findings from the literature show that unexpected or anomalous endings to strong and weak constraint sentences produce equivalently large N400s (Kutas & Hillyard, 1984), suggesting that this increased N400 has less to do with dealing with prediction error and more to do with processing the lack of featural overlap between the unexpected ending and the context. By only examining the N400, one might conclude that a violation of predictions leads to little processing cost.

An alternative explanation is that prediction costs are indexed at a different timepoint by another ERP component. Indeed, electrophysiological studies have identified other ERP components potentially related to prediction errors or costs. Federmeier, Wlotko, De Ochoa-Dewald, & Kutas (2007) observed a later (approximately 500-900 ms) frontally distributed positivity in response to unexpected but plausible completions to strong constraint sentences, but not weak constraint sentences. Thus, the response was only observed when strong predictions were violated, suggesting a possible link of this component to inhibition or revision processes. This effect has been reported in multiple studies (DeLong, Quante, & Kutas, 2014; DeLong, Urbach, Groppe, & Kutas, 2011; Kutas, 1993; Otten & Van Berkum, 2008; Thornhill & Van Petten, 2012; Wlotko, Zeitlin, & Kuperberg, 2015; Wlotko et al., 2016); however, it is not always observed, and the cases where it is not elicited provide additional information about the cognitive processes related to this component. Older adults, who are thought to utilize contextual information to generate predictions to a lesser degree than younger adults (Federmeier et al., 2002; Federmeier & Kutas, 2005; Gunter, Jackson, & Mulder, 1992; Woodward, Ford, & Hammett, 1993), do not exhibit a frontal positivity response to unexpected endings to strongly constraining sentences (Wlotko, Federmeier, & Kutas, 2012), suggesting the response may reflect prediction error, or updating of the stored contextual representation. Additionally, a frontal positivity is not observed when information is selectively presented to either the left or right hemisphere (Wlotko & Federmeier, 2007), indicating that the component

requires focused attention or interhemispheric cooperation to be elicited, perhaps suggesting a complicated mixture of processes requiring top-down control.

Results from these studies suggest several possible interpretations of the function of the frontal positivity. This component might reflect an inhibition or revision mechanism based on the context built from the preceding sentence, or perhaps just a general detection of unpredicted words. To adjudicate between error detection and contextual updating accounts, Brothers, Swaab, & Traxler (2015) had subjects read medium cloze (approximately 50%) sentences, which ended either with the highest cloze word or with a low cloze completion, and respond whether the word they saw matched the word they predicted. ERPs at the final word were then separated into medium cloze endings that were predicted, medium cloze endings that were not predicted, and low cloze endings that were not predicted. If the frontal positivity only reflects error detection, then both unpredicted conditions should be equivalent. However, if this component reflects revision, then the level of mismatch between the context and the unpredicted stimulus should affect the amplitude. Not only was the N400 amplitude modulated by both prediction accuracy and cloze probability (predicted and medium cloze < unpredicted and medium cloze < unpredicted and low cloze), but also the later frontal positivity showed a remarkably similar pattern. Thus, this component more likely indexes a revision of the established context. This is further supported by the results of Lau, Holcomb, & Kuperberg (2013), in which unassociated words in prime-target pairs did not elicit a frontal positivity, presumably due to the lack of any context to revise.

Further support for the context revision account of the frontal positivity comes from sentence processing studies that include anomalous sentence endings. The studies previously discussed used unexpected, but still plausible, endings ("Tim threw a rock and broke the *camera*"); however, unexpected and implausible endings could be used as well ("Tim threw a rock and broke the *novel*"). Anomalous sentence endings do not elicit frontal positivities, but

rather lead to a late positivity with a posterior parietal distribution (DeLong, Quante, & Kutas, 2014; Geyer, Holcomb, Kuperberg, & Pearlmutter, 2006; McCallum, Farmer, & Pocock, 1984; Pijnacker et al., 2010; Van de Meerendonk, Kolk, Vissers, & Chwilla, 2010; Wlotko et al., 2016). This difference suggests the frontal positivity may reflect a revision of the built-up context in order to integrate the unexpected but plausible information, which is not required when the word does not plausibly fit with the context. Understanding the functional role of the posterior positivity requires more investigation; it may be a P600, an ERP component also thought to be related to review or revision of a constructed mental context representation (Bornkessel, Schlesewsky, & Friederici, 2002; Brouwer, Fitz, & Hoeks, 2012; Van Petten & Luka, 2012). How the revision process indexed by the P600 differs from that of the frontal positivity is currently unclear.

Studies examining ERPs have given great insight into the mechanisms of integration and prediction that occur during language comprehension; however, ERPs are only one method of analyzing EEG data, and its reliance on time-domain averaging means certain aspects of the neural signal are lost. Examining neurocognitive processes using multiple analytical methods is important to fully understand how neural activity is related to mental operations. To this end, investigators have recently employed time-frequency analysis to assess how oscillatory signals relate to cognitive processes (Cohen, 2014). Neural communication is carried out through oscillations of neural activity, in which feed-forward and feedback connections between neurons lead neural ensembles to fire rhythmically at specific frequencies (Buzsaki, 2006). These different frequencies of oscillatory activity are thought to represent areas of the brain coordinating their processing in order to carry out certain mental operations. Using traditional ERP analyses, accurately measuring power in different frequency bands is very difficult, as non-phase locked oscillatory activity across trials will cancel out and sum to zero. By utilizing time-frequency analyses, in which EEG signals are first decomposed into a combination of sine waves at different frequencies and then averaged, researchers can examine how specific experimental

manipulations lead to changes in the ongoing rhythmic activity of the brain, also known as event-related spectral perturbations, or ERSPs (Makeig, 1993; Pfurtscheller, 1977) that are very often separate from the neural activity underlying an ERP (Fell et al., 2004).

Time-frequency analysis of EEG activity following presentation of a semantically incongruous sentence-final word, compared to the activity following a congruous word, has revealed relative power increases in the theta band (Bastiaansen & Hagoort, 2015; Braunstein et al., 2012; Hagoort, Hald, Bastiaansen, & Petersson, 2004; Hald, Bastiaansen, & Hagoort, 2006; Kielar et al., 2014; Wang, Zhu, & Bastiaansen, 2012). Similar increases in theta have been observed for syntactic incongruities as well (Bastiaansen, van Berkum, & Hagoort, 2002). Across these studies, cloze probabilities and plausibility of the items used vary, and it is unclear how much this effect relates to a prediction-related response or a general detection of anomalies. To address this, a recent study specifically examined ERSPs from expected endings to highly constraining sentences compared to unexpected words from those sentences (Rommers et al., 2016). Here, an increase in theta power for unexpected words was also found; however, the effect specifically had a frontal topography, which was not always observed in previous work. Importantly, this theta effect was not found when comparing expected and unexpected endings to weakly constraining sentences, suggesting this effect is more closely related to prediction error. Finally, frontal theta power was only moderately correlated with frontal positivity ERP amplitude, and time-frequency analysis of the ERP revealed only a very small theta effect, suggesting that the frontal theta effect reflects processing that is separate from other aspects captured in the ERP components.

The frontal topography of this theta effect may be important, as frontal theta effects have been reported in non-linguistic tasks examining prediction and cognitive control (Cavanagh & Frank, 2014). For example, frontal theta power increases linearly with increasing interference in a Stroop task (Hanslmayr et al., 2008), and increased frontal theta power has been observed

for both stimulus and response conflict in a flanker task (Nigbur, Cohen, Ridderinkhof, & Stürmer, 2012). Similar frontal theta increases in situations where top-down control is necessary have been reported in rats (Narayanan, Cavanagh, Frank, & Laubach, 2013) and primates (Phillips, Vinck, Everling, & Womelsdorf, 2013). In some cases, increases in frontal theta are predictive of learning or changes in behavior (Cohen & Donner, 2013; Van de Vijver, Ridderinkhoff, & Cohen, 2011), though these behavioral relationships are not always reported. Investigations of oscillatory activity are few and relatively recent, and the consistency of frequency band, timing, and topography of effects is not always consistent; however, current evidence suggests that the frontal theta response may reflect a domain-general response related to increased cognitive control which can, but not necessarily will be, used as an indication to modulate behavior. The overlap of the frontal theta reported by Rommers et al. (2016) with this previous work may shed light on the mechanisms engaged when linguistic predictions are violated; namely, top-down cognitive control is increased, potentially with the goal of updating behavior to account for prediction error.

To summarize, behavioral, eyetracking, and electrophysiological investigations have shown fairly convincing evidence that, during language processing, individuals utilize the ongoing linguistic context to predict upcoming information. Both semantic relatedness and predictability modulate the amplitude of the N400, likely reflecting semantic processing of the encountered stimulus. In situations where a particular word is highly predicted, an unexpected but plausible word will elicit a late frontal positivity, whereas a semantically anomalous word will lead to a posterior positivity. Both of these components have been related to revision of the mental representation of the context following an unexpected event. Finally, unexpected words also elicit an increase in frontal theta power, which may reflect a more domain-general top-down cognitive control process.

Downstream Benefits and Costs of Prediction

There is substantial evidence that prediction can lead to cognitive benefits and facilitation of processing; reading and lexical decision times are decreased, perceptual processing is enhanced, and N400 amplitudes are reduced when words are preceded by a predictive context. There appear to be costs following violations of predictions as well, as indexed by the later positivities previously discussed; however, is there behavioral evidence of these costs? Theoretically, the processing following prediction errors will take extra time or resources, leading to slowing of behavioral responses. In other tasks involving continuous responding to stimuli, responses following an error are slower than those before, a phenomenon known as post-error slowing (Rabbitt & Rodgers, 1977), which might be related to online recruitment of top-down cognitive control (Botvinick et al., 2001). These tasks potentially differ from language comprehension experiments, in that prediction may not be necessary or engaged, and participants are required to make specific decisions. However, if violations of predictions lead to a similar engagement of top-down control, then it seems reasonable to hypothesize that a prediction error during language comprehension could lead to a similar slowing of behavioral responses.

Behavioral evidence in support of this hypothesis has been mixed. In lexical decision tasks, identification of predicted words is faster than unpredicted words, but it is unclear if unpredicted words are slower than a "neutral baseline". In many cases, no neutral baseline condition is tested, meaning facilitation cannot be separated from inhibition (Schuberth, Spoehr, & Lane, 1981). Small slowing effects are found when compared to presentation of the unpredicted word without a sentence context (Schuberth & Eimas, 1977). When each word in sentence contexts are replaced by X's, unexpected words might lead to slowing (Fischler & Bloom, 1979), or might show a comparative facilitation (Schwanenflugel & Shoben, 1985), but the use of X's as a neutral context is questionable as a comparison to natural reading of

sentences (de Groot, Thomassen, & Hudson, 1982). Compared to the neutral condition of three presentations of "the" before the critical word, lexical decision times of neutral and unexpected words do not significantly differ, unless the response-stimulus interval is increased to 800 ms (Stanovich & West, 1981); however, it is also unclear how reading three the's in a row, a rather anomalous occurrence, could be considered neutral, and 800 ms is much longer than natural reading times (Rayner, 1998). Finally, one study using highly unconstrained sentences ("The next word is ___") for a baseline found a comparative slowing for unexpected but related words in strong constraint sentences (Schwanenflugel & LaCount, 1988). However, this effect was not replicated in a second follow-up experiment. Thus, there is only moderate evidence that unexpected words produce slowing of lexical decisions.

Behavioral evidence of prediction costs has been similarly mixed in research with other paradigms. In classic studies of semantic priming, lexical decisions are made faster when words are preceded by semantically related words, slower when preceded by a series of X's, and even slower when preceded by a semantically unrelated word (Neely, 1976). Similar criticisms as before can be applied here, as it is unclear how effective X's are as a baseline measurement. When relatedness proportion is increased, expected words are processed faster; however, unrelated words are processed at the same speed regardless of relatedness proportion (Tweedy, Lapinski, & Schvaneveldt, 1977). Inhibition in semantic priming seems to be largely influenced by the semantic similarity of items, raising questions of how much top-down controlled processes contribute to observed slowing effects (Thompson-Schill, Kurtz, & Gabrieli, 1998). Recent work, in which subjects read sentences at their own pace while eye movements were tracked, reported no evidence of slowing or increase in re-reading for unexpected words (Luke & Christianson, 2016). A similar study with higher cloze materials found no cost for unexpected words as well, and in fact reported facilitation when the unexpected word was semantically similar to the expected word (Frisson, Harvey, & Staub, 2017). Across multiple behavioral paradigms of language processing, evidence of prediction costs has been lacking.

Recently, some researchers have conceptualized the later frontal positivities following unpredicted words as costs of prediction (Kutas, DeLong, & Smith, 2011; Van Petten & Luka, 2012). These ERP components might reflect some "additional processing", such as updating or inhibition, which is required due to the prediction error. However, is it reasonable to consider this as a cost? While potential ties have been made between the frontal positivity and cognitive control processes, these are only hypothetical and have not been thoroughly researched. Additionally, no study to date has found brain-behavior relationships between the amplitude of the frontal positivity and any behavioral measure, such as comprehension of the sentence or reading time. The frontal positivity, as well as the posterior positivity, may be a natural consequence to encountering unpredicted information, but calling it a cost implies some loss or downside, which has yet to be shown.

In the research on prediction in language comprehension described above, behavioral and electrophysiological effects of prediction were predominantly measured at the time of encountering the predicted or unpredicted stimulus. While this has been useful for identifying the immediate effects of prediction, it is worthwhile to consider what downstream effects confirmed or disconfirmed predictions might have on later cognition. It may be the case that costs of prediction do not emerge so immediately, and instead will appear in behavioral measures following the unexpected word, as in the previously described post-error slowing literature, where effects emerge on trials following the error event.

In order to investigate downstream effects of prediction, I tested long-term memory for sentence final words following reading of high constraint and low constraint sentences that ended with either a predicted or unpredicted word. Based on previous evidence of prediction leading to facilitation, an individual's memory for predicted words might be enhanced compared to unpredicted words. Additionally, if the frontal positivity does reflect some revision process, then the mental representation of the unexpected information or event would potentially be

altered, which could change the individual's memory retrieval of that event and potentially lead to impairments in remembering this information. By also including a constraint manipulation, I assessed how predictability affected memory retrieval – for example, impairments to memory might only emerge for words from high constraint sentences. Finally, ERPs were recorded during the experiment, which allowed for evaluation of differences in memory retrieval processes engaged to recognize information. Experiment 1 contained unpredicted but semantically congruous words, while Experiment 2 included unpredicted congruous as well as unpredicted incongruous words.

Predictive Processing Before the Critical Word

Research in the literature of language comprehension has been largely successful in demonstrating that individuals predict upcoming information, and that this process can be engaged flexibly depending on demands of the environment. Less research has focused on the nature and timecourse of the mechanism of prediction itself. For instance, a large body of work has identified what occurs when a highly predicted word is encountered, or when predictions are violated, but little is known about processing occurs beforehand that might be related to effects observed at the critical word. Eyetracking studies have identified that eye movements to critical targets can occur before hearing the target, but these effects have not been linked to specific neural processes (Altmann & Kamide, 1999). If the brain is able to use contextual cues to generate predictions of upcoming information, then there should be some differences in neural processing between high and low cloze sentences even before the critical final word is encountered. Such a finding would also provide stronger evidence against an integration account of prediction effects – if the semantics or meaning of the information prior to the critical word is relatively controlled, such that semantic integration should not differ between

conditions, but the predictability of the upcoming critical word varies, then observed effects would likely be due to the differences in predictive processing occurring in the brain.

Some electrophysiological studies have investigated this hypothesis. DeLong, Urbach, & Kutas (2005) presented subjects with sentences that ended with either expected or unexpected indefinite article / noun pairs (e.g. "the boy went outside to fly *a kite* / *an airplane*"). The N400 response pattern at the noun mirrored previously observed effects, with smaller N400s to expected nouns; however, a similar, albeit smaller N400 effect was also observed at the article. Both "a" and "an" have identical meaning, so such an effect is unlikely to be due to greater difficulty of integration, and is more likely due to predictive processing. A later, more frontal ERP difference was observed for congruent vs. incongruent articles in Spanish (Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2004), and a posterior negativity was elicited by Polish adjectives carrying a prediction-inconsistent suffix (Szewczyk & Schriefers, 2013). Thus, while the timing and topography of the effect differs, there seem to be ERP differences elicited by incongruent words immediately preceding a critical predicted.

While these studies provide interesting results, they primarily show that an unpredicted stimulus prior to a critical word that is not necessarily semantic anomalous can still produce similar ERP effects to an unpredicted critical word. Does this really tell us about anticipatory processing based on the predictability of information? Using the DeLong study as an example, one account is that individuals might not predict just a word ("kite"), but also its preceding article ("a kite"), and so when an unexpected article is encountered, the usual ERP effects are elicited. Another account is that, for whatever reason, the article is more difficult to integrate, and thus similar ERP effects are elicited, which sidesteps a prediction account entirely. N400 amplitudes to the article showed the usual linear relationship with the cloze of the article; if cloze in fact reflects ease of integration, then the same argument holds for articles as for nouns. Essentially, examining prediction violations may not be particularly informative about

understanding mechanisms of anticipatory processing – even a comparison of low cloze endings (e.g. 30%) to unexpected endings show nearly identical N400 effects as in the high cloze comparison.

Other similar work investigating neural activity and prediction during comprehension of spoken Swedish words has focused on activity during the word-initial fragment (WIF). In Swedish, initial segments of words (the WIFs) have prosodic accents that are indicative of certain words as opposed to other words (i.e. accent 1 can lead to certain suffixes, resulting in certain words, while accent 2 of the same WIF will lead to different words). In ERP studies, accent 1 WIFs elicit an early (140 ms onset) left anterior negativity compared to accent 2 WIFs, and suffixes that are invalidly cued by WIFs will elicit a P600 (Roll, Horne, & Lindgren, 2010; Roll, Söderström, & Horne, 2013; Roll et al., 2015). This early frontal ERP, dubbed the "pre-activation negativity" or PrAN, has been proposed to index pre-activation of upcoming phonemes or suffixes. To more directly test this, ERPs were recorded to WIFs that varied in their number of possible completions, where a high number of possible completions would potentially lead to less certainty about the upcoming suffix (Söderström, Horne, Frid, & Roll, 2016). The amplitude of the PrAN was found to be linearly related to number of completions, supporting the notion that the amplitude of the component could be related to active prediction of an upcoming suffix.

This work provides important data showing that neurophysiological activity can be modulated based on the predictability of upcoming information even when predictions aren't violated in any way. Additionally, it leads to interesting hypotheses about the timing of predictive mechanisms – potentially, predictions about upcoming stimuli are generated rapidly, within 200 ms, following a cue. However, it is unclear if WIFs with fewer possible completions are also easier to integrate with the current context, which could explain the ERP differences. Additionally, Swedish accent tones present an interesting case with which to study prediction,

but there may be different mechanisms that are engaged for processing these tones that are not engaged in other situations of language comprehension. Essentially, it is unclear if this effect generalizes to other areas of prediction.

A few other recent studies have examined differences in electrophysiological activity based on the predictability of upcoming information. Rommers et al. (2016) examined time-frequency effects prior to the onset of the sentence final word in high vs. low constraint sentences. High-constraint endings were preceded by a decrease in occipital alpha and beta power relative to low-constraint sentence endings, and the magnitude of this effect was positively correlated with the magnitude of the frontal theta effect following disconfirmations. Maess et al. (2016) presented participants with sentences containing verbs that were highly predictive for a particular noun ("he conducts the orchestra"), or were not very predictive ("he leads the orchestra") while MEG was recorded. An N400 reduction was found for more highly predictive nouns, but the opposite effect was found at the verb – highly predictive verbs elicited larger N400s. Additionally, these effects were negatively correlated in several areas of left temporal cortex, a brain area implicated in semantic processing (Lau, Phillips, & Poeppel, 2008). Finally, Freunberger & Roehm (2017) examined ERPs to adverbs preceding target words, and found that adverbs that increased the predictability of the target word elicited more negative N400 responses, similar to the pattern reported by Maess et al. Thus, neural processing prior to a sentence final word differs based on the predictability of that word, and that activity could be related to or affect the processing that occurs when the word is encountered.

One potential criticism is that sentence reading may not be the best paradigm for the investigation of anticipatory processing in the brain due to its multifaceted nature. For example, in the study using predictive verbs, differences in neural activity may be due to prediction, but could potentially be partially explained by differences in information contained by nouns as

opposed to verbs, as these different word classes are processed differently (Pulvermüller, Lutzenberger, & Preissl, 1999). Sentences vary in length, and words in sentences vary in frequency and other lexical characteristics, making it difficult to separate aspects of predictive processing from processing of the word or sentence up to that point. Additionally, it remains unclear what makes a particular word in a sentence very predictable – in some cases, semantic relationships between words in the sentence seem a probable explanation (e.g. "The piano was out of *tune*", cloze 74%), but in other cases, other factors, perhaps the frequency of a particular phrase, seem at work (e.g. "She thought she had seen a *ghost*", cloze 69%). Anticipatory processes may differ depending on what aspects of the context are informative or predictive, and controlling for these aspects would be difficult, if not impossible. A simpler paradigm might be more effective for investigating anticipatory processing.

To this end, I report results from Experiment 3, in which I utilized a word stem completion paradigm (Graf & Mandler, 1984) to investigate anticipatory predictive processing in the brain. While word stem completion tasks are normally used to test implicit memory, I focused on the generative aspect of the task – essentially, completing a word stem might rely on similar processes as completing the final word of a sentence, and like sentences, certain word stems might be more predictive of certain completions. However, in this case, all of the cues to generate information were 3 letter word stems; thus, there were likely little differences in semantic content or contextual facilitation of the stems themselves. By presenting participants with stems that were normed so that the predictability of their completions are known, I was able to examine differences in neural activity at the time of the *stem* presentation based on the probability of the future *completion* to make inferences about differences in anticipatory processing. Additionally, participants were subsequently presented with either a high probability completion, a low probability completion, or a pseudoword following the presentation of the stem. ERPs to these post-stem words, particularly the N400, should essentially replicate the pattern seen for sentence ending words that are expected or unexpected

based on the probability of the completion. Such a cross-task agreement in results would lend credence to the notion that similar processes occur in the word-stem task as in natural reading tasks in the literature.

In summary, the current body of evidence strongly suggests that individuals utilize contextual information to generate predictions about upcoming language information, but it remains unclear what the precursors and downstream costs of having predictions confirmed or disconfirmed are. My experimental data not only provide further support for the notion that processes of prediction are utilized in language comprehension, but also set the stage for exploring these mechanisms of prediction across different populations and task demands in a novel way.

CHAPTER 2: BEHAVIORAL AND ELECTROPHYSIOLOGICAL CONSEQUENCES OF PREDICTION ON RECOGNITION MEMORY

As discussed above, Experiment 1 was designed to investigate downstream benefits and costs of predicting information during language comprehension. Having predictions disconfirmed could lead to behavioral costs or changes in neural processing that might not be apparent immediately at the time of the disconfirmation. In Experiment 1, participants read high or low constraint sentences that ended either with the expected or an unexpected word, and they were then were given a recognition memory test containing sentence ending words. The memory test allows for investigation of downstream effects of prediction; namely, predicting information, or having predictions disconfirmed, could lead to enhancements or deficits in long-term memory retrieval of the information.

Why test recognition memory? While many cognitive tasks could be tested, I chose to test participants' memory in an attempt to understand the state of the mental representations of encoded words that were strongly predicted, weakly predicted, or completely unpredicted. Context-driven prediction could change the encoding of information into long-term memory, for instance, by modulating the level of attention given to the predicted or unpredicted information that is encoded (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996). Essentially, paying more attention to certain stimuli could modulate the depth or level of processing, leading to a more stable and persistent memory representation (Craik & Lockheart, 1972; Craik & Tulving, 1975). Whether expected or unexpected stimuli will receive greater depth of processing remains an open question. For instance, stimuli that appear in predicted spatial locations are detected more rapidly and accurately than when no predictive cue is given (Posner, Snyder, & Davidson, 1980), suggesting more attention is given to predicted information. However, in eyetracking experiments with natural reading, individuals spend less time looking at and exhibit less

regressions to predicted words (Ehrlich & Rayner, 1981), suggesting they may in fact pay less attention to them. In the case of unpredicted 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), and unexpected or error-related events modulate early attention-related ERPs (Wills, Lavric, Croft, & Hodgson, 2007), suggesting distinctive, unpredicted events might be more attended to and more easily remembered. I hypothesized that the distinctiveness of unexpected words would lead to higher hit rates than for expected words; I also predicted that strongly constraining expected words would show lower hit rates than weakly constraining words, as their expectedness would make them less distinctive and thus would be less deeply encoded.

Testing recognition memory also allowed for separately investigating the effects of prediction and observation of a stimulus on encoding. Participants were tested on sentence ending words that were expected (predicted and observed), unexpected (unpredicted and observed), expected when an unexpected word was presented during sentence reading (predicted and unobserved), and completely new (unpredicted and unobserved). By including the predicted but unobserved stimuli, which I am calling *lures*, the experiment also tested for participants' false memories (Brainerd & Reyna, 2005), a phenomenon in which individuals recall or recognize events that did not actually occur. Classic studies using the "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", etc.), suggesting that the representation of the lure was activated and erroneously selected during retrieval (Deese, 1959; Roediger & McDermott, 1995; Steffens & Mecklenbräuker, 2007). In the current experiment, I hypothesized that prediction during sentence comprehension would lead to activation of the representation of the lure, which would lead to greater false alarms to lures on the memory test as compared to completely new items.

Lastly, testing recognition memory is important because there is a large overlap in the neurocognitive processes that are utilized in language comprehension and episodic memory, and thus understand the processing that occurs during the memory test will be informative as to the processing that occurs during language comprehension. Classic conceptualizations of episodic memory have changed; rather than encode entire episodes, certain neural systems are thought to encode specific features of the environment (e.g. the perirhinal cortex encodes information about objects, whereas the parahippocampal cortex encodes spatial and contextual features; Ranganath, 2010), while the hippocampus encodes the temporal, spatial, and other meaningful relational information between these features (Konkel & Cohen, 2009). Contextual information's impact on memory encoding and retrieval has been emphasized (Polyn, Norman, & Kahana, 2009), and neural systems involved in memory retrieval also show involvement in thinking about the future or simulating events (Buckner, 2010; Schacter, Addis, & Buckner, 2007). Thus, the processes involved in encoding and retrieving episodic information – computing relations between features, utilizing contextual information, and generating simulations – are likely involved in language processing as well. Indeed, recent work has shown that neural systems related to episodic memory, namely the hippocampus and surrounding medial temporal lobe, also play a role in certain aspects of language processing, and that amnesiacs with damage to these areas will show specific behavioral deficits in language tasks (Duff & Brown-Schmidt, 2012; Kurczek, Brown-Schmidt, & Duff, 2013). While the MTL has not yet been specifically implicated in language prediction, it shows involvement in statistical learning and extracting temporal regularities from streams of information, which may be a critical component of generating predictions (Schapiro et al., 2014; Shohamy & Turk-Browne, 2013).

Recording ERPs during the recognition test segment of the experiment may also reveal differences in processes involved in retrieving information that was predicted or unpredicted. In ERP studies of recognition memory, previously encountered words or stimuli (*old* items)

elicit different patterns of neural activity than stimuli that weren't previously presented (*new* items). This ERP difference has been called the old-new effect (Paller, 2001; Rugg, 1995), which, over years of research, has been separated into multiple components that may reflect different levels of cognitive processing. These different levels have primarily been mapped onto the dual-process theory of recognition memory (Mandler, 1991; Rugg & Curran, 2007; Yonelinas, 2002), which states that recognition of information is supported by the processes of familiarity and recollection. Familiarity is a feeling that the stimulus has been encountered before without explicit details of the event (e.g. "That person looks familiar but I can't remember who they are"; Mandler, 1980) that varies quantitatively in strength, and has been related to conceptual implicit memory (Wang, Ranganath, & Yonelinas, 2014; Wang & Yonelinas, 2012). Recollection refers to more explicit retrieval of details or "qualitative" information about the study episode, and supports highly confident recognition and associative recognition.

In ERP studies, familiarity has been correlated with an old-new effect from 300-500 ms with a frontal-central topography known as the FN400 (Curran, 2000; Curran & Hancock, 2007; Duarte et al., 2004). The amplitude of the FN400 old-new effect increases with increasing confidence (Yu & Rugg, 2010), is insensitive to depth of encoding (Paller & Kutas, 1992), and remains even when attention is divided during encoding (Curran, 2004), suggesting a link to familiarity. However, familiarity is a somewhat vaguely defined process, and clearly other forms of implicit processing can contribute to feelings of familiarity for a stimulus (Verfaellie & Cermak, 1999; Voss, Schendan, & Paller, 2010; Wang et al., 2015; Wolk et al., 2004). Some researchers argue that the FN400 is functionally identical and reflects the same processing as the N400 (Bermudez-Margaretto, Beltrán, Domínguez, & Cuetos, 2015; Voss & Federmeier, 2011), while others debate that the two components are functionally distinct (Bridger et al., 2012; Stenberg, Hellman, Johansson, & Rosén, 2009; Stróżak, Abedzadeh, & Curran, 2016). Other evidence suggests that the FN400 old-new effect can be modulated by contextual and task-related information (Guillaume & Etienne, 2015; Guillaume & Tiberghien,

2013; Tsivilis, Otten, & Rugg, 2001), opposing a strictly stimulus-driven familiarity explanation of the component. Here, I take the stance that many sources of information can impact familiarity for a stimulus, including the pre-activation of the stimulus in semantic memory – essentially, the process indexed by the N400. Given the strong evidence of similar contextual manipulation of the N400 effect, it seems reasonable to believe that the FN400 is largely made up of an N400 and indexes similar processing.

Recollection has been related to a later ERP component, from roughly 500-800 ms, over left parietal scalp areas known as the late positive complex or LPC (Curran, 2000). The LPC old-new effect is more positive following deep vs. shallow encoding (Rugg et al., 1998), for explicit judgments of recollection (Duzel et al., 1997; Woodruff, Hayama, & Rugg, 2006), for highly confident memory judgments (Yu & Rugg, 2010), and in some cases when source information is successfully retrieved as well (Senkfor & Van Petten, 1998; Wilding, 2000; Wilding & Rugg, 1996), supporting the idea that the LPC indexes explicit retrieval of qualitative details. Compared to the FN400, there is far less debate as to the functional underpinning of the LPC – one argument has been made that the LPC may reflect decisional factors (Finnigan, Humphreys, Dennis, & Geffen, 2002), though separating memory strength from memory decisions is difficult. While there is strong overlap between the FN400 of the memory literature and the N400 of the language literature, it remains unclear if there is a similar overlap between the function of the LPC and the P600 (Van Petten & Luka, 2012).

By examining ERPs during the memory test, inferences can be made about the processes involved in successfully recognizing, or false alarming, to predicted and unpredicted words. Based on the hypotheses outlined earlier, I predicted that words that were previously unexpected during sentence reading would elicit larger LPCs during the memory test, due to the greater attention and thus depth of encoding for these items. I also hypothesized that weakly predicted matches would elicit larger LPC amplitudes than strongly predicted matches, though

not as large as unexpected matches. Essentially, I expected the LPC responses at test to mirror the N400 effects found at study. Results are mixed on whether false recognition effects are found in the N400 (Chen, Voss, & Guo, 2012; Geng et al., 2007) or in the LPC (Beato, Boldini, & Cadavid, 2012; Curran, Schacter, Johnson, & Spinks, 2001; Wolk et al., 2006). Here, I expected false alarm effects to lures to be found in the N400, not the LPC; essentially, prediction leads to pre-activation of semantic features, and I expected this would lead to greater conceptual priming for the predicted information.

To summarize, participants in Experiment 1 read high and low constraint sentences with expected and unexpected endings, and were later given a memory test containing sentence ending words that either matched what was read, were novel, or were predicted but never read (a lure). If prediction strongly pre-activates the expected word, then participants should be more likely to false alarm to the lures compared to the novel items. ERPs were recorded during the experiment to investigate retrieval mechanisms. Recognition of predicted and unpredicted words may be supported by different neurocognitive mechanisms.

Previous studies have used a similar paradigm, albeit to investigate different theoretical questions. In an earlier ERP study, participants read a sentence ("A type of bird") followed by a word that fit ("robin") or did not ("nail"), and later were given a memory test on the final words (Neville, Kutas, Chesney, & Schmidt, 1986). Old-new effects were found, but there were no ERP differences between recognized congruous and incongruous words at test. However, congruous is not equivalent to predicted, and ERP differences may still emerge in the current experiment. Corley, MacGregor, & Donaldson (2007) tested individuals on predicted and unpredicted sentence ending words; while their focus was on the effect of disfluencies, they reported better recognition memory performance for unpredicted words, supporting the idea that predicted words are not encoded as strongly. Finally, multiple 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-close or unexpected ending (Hartman & Hasher, 1991; Hasher, Quig, & May, 1997; Lorsbach, Wilson, & Reimer, 1996). These studies have focused mainly on inhibition and control processes; however, they have demonstrated that individuals can retain the expected but disconfirmed endings, consistent with my hypothesis for the lures. Thus, Experiment 1 utilized a relatively novel paradigm to provide unique data investigating downstream consequences of prediction on memory.

Methods

Participants

33 right-handed, native speakers of English with normal or corrected vision from the University of Illinois, Urbana-Champaign participated in the experiment and were paid $10 an hour for their participation. All participants had no 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. 2 participants were excluded due to excessive noise artifacts in the EEG signal, leaving a total of 31 subjects whose data were included for analysis. Note that the *a priori* number of subjects to run was 32; mid-way through data collection, a participant's recorded data was noisy, and thus an extra subject was run. Final analysis led to removal of another subject's data, leading to a final count of 31 participants.

Materials

The stimuli were comprised of 192 English sentences, a subset of the sentences used in Federmeier, Wlotko, De Ochoa-Dewald, & Kutas (2007). The cloze probabilities of the endings of the sentences were previously determined in a norming study conducted at the University of

California, San Diego, 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), 48 with unexpected endings (SCU), 48 weakly constraining sentences with expected endings (WCE), and 48 with unexpected endings (WCU). These stimuli were evenly split into 8 blocks (6 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 statistically significant differences between variables across conditions, though expected items were on average higher frequency than unexpected items (E = 101, U = 81).

After each block of sentence reading, participants had a memory test. For the recognition memory test, participants were presented not with the whole sentence that they had previously seen, but single words, the majority of which were words that had ended the previously read sentences. Test items included "Matches", or words that were sentence endings previously seen (either expected or unexpected); "New" words, or words not previously seen; "Lures", or sentence ending words that were expected, but an unexpected word had been presented; and sentence medial words that were presented to ensure that subjects were paying attention and encoding the entire sentence, not just the endings. As an example of a "Lure" item, during the encoding phase a participant might read the sentence "The prisoners planned for their *party*", where *party* is an unexpected ending, and in the test phase read the word

"*escape*", the expected ending of the 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 1 provides an overview of the different types of test items, the number of each type, and examples of each type. Lexical properties (word frequency, concreteness, imageability, familiarity) were controlled such that New items did not differ from Lure items; however, expected Matches were higher frequency than unexpected Matches (E = 104, U = 62), and this difference was significant when log transformed (t = 3.8228, p = 0.0003). Frequency differences between these conditions are plotted in Figure 1.

<table>
<tr><th colspan="2"><u>Encoding</u></th><th><u>Recognition</u></th></tr>
<tr><td rowspan="4"><u>Sentence (SC/WC)</u>

"Tim threw a rock and broke the…"

(96 SC / 96 WC)</td><td rowspan="2"><u>Expected (E)</u>

"window"

(48 SC / 48 WC)</td><td><u>Match (M)</u>
"window"
(24 SC / 24 WC)</td></tr>
<tr><td><u>New (N)</u>
"leather"
(24 SC / 24 WC)</td></tr>
<tr><td rowspan="2"><u>Unexpected (U)</u>

"camera"

48 SC / 48 WC)</td><td><u>Match (M)</u>
"camera"
(24 SC / 24 WC)</td></tr>
<tr><td><u>Lure (L)</u>
"window"
(24 SC / 24 WC)</td></tr>
</table>

Table 1: Overview of experimental materials of Experiment 1.

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. This is difficult to accomplish in an experiment containing multiple sentences, each containing multiple words that could be a sentence ending word in another sentence. 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". If participants read both sentences before being tested, they might have better memory for the word "dog", or perhaps interference from seeing the word in two different sentences. Thus, the stimulus list was structured such that any sentence containing a critical test item in the middle of it was presented after the test item had already been tested. Unfortunately, this allowed for very little counterbalancing of materials. All participants read the same list of stimuli; while the order of presentation of each stimulus within blocks was randomized, the order of presentation of the blocks was not.

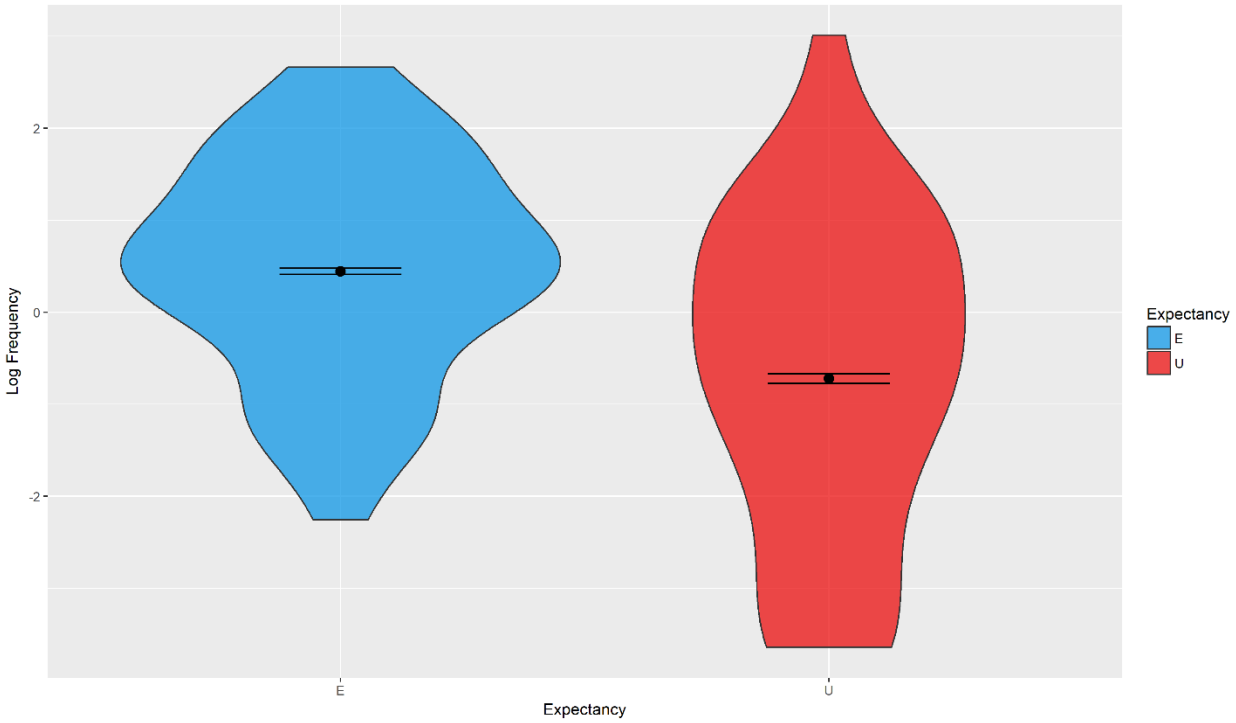


Figure 1: Violin plot of Log transformed word frequency values between Expected and Unexpected Match test items. Expected Matches were significantly higher frequency than Unexpected Matches.

Procedure

Subjects first agreed to participate in the experiment by signing an IRB consent form, as well as filled out a demographic information form (which contained questions regarding vision, English proficiency, and psychiatric medications) and an adapted form of the Edinburgh Handedness Inventory to ensure right handedness. Afterwards, 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 themselves 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 8 study-test blocks, in which participants first studied a set of sentences, and then were tested on their memory for critical words. Between each block, participants could take a short break if they felt they needed to. During the encoding phase, participants were instructed to read the sentences to themselves, and to try to remember what they read, as their memory would be tested. Sentences were presented on the screen in rapid serial visual presentation (RSVP), meaning one word at a time in fairly rapid succession. Each word appeared in the center of the screen for 200 ms, and the interstimulus interval was 300 ms. After the last word of the sentence was presented, a blank screen was presented for 500 ms, followed by a fixation cross for 1000 ms. In an attempt to minimize contamination of the EEG data, 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 seconds. The math consisted of two addition problems, one on the left side of the screen and

one on the right, each involving adding two numbers together. Participants were instructed to press the "1" key if the sum on the left was larger, the "3" key if the sum on the right was larger, and the "2" key if the two sums were equal. The numbers in the equations were randomly generated. The math problems were 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. Here, they were given a recognition test on sentence ending words that appeared in the sentences they had read during the encoding phase. Each trial began with a fixation cross in the center of the screen for 1000 ms, which was replaced by a test word. After 1000 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 4 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 that if they felt like they were guessing or unsure, to use the "Maybe" option. Lastly, 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, 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 was recorded from 26 Ag/AgCl electrodes arranged in a geodesic array and embedded into a flexible elastic cap, which was placed on the participant's head. A schematic of the electrode locations is shown in Figure 2. 5 additional electrodes were attached, including one on each mastoid bone behind the ear, one adjacent to each outer canthus of the eye, which were

used for monitoring of the electro-oculogram (EOG), and one below the lower eyelid of the left eye, which was used for monitoring of blinks. Scalp to electrode connection was made by applying electrode gel to each electrode site. Electrode impedences were kept below 5 kΩ by moving obtrusive hair, adding more gel, and abrading the scalp. Signals were amplified by a BrainVision BrainAmp Standard, with a 16-bit A/D converter, an input impedence of 10 MΩ, a bandpass filter of 0.016-100 Hz, and a sampling rate of 1 kHz. Electrodes that became noisy, drifted, or flatlined during recording were fixed or replaced as quickly as possible. 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.

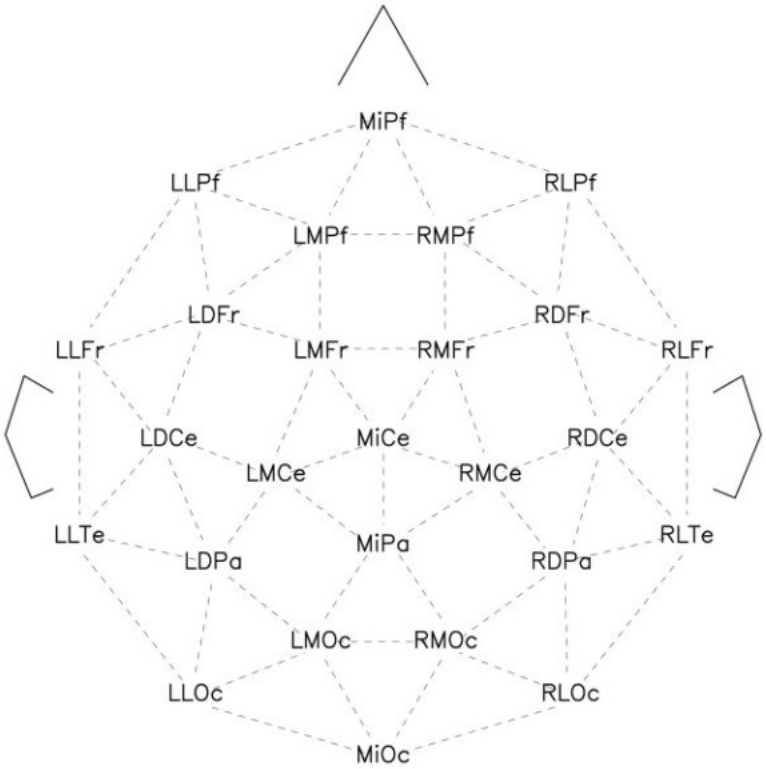


Figure 2: Schematic of the electrode layout used in Experiments 1-3.

Following collection, each raw EEG timeseries was passed through a 0.1-30 Hz Butterworth filter. ERPs were computed with a 200 ms baseline time period prior to the onset

and a 1000 ms epoch following the onset of each sentence ending word during encoding and each test item during the test phase. The average of the pre-stimulus baseline was subtracted from the post-stimulus data for each trial. Additional ERPs were computed to examine Dm effects by binning epochs at encoding based on whether the decision for that item at test was correct or incorrect. Following artifact correction (described below), epochs for each bin were averaged together to create an ERP for each subject. 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 then scanned with a sliding window step function to detect blinks. If the number of epochs containing blinks was less than 10%, the epochs were simply discarded and artifact correction moved on to the next stage. If the number of epochs was greater than 10%, the data was run through AMICA (Palmer, Kreutz-Delgado, & Makeig, in prep), an ICA decomposition algorithm that generalizes Infomax and multiple mixtures approaches adaptively. AMICA has been shown to outperform other ICA algorithms in terms of removing artifacts from EEG data (Coffman, et al., 2013; Leutheuser, et al., 2013), as well as in terms of maximal separation of independent components (Delorme, et al., 2012). 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 correlation higher than 0.6 were removed from trials marked as containing blinks. The threshold of 0.6 was chosen based on qualitative assessment with various thresholds and subjects data, and was generally conservative so as not to remove components with good data. The remaining components were then recombined to reconstruct the EEG data, which was then scanned with an additional sliding window amplitude threshold, and finally manually checked by the experimenter for any additional artifacts.

Statistical analyses were performed on channel clusters as opposed to single channels to improve signal to noise. Channel clusters were decided based on previous research findings and experimental hypotheses. These clusters were: a Frontal Cluster (MiPf / LLPf / RLPf / LMPf / RMPf), a Central Cluster (MiCe / LMCf / RMCf / MiPa), a Left Parietal Cluster (LDCe / LLTe / LDPa / LLOc / LMOc), and a Right Parietal Cluster (RDCe / RLTe / RDPa / RLOc / RMOc). Time windows of analyses are outlined in each Results subsection. Plotted ERPs were filtered with a 10 Hz lowpass filter simply for clarity.

Results

Behavioral Data

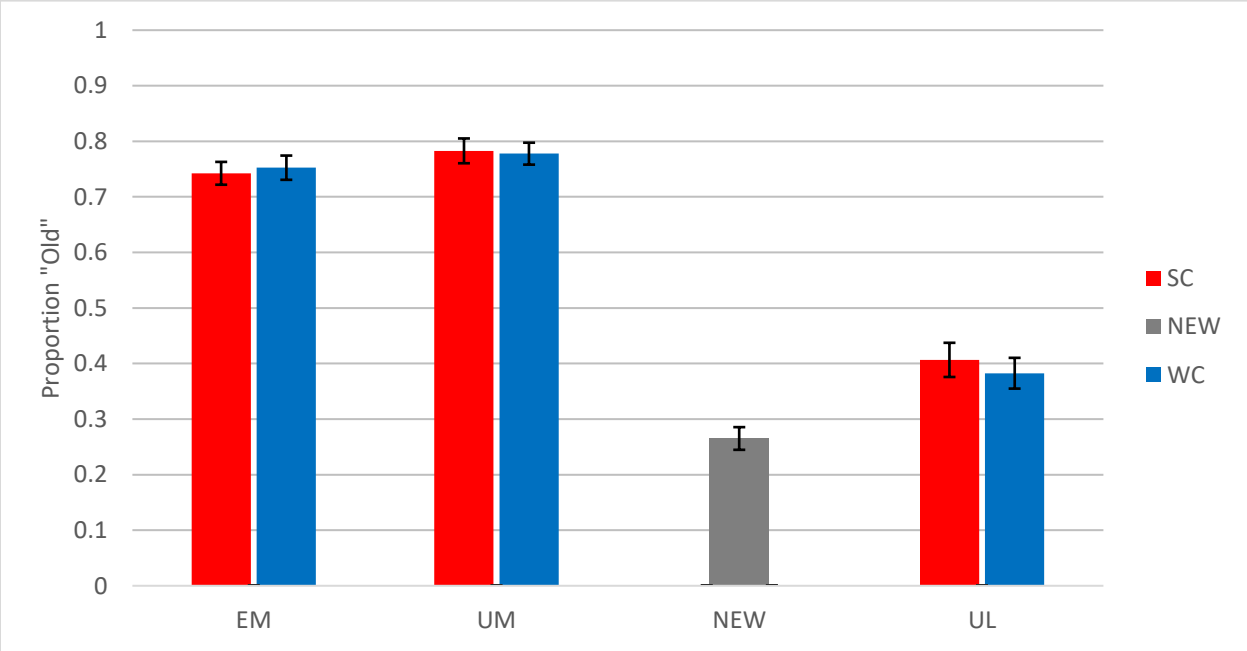


Figure 3: Recognition memory accuracy in Experiment 1. Proportion "Old" responses are plotted on the Y axis. SC = strong constraint, WC = weak constraint, EM = expected match, UM = unexpected match, UL = unexpected lure.

Proportion "Old" responses is plotted in Figure 3. For Matches, "Old" was a correct response, while 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. Accuracy in each condition was above chance performance. Recognition accuracy between Expected and Unexpected Matches did not seem to differ, whereas performance decreased (participants false alarmed more) to Lures compared to New items.

To assess this statistically, behavioral responses on each trial (Old or New) were submitted to a mixed-effects logistic regression model fit by maximum likelihood using the lme4 package in R. Mixed logistic regression models deal with several serious issues when calculating statistics on categorical data using ANOVAs, and also allow for modeling of random effects (Jaeger, 2008). Here, random factors included intercepts for items and slopes and intercepts for participants. Correlations between random factors were not calculated to ease convergence of the models. In logistic regression, Wald's z-scores can be computed for each coefficient to test for significance. Additionally, to assess significance of particular fixed effects, for each model a second model was fit that removed the fixed effect of interest, and these two models were compared with a chi-square test to determine which better fit the data.

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 ($z = -3.184$, $p = 0.00145$), but accuracy did not differ between SC Lures and WC Lures ($z = -0.417$, $p = 0.67683$). There was a significant difference between the model containing a factor for condition (Lure vs. New) and the model without the factor ($X^2 = 39.115$, $p < 0.001$); thus, adding the factor gave a better fit to the data.

Despite the fact that lexical variables (frequency, concreteness, etc) of stimuli were controlled for across New items, it could be the case that a subset of the Lures were more frequent than other Lures or New items, and this contributed substantially to the false alarm effect. To assess this, a second model was fit with Condition and Frequency as fixed effects. Frequency had a significant effect on response ($z = 4.719$, $p < 0.001$), but recognition accuracy still differed between Lures and New items ($z = -2.913$, $p = 0.00279$). Model comparison confirmed a significant contribution of word frequency to the model ($X^2 = 20.141$, $p < 0.001$). Thus, word frequency affected memory responses, but this did not explain the false alarm effect observed.

The next model assessed responses for Matches by modeling responses to match items with Constraint, Expectedness, and the interaction (C*E) as fixed factors. Ultimately, none of the coefficients, Constraint ($z = 0.060$, $p = 0.9520$), Expectedness ($z = 1.261$, $p = 0.207$), or the interaction ($z = -0.267$, $p = 0.790$), returned significant z-scores. Follow-up model comparison chi-square tests were also non-significant, meaning adding the factors gave no additional benefit to fitting the response data. Including word frequency in the model (C*E*F) did not change previous results, though word frequency was trending toward significance ($z = -1.936$, $p = 0.0529$).

Sentence Final Word ERPs

First, ERPs to sentence final words were analyzed to determine if N400 / cloze probability relationship was replicated. Grand average ERPs (across all participants) at the sentence final word at the Central Cluster are plotted in Figure 4. The N400 shows a response graded by expectancy. To assess effects statistically, linear mixed effects models were utilized (Baayen, Davidson, & Bates, 2008), using the lme4 and lmerTest packages in R, with N400 amplitudes from the Central Cluster on each trial as the dependent variable. 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).

A linear mixed effect model was run comparing N400 amplitudes between weakly constrained expected (WCE) endings and strongly constrained expected (SCE) endings, as well as WCE and combined strongly and weakly constrained unexpected (U) endings. There were significant differences in N400 amplitude between WCE and SCE endings ($t = 2.852$, $p = 0.006$), as well as between WCE and U endings ($t = -4.386$, $p = 0.00004$). When word frequency was added as a fixed effect, a significant p value was found for word frequency ($t = 3.042$, $p = 0.004$), but the condition effects remained significant. Thus, the graded N400 effect was replicated in this experiment.

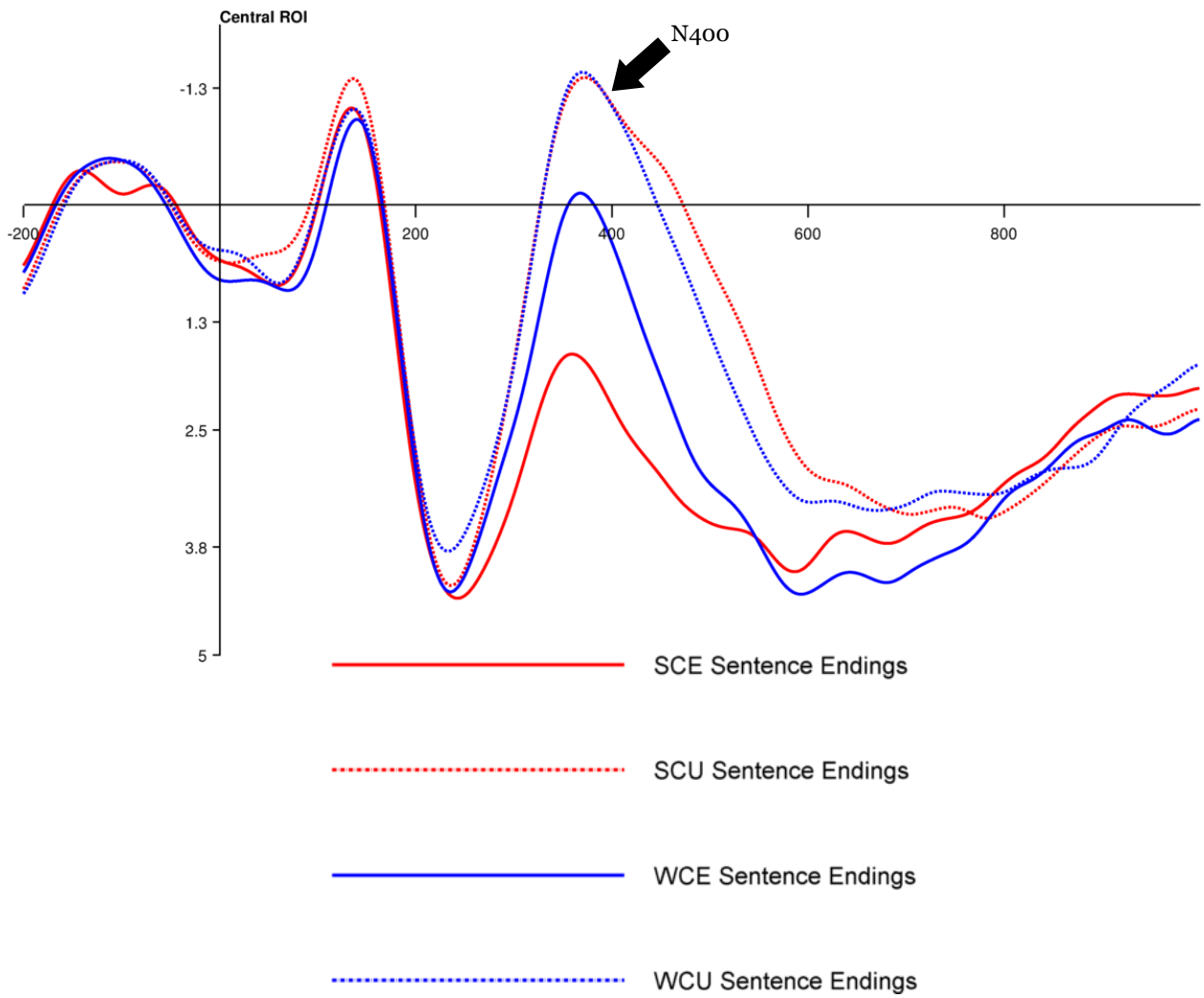


Figure 4: Grand average ERP waveforms for expected and unexpected endings to strongly and weakly constraining sentences at the Central Cluster of channels. Negative is plotted up. The amplitude between 300-500 ms (the N400) is graded based on expectancy.

For consistency with previous experiments, an additional repeated measures ANOVA with two levels of Constraint (strong and weak) and Expectancy (expected and unexpected) was conducted on mean amplitudes between 300-500 ms at the Central Cluster. The main effect of Expectancy [$F(1,31) = 55.33$, $p < 0.001$] and the interaction of Expectancy and Constraint [$F(1,31) = 8.85$, $p = 0.006$] were significant; the main effect of Constraint [$F(1,31) = 3.97$, $p = 0.055$] was not. However, follow-up planned comparisons revealed significant differences in N400 amplitudes between expected and unexpected endings in both strong [$F(1,31) = 43.16$, $p < 0.001$] and weak [$F(1,31) = 23.47$, $p < 0.001$] constraint sentences, as well as differences

between expected endings from strong constraint sentences and expected endings from weak constraint sentences [$F(1,31) = 12.02$, $p = 0.002$].

Next, ERPs to sentence final words in the Frontal Cluster were analyzed to determine if the reported frontal positivity to SC-U endings was replicated. Grand average ERPs at the Frontal Cluster are plotted in Figure 5. A positivity is apparent for SC-U endings; however, there seems to be a positivity for WC-U endings as well. Based on time windows found in DeLong, Quante, & Kutas (2014), frontal positivity effects were assessed by submitting mean amplitudes from 700-1000 ms at the Frontal Cluster to a linear mixed effects model, as was done to examine N400s. The frontal positivity has been conceptualized as a difference between SCU and WCU endings; therefore, the model compared WCU endings to SCU endings, as well as a combination of expected endings (E) to SCU endings to determine if unexpected endings at least differed from expected endings. There were no significant differences in frontal positivity amplitudes between SCU and WCU conditions ($t = -0.944$, $p = 0.348$), and the difference between SCU and E conditions was only trending toward statistical significance ($t = -1.977$, $p = 0.056$). A comparison of WCU and E conditions showed no significant differences ($t = -1.035$, $p = 0.31$). A follow-up ANOVA with levels of Constraint (strong and weak) and Expectancy (expected and unexpected) also found no significant main effect of Expectancy [$F(1,31) = 2.65$, $p = 0.11$], Constraint [$F(1,31) = 0.08$, $p = 0.77$] or a significant interaction [$F(1,31) = 1.35$, $p = 0.25$]. Thus, the results suggest more positive amplitudes for SCU compared to WCU endings, but the frontal positivity effect was not replicated.

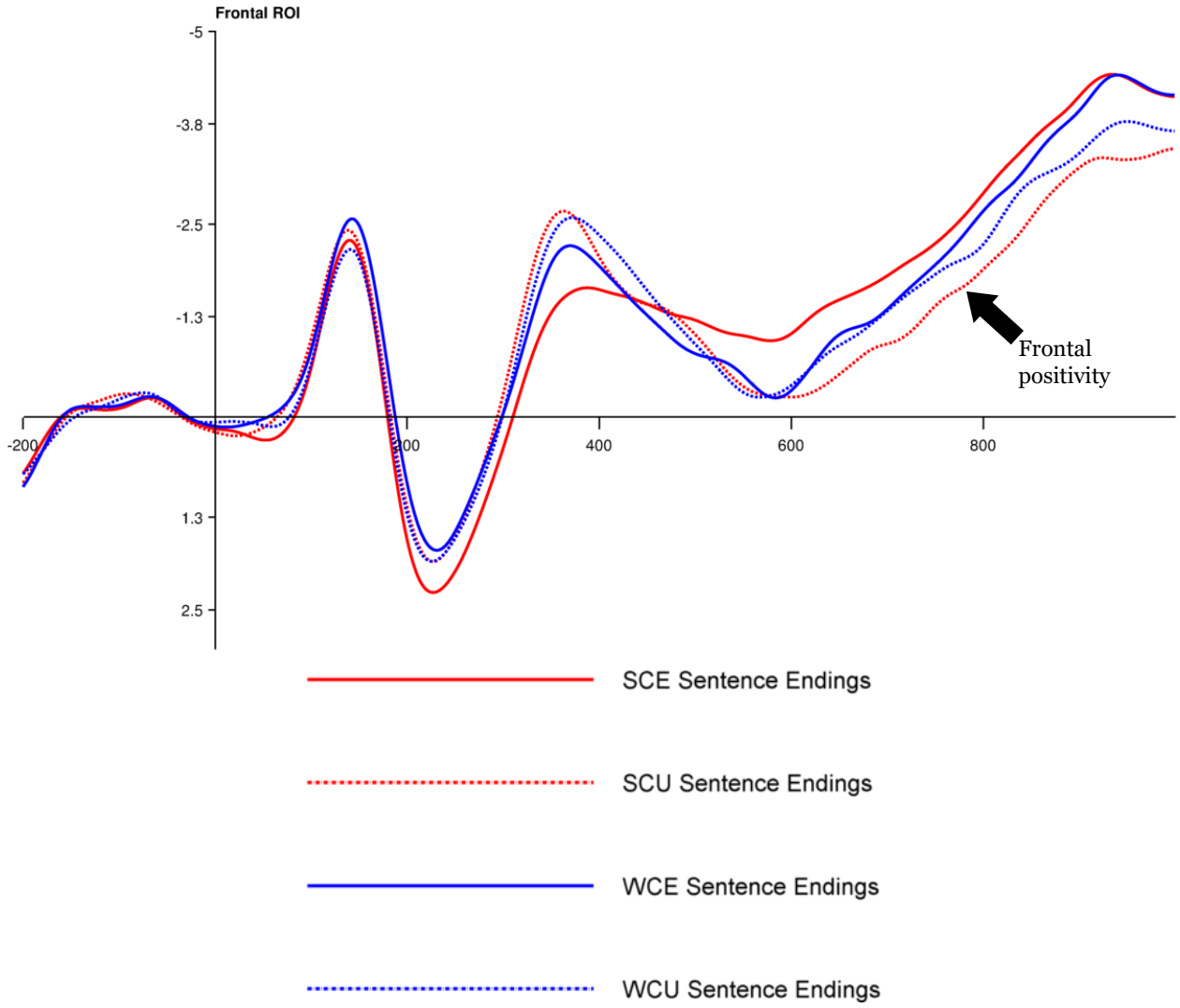


Figure 5: Grand average ERP waveforms for expected and unexpected endings to strongly and weakly constraining sentences at the Frontal Cluster of channels. The amplitude between 700-900 ms (the frontal positivity) is more positive for SCU, but also WCU, sentence endings.

Recognition Memory ERPs - Matches

Next, ERPs to 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 6. ERPs are time-locked to the onset of the test item. Visual inspection suggested an Expectancy effect on the N1 component, no amplitude

difference between conditions at the N400, and an interesting pattern of effects in the time window of the LPC, 500-800 ms.

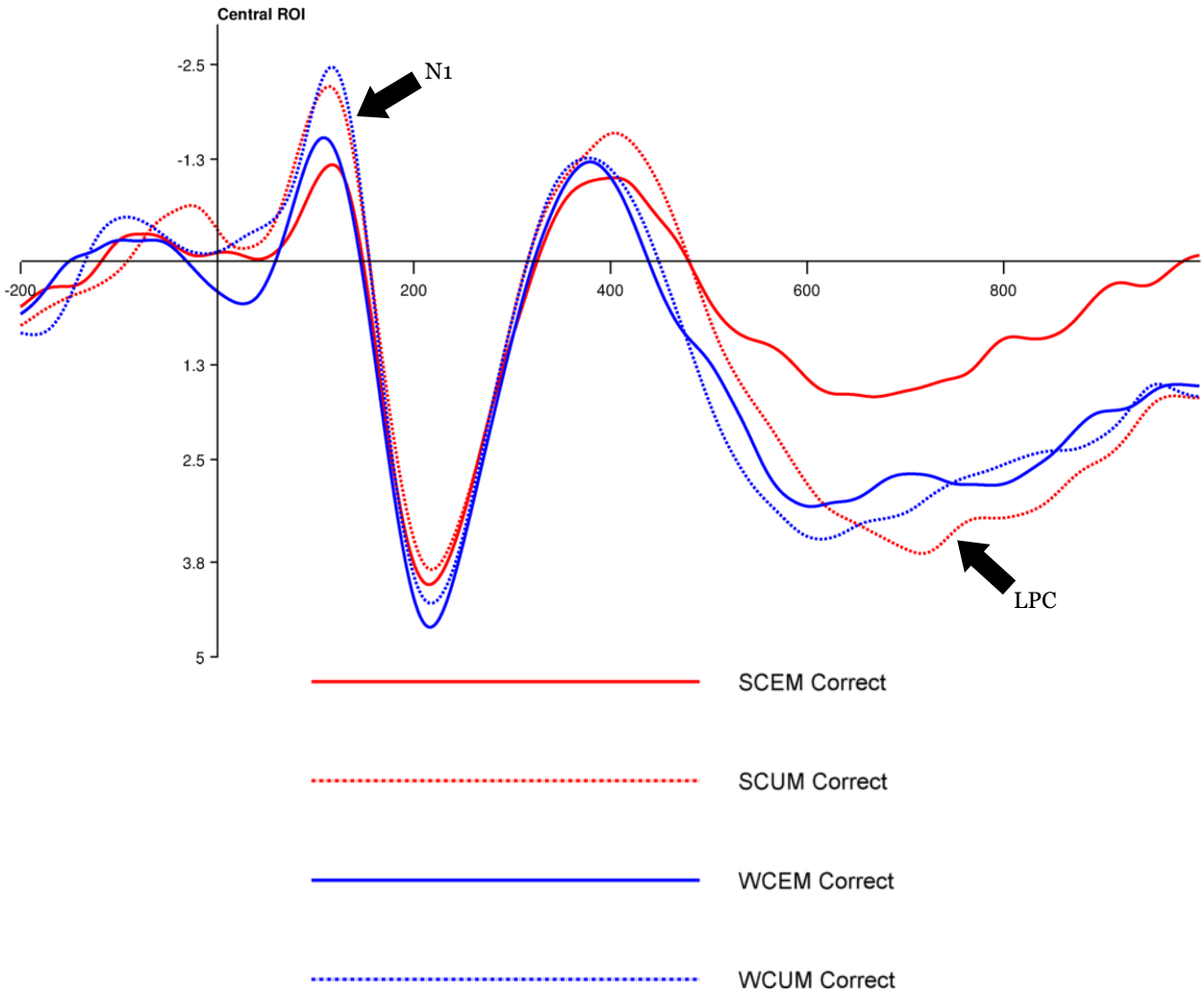


Figure 6: Grand average ERP waveforms to Match items during the memory test. SC = strong constraint, WC = weak constraint, E = expected, U = unexpected. An early N1 expectancy effect and a later LPC effect are observed.

The N1 can be fairly broadly distributed, and most often onsets earlier over frontal sites than over posterior sites (Mangun & Hillyard, 1991); thus, choosing a time window for analysis can be difficult. Here, the effect was assessed statistically by adapting the method used by Vogel & Luck (2000); namely, a 75-125 ms time window at the Central Cluster. Mean amplitudes in

this time window were submitted to a linear mixed effects model with a Fixed Effect of Expectancy (E vs. U). Constraint was collapsed across to increase statistical power. The reported effect was trending toward significance (t = -1.930, p = 0.057). A follow-up repeated measures ANOVA with two levels of Constraint (strong and weak) and Expectancy (expected and unexpected) found a main effect of Expectancy [$F(1,31) = 5.43$, $p = 0.027$], while Constraint [$F(1,31) = 0.59$, $p = 0.45$] and the interaction [$F(1,31) = 0.70$, $p = 0.41$] were not significant. Thus, there was some evidence that unexpected Match items elicited more negative N1 amplitudes. Violin plots of N1 amplitudes are presented in Figure 7.

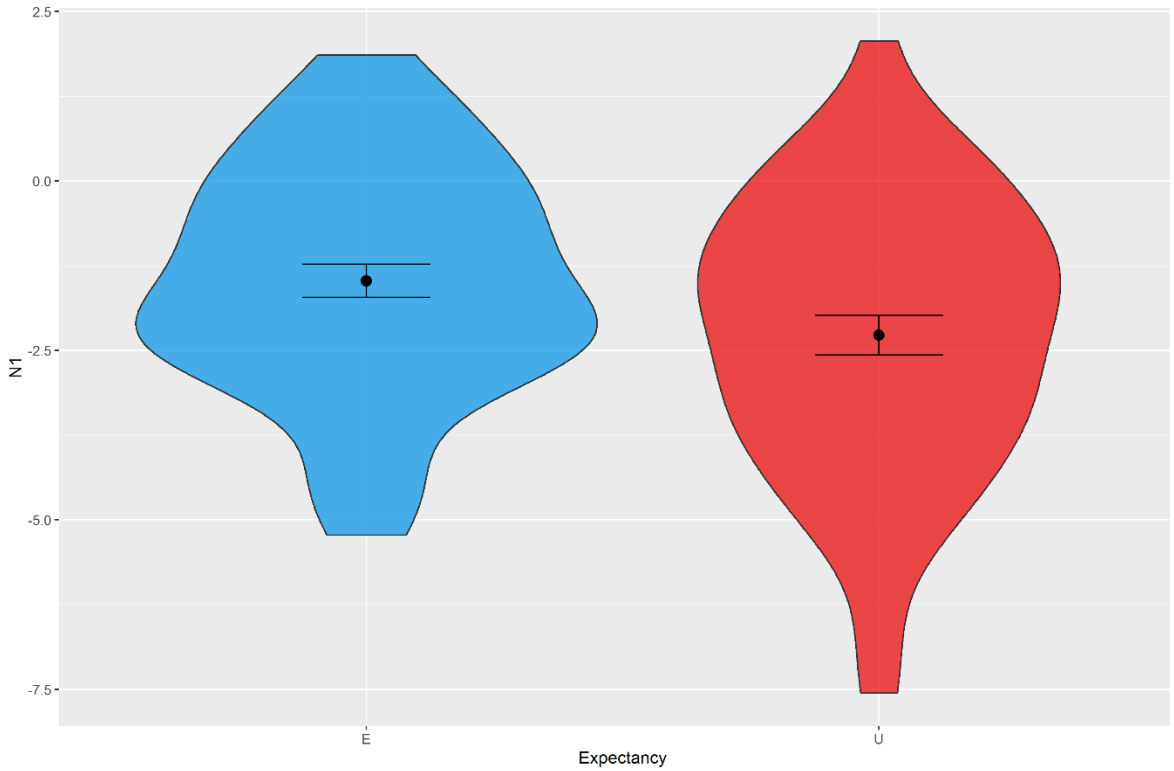


Figure 7: Violin plot of average N1 amplitudes for Expected and Unexpected Match test items. A difference is observed, though the effect is statistically trending .E = Expected, U = Unexpected.

Visually, there appeared to be no differences in N4 amplitudes between conditions. A repeated measures ANOVA with two levels of Constraint and Expectancy on mean amplitudes from 300-500 revealed no significant effects (all $p > .05$), confirming a lack of effects on the N400 between Match stimuli.

LPC mean amplitudes from 500-800 at the Left Parietal cluster (where LPCs are maximal) were submitted to a linear mixed effect model with condition (SCE, SCU, WCE, and WCU) as a factor. LPC amplitudes for SCE matches did not differ significantly from WCE matches ($t = 1.412$, $p = 0.1625$), but did significantly differ from SCU ($t = 2.439$, $p = 0.0174$) and WCU ($t = 2.450$, $p = 0.0174$) matches. However, when word frequency was included in the model, significant condition effects vanished and only frequency was significant ($t = -3.075$, $p = 0.003$). Importantly, in a mixed-effect analysis with N1 amplitude, word frequency was not significant ($t = -0.917$, $p = 0.36$), and the condition effect remained. To visualize word frequency effects, linear regressions between LPC amplitude and word frequency, as well as N1 amplitude and frequency, were plotted in Figure 8. A significant negative relationship ($t = -4.866$, $p = .000005$) between LPC amplitude and log frequency is observed, while a non-significant relationship ($t = -0.204$, $p = 0.8$) is found with N1 amplitudes. Thus, LPC amplitudes appeared to follow condition effects, but this effect was in fact driven by differences in word frequency between expected and unexpected test items. However, the previously described N1 effect was not driven by word frequency.

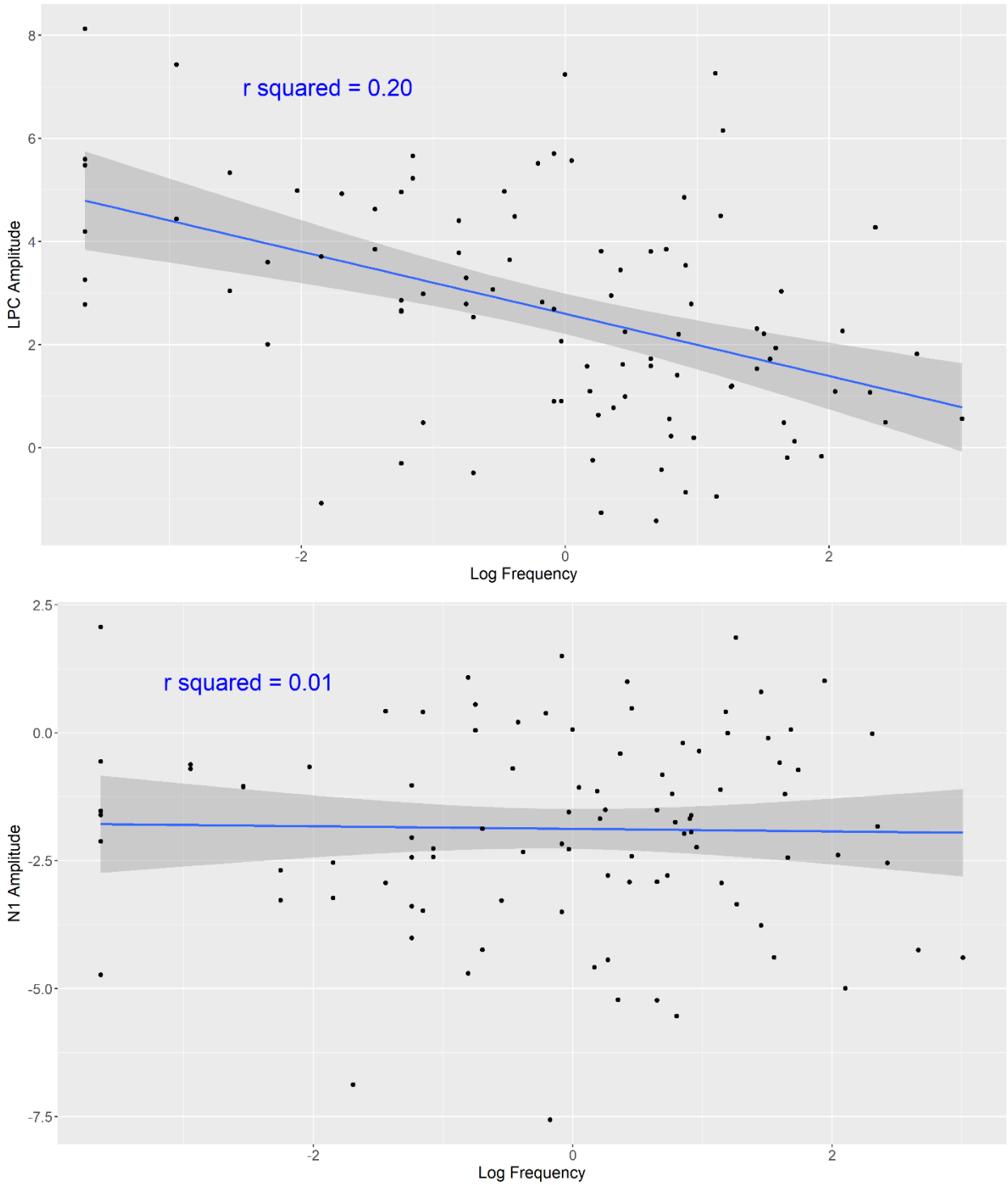


Figure 8: Linear fits of LPC amplitude (top) and N1 amplitude (bottom) to log word frequency. A significant negative relationship between LPC and frequency is observed, whereas there is no relationship between N1 amplitude and frequency.

Recognition Memory ERPs – Lures

The analysis of lures is more complicated, as there are several comparisons that can be made between different conditions. I chose to focus my analyses on two questions. First, does false alarming to a lure elicit the same neural responses as making a correct response to an old stimulus? To answer this, I compared false alarms to Lures with hits to Matches, as well as to correct rejections of New items, to determine if old-new effects were similar between Matches and Lures. Lures were compared with expected Matches, as the Lures were essentially expected Matches that weren't seen. Second, does false alarming to Lures vs. correctly rejecting them differ based on constraint? For this question, I compared correct and incorrect responses to strong constraint Lures with correct and incorrect responses to weak constraint Lures.

For the first analysis, I plotted grand average ERPs to expected Matches, Lures, and New items separately by constraint. The strong constraint ERPs are plotted in figure 9. Visual inspection suggests there was an old-new effect at the N400, but not the LPC; that false alarms to Lures did not differ from hits to Matches at the N400 or the LPC; and there was a surprising P2 difference between false alarms to Lures and hits to Matches. To assess the P2 effect statistically, mean amplitudes from 200-275 ms in the Right Parietal Cluster (where the effect was maximal) for Matches and Lures were submitted to a paired t-test. P2 amplitudes for Lures were more positive than for Matches, though the effect was small; $t(31) = 2.06$, $p = 0.048$.

Weak constraint ERPs are plotted in figure 10. The observed effects were substantially different from the strong constraint comparison – no P2 effect was apparent, and the pattern of effects differed at the N400 and the LPC. To assess these effects, 4 one-way ANOVAs with three levels of item type (Match, Lure, New) were computed. One tested amplitudes from 300-500 ms at the Central Cluster (the N400) for SC items, and another for WC items. The other 2 tests were performed on amplitudes from 500-800 ms at the Left Parietal Cluster (the LPC). Follow-up comparisons of significant ANOVAs were computed using Tukey's HSD test.

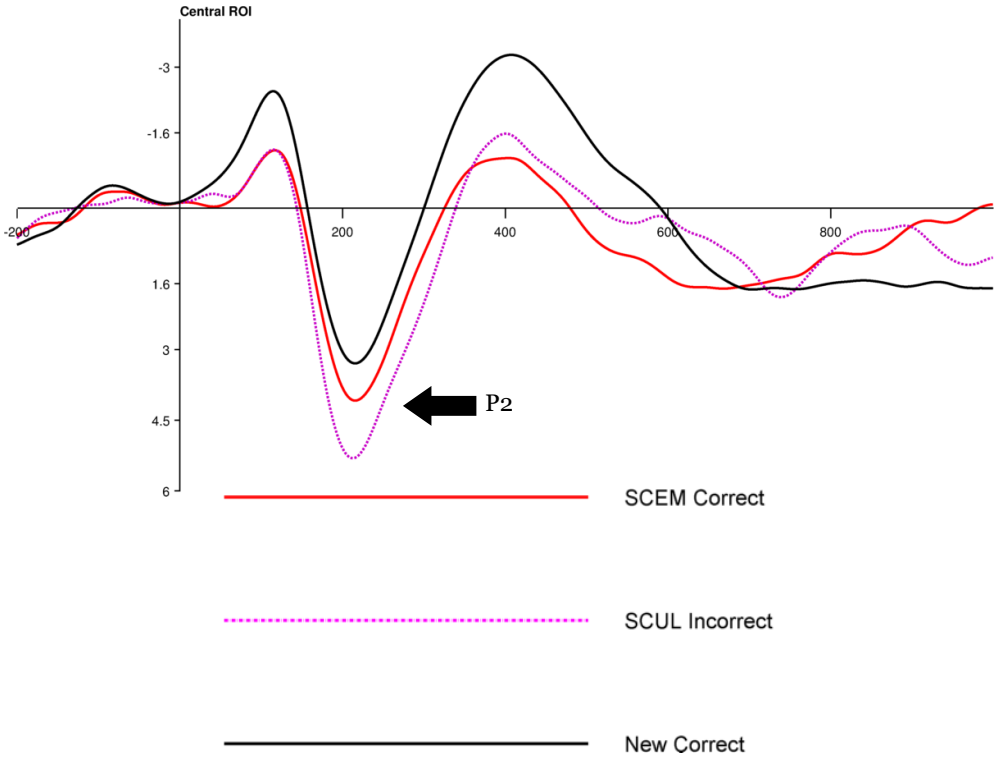


Figure 9: Grand average ERP waveforms during the memory test to Match hits, Lure false alarms, and New correct rejections from strong constraint sentences. SCEM = strong constraint expected Match, SCUL = strong constraint unexpected Lure.

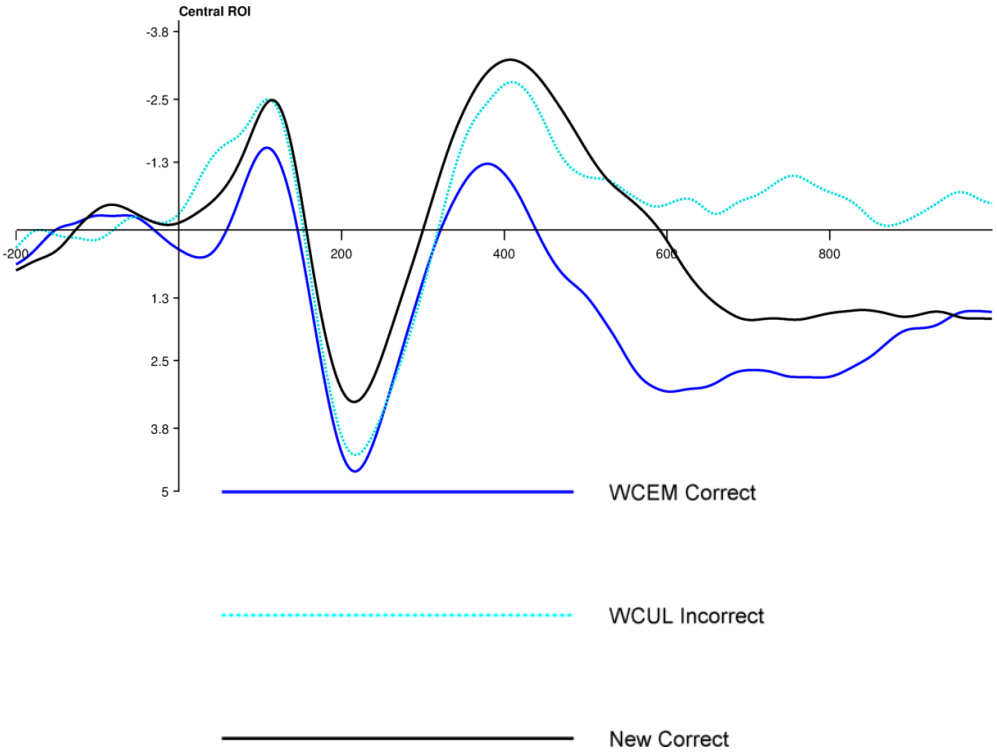


Figure 10: Grand average ERP waveforms during the memory test to Match hits, Lure false alarms, and New correct rejections from weak constraint sentences. WCEM = weak constraint expected Match, WCUL = weak constraint unexpected Lure.

	Strong Constraint			Weak Constraint		
	Match-New	Match-Lure	Lure-New	Match-New	Match-Lure	Lure-New
300-500 ms	3.09**	0.15	2.95**	3.77**	2.59*	1.18
500-800 ms	2.52*	0.62	1.9	5.23**	5.62**	-0.38

Table 2: Results from follow-up Tukey's tests in *z* values. * = *p* < .05, ** = p < .01.

All 4 of the one-way ANOVAs were significant; thus, post-hoc Tukey's HSD tests were run for each. The results of these tests are summarized in Table 2. There were old-new differences for both strong and weak constraint Matches and New items in the N400 and the LPC time windows. The Lures presented a different pattern; strong constraint Lures did not differ from Matches in either the N400 or LPC window, and showed an old-new effect in the N400 time window. In contrast, weak constraint Lures differed from Matches in both the N400 and LPC time windows, and did not significantly differ from New items in either window. Essentially, strong constraint Lures were more similar to Matches, whereas weak constraint Lures were more similar to New items.

Next, correct rejections and false alarms to strong constraint Lures were compared to weak constraint Lures. This comparison is plotted in figure 11 at the Right Parietal Cluster, where effects were maximal. There is an apparent correct/incorrect N400 difference for strong constraint Lures that is absent for weak constraint Lures, and an LPC difference for weak constraint Lures but not for strong constraint Lures. The N400 difference was tested by submitting mean amplitudes from 300-500 at the Right Parietal Cluster to a repeated measures ANOVA with 2 levels of Constraint (strong and weak) and Accuracy (correct and incorrect). The main effect of Constraint was not significant [$F(1,31) = 0.71$, $p = 0.41$]; however, Accuracy [$F(1,31) = 4.29$, $p = 0.047$] and the interaction [$F(1,31) = 5.47$, $p = 0.03$] were. A specific comparison of correct rejection and false alarms to strong constraint Lures was significant; $t(31) = 2.82$, $p = 0.008$.

A second repeated measures ANOVA was conducted on mean amplitudes from 500-800 at the Right Parietal Cluster. This test revealed no significant main effect of Accuracy [$F(1,31) = 1.30$, $p = 0.26$] or Constraint [$F(1,31) = 0.003$, $p = 0.96$], but a significant interaction [$F(1,31) = 4.81$, $p = 0.04$]. A follow-up comparison of correct rejections and false alarms for weak constraint Lures was significant; $t(31) = -2.7$, $p = 0.01$. Thus, there was a significant decrease in N400 amplitude when participants false alarmed to strong constraint Lures, whereas false alarming to weak constraint Lures was associated with smaller LPCs.

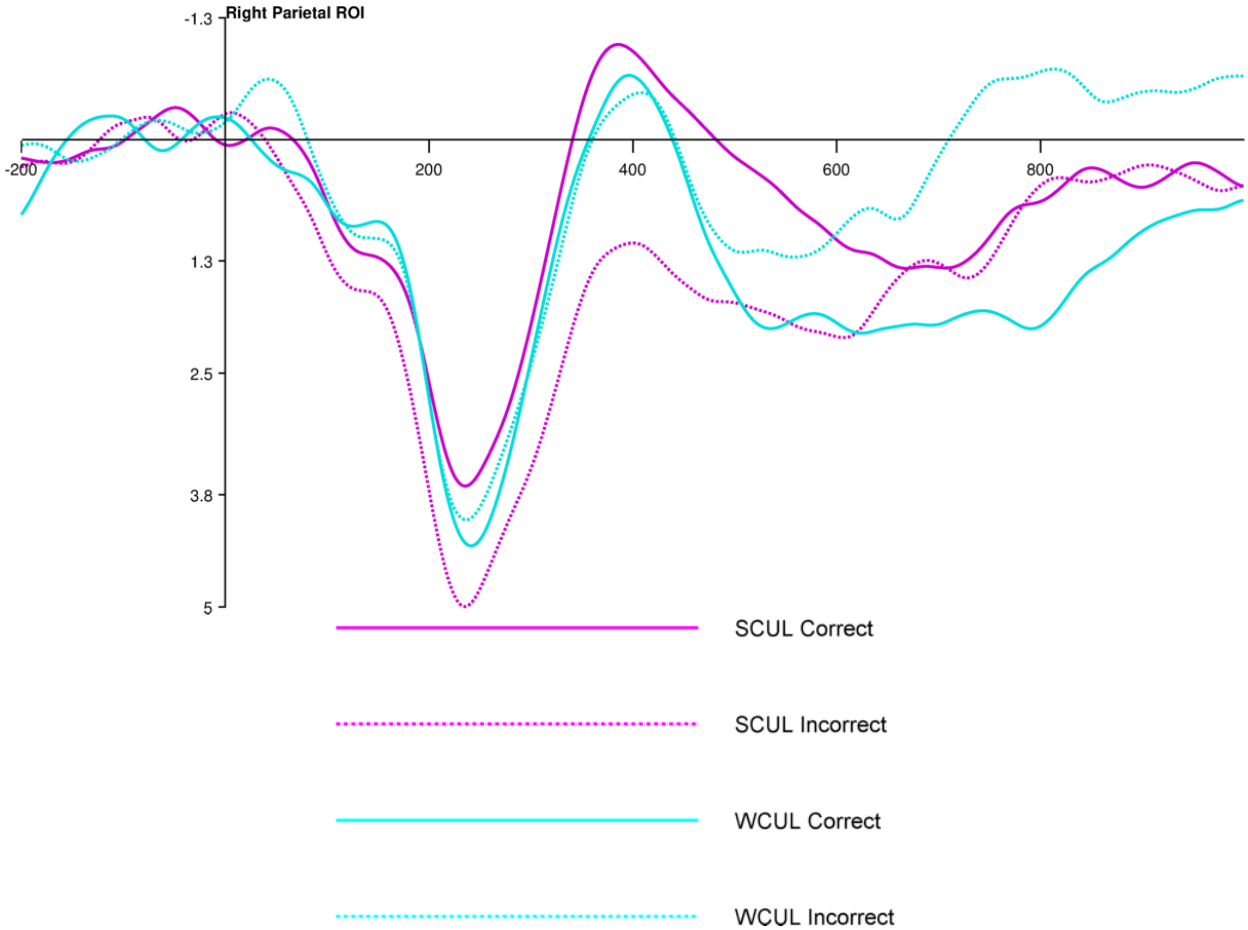


Figure 11: Grand average ERP waveforms elicited by false alarms and correct rejections to strong constraint and weak constraint Lures, plotted at the Right Parietal Cluster. False alarms to strong constraint lures elicited a reduced N400, whereas false alarms to weak constraint lures elicited a reduced LPC.

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 predicted endings that were never seen (Lures). Behaviorally, individuals were significantly more likely to false alarm to these Lures than to new items that weren't studied. This effect mirrors previous 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, in those tests, the lure items are usually closely related to an entire list of words. Here, predicted sentence endings were related only to one sentence in a block, and arguably less strongly semantically related than in the DRM case. Yet, a robust increase in false alarming to the lures compared to new items was found, suggesting prediction has fairly substantial downstream effects on memory.

This data also adds fairly strong evidence in favor of a predictive account, as opposed to an integrative account, of facilitation effects during language comprehension. The Lure items were never seen during the study period, yet participants false alarmed more to them, suggesting some level of activation of these items. This activation without observation could only occur if the sentence context led to a pre-activation of the word or features of the word. An alternative argument is that the facilitation to integration that would have occurred if the word had been read at the end of the sentence simply occurred when the word was read during the memory test, and this facilitation to integration led to increased false alarms. However, there would be multiple sentences, as well as a series of math problems, in between the predicted occurrence of the word and its occurrence on the memory test. Individuals would have to retain

the contextual information of many different sentences over this period in order to integrate the words on the test into this contextual representation, which seems unlikely.

Another explanation of this luring effect is that participants might have used the word presented during the test as a cue to perform a retrospective search through memory for a sentence that preceded it. When a Lure was presented, subjects were able to retrieve the associated sentence, and thus more false alarming occurred. Similar to the previously discussed Neely & Keefe (1989) hybrid prospective-retrospective processing theory, such an account would sidestep the necessity of pre-activation, as this retroactive search could be performed regardless of any pre-activation of the test item. However, in the case of the Lures, the associated sentence was completed by an unexpected word; for a retroactive search strategy to work, this unexpected word would have needed to be overwritten or ignored. Alternatively, individuals could have extracted some message-level "gist" from the sentences during encoding, and searched for this gist in memory during retrieval. Indeed, earlier work has argued for a Gestalt hypothesis, in which sentences are not represented in memory as a collection of independently associated items or words, but as a holistic concept or proposition (Anderson, 1974; Goetz, Anderson, & Schallert, 1981). While the gist encoded into memory would have to be relatively unaffected by the unexpected ending (and it is unclear why this would be), this account is difficult to rule out with the current data. A convincing counter to this explanation would be a lack of increase in false alarms to words that were not read, but were plausible completions to presented sentences (essentially, synonyms to the expected sentence endings), as these words would likely fit just as well to the gist in a retroactive search strategy.

There was no significant statistical evidence of a frontal positivity following unexpected endings to strongly constraining sentences compared to weakly constraining sentences. Additionally, highly unexpected endings were only trending towards significant difference from expected sentence endings. One interpretation is that frontal positivities simply were not

elicited in this study. Alternatively, in looking at the ERPs at the frontal locations, there seemed to be evidence of a frontal positivity to unexpected endings to weakly constraining sentences as well. This pattern differs from previous studies, and may be a result of differences in processing due to the nature of the task. When participants know they will be tested on their memory, they may engage in more predictive processing than in other language processing situations. Thus, individuals may have been encouraged to predict upcoming information even when reading weak constraint sentences, leading to a frontal positivity for these items. In support of this idea, Brothers, Swaab, & Traxler, (2017) reported larger frontal positivites to unexpected words when participants were instructed to predict upcoming information compared to when they simply read for comprehension. This could also explain the lack of constraint differences in behavioral performance; for both matches and lures, no constraint differences were found.

Numerically, hit rates were higher for unexpected Matches than for expected Matches, though this comparison was not statistically significant. A higher hit rate for unexpected words is in line with pre-experimental hypotheses given the distinctiveness of this information, and this finding was reported in a follow-up recognition memory test after reading sentences with expected and unexpected endings (Federmeier, Wlotko, Ochoa-Dewald, & Kutas, 2007). The distinctiveness of these items may have led to the surprising N1 ERP effect found during the memory test, with unexpected Matches eliciting greater (more negative) N1 amplitudes, regardless of constraint. N1 amplitude modulations are not often found or reported in electrophysiological studies of recognition memory. In one study, greater N1 amplitudes were elicited during incidental retrieval of words as compared to intentional retrieval (Kompus, Eichele, Hugdahl, & Nyberg, 2011), but in the current experiment subjects were explicitly instructed to remember items and try to recall them correctly. As discussed previously, unexpected sentence endings may have received greater depth of processing during encoding; however, ERP studies examining retrieval of words that were deeply or shallowly encoded have

not found modulations of the N1 (Allan, Robb, & Rugg, 2000; Rugg, Allan, & Birch, 2000; Rugg et al., 1998).

An alternative explanation of the reported N1 effect comes from experimental studies of visual attention and load. Previous work suggests that the N1 indexes allocation of attentional resources (Hillyard & Anilo-Vento, 1998; Mangun & Hillyard, 1991), and may reflect an early attention-dependent visual discrimination process that is sensitive to category membership (Hopf et al., 2002; Vogel & Luck, 2000). In one study (Curran, Tanaka, & Weiskopf, 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. Thus, during sentence reading, participants may have implicitly separated expected and unexpected endings into separate categories during sentence reading, and discriminated between these categories during recognition testing. Interestingly, Curran and colleagues found more negative N1 amplitudes for "in-category" stimuli (closer to the prototype) than for "out-category" stimuli; similarly, larger N1 amplitudes are found for stimuli that experts have expertise in (e.g. birds) compared to other stimuli (Tanaka & Curran, 2001). Here, larger N1s were found for unexpected stimuli. A potential explanation is that during sentence reading, participants essentially trained themselves on detecting unexpected sentence endings, leading to increased N1 amplitudes during the recognition test.

LPC amplitudes differed between strongly and weakly constrained expected Matches, with more positive LPCs for weak constraint Matches. This occurred despite any behavioral differences in hit rate for these items. Additionally, LPCs were more positive for weakly and strongly constrained unexpected Matches compared to strongly constrained expected Matches. One explanation for this finding is that strongly predicting a word actually leads to less encoding

and results in an impoverished representation downstream, as individuals devote less attentional resources to predicted stimuli, while weakly predicted words are in fact encoded more deeply. A recent ERP repetition study supports this account (Rommers & Federmeier, 2016); namely, expected sentence endings from strongly constraining sentences in fact show reduced ERP repetition effects compared to expected sentence endings from weakly constraining sentences.

However, mixed effect modeling and regression analyses revealed a relationship between LPC amplitude and word frequency. Unexpected match test items were unfortunately significantly lower frequency words than Expected match items. This difference in word frequency could have contributed to the numerical difference in hit rates between these conditions, as well as the observed LPC amplitude effects. This would be in line with the "mirror effect" identified in classic behavioral studies of memory, in which participants show an advantage for both hit rates and false alarms for lower frequency words compared to higher frequency words (Glanzer & Adams, 1985, 1990; Glanzer & Bowles, 1976). Previous ERP studies are in agreement with the LPC results reported here, with lower frequency words eliciting more positive LPCs during recognition memory testing (Bridger, Bader, & Mecklinger, 2014; Rugg, 1990; Rugg, Cox, Doyle, & Wells, 1995), potentially suggesting greater recollection. Thus, differences in word frequency could potentially explain the observed effects.

Even if the observed memory effects are in fact driven by word frequency, the results provide interesting insight into language comprehension and memory processes. Namely, several studies have demonstrated that when reading single words or the first words of a sentence, the amplitude of the N400 is modulated by word frequency, but as a sentence progresses, word frequency effects are reduced (Payne, Lee, & Federmeier, 2015; Van Petten & Kutas, 1990, 1991). As a semantic context is built over the course of the sentence, these message-level contextual constraints override lower level lexical properties. However, effects of

lexical information return when individuals observe the same words during a memory test. Word frequency effects in comprehension have been shown to be task-dependent (Fischer-Baum, Dickson, & Federmeier, 2014), and according to one view, individuals may extract information that aids performance in whatever task is currently underway (Norris, 2006). Attending to word frequency may not be informative during language comprehension, but participants in the current experiment may have utilized word frequency information to guide decisions during the memory test.

Statistically, word frequency appeared to explain the observed effects; however, it is possible that Expectancy of the words still played a role. Word frequency has also been shown to have very early effects on ERP amplitudes – namely, in the N1 time window (Hauk & Pulvermüller, 2004; Penolazzi, Hauk, & Pulvermüller, 2007; Strijkers, Costa, & Thierry, 2009). However, there was no statistical relationship observed between N1 amplitudes and word frequency here. Additionally, word frequency differences in previous studies are generally much larger (e.g. low frequency words < 10, high frequency words > 200). It seems surprising that the smaller differences in word frequency in the current experiment would still produce large magnitude ERP differences in the LPC, and no differences in the N1. However, this question can only be answered with certainty by replicating the experiment with more tightly controlled stimulus materials.

ERPs to strong and weak constraint Lures also revealed neurocognitive differences, despite a lack of a behavioral effect. False alarms to strong constraint Lures led to similar neural activity as a hit to an expected Match, except for a larger amplitude P2 response, which was not observed for weak constraint Lures. P2 effects have been observed in repetition studies, and may reflect perceptual priming (Rugg, Doyle, & Melan, 1993; Rugg, Doyle, & Wells, 1995); however, they are not always observed in recognition memory ERP experiments. This may be due to the length of study / test blocks, and experiments with fewer items per test may be more

likely to find P2 effects (Curran & Dien, 2003). However, P2 effects have also been found for recognition, but not lexical decision, of pseudowords (Curran, 1999); additionally, previously reported P2 effects have had a frontopolar topography, whereas the topography in Experiment 1 was fairly distributed. It is unclear exactly what mechanism this effect indexes; however, a visual priming account would be interesting, given that participants did not see the lure during sentence reading.

Comparing correct rejections and false alarms to strong and weak constraint Lures revealed different mechanisms potentially drove false alarming across constraint. Namely, false alarming to strong constraint Lures was associated with an N400 effect, while false alarming to weak constraint Lures was associated with an LPC effect. This could be viewed as evidence that strongly predicting a sentence ending word led to pre-activation of semantic features of this word, which led to increased conceptual priming, leading to endorsement on the memory test. This pre-activation did not occur as strongly in weakly constraining contexts, leading to less conceptual priming and more reliance on recollection. This result suggests that, under certain conditions, a dissociation in false alarming could be found. For example, if the recognition test were speeded, or if attention were divided during encoding, then participants might false alarm less to weak constraint Lures, but still false alarm to strong constraint Lures.

Overall, these results demonstrate that prediction during language comprehension has important downstream effects on recognition memory. Individuals had slightly enhanced memory for unpredicted information, as evinced by somewhat higher hit rates and an N1 and LPC effect during recognition testing, though the differences in word frequency raise questions about the LPC effects. Additionally, participants were more likely to false alarm to predicted, but never observed words compared to unpredicted and unstudied words. Behaviorally, there was no difference in hit or false alarm rates by constraint, perhaps due to task demands

encouraging prediction; however, ERPs revealed differences in neurocognitive mechanisms involved in false alarming to Lures.

CHAPTER 3: INVESTIGATING CONSEQUENCES OF PREDICTION ON RECOGNITION MEMORY FOR ANOMALOUS STIMULI

Experiment 1 demonstrated that individuals were more likely to false alarm to sentence ending words that were predicted but never observed, suggesting that information that was predicted was encoded into memory to some extent. An open question is whether this false alarm effect was driven solely by the pre-activation of the expected word, or if the presentation of the unexpected but plausible sentence ending led to some re-activation of the expected word. The frontal positivity following an unexpected word is thought to reflect a revision of the contextual representation mentally constructed up to that point. This revision process may take some time and effort, and the expected but unobserved word may become activated and maintained in working memory during this process, leading to some long-term encoding. In cases of syntactic ambiguities that are disambiguated late in the sentence, only individuals with high working memory span show a late positivity reflecting some revision process, suggesting working memory may be a critical component for ambiguity resolution (Friederici, Steinhauer, Mecklinger, & Meyer, 1998). A similar situation may exist revision following semantic violations, and this involvement of working memory in the revision process could lead to deeper long-term memory encoding. Indeed, maintenance of information in working memory has been shown to contribute to episodic memory formation (Ranganath, Cohen, & Brozinsky, 2005).

In contrast to unexpected but plausible sentence endings, the frontal positivity is not elicited by anomalous, implausible sentence endings – rather, a posterior positivity is elicited. This posterior positivity may be related to a P600, which is also thought to reflect a revision process, but it remains unclear how the posterior positivity is related to the frontal positivity, and the revision that occurs following these two types of stimuli may differ. One possibility is that anomalous endings lead to less revision than plausible words, as the word violates the

preceding context and can essentially be disregarded. On the other hand, an unexpected but plausible word requires reconsideration of the preceding context to arrive at a different meaning, leading to more effort and processing. The potential result is that observing an unexpected but plausible word will lead to false alarming to the expected word, whereas observing an anomalous word will reduce or eliminate the false alarm effect.

An alternative hypothesis is that anomalous words lead to equivalent or greater revision than unexpected words. Individuals might always perform some revision process if the expected stimulus is not encountered, regardless of the content of the unexpected stimulus. If this were the case, the false alarm rates between unexpected and anomalous words would not differ. Or, if the amount of revision undergone depends on the level of unexpectedness or surprisal, than more revision might occur for anomalous stimuli compared to unexpected ones. For instance, the amplitude of the N400 is larger for anomalous sentence endings compared to unexpected sentence endings (Kutas & Hillyard, 1980a); if this increased semantic processing led to greater revision, than encountering anomalous sentence endings would lead to a higher proportion of false alarms to the expected word compared to unexpected endings.

In Experiment 2 I investigated this question by presenting participants with strongly and weakly constraining sentences containing expected, unexpected but plausible, or anomalous sentence endings ("After dinner they washed the *dishes* / *dog* / *cigarette*"), and then testing their recognition memory for sentence ending words and predicted lures. This design not only allowed for replication of the first experiment, but also for testing if the same effects occur for anomalous stimuli. The critical comparison was between the false alarm rates for expected words when an unexpected but plausible word was read, compared to the false alarm rates when an anomalous word was read. Of additional interest was whether constraint would lead to behavioral differences in false alarm rates for expected words when anomalous words was read. This pattern of results for the unexpected but plausible words was not found in Experiment 1;

however, it is possible the process indexed by the posterior positivity differs by constraint. If the unexpected word is plausible, some revision may always be required, but anomalous words may only trigger some form of revision if a strong prediction was in place.

The hit rates for anomalous matches was also assessed. In Experiment 1, unexpected matches were somewhat more often correctly recognized than expected matches, though this difference was not statistically significant. Based on previous ERP and reading time data, semantically anomalous words could be considered even more distinct and attention-grabbing than unexpected but plausible words (Payne, Stites, & Federmeier, 2016). Thus, I predicted that hit rates for anomalous matches would be greater than for unexpected or expected matches. Additionally, by adding a more distinctive condition to the experiment, participants may pay less attention to unexpected but plausible endings. Accordingly, I predicted that the hit rates between unexpected but plausible and expected endings would not differ, and that the previously found N1 effect during test would be diminished for unexpected matches but would appear for anomalous matches.

ERPs were also recorded during the test phase as in Experiment 1. Like before, the N400 and LPC responses during the test phase were analyzed to assess which neurocognitive mechanisms supported successful recognition memory. Based on the results from Experiment 1, I hypothesized that anomalous matches would lead to even larger LPC responses than the unexpected matches. The amplitude of the LPC is thought to be graded based on the quality or number of details that are recollected (Leynes & Phillips, 2008; Vilberg, Moosavi, & Rugg, 2006). I predicted that participants would devote more attention to anomalous stimuli, leading to deeper encoding and greater LPC amplitudes during retrieval. However, the stimuli of Experiment 2 were constructed from the same pool as Experiment 1, and thus word frequency effects could still play a role in the results.

Methods

Participants

41 right-handed, native speakers of English with normal or corrected vision from the University of Illinois, Urbana-Champaign participated in the experiment and were paid $10 an hour for their participation. All participants had no 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. 3 participants were removed due to excessive EEG artifacts, leaving 38 participants in the final analysis.

Materials

The stimuli were comprised of 240 English sentences, half of which were strongly constraining (cloze > 0.68), while the other half were weakly constraining (cloze < 0.42). A third of the sentences ended with the expected or highest cloze sentence ending, a third ended with an unexpected but plausible sentence ending (cloze approximately 0), and the final third ended with a semantically anomalous ending word. Anomalous words were not cloze normed, but were presumably all cloze of 0, as these words were never produced during the original norming task, and were semantically incongruous with the preceding sentence. Thus, participants read 40 strongly constraining sentences with expected endings (SCE), 40 with unexpected endings (SCU), and 40 strongly constraining sentences with anomalous endings (SCA); this was also the case for the weakly constraining sentences (40 WCE, 40 WCU, 40 WCA). These stimuli were evenly split into 10 blocks (4 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.

As in Experiment 1, participants were tested on their memory following sentence reading. The stimuli for the memory test were sentence ending words from the study phase, as well as sentence medial words to ensure participants were paying attention. Table 3 provides an overview of the different types of test items, the number of each type, and examples of each type. Lexical properties (word frequency, concreteness, imageability, familiarity) were mostly controlled across test items; however, Unexpected test items did significantly differ from Expected test items in word frequency, as in Experiment 1 ($E = 107$, $U = 85$, $t = 1.9846$, $p = 0.049$). Anomalous stimuli did not significantly differ from Expected or Unexpected stimuli in word frequency. Similarly to Experiment 1, the stimulus list was constructed to avoid repetition of critical items, and thus presentation of stimuli was randomized within blocks, but not across blocks.

<table>
<tr><th colspan="2"><u>Encoding</u></th><th><u>Recognition</u></th></tr>
<tr><td rowspan="6"><u>Sentence (SC/WC)</u>

“Tim threw a rock and broke the…”

(120 SC / 120 WC)</td><td rowspan="2"><u>Expected (E)</u>

“window”

(40 SC / 40 WC)</td><td><u>Match (M)</u>
“window”
(20 SC / 20 WC)</td></tr>
<tr><td><u>New (N)</u>
“leather”
(20 SC / 20 WC)</td></tr>
<tr><td rowspan="2"><u>Unexpected (U)</u>

“camera”

40 SC / 40 WC)</td><td><u>Match (M)</u>
“camera”
(20 SC / 20 WC)</td></tr>
<tr><td><u>Lure (L)</u>
“window”
(20 SC / 20 WC)</td></tr>
<tr><td rowspan="2"><u>Anomalous (A)</u>

“novel”

40 SC / 40 WC)</td><td><u>Match (M)</u>
“novel”
(20 SC / 20 WC)</td></tr>
<tr><td><u>Lure (L)</u>
“window”
(20 SC / 20 WC)</td></tr>
</table>

Table 3: Overview of experimental materials of Experiment 2.

Procedure

The experimental procedure of Experiment 2 was essentially identical to the procedure of Experiment 1. The main difference was that Experiment 2 had 10 study-test blocks. The paradigm in each block – sentence reading, followed by math, followed by a memory test – remained the same.

EEG Recording and Processing

The EEG recording procedure of Experiment 2 was identical to that of Experiment 1. Eyeblinks were once again corrected using AMICA, and artifacts were removed afterwards with an additional scan. Due to the difficulty of including expected, unexpected, and anomalous trials in one experiment, while not fatiguing participants with an exceptionally long procedure, the trial numbers of each bin (20) were somewhat low for an ERP experiment. In an attempt to remedy this, a higher number of subjects were run than in Experiment 1; however, the signal to noise ratio of ERPs may be lower than desired.

Results

Behavioral Data

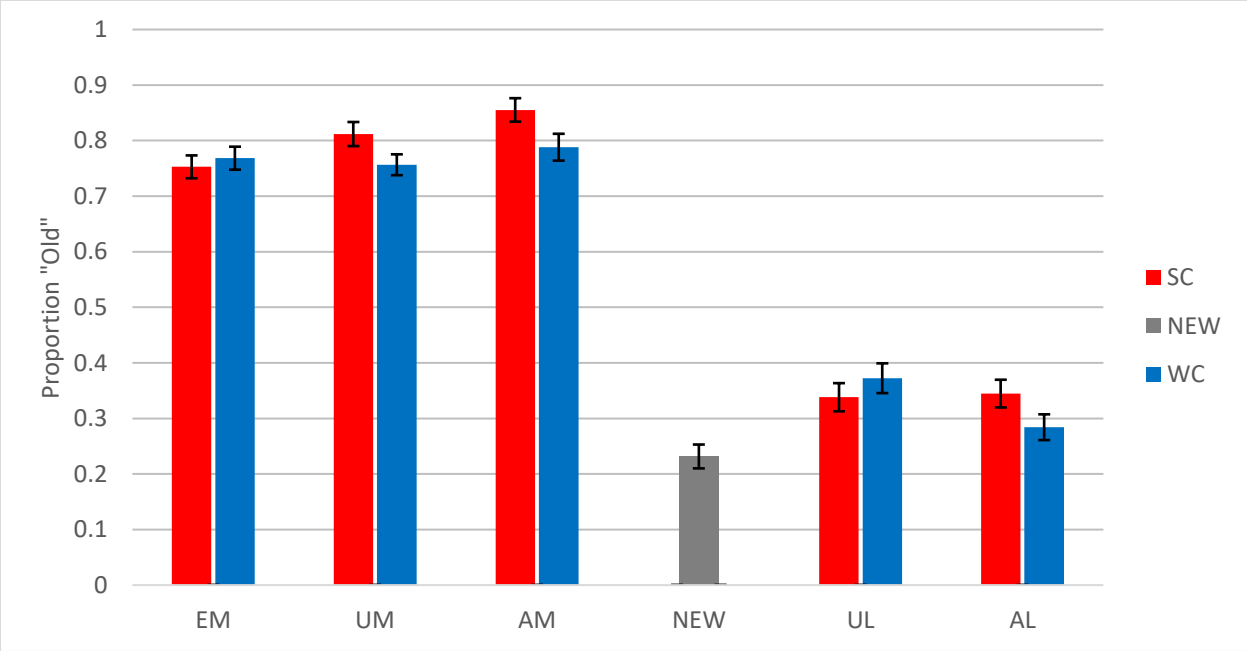


Figure 12: Recognition memory accuracy in Experiment 2. Proportion "Old" responses are plotted on the Y axis. SC = strong constraint, WC = weak constraint, EM = expected match, UM = unexpected match, AM = anomalous match, UL = unexpected lure, AL = anomalous lure.

Proportion "Old" responses is plotted in figure 12. As in Experiment 1, individuals successfully performed above chance. Two intriguing patterns are apparent: first, the increased rate of false alarms to Lures found in Experiment 1 was replicated. However, an interesting difference emerged for Lures following an anomalous ending, in that there was greater false alarming for strong constraint items compared to weak constraint items. Second, a linear increase in accuracy was found for strong constraint Matches, with anomalous Matches > unexpected Matches > expected Matches. No apparent differences in accuracy were observed across the 3 item types for weak constraint.

As in Experiment 1, behavioral performance was assessed statistically with mixed logistic regression models. The first model compared responses to New items to responses to each of the 4 different types of Lure items. Participants significantly false alarmed more to SCUL ($z = 2.635$, $p = 0.008$), WCUL ($z = 3.036$, $p = 0.002$), and SCAL items ($z = 2.864$, $p = 0.004$), but not to WCAL items ($z = 1.708$, $p = 0.09$). Model comparison identified a significant contribution of Condition to the model ($X^2 = 48.768$, $p < 0.001$). Thus, the false alarm effect for Lures found in Experiment 1 was replicated in Experiment 2, and also extended to SCAL items. False alarms between New items and WCAL items did not differ significantly; however, when word frequency was included in the model, this difference did become significant ($z = 3.095$, $p = 0.002$), with a significant effect of word frequency ($z = 6.565$, $p < 0.001$) and a significant difference in the model comparison for word frequency ($X^2 = 38.041$, $p < 0.001$). Clearly, word frequency contributed to memory responses as expected, and potentially reduced false alarms for WCAL items.

A second model compared responses for Match items with Constraint, Expectedness, and the interaction (C*E) as fixed factors. Expectedness was modeled with contrast coding, i.e. Anomalous matches were coded as 0.5, and Expected and Unexpected matches were coded as -

0.25. Significant effects were found for Expectedness (z = 2.968, p = 0.003) and Constraint (z = -2.126, p = 0.034), with no significant effect of the interaction (z = -1.234, p = 0.2). Model comparisons revealed significant contributions of Expectedness (X^2 = 8.7124, p = 0.003) and Constraint (X^2 = 4.4309, p = 0.036). When word frequency was added to the model, the previous effects of Expectedness and Constraint remained significant, and the fixed effect of frequency was also significant (z = -2.009, p = 0.045); however, model comparison did not show a significant improvement to the model when frequency was added (X^2 = 5.3267, p = 0.2554). Responses to Match items are plotted in Figure 13. Proportion "Old" responses increase linearly for SC Matches, but not for WC Matches, leading to the significant main effect of Constraint. Responses also increased for Anomalous Matches across both levels of Constraint.

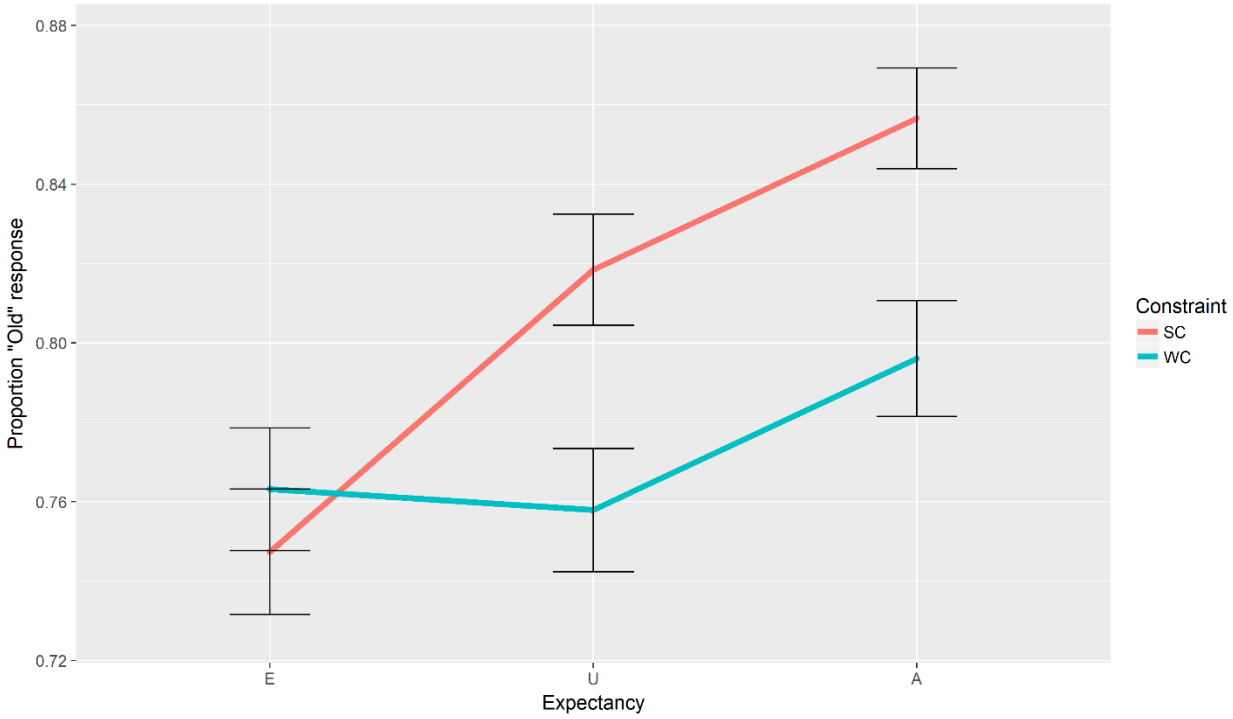


Figure 13: Proportion "Old" response for Match items by Expectancy and Constraint. Old responses monotonically increase with decreasing Expectancy for Strong Constraint Match items. SC = Strong Constraint, WC = Weak Constraint, E = Expected, U = Unexpected, A = Anomalous.

Sentence Final Word ERPs

Grand average ERPs to sentence endings at the Central Cluster are plotted in figure 14. The N400 is graded by expectancy, and anomalous endings elicit greater N400s than unexpected endings, replicating previous findings. To analyze these effects statistically, N400 amplitudes (averaged activity from 300-500 ms at the Central Cluster) were submitted to a mixed effect model comparing SCE vs. WCE, WCE vs. Unexpected (SCU + WCU), and WCE vs. Anomalous (SCA + WCA) endings as a Fixed Effect, with random intercepts and slopes for participants and random intercepts for items. Model comparison showed a significant contribution of Condition to the model ($X^2 = 44.633$, $p < 0.001$). There were significant differences between WCE and U endings ($t = -4.253$, $p < 0.001$), as well as WCE and A endings ($t = -7.292$, $p < 0.001$). Curiously, there was no significant difference between SCE and WCE endings at the N400 ($t = 0.868$, $p = 0.39$). This is a robust difference in the literature, and presumably this difference will become significant as more data is collected. Thus, N400 amplitudes were graded by expectancy, as has been previously found in the literature.

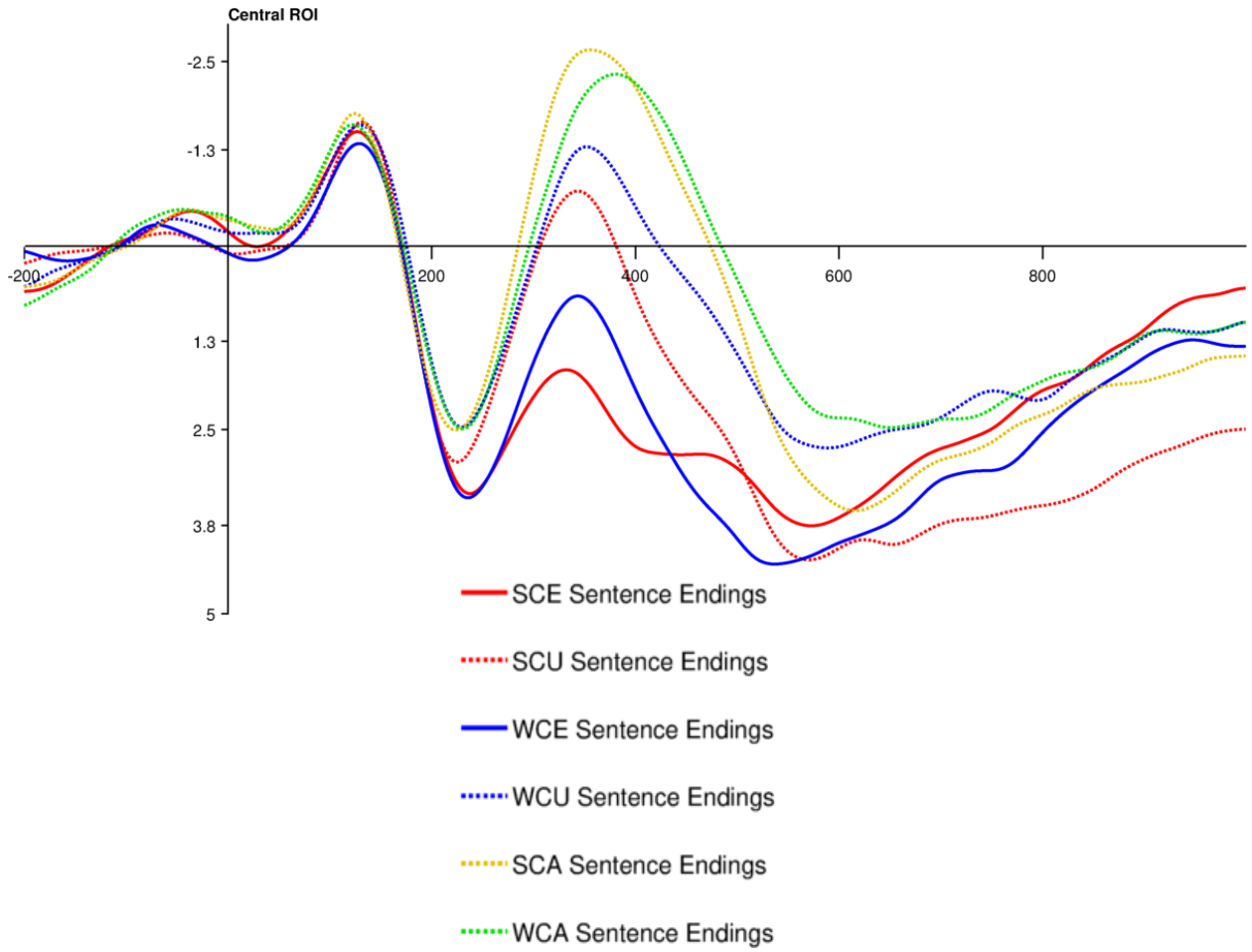


Figure 14: Grand average ERP waveforms for expected, unexpected, and anomalous endings to strongly and weakly constraining sentences at the Central Cluster of channels. The N400 is graded based on expectancy.

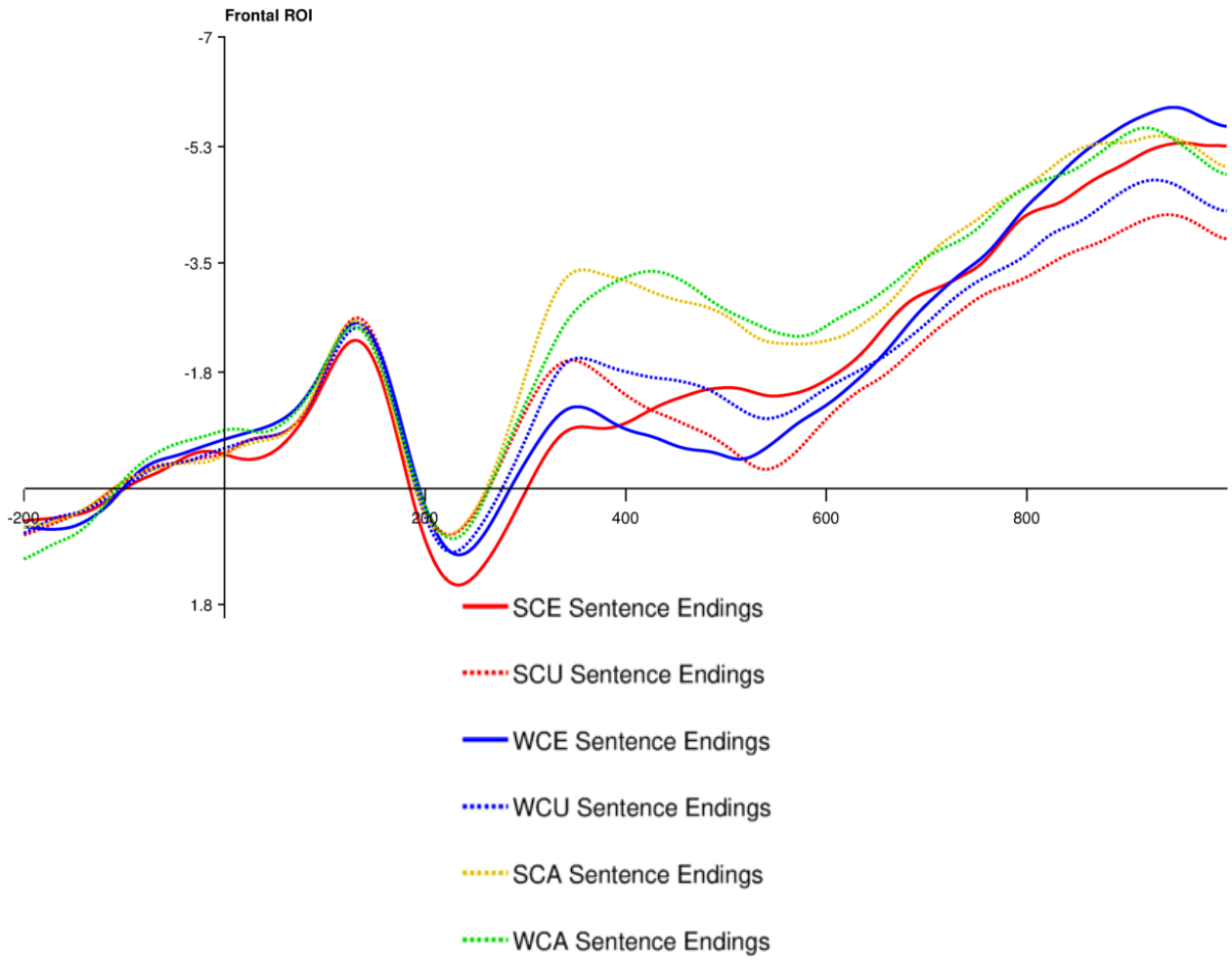


Figure 15: Grand average ERP waveforms for expected, unexpected, and anomalous endings to strongly and weakly constraining sentences at the Frontal Cluster of channels. Only a small frontal positivity effect is apparent.

Sentence ending ERPs at the Frontal Cluster are plotted in Figure 15. As in Experiment 1, there was no evidence for a larger positivity for SCU sentence endings compared to WCU sentence endings, though once again WCU endings appeared to display a similar amplitude positivity. A linear mixed effect model found no statistically significant difference in frontal positivity amplitudes between SCU and WCU sentence endings ($t = 0.688$, $p = 0.49$).

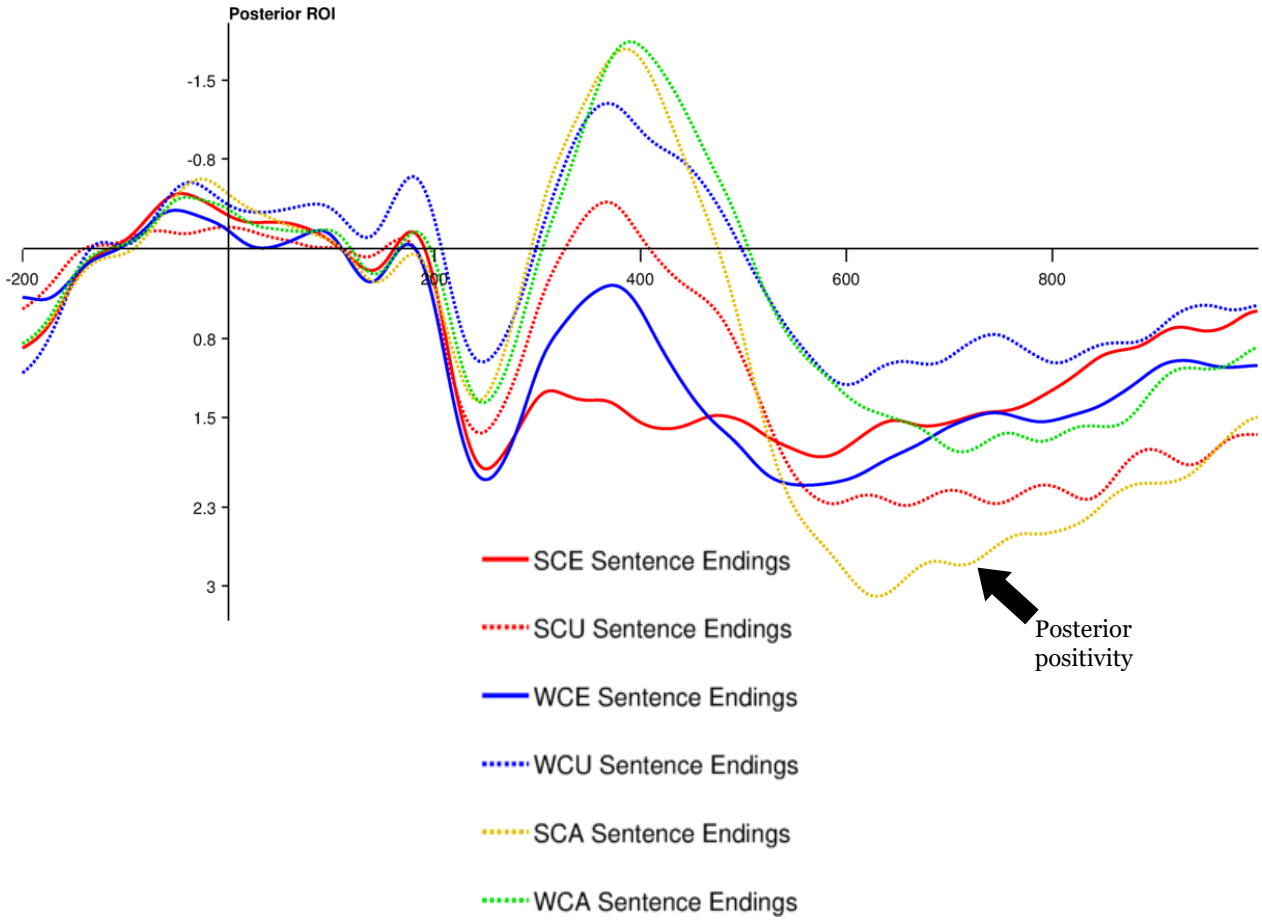


Figure 16: Grand average ERP waveforms for expected, unexpected, and anomalous endings to strongly and weakly constraining sentences at channel MiOc. A posterior positivity to SCA items is apparent.

Figure 16 presents the grand average ERPs for sentence endings at channel MiOc, the most posterior channel. A posterior positivity is present for SCA sentence endings, replicating previous findings. A linear mixed effect model confirmed a significant difference between SCA and WCA sentence endings (t = 3.141, p = 0.002), as well as a significant contribution of Condition to the model (X^2 = 11.993, p = 0.007).

Recognition Memory ERPs - Matches

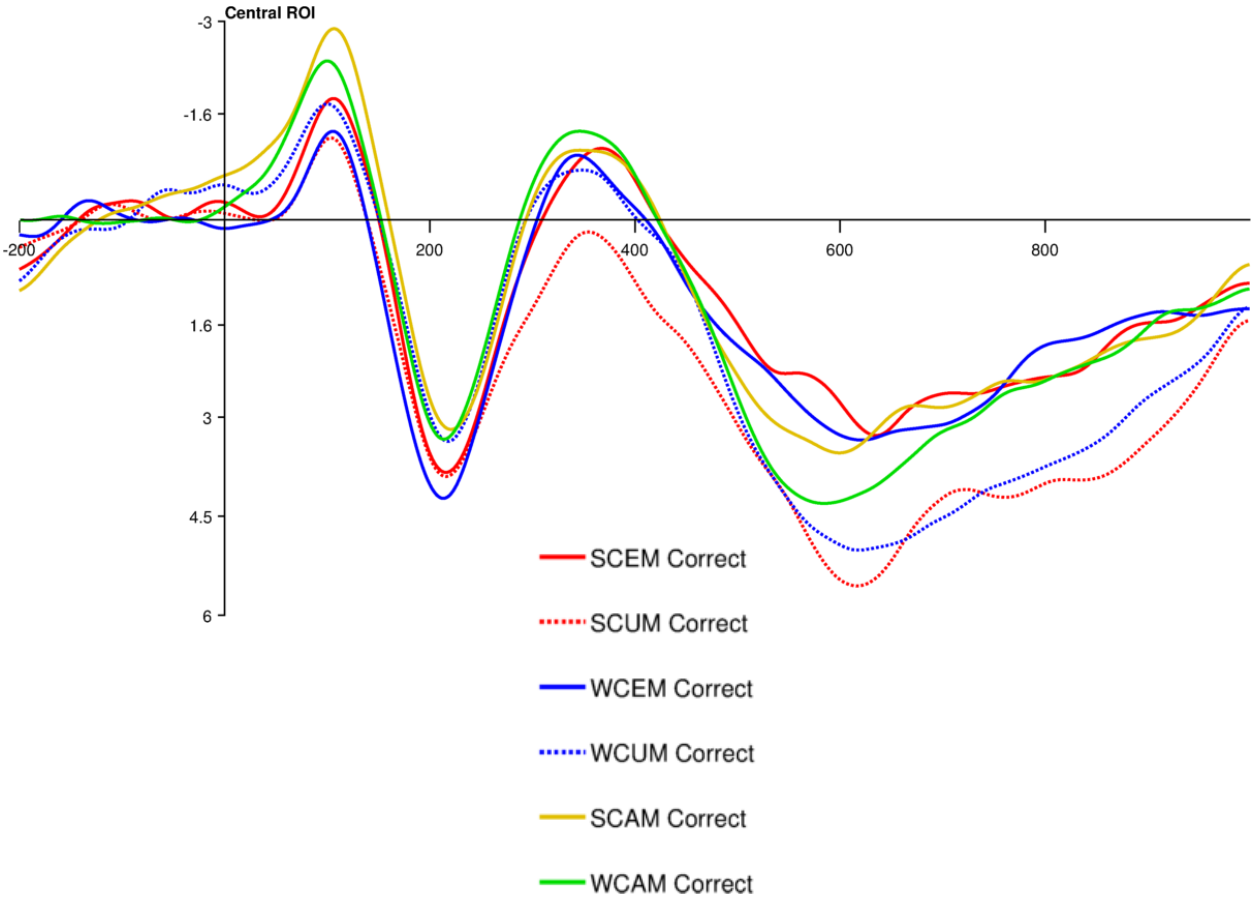


Figure 17: Grand average ERP waveforms to Match items during the memory test. SC = strong constraint, WC = weak constraint, E = expected, U = unexpected, A = anomalous. An N1 effect is present for anomalous matches, and an LPC effect for unexpected matches are observed.

Grand average ERPs to Match items during the memory test are plotted in figure 17. As predicted, the N1 effect previously found for unexpected Matches in Experiment 1 was essentially diminished, while an N1 effect for anomalous Matches emerged. N1 amplitudes (100-175 ms at the Central Cluster) were submitted to a repeated measures ANOVA with factors of Constraint (SC and WC) and Expectancy (Expected, Unexpected, Anomalous). There was a significant main effect of Expectancy ($F = 5.98$, $p = 0.004$), with no significant effect of Constraint ($F = 0.915$, $p = 0.25$) and no significant interaction ($F = 2.48$, $p = 0.09$). Thus, N1s

elicited by Anomalous Matches significantly differed from N1s from Expected and Unexpected Matches.

As in Experiment 1, LPC amplitudes were submitted to a linear mixed effect model. As before, this test revealed significant differences between SCEM amplitudes and SCUM ($t = 2.152$, $p = 0.03$) and WCUM ($t = 2.482$, $p = 0.015$) amplitudes; however, there was no significant difference between SCEM and WCEM amplitudes ($t = 0.290$, $p = 0.77$), and no difference between SCEM and either SCAM ($t = 0.499$, $p = 0.62$) or WCAM ($t = 0.951$, $p = 0.34$) amplitudes. Importantly, when word frequency was added to the model, significant condition differences vanished and only word frequency was significant ($t = -4.261$, $p < 0.001$).

Discussion

The results of Experiment 2 replicate and extend the findings from Experiment 1. Namely, participants once again showed greater false alarm rates to sentence ending words that were predicted, but never observed. This effect was found both when the sentence ending word that was read during encoding was unexpected but plausible, as well as when the word was semantically anomalous. These results suggest this false alarm effect is fairly robust across experimental designs and stimuli.

An important difference emerged in Experiment 2, in which there was greater false alarming to Lures following anomalous endings to strong constraint sentences vs. weak constraint sentences. In fact, when word frequency was not included in the model, false alarm rates between WCAL items and NEW items did not statistically significantly differ. Clearly, the plausibility of unexpected information played an important role in the future memory for the expected information. Additionally, the observed pattern of false alarms aligned with the observed post-N400 positivity ERP pattern; namely, while both strong constraint and weak

constraint unexpected endings elicited frontal positivities and produced luring behavior, only strong constraint anomalous endings elicited posterior positivities, and produced greater luring behavior compared to weak constraint anomalous endings. These results support the proposition that post-stimulus revision processes indexed by late positivities lead to encoding of predicted information into long-term memory. However, participants did still exhibit greater false alarms to WCAL items compared to NEW items, suggesting a pre-activation account of the false alarm effect cannot be ruled out. Potentially, both mechanisms – pre-activation and revision – could play a part in encoding predicted but unobserved information.

While the pattern of observed results seem to agree, in other ways they are in conflict. One argument based on the frontal positivity results is that the context of the memory task encouraged participants to form predictions of upcoming information, even in weakly constraining sentences. However, if the memory task encouraged individuals to predict in weak constraint sentences, then unexpected but plausible endings should have elicited a frontal positivity – which they did to an extent – and anomalous endings should have elicited a posterior positivity – which they did not. An as of yet unconsidered account is that the mechanisms reflected by the late positivities are not obligatory and not specifically post-prediction processes; that is, individuals may not have generated strong predictions when reading weakly constraint sentences, but the memory task may have encouraged participants to engage revision processes following unexpected information, and thus they elicited a frontal positivity. These components may reflect top-down processes which are engaged based on certain goals or task demands, which may not always be the critical focus of an experiment, leading to varied observations of the components; indeed, there is substantial variability in the appearance of these components across the literature (Van Petten & Luka, 2012). Further studies utilizing various task manipulations will need to be performed to determine what experimental demands lead to late positivity effects, and whether violations of strong predictions are necessary for them to occur. Alternatively, if the late positivities reflect revision

processes not specifically related to linguistic prediction, but are related to general prediction mechanisms, then they should potentially be observed in ERP experiments in other domains where predictions are violated.

The linear increase in hit rates by lack of fit to the sentence for strong constraint Matches gives further argument in favor of a predictive account as opposed to an integrative one. Anomalous words should be equally difficult to integrate following a strong or weakly constraining sentence, and thus should be equally distinctive or surprising. However, if individuals had strong predictions in place, then an anomalous ending would lead to greater prediction error, which would not occur if weak predictions, or no prediction at all, had been generated. It is unclear why the increased hit rate for strong constraint unexpected but plausible Matches compared to weak constraint was not found in Experiment 1. One explanation is simply a statistical one; the effect was present, but due to noise was not observed. However, the inclusion of anomalous items in Experiment 2 may have changed participants' behavior. The proportion of items that were expected was lower in Experiment 2 compared to Experiment 1, which could have led participants to predict more strongly in strong constraint sentences, but adopt more of a "wait-and-see" approach for weak constraint sentences.

In looking at ERPs at test, a similar pattern emerged as in Experiment 1, though there were interesting differences – namely, an N1 effect was found, but now for Anomalous matches as opposed to Unexpected matches. Thus, across experiments the effect on the N1 tracked the most incongruous, and potentially most salient, stimuli. The results here are in accord with the previous explanation of N1 effects – specifically, that individuals categorized stimuli during encoding, and utilized this categorization during encoding. That Anomalous matches were now the most negative would mean that participants trained themselves on categorizing Anomalous vs. not Anomalous stimuli. This gives insight into task-relevant effects, in that introducing Anomalous stimuli into the experiment very well could have changed processing for the

Unexpected and Expected stimuli, even though those stimuli were not changed across experiments.

The LPC pattern of Experiment 2 were also somewhat similar to Experiment 1, with important differences. As before, expected Matches from strongly constraining sentences elicited lower amplitude responses compared to unexpected Matches from either constraint; however, unlike in Experiment 1, expected Matches from weakly constraining sentences did not differ from those from strongly constraining sentences. However, as before, including word frequency into the model abolished significant condition effects. This was somewhat puzzling, as the weak constraint items only differed slightly, in that there were 4 less trials in Experiment 2 compared to Experiment 1. One possibility is that this result was due to a statistical error in model fitting, leading to all variance being accounted for by frequency, but is unclear why this would happen. Currently, the results are difficult to explain without a replication with more carefully controlled experimental materials.

CHAPTER 4: ELECTROPHYSIOLOGICAL PRECURSORS OF PREDICTION

The results of Experiments 1 and 2 demonstrated that prediction during language comprehension can have downstream consequences on cognition. In Experiment 3, I investigated a different question – whether predictive processing can be measured or identified prior to the onset of the predicted information. Given that prediction can be defined as anticipatory pre-activation of some aspects of upcoming information, it seems logical that there could be neural evidence of this pre-stimulus predictive processing. Finding behavioral evidence could potentially be difficult, as having participants give a response or make a judgment prior to the predicted stimulus could disrupt or alter predictive processing. Therefore, EEG was once again used to measure neural activity without behavioral output. Additionally, as described above, sentences are useful for providing a predictive context, but are also multifaceted and complex, and potentially lead to different forms of prediction depending on the information in the sentence. Thus, a simpler paradigm of word stem completion was used in Experiment 3 to specifically investigate generating predictions.

The word stem completion task has primarily been used to investigate implicit memory, or memory for events not available to conscious awareness (Graf & Mandler, 1984; Roediger, 1990). A common implementation is to have participants study a list of words, and later give them a list of stems (the first few letters of a word, e.g. "tru___") with the instruction to complete the stem with the first word they can think of. Some proportion of these items will be stems for words that were previously studied, and generally participants will be more likely to complete those stems with the studied words, even if they are not consciously aware that they completed the stem with a word they had studied. A plethora of behavioral and electrophysiological studies have utilized this paradigm for the study of implicit memory (Allan, Doyle, & Rugg, 1996; Bassili, Smith, & MacLeod, 1989; Fay, Isingrini, & Clarys, 2005; Fay,

Pouthas, Ragot, & Isingrini, 2005; Geraci, Hamilton, & Guillory, 2015; Paller, 1990; Roediger, Weldon, Stadler, & Riegler, 1992).

In Experiment 3, I focused on the generation aspect of word stem completion. Essentially, word stem completion is a language production task – individuals must produce some word based on a cue that they are given. In prior work, the goal was to see whether the words that participants produced to complete stems matched words that were previously studied, which would be evidence for implicit memory of the studied words guiding language production. Here, the goal was to first examine what words individuals chose to complete stems "spontaneously", or without a previous study period, in order to have an idea of how probable certain stem completions are. A norming task was given to a group of individuals, similar to the Cloze task described earlier – the only instruction was to complete a series of stems with the first word they could think of. Essentially, if the majority of individuals gave a particular completion to a certain stem, then that stem served as a good cue for that word, just as a certain sentence context might serve as a good cue for a particular ending. This data quantified the predictability of upcoming completions based on the stem cue. With this information, I then tested how neural activity at the time of presentation of the *stem* differed based on the predictability of an upcoming *completion* that was presented, thus allowing for investigation of mechanisms of predictive processing prior to the onset of the predicted information.

Why use a word stem completion task to investigate mechanisms of prediction? As mentioned above, this theoretically allowed for tighter control of potential extraneous factors that could play a role at the time of the cue than in the sentence processing case. For example, neural activity might differ at a medial word in a sentence prior to an upcoming ending word based on the predictability of the final word, but that activity might additionally be modulated by the frequency, class, or predictability of the medial word, the length of the sentence up to that point, the semantic associations between words in the sentence, or other potential factors.

While it is unknown if processing of a stem is heavily influenced by its frequency or familiarity, it seems unlikely that semantic associations between words would play a large role, given only the stem is presented without any other associated context. Additionally, stems essentially do not have a word class, and without any preceding the context the presented stems would not differ in predictability. Thus, in this experiment there were theoretically cleaner delineations between activity related to encoding of the stem, and activity related to pre-activating upcoming completions, than in a sentence processing experiment. Differences in neural activity following the presentation of the stem based on probability of the completion was likely to be related to predictive processing, not other extraneous factors.

The predictabilities of word completions for each stem were calculated and used to quantify the constraint of each stem with entropy, an information theoretic measure that summarizes the information of a distribution of events. Entropy has previously been used as a measure of sentential constraint in the psycholinguistics literature (Yun, Mauner, Roland, & Koenig, 2012). To illustrate the concept of entropy with an example, a fair coin toss has a 0.5 probability of heads and a 0.5 of tails, leading to an entropy of 1. A rigged coin that always landed heads would have an entropy of 0, while a somewhat rigged coin with a 0.75 probability of heads would have an entropy of approximately 0.8. Thus, there is an inverse relationship between entropy and predictability – if the entropy is 0, the outcome is perfectly predictable, while if the entropy is very high, the predictability is very low. The measure of entropy provides different information the probability of the most probable event (Cloze probability), in that it takes into account the number and probabilities of other possible events, which are potentially important. For example, a stem completion with a generation probability of 0.5 may be generated differently if there is one other completion with a probability of 0.5, or many other completions with much lower probabilities.

In this experiment, the paradigm used was to present participants with a stem with the instructions to generate a completion, and then later present them with the most probable completion of the stem, an improbable completion of the stem, or a pseudoword that started with the same letters as the word stem. This aspect of the experiment paralleled the previously described sentence processing experiments, in which an expected, unexpected, or anomalous word was presented following a context. Here, instead of a sentence context, a simpler cue in the form of a word stem was used; however, the structure remained essentially the same. By constructing the experiment in this way, I could determine if ERP effects at the presented word mirrored observed effects at sentence final words in sentence processing experiments. A stem with a low entropy could be compared to a highly constraining sentence, as the outcome was fairly predictable, while a stem with a high entropy could be compared to a weakly constraining sentence. If the "constraint" of the stem is similar to that of sentences, then N400 amplitudes should be lowest to highly probable words from low entropy stems, greater for highly probable words from high entropy stems, and greater and equal for low probability words from high and low entropy stems. Such a finding would suggest individuals were processing words following stem cues similarly to sentence ending words following sentence contexts, and potentially engaging in similar processes between the two tasks, giving credence to using the word stem paradigm to investigate language comprehension and production without the usage of sentences.

To further investigate the similarity in mechanisms across tasks, time-frequency analysis was performed on EEG data following presentation of the probable completions, improbable completions, and pseudowords. Previous results (Rommers et al., 2016) identified an increase in frontal theta power for unexpected sentence endings to high constraint sentences compared to expected sentence endings. Here, I examined if similar frontal theta effects would be observed following improbable completions to high constraint (low entropy) stems compared to probable completions. I additionally analyzed ERPs following improbable completions to

determine if a late frontal positivity was present. Considering the frontal positivity has been cast as indexing a process of revision, unexpected word stems may not elicit frontal positivities, as there is little contextual or message-level meaning to revise. However, frontal theta effects have appeared in other studies of cognitive control and conflict, and thus an increase in frontal theta may be elicited by unexpected stem completions.

ERPs were also measured at the time of the stem, and binned based on the associated entropy values. In this way, brain activity following stems with highly predictable outcomes could be compared to stems with less predictable outcomes. This analysis was somewhat exploratory, as it was unclear when exactly these effects would emerge – based on previous findings, differences could be found in N400 amplitudes (DeLong, Urbach, & Kutas, 2005). This finding would be interesting, as it could suggest information about upcoming stimuli is also integrated during the N400. However, prediction mechanisms can be flexibly engaged at different timescales, even quite rapid ones (Wlotko & Federmeier, 2015), which means differences could be observed prior to the N400. If mechanisms are engaged rapidly after a cue, as in the case of the PrAN, then effects may emerge within 200 ms after the stem. Lastly, if individuals use consistent temporal information to engage prediction right before the onset of the stimulus, as in the alpha decrease prior to high constraint endings (Rommers et al., 2016), then effects may be found just prior to the stimulus.

To summarize, Experiment 3 consisted of an online norming study and an ERP experiment. The goal of the norming study was to determine probability distributions for completions of word stems. While previous norming studies completions to word stems have been performed (Graf & Williams, 1987; Martin et al., 2009; Migo, Roper, Montaldi, & Mayes, 2010; Shaw, 1997; Soler, Dasí, Ruiz, & Cervera, 2017), not all of these studies have been in English, and English norms may be somewhat outdated. This normed data was used to calculate entropy values for each word stem, giving an index of predictability. These word stems

were then used in an ERP experiment, in which participants were presented with a word stem, followed by the most probable stem completion, a low probability stem completion, or a pseudoword.

Online Word Stem Norming Study

Methods

Participants & Materials

330 word stems were selected based on previous work by finding 3-letter stems in English that could be completed by at least 5 fairly common English words. The word stems were split into 3 groups of 110 stems. These 3 blocks of stems were entered into separate Amazon's Mechanical Turk (MTurk) experiments, each of which were completed by 103 participants. Since the blocks were separate experiments, it was not the case that the same 103 participants completed all 330 word stems. After filtering for participants that participated in multiple blocks, there were roughly 280 unique participants across blocks. Pre-requisites were set to collect data from highly rated Turk workers located in the United States. However, information on age and gender were not collected. All participants gave informed consent and were compensated for their participation.

Procedure

Individuals participated in the experiment through MTurk, a website where participation in short surveys or experiments is compensated, allowing for rapid data collection from numerous individuals at fairly low cost. The MTurk experiment itself redirected participants to Qualtrics, a website for giving surveys and simple experiments, where they completed the word stem

completion task. Once the task was completed, they were given a completion code to enter into MTurk, at which point they were compensated once their participation was approved. The completion code also verified that participants had actually completed the task instead of skipping to the end.

Each experiment contained 110 word stems to complete. The stems were listed vertically on the webpage in capital letters, e.g. "SCH___". Participants were instructed to fill in each stem with the first word completion that came to mind, and were given an example using a stem not in the list. Additionally, individuals were instructed to feel free to use proper names or places if that was what came to mind first. Participants were encouraged to put down words even if they weren't sure how to spell them, and to emphasize speed and putting the first word down quickly over getting the word exactly right.

Following data collection, the resulting collection of completions were checked for misspellings, non-words, non-answers (blanks), and incorrect completions (e.g. "alligator" for "ALI___"). To avoid biasing certain completions, a conservative cleaning approach was taken. Only obvious misspellings were corrected to the intended word; if the intended word was unclear, the completion was simply removed. Ultimately, only a small amount of cleaning was necessary, and roughly 100 observations per stem remained.

One the data were cleaned, entropy values for each stem were computed. Entropy was calculated with the following formula:

$$H(X) = -\sum_{i} P(X_i) \log_2 P(X_i)$$

Where $P(X_i)$ is the probability of a completion for a particular stem. Probabilities for each completion were calculated by dividing the number of observations the completion was produced by the total number of observations for the stem.

As was done in previous norming studies, exploratory analyses were conducted to examine what lexical properties of words are predictive of a higher probability of producing a certain completion. Previous work has demonstrated a relationship between probability of production and word length and frequency (Graf & Williams, 1987). Here, the lexical variables considered were word length, frequency, familiarity, concreteness, imageability, orthographic neighborhood size and frequency, and phonological neighborhood size and frequency. Each of these variables for each completion were correlated with the log probability of producing that completion.

Results and Discussion

An average of ~18 unique completions were produced for each word stem (range 6-59). The average entropy across stems was 3.12, with a range of 0.73 to 5.41. The distribution of entropy values across stems, as well as an example of two distributions of completions with similar number of completions but very different entropies, is shown in figure 18 below.

Correlations between probability of completion and lexical variables are plotted in figure 19. The probability values used were –log(prob), which leads to an inverse relationship (e.g. –log(0.9) = 0.05, -log(0.1) = 1); correlations with probabilities should be inversely interpreted. Figure 19 shows a replication of previously found effects; shorter words and more frequent words were associated with higher probabilities of completion. Concreteness, imageability, and metrics of neighborhood size were not related to probability of completion.

To summarize, the norming study resulted in a database of completions for word stems with a range of entropy values. The reported relationship between lexical variables and probability of completion was in line with previous work, suggesting the data from MTurk were

not aberrant or strongly biased. The stems and entropy values were used for the ERP portion of Experiment 3.

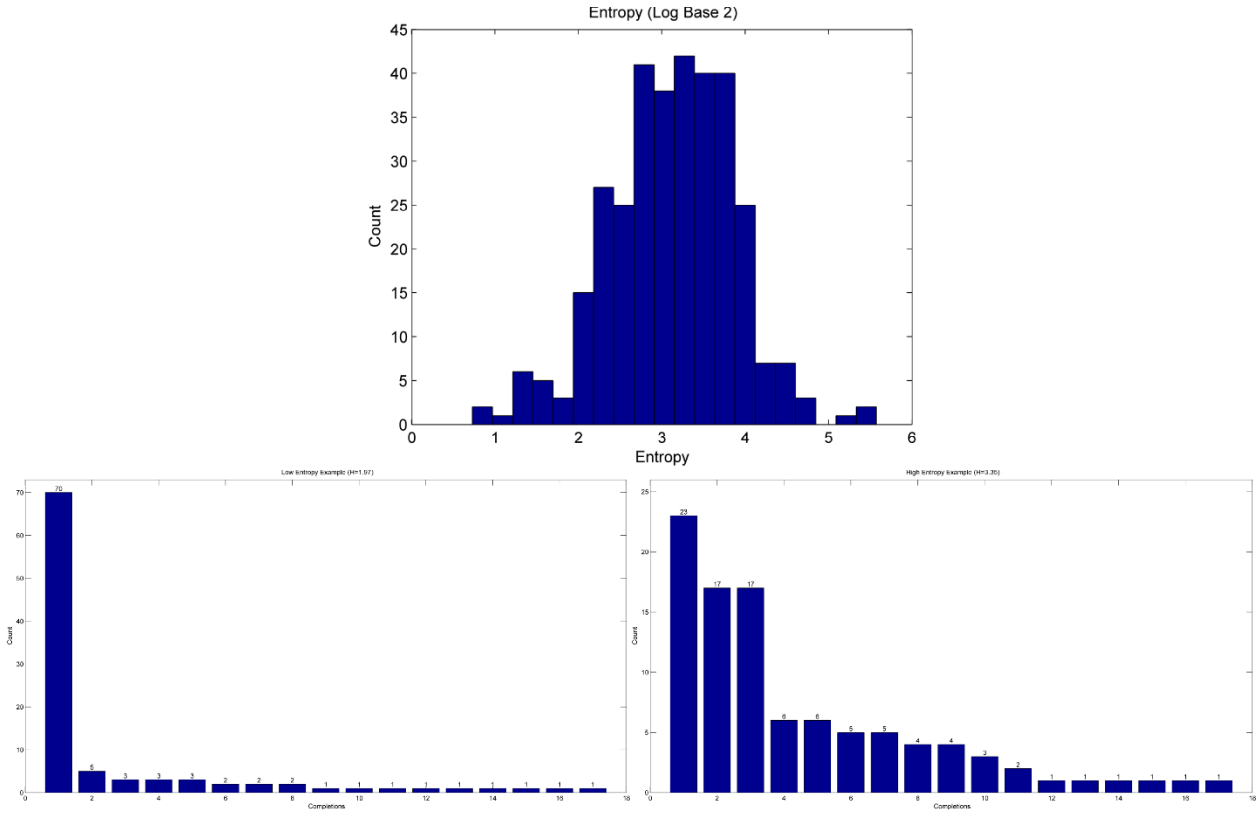


Figure 18: Entropy of word stems. Top figure shows the distribution of entropies across all stems (average = 3.12). Bottom left shows a distribution of completions (17 unique) for a stem with entropy of 1.97. Bottom right shows a distribution with entropy of 3.35.

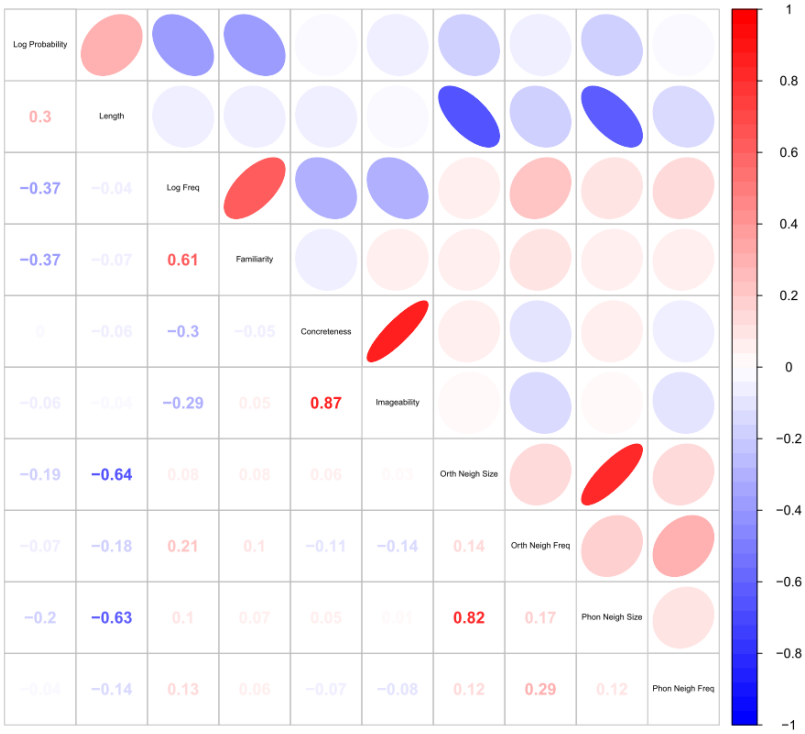


Figure 19: Correlations of completion probability (-log(prob)) with lexical variables. Word frequency and length effects are replicated.

Word Stem Completion ERP Study

Methods

Participants

26 right-handed, native speakers of English with normal or corrected vision from the University of Illinois, Urbana-Champaign participated in the experiment and were paid $10 an hour for their participation. All participants had no 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. 2 participants were removed due to excessive EEG artifacts, leaving 24 participants in the final analysis. For adequate power and complete counterbalancing, the *a priori* number of subjects to run was 24; following removal of 2 participants, 2 extra subjects were run.

Materials

The stimuli used in the experiment were the 330 word stems with entropy values from the MTurk experiment. For each stem, the most probable completion, an improbable completion, and a pseudoword were selected. The improbable completions were chosen such that the associated probability was close to the minimum for that stem, the completion was not a plural or different tense of a more probable completion, and the lexical variables (frequency, length) were as close to the most probable stem as possible. Despite these efforts, the average word frequency of the probable completions was higher than the improbable completions (Probable = 204, Improbable = 21).

Pseudowords were generated using the program Wuggy. a multilingual pseudoword generating program (Keuleers & Brysbaert, 2010). Wuggy breaks input words into subsyllabic elements, which are edited and recombined while segment length and transition frequencies

between elements are controlled for. The program also automatically calculates orthographic Levenshtein distance between each generated pseudoword and its 20 most similar words (OLD20), as well as the number of orthographic neighbors made by editing a single letter (NED1). More closely matched pseudowords can be selected by minimizing the difference in these values between the pseudoword and the template word. Here, all completions for each stem were entered into Wuggy, and the pseudowords with the lowest differences in OLD20 and NED1 were selected.

During the experiment, participants saw all 330 stems, 110 of which were completed with a probable completion, 110 with an improbable completion, and 110 with a pseudoword. 6 lists were created to counterbalance across subjects which item type completed each stem. Each set of 110 items for each item type were also split into fifths by entropy values, such that there were 22 items from very high entropy stems, 22 from high entropy stems, 22 from medium entropy stems, 22 from low entropy stems, and 22 from very low entropy stems. These entropy delineations were created by splitting the rank ordered entropy values across all stems, such that the top 66 entropy values were considered to be very high entropy, the next 66 were high entropy, etc.

Procedure

After giving informed consent to participate, each subject completed 10 blocks of 33 trials each. On each trial, subjects were first presented with a word stem for 1000 ms, and were instructed to mentally complete the stem with the first word that came to mind. After a delay of 2000 ms to give subjects time to generate a word, a completion to the previously presented stem was displayed for 1000 ms. This completion was the most probable word, an improbable word, or a pseudoword. A "?" was then displayed, cueing subjects to respond whether the word displayed

matched the word they had mentally generated. There was no time limit for the subject to make this response.

EEG Recording and Processing

The EEG recording procedure of Experiment 3 was identical to that of Experiments 1 and 2. Eyeblinks again corrected using AMICA, and trials were removed with a follow-up artifact scan.

To assess frontal theta effects, time-frequency analysis was performed on EEG epochs. A short-time Fast Fourier Transform (FFT) approach was used, with a time window size of 400 ms that shifted by ~17 ms. Each 400 ms period was windowed with a Hanning function, zero-padded to 512 ms to increase frequency resolution, and finally Fourier transformed. Additionally, a spectrogram of the ERP was subtracted from each trial to focus on non-phase-locked power and remove any potential late frontal positivity. Finally, relative power values were computed by dividing each post-stimulus time bin by a 400 ms pre-stimulus baseline period, and then converting the result to decibels [log_{10}(activity/baseline)]. The resulting power values were averaged across trials for each participant, and averaged across participants for a grand average.

Results

Word Completion ERPs

ERPs at the Central Cluster time-locked to the presentation completions of word stems, collapsed across entropy are plotted in the top half of Figure 20. There appears to be an N400 modulation based on the "expectancy" of the presented completion: highly probable completions elicited the smallest N4s, followed by improbable but essentially plausible

completions, and finally followed by Pseudowords. Indeed, a linear mixed effects model comparing N400 amplitudes revealed significant differences between high and low probability completions ($t = 4.252$, $p < 0.001$), as well as between low probability completions and Pseudowords ($t = -4.453$, $p < 0.001$). However, it is possible this modulation may be driven in part by having expectations matched. If subjects consider the word they came up with to be a "target", then neural responses to observing a target would differ from when a "non-target" was observed, especially if this target had a low probability of occurrence (Duncan-Johnson & Donchin, 1977). High probability completions are much more likely to match expectations, and thus ERPs to high probability completions could be a mixture of "target", or "match", and "non-target", or "miss" responses. On the other hand, low probability completions are less likely to match the word the participant generated, and thus ERPs are less likely to be made up of "target" or "match" responses.

To assess this, ERPs were binned based on participant responses of the completion matching or not. This is plotted in the bottom half of Figure 20. Matching expectations clearly led to a much greater positivity, and amplitudes did not differ when the completion did not match. Both matches and misses showed lower amplitudes compared to pseudowords, suggesting there was still some facilitation of processing even for misses. These N400 amplitudes differed statistically significantly in a linear mixed model; "Matches" significantly differed from "Misses" ($t = -7.881$, $p < 0.001$), and "Misses" significantly differed from Pseudowords ($t = -4.432$, $p < 0.001$). When the presented word matched the word the participant generated, a much larger positivity was generated than when the word did not match; however, words that did not match still elicited more positive amplitudes than Pseudowords.

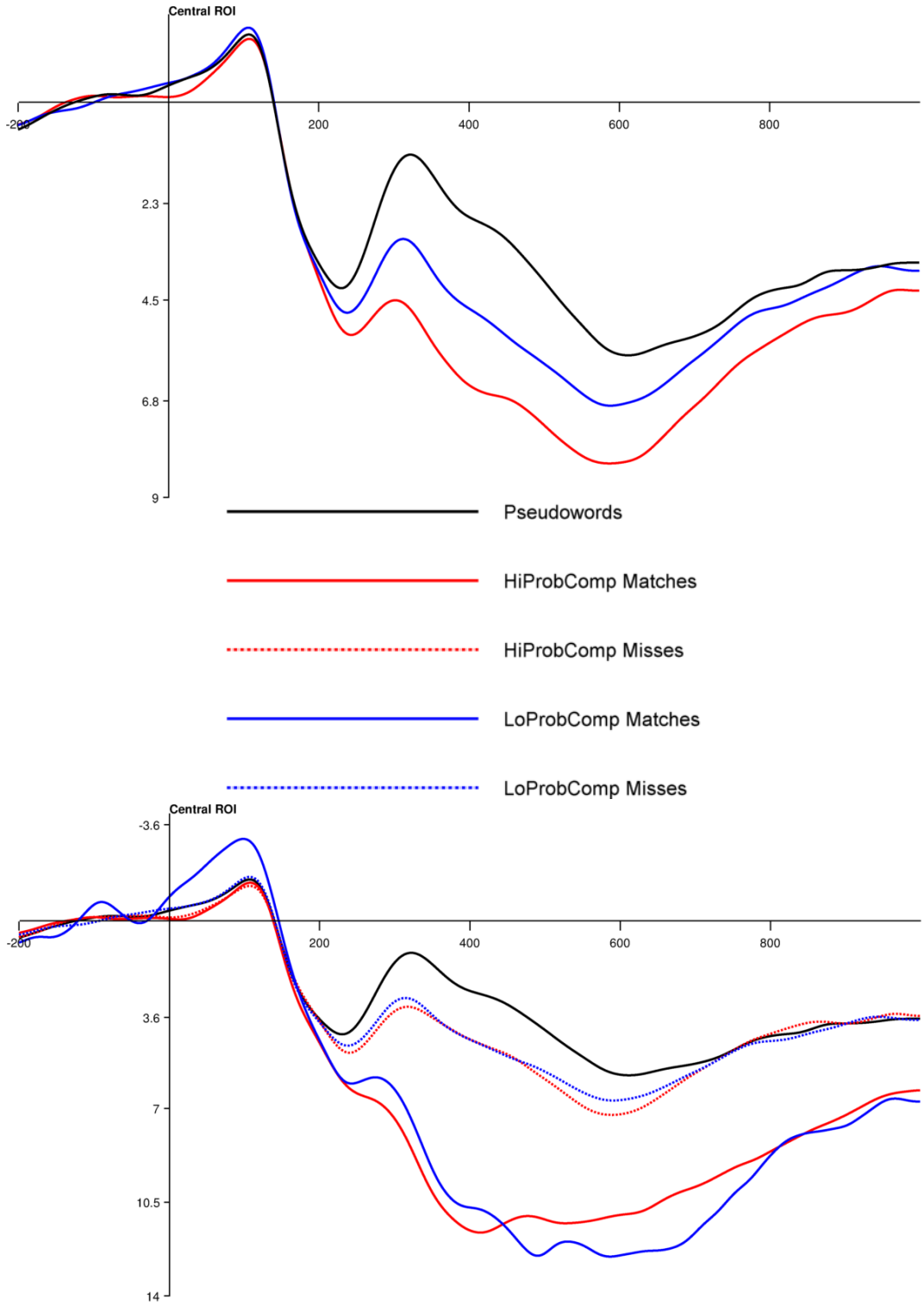


Figure 20: Grand average ERPs to presented completions at the Central Cluster. Top figure: completions collapsed across participant response. Bottom figure: completions subdivided by participant response. Both matches and misses differ in amplitude from pseudowords, but completions do not differ in amplitude based on probability of completion.

Despite differences in N400 amplitude between words that matched what was generated and words that did not, there did not appear to be differences in amplitude based on the probability of the completion, e.g. high probability matches did not differ from low probability matches. Indeed, a model with factors of Match (Match, Miss) and Probability (High, Low), as well as the interaction, found only a significant effect of Match ($t = 7.244$, $p < 0.001$); Probability ($t = 0.547$, $p = 0.584$) and the interaction ($t = 0.066$, $p = 0.947$) were not significant. However, these responses were collapsed across stem entropy (constraint); differences could emerge when stems are more constraining towards a particular word completion, as in the sentence ending experiments.

To assess this, Figure 21 presents a comparison of high probability Matches from low and high entropy stems, low probability Misses from low and high entropy stems, and Pseudowords. This plot can be compared to the N400 pattern seen in Figure 14 of Experiment 2. The high probability, low entropy Match is comparable to a strong constraint expected ending, as both are the most probable ending to a predictive cue. Similarly, the high probability, high entropy Match is comparable to a weak constraint expected ending, as both are the most probable ending to a non-predictive cue. The Misses are comparable to unexpected sentence endings, and the Pseudowords comparable to semantically anomalous endings. Statistical analysis was carried out in a manner similar to Experiments 1 and 2 – namely, N400 amplitudes were submitted to a linear mixed effects model comparing low entropy and high entropy Matches, as well as comparing high entropy (weak constraint) Matches to Misses. High entropy Matches differed from Misses ($t = -2.489$, $p = 0.02$); however, the Match conditions did not differ by entropy ($t = 1.482$, $p = 0.14$). Thus, the N400 pattern observed in Experiments 1 and 2 was not entirely replicated here, as constraint had no significant effect.

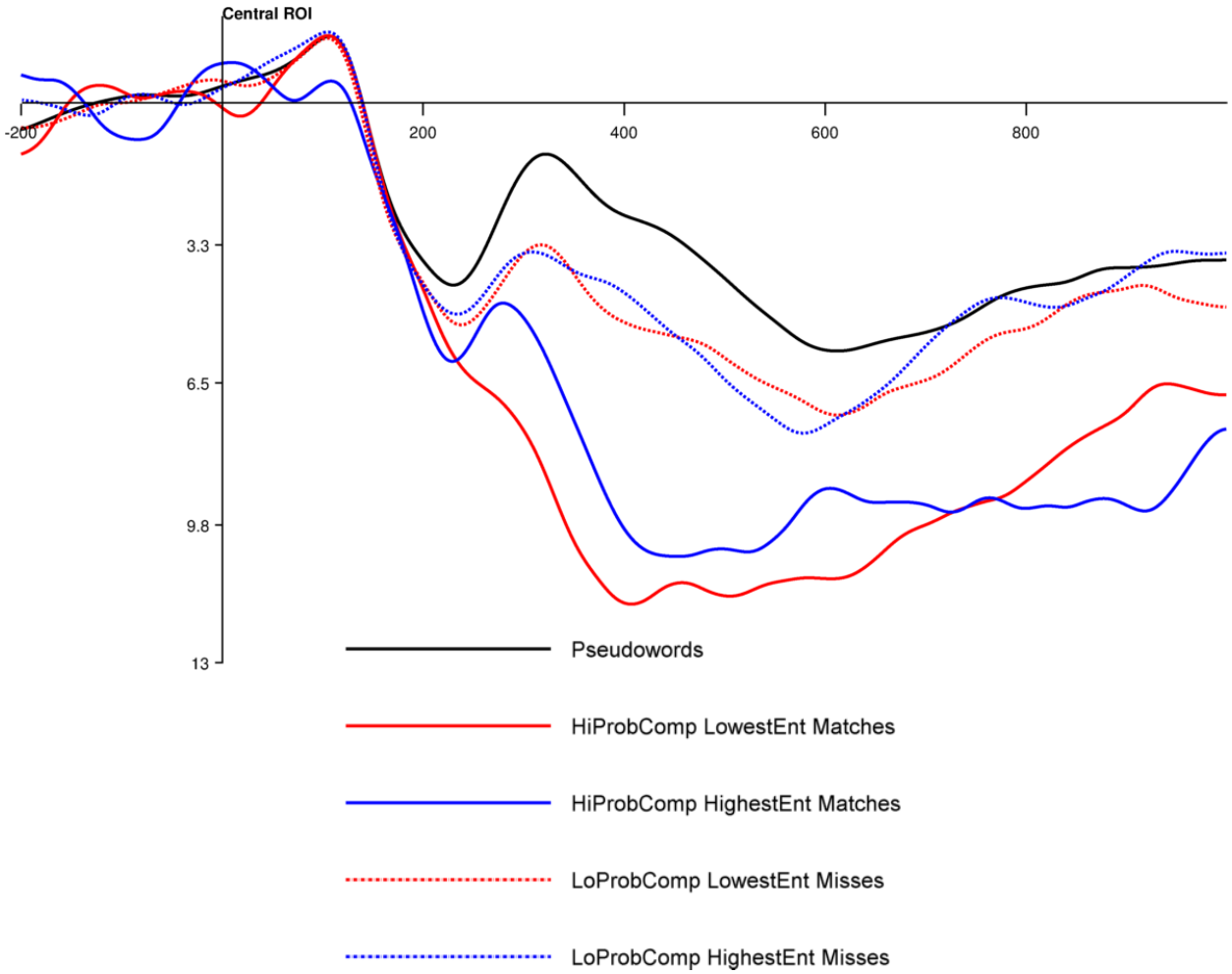


Figure 21: Grand average ERPs at the Central Cluster showing a comparison of high and low probability completions to high and low entropy stems, as well as pseudowords. The N400 pattern here mirrors the pattern found for strongly expected, weakly expected, unexpected and anomalous sentence endings.

For a closer look at any possible effects of constraint on the N400, ERPs to high probability Matches across levels of entropy was plotted in figure 22. Note that as entropy increased, the likelihood of the presented word decreased, and thus trial numbers of the ERP decreased as well. In figure 22, there appears to be N400 amplitude modulation based on entropy; however, there is not a clear linear pattern. The visible differences could potentially be due to random variability, which could be inflated by the low signal to noise ratio caused by low trials numbers.

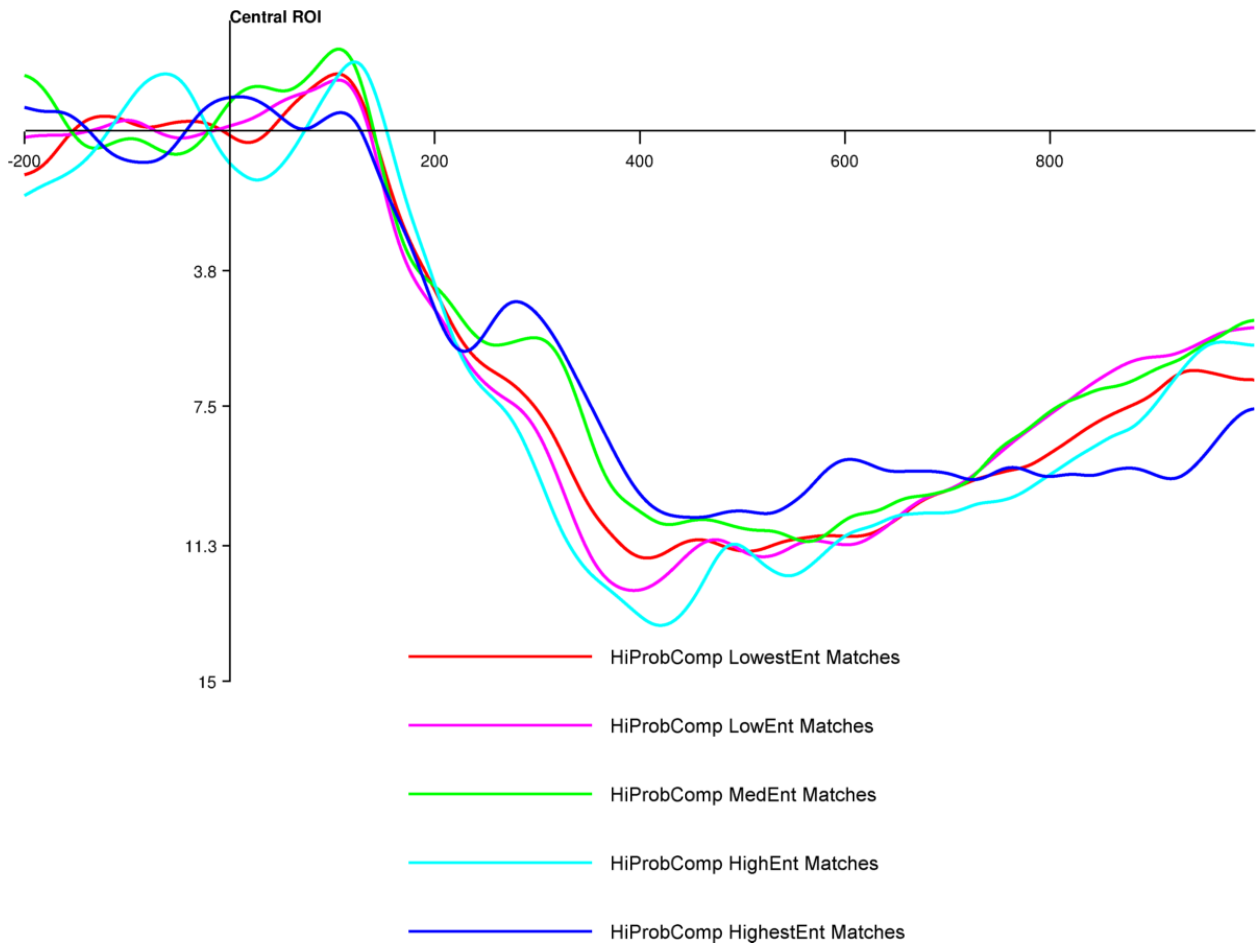


Figure 22: Grand average ERPs to high probability Matches at the Central Cluster. Warmer colors = lower entropy (higher predictability), cooler colors = higher entropy (lower predictability). ERPs show a somewhat graded response based on the predictability of the completion.

One possibility is that entropy may not be the best metric for analysis of stem completion ERPs. Entropy takes into account the probability of all completions, but once a participant has generated a completion, these other probabilities may not matter. Cloze probability – the probability of the most common completion – may show a better fit to N400 amplitudes to presented word completions. To analyze this statistically, two linear regressions were fit predicting N400 amplitudes of high probability Matches – one using entropy, and one using cloze probability. These regressions are plotted in Figure 23. The coefficient of entropy was not statistically significant ($t = -1.538$, $p = 0.125$, $R^2 = 0.005$), whereas the coefficient of cloze

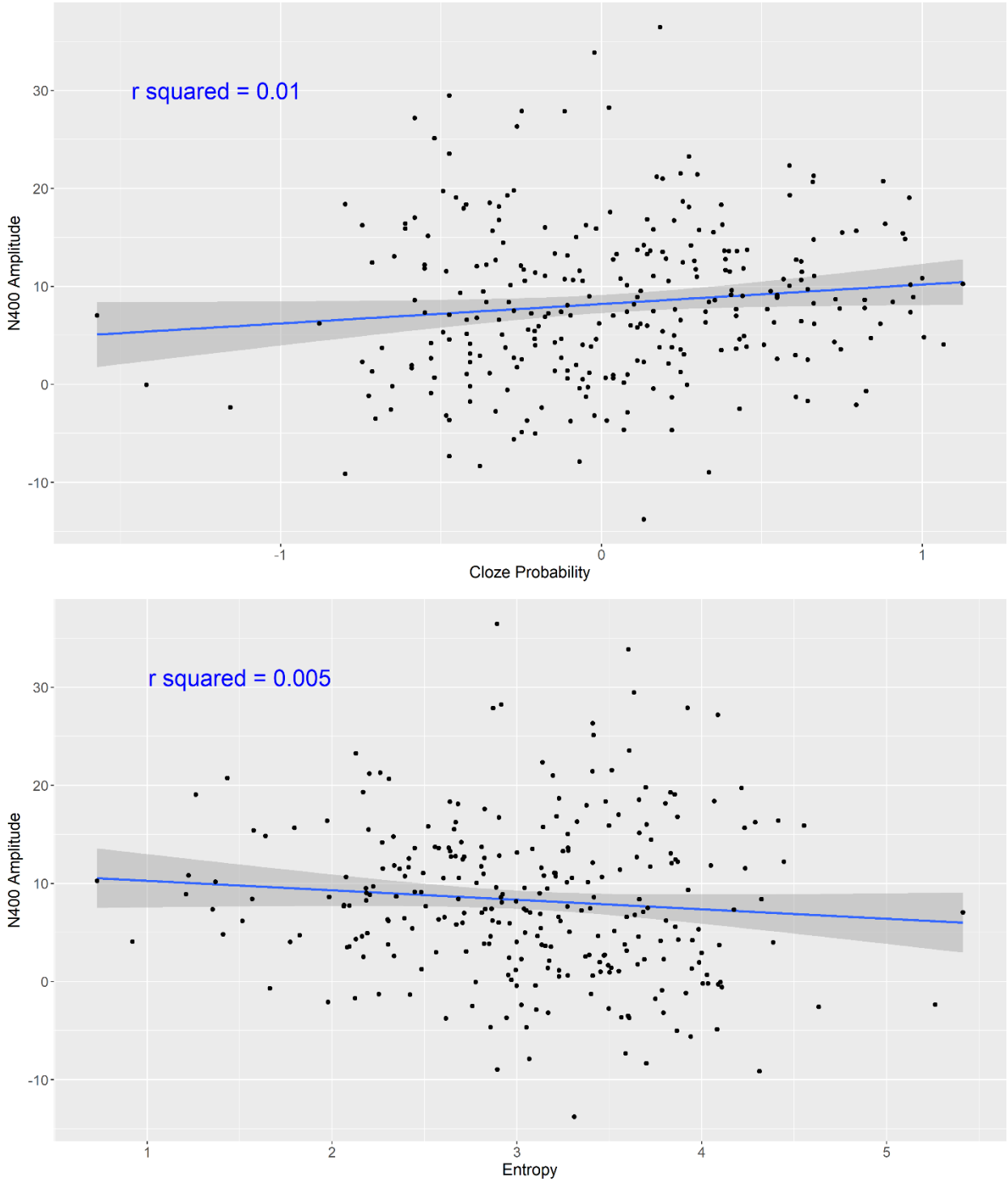


Figure 23: Linear fits of N400 amplitude elicited by high probability Matches to Cloze probability (top) and entropy (bottom). A significant positive relationship between N400 amplitude and cloze probability is observed, whereas there is no significant relationship between N4 amplitude and entropy. However, both R^2 values are quite small.

significantly predicted N400 amplitude ($t = 1.985$, $p = 0.048$, $R^2 = 0.01$). However, note that the R^2 for the cloze model was rather small; additionally, when cloze probability was entered

into a mixed effect model to predict N400 amplitude with random slopes and intercepts, the effect of cloze was not statistically significant ($t = 1.096$, $p = 0.29$). Thus, there was only limited evidence that the cloze probability of the presented high probability completion was predictive of its elicited N400 amplitude.

To examine the frontal positivity following unexpected stem completions, high probability Matches from low and high entropy stems, low probability Misses from low and high entropy stems, and Pseudowords were plotted at the Frontal Cluster in Figure 24. If processing mirrored what is observed in sentence comprehension experiments, then a positivity would be expected to be observed to the low probability Misses from low entropy stems compared to those from high entropy stems. This pattern is clearly not observed in Figure 24, where there appears to be no difference in amplitude between these conditions. Indeed, these conditions did not statistically differ ($t = -0.608$, $p = 0.5$). As hypothesized, a frontal positivity to unexpected completions was not observed in this experiment.

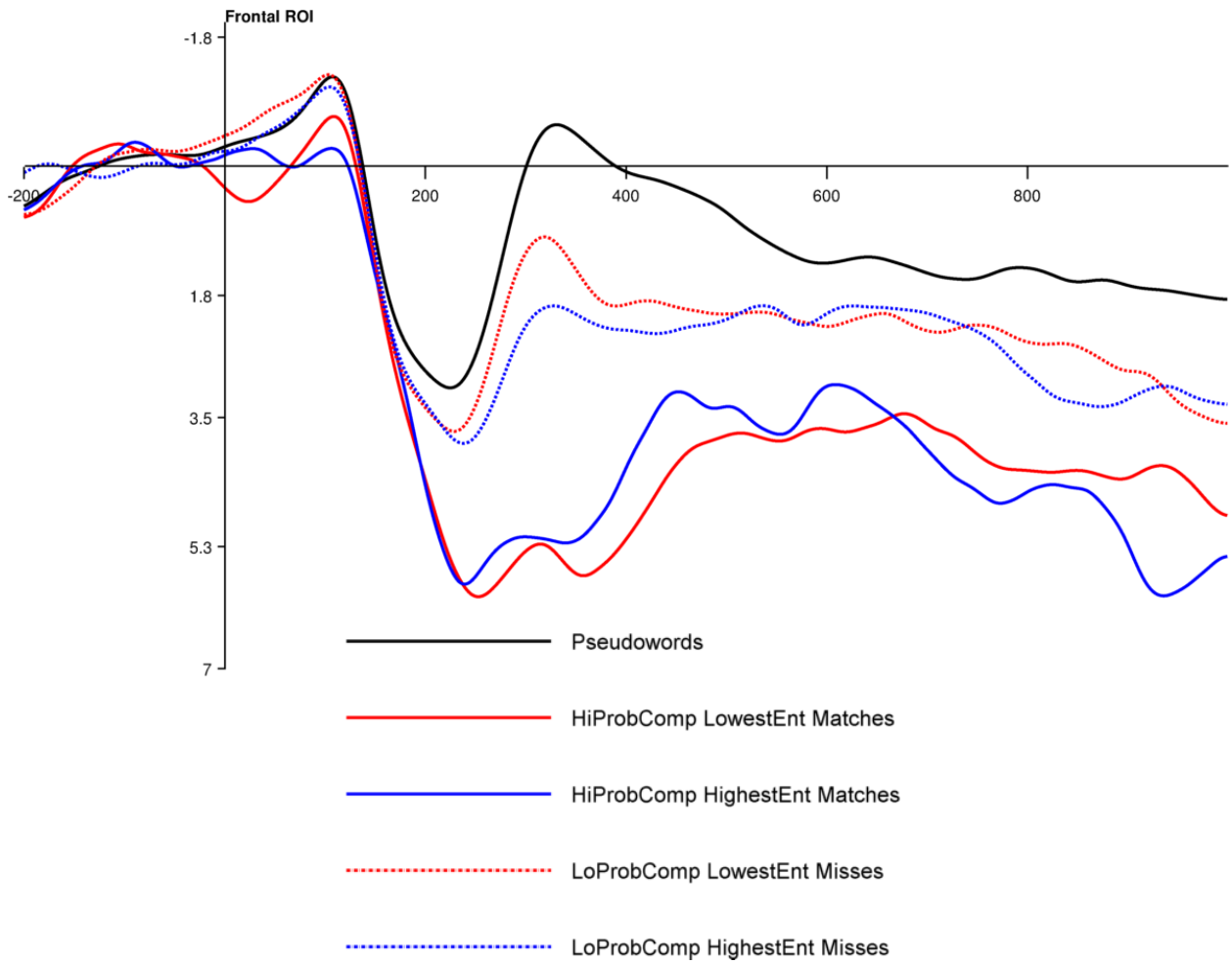
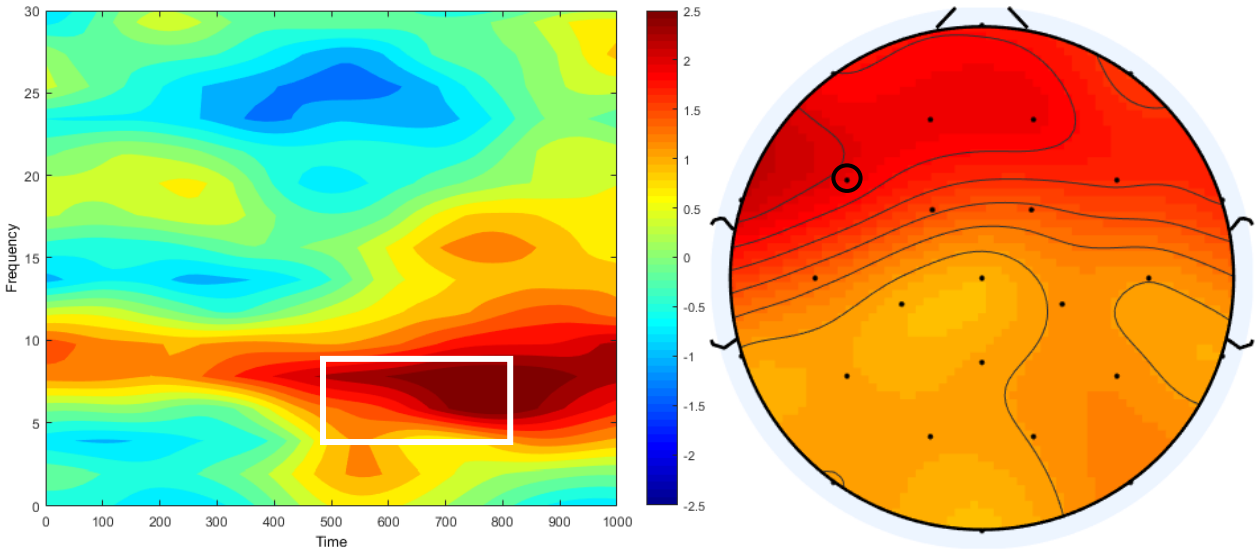


Figure 24: Grand average ERPs at the Frontal Cluster showing a comparison of high and low probability completions to high and low entropy stems, as well as Pseudowords. There is no apparent frontal positivity to the low probability completions to low entropy stems.

Despite the lack of a frontal positivity, an increase in frontal theta power may have been elicited by unexpected stem completions compared to expected stem completions. Figure 25 displays the comparison of theta power for low probability, lowest entropy (high constraint unexpected) completions to high probability, lowest entropy (high constraint expected) completions. The time-frequency plot on the left side (channel LDFr) shows a clear increase in power from approximately 500-1000 ms in the 4-12 Hz range, and the topography of power in the 4-8 Hz range from 500-800 ms shows a frontal topography. To increase statistical power and avoid multiple comparisons, a t-test was run comparing the two conditions of interest by averaging across 4-8 Hz and 500-800 ms at channel LDFr for each participant's data. These

windows were chosen *a priori* based on Rommers et al. (2016). The test was statistically significant ($t = 4.2453$, $p < 0.001$), suggesting the previously observed frontal theta effect was replicated in this experiment.

Low probability completion / lowest entropy Miss vs. High probability completion / lowest entropy Match



Time-frequency plot, Channel LDFr

Scalp topography, 4-8 Hz, 500-800 ms

Figure 25: Grand average time-frequency plots of the contrast of low probability, lowest entropy Misses (unexpected items) vs. high probability, lowest entropy Matches (expected items). The left plot shows a time-frequency plot at a single channel (LDFr) with the theta effect (4-8 Hz) highlighted. X axis = time, Y axis = frequency. The right plot shows the scalp topography of the 4-8 Hz effect from 500-800 ms. Channel LDFr is highlighted.

An important aspect of the previously found frontal theta effect was its specificity to unexpected sentence completions to strongly constrained sentences, as opposed to weakly constrained sentences. To assess if there was similar specificity in the current experiment, the comparison of theta power for low probability, low entropy completions to high probability, low entropy completions was plotted in figure 26. This is essentially the same figure as figure 25, only focused on the next highest entropy bin ("low" as opposed to "lowest"). There was no appearance of a frontal theta effect. A t-test in the same window as before (4-8 Hz, 500-800 ms, channel LDFr) was not significant ($t = 0.5435$, $p = 0.6$). Qualitatively, there was no frontal theta

effect observed for the other entropy bins as well. Thus, the frontal theta effect was replicated, and was specific to only the unexpected completions with the lowest entropy.

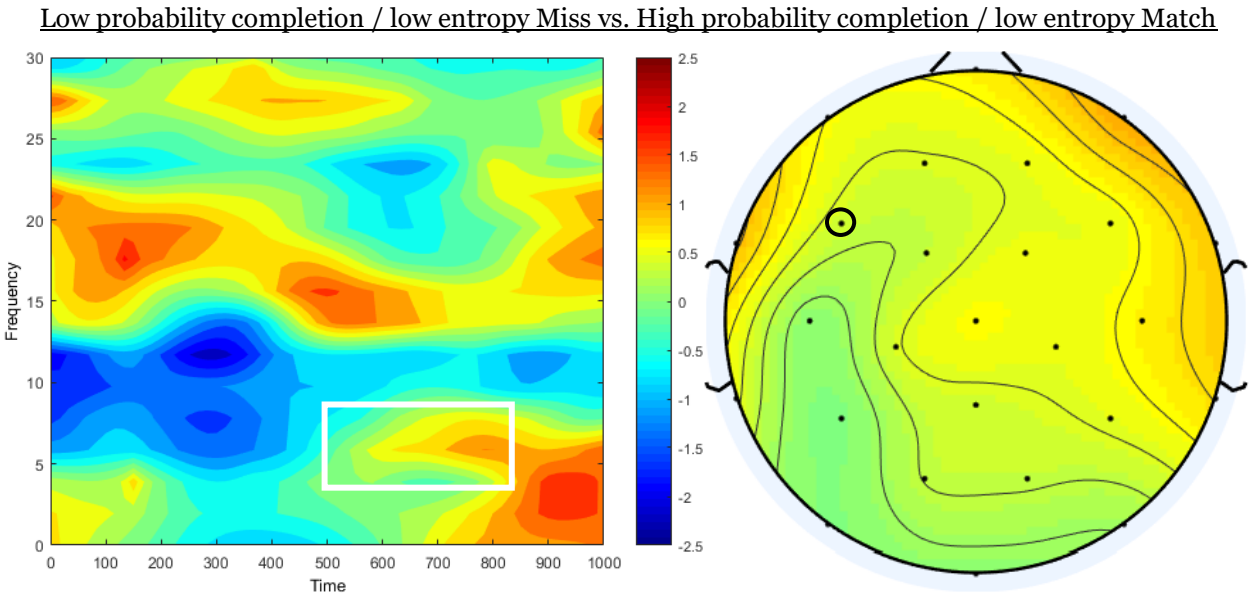


Figure 26: Grand average time-frequency plots of the contrast of low probability, low entropy Misses (unexpected items) vs. high probability, low entropy Matches (expected items). Essentially the same plot as Figure 23, but with higher entropy (lower constraint) items. Clearly the frontal theta effect previously seen with lowest entropy items is not present.

Word Stem ERPs

ERPs following the word stem were examined for a pattern of activity linearly graded by entropy of the stem. Given the novelty of the paradigm, it was unclear at what time and location any effects might occur, and thus a more exploratory approach was taken. First, the difference between ERPs from the lowest entropy stems and ERPs from the highest entropy stems – the two extremes - were plotted in figure 27. This difference was used to guide selection of time windows for repeated measures ANOVAs analyses. 4 channel clusters were constructed by dividing the scalp into four quadrants: Left Frontal, Right Frontal, Left Posterior, and Right Posterior. ANOVAs were then run on each selected time window, with factors for Hemisphere (Left, Right), Direction (Frontal, Posterior), and Condition (Low Entropy, High Entropy). If the

ANOVA had significant condition effects, the time window was submitted to the second step of the analysis.

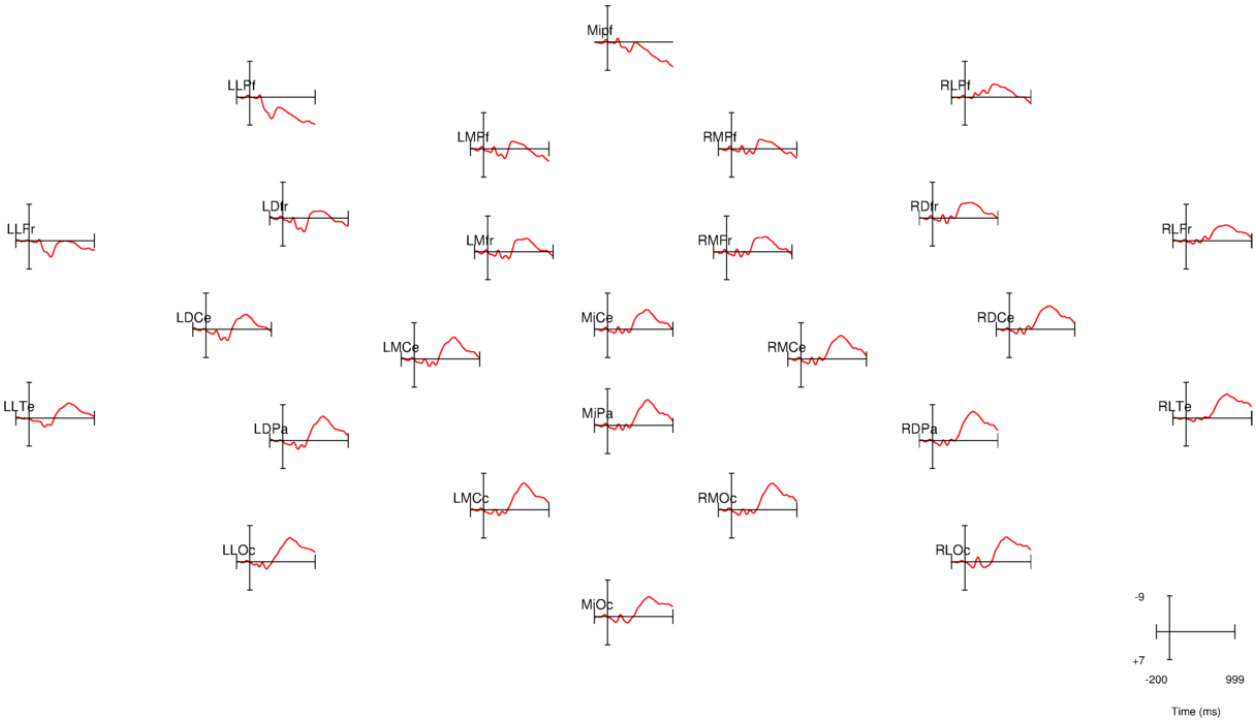


Figure 27: Difference waves for ERPs following highest entropy stems vs. ERPs following lowest entropy stems at each channel.

ANOVAs were run on 3 separate time windows: the P2 time window (200-275 ms), the N400 time window (300-500 ms), and a late time window (500-800 ms). The ANOVA on the P2 time window revealed a significant main effect of Condition ($F = 5.3$, $p = 0.03$), a significant interaction between Direction and Hemisphere ($F = 28.15$, $p < 0.001$), and a significant three way interaction between Condition, Direction, and Hemisphere ($F = 5.2$, $p = 0.03$). The ANOVA on the N400 time window revealed no significant condition effects. Finally, the ANOVA on the late time window revealed a significant main effect of Condition ($F = 7.4$, $p = 0.01$), as well as an interaction between Direction and Hemisphere ($F = 20.9$, $p < 0.001$). For visualization, ERPs are plotted at the Central Cluster and Posterior Cluster in figure 28. The P2 at the Central Cluster, and the late component at the Posterior Cluster, appear graded by entropy.

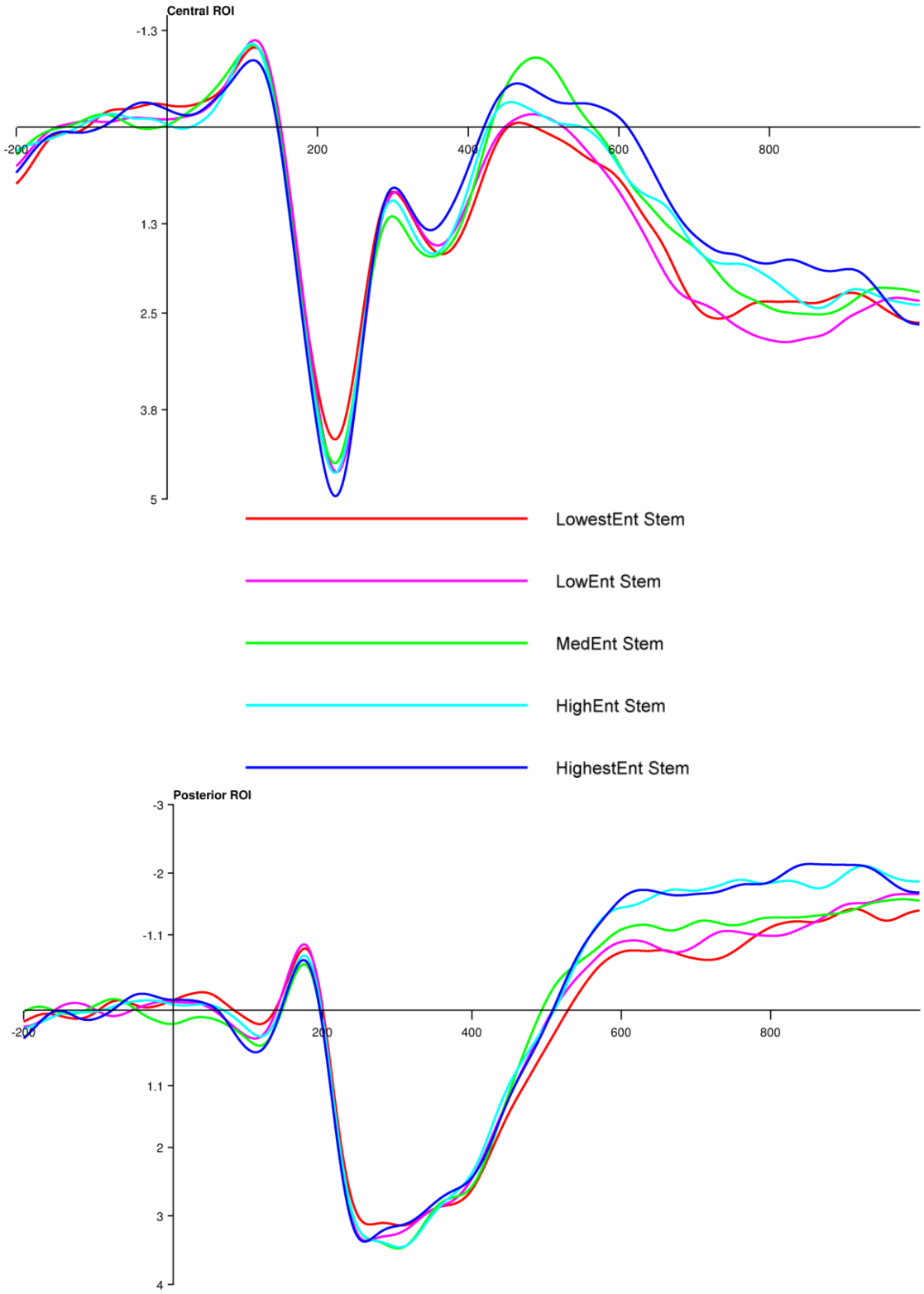


Figure 28: Grand average ERPs to word stems. The top plot shows the Frontal Cluster, the bottom the Posterior Cluster. Warmer colors = lower entropy (higher predictability), cooler colors = higher entropy (lower predictability). The P2 response at the Central Cluster appears fairly graded by entropy, while a later graded response (600-900 ms) is found in the Posterior Cluster.

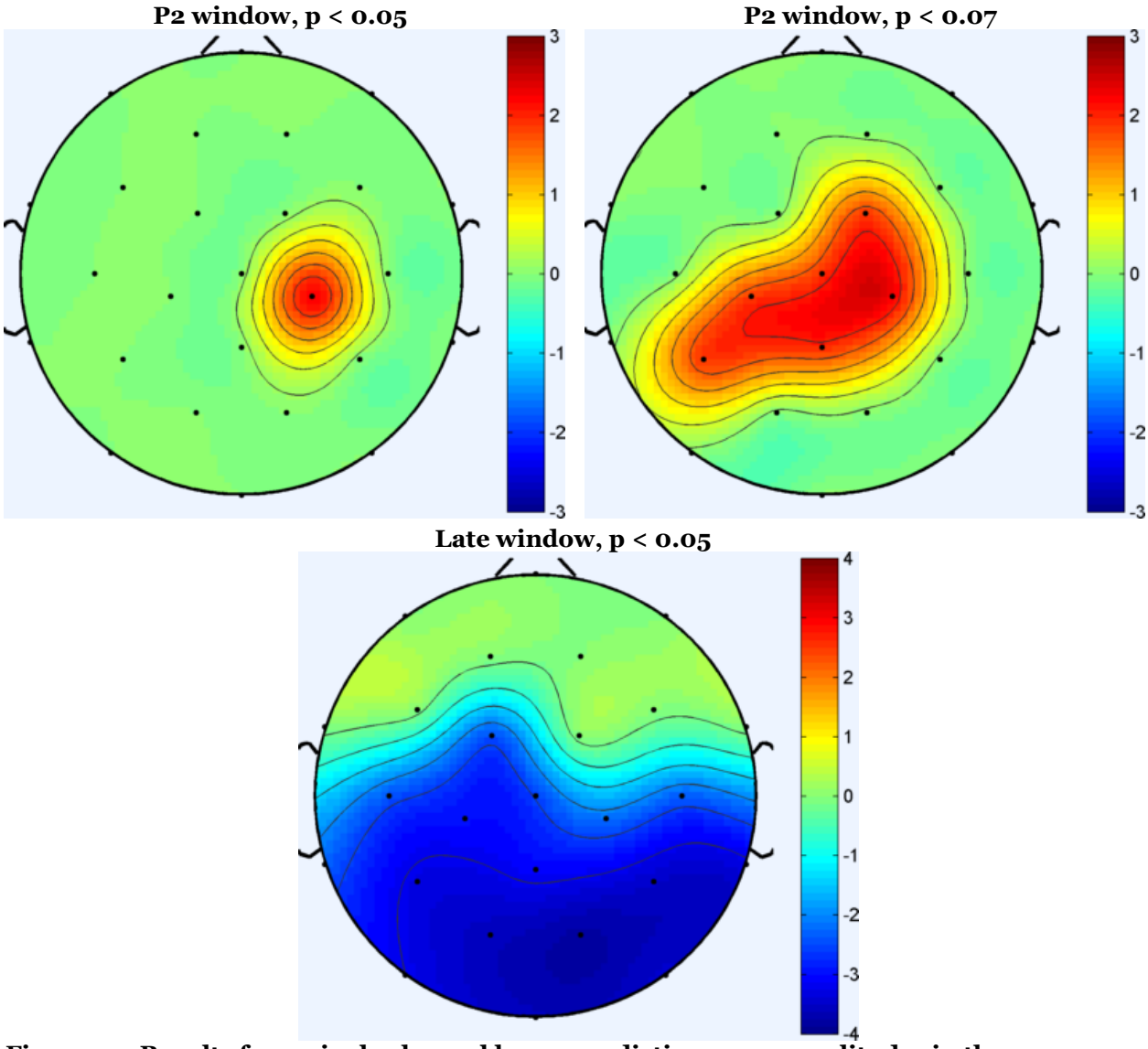


Figure 29: Results from single channel lmers predicting mean amplitudes in the designated time window with stem entropy. The P2 effect is only significant at 1 channel with a p threshold of 0.05, but is broader with a more liberal threshold. The late window effect is broadly distributed and maximal at posterior sites.

For the second step of the analysis, mean amplitudes for the specified time window at each trial were submitted to a linear mixed effect model predicting amplitude with entropy. A model was run at each channel, and the resulting t value was plotted on a topography plot if its p value was less than 0.05, leading to a statistically thresholded mask. This analysis was run for the P2 time window and the late component time window. The resulting topographies are plotted in Figure 29. Note that for descriptive purposes, two topographies are plotted for the P2

window - with a threshold of 0.05, only one channel reached significance, but by changing the threshold only slightly (0.07), the effect broadened significantly. These results showed two ERP components following presentation of a word stem that differed based on the upcoming information – an early component (the P2) and a late component.

Discussion

Experiment 3 provided interesting and novel results regarding mechanisms of prediction and language comprehension by implementing a word stem completion paradigm with stimuli normed for entropy. The experiment was constructed in such a way to not only allow comparison to sentence comprehension ERP studies, but also to investigate neural mechanisms of prediction prior to the onset of an expected stimulus. The reported results suggested that similar processing occurred during reading of stem completions as during sentence ending words, in that a similar N400 pattern was observed, as well as a comparable increase in frontal theta to highly unexpected completions. Additionally, in looking at neural activity following stem presentation, two components were found to differ in amplitude based on entropy, potentially suggesting an engagement of multiple mechanisms of prediction at different timescales.

Of course, the experiment was not without its drawbacks. Instructing individuals to generate a completion and respond whether the presented word matched the word they generated potentially altered their experimental goals, biasing them to focus on word matches. Theoretically, this could have even encouraged participants to attempt to think of more modal or predictable responses in order to increase number of matches; however, participants were pointedly instructed to not explicitly try to match their generation with the presented word, and just think of whatever word came to mind first. Still, the large positive increase in amplitude for matches compared to misses was reminiscent of the P3 ERP component, a component indexing

information processing or context updating following perception of target stimuli (Donchin & Cohen, 1967; Donchin & Coles, 1988; Duncan-Johnson & Donchin, 1982; Nieuwenhuis, Aston-Jones, & Cohen, 2005; Pritchard, 1981). The amplitude of the P300 is modulated not only by demands of the task, but also by the subjective probability of occurrence of the task-relevant stimuli (Duncan-Johnson & Donchin, 1977; Polich, 1986, 1987; Sutton, Braren, Zubin, & John, 1965). In the current experiment, participants were instructed to respond if the presented word matched what was generated, making the stimulus task-relevant, and given the presence of low probability completions and pseudowords, matches were uncommon occurrences with low subjective probability; thus, matches generated large amplitude P300 responses, potentially masking other more subtle electrophysiological changes. A useful follow-up study would be to give participants a more incidental task, i.e. lexical decision, to perform when viewing the stem completions.

While the magnitude of the difference between matches and misses was large, the pattern of this effect does in some sense replicate those seen in sentence processing experiments: namely, expected sentence endings elicit more positive N400 amplitudes than unexpected sentence endings. However, N400s to word completions of stems that were low in entropy, and thus more constraining, did not differ from N400s to completions of stems that were high in entropy. This differs from the pattern seen in sentence processing experiments, where strongly and weakly constraining expected sentence endings differ in elicited N400 amplitudes. There are multiple possible explanations for this outcome. Firstly, a difference in N400 amplitude by constraint may in fact have been present, but was simply not detected statistically due to low signal to noise. Word completions to high entropy stems, despite being the most probable completion to the stem, were often still fairly low probability, and thus did not often match participants' generations. This led to reduced trial numbers and noisier ERPs, which potentially obscured effects.

A second explanation is that the range of entropy and cloze probability values utilized in this experiment was not large enough for detection of electrophysiological effects. Previous sentence processing experiments have often utilized stimulus lists with numerous items with cloze probabilities above 0.8 or below 0.2 for strong and weak constraint sentences, respectively. Here, most completions had cloze probabilities in the 0.3-0.5 range, and thus the list of stimuli could be considered more of a comparison within medium constraint than between high and low constraint. This likely occurred due to the way the stimulus list was constructed, in which there had to be at least common 5 completions for each stem. However, this constraint need not necessarily be applied; for example, the stem "KIW___" has essentially one common completion, "KIWI", and thus would be a very highly constraining stem. A follow-up experiment using more highly constraining and weakly constraining stems may reveal differences in the N400.

Alternatively, entropy may not describe the pattern of N400 responses to completions as well as the cloze probabilities of the completions. In sentence completion tasks, constraint and cloze are somewhat conflated; sentences are considered high constraint when they produce completions above a certain threshold of cloze probability, and weak constraint when below a certain threshold. Entropy theoretically takes more information into account; however, this information may be extraneous once the stimulus has been encountered. In other words, when individuals are generating predictions, the probabilities of multiple outcomes may be important information, but once the outcome has been observed, the response to that outcome may be better represented by the probability of that outcome, and not the entire space of probabilities. Cloze probabilities, despite conveying less information, may be closer to individuals' mental probabilistic models of language input (Smith & Levy, 2011). Indeed, the previously described regression analyses showed a significant relationship between N400 amplitude and cloze probability, and no significant relationship with entropy, though both produced noticeably small R^2 values.

An interesting alternative explanation is that there was in fact no difference between N400 amplitudes due to constraint, and differences by constraint emerge in other experiments due to combining prediction matches and misses. To clarify, consider that an individual reading or hearing a sentence, regardless of constraint, might generate a prediction of what the upcoming words will be. In some instances, the word that is read or heard will match what was generated, while in other cases it will not; however, there will likely be more matches when constraint is high compared to when it is low. By this account, the difference in N400 amplitude between strong and weak constraint expected sentence endings would be due to a greater proportion of matches in strong constraint ERPs compared to weak constraint ERPs. In sentence processing experiments, participant responses of whether a word matched or not are generally not recorded, as this would essentially encourage active prediction by the participant, which may not be what happens during natural reading; thus, it remains an open debate whether individuals generate single predictions (Kleinman, Runnqvist, & Ferreiera, 2015) or multiple predictions, or whether this differs based on task demands. The proposed account would somewhat argue against an account that N400 amplitudes are graded by expectancy, as has been shown with item-level regression analyses (Wlotko & Federmeier, 2012); however, these item-level analyses collapse across participants, and different participants may generate different predictions, leading to the same issue. More complicated analyses may be required to adjudicate between these possibilities.

In potential opposition to this account, participants elicited more positive N400 amplitudes to unexpected completions compared to Pseudowords. Note that words do not obligatorily elicit more positive amplitude N400s compared to Pseudowords; in fact, other studies have found these amplitudes to be roughly equivalent (Laszlo & Federmeier, 2009), suggesting some facilitation of processing for unexpected completions. Participants may have pre-activated features that overlapped with multiple outcomes as opposed to just the most prominent word, leading to some facilitation for unexpected completions. However, task

demands must also be taken into account – since subjects were potentially searching for words that matched or did not match the generated word, a pseudoword would be a clear mismatch, and this could lead to differences in neural activity. For instance, Azizian, Freitas, Watson, & Squires (2006) found that P3 amplitudes were larger for non-target stimuli that were perceptually more similar to targets than for those that were not similar. This effect could theoretically extend into the semantic domain as well, leading to amplitude differences between unexpected words and Pseudowords simply due to task goals.

The results reported here also demonstrated a lack of any late frontal positivity to unexpected word completions to low entropy stems, but the presence of increased frontal theta power for these items compared to expected word completions to low entropy stems. This result not only supports the notion that the frontal positivity and frontal theta are dissociable effects that index separable information processing mechanisms, but also that receiving an unexpected stem completion leads to engagement of similar top-down control mechanisms as when a prediction is violated in other cognitive domains, but only when the expected stem was very predictable. It is possible the frontal positivity simply was not observed due to statistical or noise issues; however, the results arguably highlight the relationship between the frontal positivity and contextual re-analysis. Here, stems provided little semantic context for the following completions, and thus little re-analysis was required. Following from this, an interesting hypothesis is that semantically congruous sentence endings that violate some aspects of the message-level context should produce a frontal positivity. A recent study found that when a white lie was expected (e.g. "the meat sauce was *tasty*"), but a blunt truth was received instead ("the meat sauce was *overcooked*"), a frontal positivity without an accompanying N400 modulation was elicited (Moreno, Casado, & Martin-Loeches, 2016). Follow-up studies should investigate manipulations of context without expectancy to further explore the process indexed by the frontal positivity.

Finally, the current study produced novel results regarding mechanisms of prediction engaged prior to the target stimulus – namely, two ERP components, an early P2 response and a later posterior component, that differed in amplitude based on the entropy of the stem. These preliminary effects potentially reflect mechanisms of prediction or generation – however, further analyses will need to be performed to rule out other possibilities. For example, these effects may reflect integration or processing of the stem itself, which could differ based on its associated entropy. While it seems unlikely that processing of stems would differ dramatically, there have been very few studies examining ERPs following stems, rendering it difficult to know what factors affect neural responses. For instance, one previous study investigating ERP activity following word stems (Klonek, Tamm, Hofmann, & Jacobs, 2009) found that frontal amplitudes roughly in the N400 time window were inversely related to number of stem completions, with the largest N400 responses to stems with only one possible completion. Klonek et al. argued that stems with higher completions essentially had higher frequency, as they appear in more words and thus are encountered more, and thus processing was facilitated for stems with lots of completions; however, this was not shown empirically. Additionally, this N4 result was not replicated here; a linear regression between N400 amplitudes and number of stem completions revealed no significant relationship ($t = 0.156$, $p = 0.88$). However, more work must be done to determine what neural activity is related to processing of the stems themselves versus processing of upcoming information.

The late posterior component reported here could also be related to maintenance of information in working memory. Previous work has shown a posterior ERP component, the contralateral delay activity (CDA), that persists over a delay and scales in amplitude with increasing number of items to remember (Ikkai, McCollough & Vogel, 2010; Vogel, McCollough, & Machizawa, 2005). Individuals could be retaining information regarding multiple completions over the delay period between the stem and the upcoming presented word completion. While this does not argue against a pre-activation account – individuals must

activate information regarding the completions to maintain it in working memory – it brings into question the timescale of this information activation. Does the onset of the late component reflect the onset of pre-activation, or is this information activated and processed beforehand and the late component reflects a maintenance process? The observed effect during the P2 time window could reflect the activation of stem completions. An interesting follow-up would be to give participants a shorter deadline for generating a completion and see if these effects are still observed.

An additional method to test if the ERP effects seen at the stem reflect some form of pre-activation would be to correlate the ERP effects observed at the stem with those seen at the completion. Theoretically, it is the generation of information beforehand that drives the difference in amplitude based on entropy at the presentation of the completion (as in figure 22). This method could also determine whether the early and late effects reported are the same underlying activity, or index different processes. For example, the early P2-like effect could reflect processing of the stem itself, while the later posterior effect could be related to activation of upcoming information. In this case, the amplitude of the posterior effect would be a better predictor of the amplitude of the N400 at the completion.

However, even in a correlation was found, an alternative explanation to a pre-activation account is a depth of encoding account. The effect seen at the time of the stem could relate to how deeply encoded the stem was, and the difference in activity at the time of completion could relate to memory retrieval of the stem. A correlation between encoding and retrieval activity that was related to successful memory performance has previously been reported (Chen, Lithgow, Hemmerich, & Caplan, 2014). Here, the completion could serve as a retrieval cue for the stem, perhaps due to the orthographic overlap. If this is the case, then a similar correlation between activity following the stem and activity following an expected completion should be observed for post-stem activity and activity following a pseudoword, which has the same

orthographic overlap. This finding would argue against a pre-activation account, as it is unlikely participants would generate a pseudoword as a completion.

Overall, these promising results show that the word stem completion experimental design can be a useful paradigm for understanding the mechanisms involved in predicting information. Of course, a question is whether the predictive mechanisms engaged in a word stem completion task are the same as those involved in language comprehension. One answer is that word stem completion is a language production task, and prediction *is* production (Dell & Chang, 2014). There may be different informational cues (e.g. orthographic, semantic, or discourse-level) that guide expectations, but it seems unlikely that the mechanism of generating upcoming information would differ largely across tasks. If one of the brain's core mechanisms is to predict information, then it seems reasonable to believe that this process would be domain-general and not task-specific.

CHAPTER 5: CONCLUSION

The results of the 3 experiments presented here not only further support the theory that individuals predict during language comprehension, but also provide new evidence in support of two claims. First, prediction during language comprehension can lead to costs for downstream cognition; namely, predicted words that are not observed are more likely to be falsely remembered than unpredicted words. Importantly, this effect differs depending on what was encountered instead of the predicted word. If predictions are not very strong, and the encountered information is implausible, then false alarming will not be as likely. The neurocognitive mechanisms by which false alarms occur also seem to differ based on the strength of the prediction, with more false alarms to more strongly predicted words being driven by conceptual priming, compared to recollective or decisional processes for weakly predicted words.

Second, mechanisms of prediction or generating upcoming information can be observed prior to the onset of the predicted stimulus or prediction violation. Neural activity following the presentation of a word stem differed based on the predictability of possible completions to the stem. This effect was close to linear – activity following stems with more predictable completions was lower than stems with less predictable completions in a graded fashion. The electrophysiological responses following confirmation or disconfirmation of completions were similar to those observed when sentence endings were confirmed or disconfirmed, potentially suggesting that the mechanisms of generating completions and sentence endings have some overlap. Thus, there may be similar evidence of pre-activation based on predictability in language processing experiments as well.

REFERENCES

Allan, K., Doyle, M. C., & Rugg, M. D. (1996). An event-related potential study of word-stem cued recall. *Cognitive Brain Research*, 4(4), 251-262.

Allan, K., Robb, W. G., & Rugg, M. D. (2000). The effect of encoding manipulations on neural correlates of episodic retrieval. *Neuropsychologia*, 38(8), 1188-1205.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247-264.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30(3), 191-238.

Anderson, R. C. (1974). Substance recall of sentences. *The Quarterly Journal of Experimental Psychology*, 26(3), 530-541.

Azizian, A., Freitas, A. L., Watson, T. D., & Squires, N. K. (2006). Electrophysiological correlates of categorization: P300 amplitude as index of target similarity. *Biological Psychology*, 71(3), 278-288.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390-412.

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

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11(7), 280-289.

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1521), 1235-1243.

Bassili, J. N., Smith, M. C., & MacLeod, C. M. (1989). Auditory and visual word-stem completion: Separating data-driven and conceptually driven processes. *The Quarterly Journal of Experimental Psychology*, 41(3), 439-453.

Bastiaansen, M., & Hagoort, P. (2015). Frequency-based Segregation of Syntactic and Semantic Unification during Online Sentence Level Language Comprehension. *Journal of Cognitive Neuroscience*, 27(11), 2095.

Bastiaansen, M. C., Van Berkum, J. J., & Hagoort, P. (2002). Syntactic processing modulates the θ rhythm of the human EEG. *Neuroimage*, 17(3), 1479-1492.

Beato, M. S., Boldini, A., & Cadavid, S. (2012). False memory and level of processing effect: an event-related potential study. *Neuroreport*, 23(13), 804-808.

Bermúdez-Margaretto, B., Beltrán, D., Domínguez, A., & Cuetos, F. (2015). Repeated exposure to “meaningless” pseudowords modulates LPC, but not N (FN) 400. *Brain Topography*, 28(6), 838-851.

Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2002). Beyond syntax: Language-related positivities reflect the revision of hierarchies. *NeuroReport*, 13(3), 361-364.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624.

Brainerd, C. J., & Reyna, V. F. (2005). *The Science of False Memory*. Oxford University Press.

Braunstein, V., Ischebeck, A., Brunner, C., Grabner, R. H., Stamenov, M., & Neuper, C. (2012). Investigating the influence of proficiency on semantic processing in bilinguals: An ERP and ERD/S analysis. *Acta Neurobiol Exp*, 72, 421-438.

Bridger, E. K., Bader, R., Kriukova, O., Unger, K., & Mecklinger, A. (2012). The FN400 is functionally distinct from the N400. *NeuroImage*, 63(3), 1334–42.

Bridger, E. K., Bader, R., & Mecklinger, A. (2014). More ways than one: ERPs reveal multiple familiarity signals in the word frequency mirror effect. *Neuropsychologia*, 57, 179-190.

Brothers, T., Hoversten, L. J., & Traxler, M. J. (2017). Looking back on reading ahead: No evidence for lexical parafoveal-on-foveal effects. *Journal of Memory and Language*, 96, 9-22.

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.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, 93, 203-216.

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.

Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, 4, 25.

Buckner, R. L. (2010). The role of the hippocampus in prediction and imagination. *Annual Review of Psychology*, 61, 27-48.

Buzsaki, G. (2006). *Rhythms of the Brain*. Oxford University Press.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, 18(8), 414-421.

Chen, Y. Y., Lithgow, K., Hemmerich, J. A., & Caplan, J. B. (2014). Is what goes in what comes out? Encoding and retrieval event-related potentials together determine memory outcome. *Experimental Brain Research*, 232(10), 3175-3190.

Chen, H., Voss, J. L., & Guo, C. (2012). Event-related brain potentials that distinguish false memory for events that occurred only seconds in the past. *Behavioral and Brain Functions*, 8, 36.

Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibáñez, A., Owen, A.M., & Bekinschtein, T. A. (2013). Expectation and attention in hierarchical auditory prediction. *The Journal of Neuroscience*, 33(27), 11194-11205.

Cheung, O. S., & Bar, M. (2012). Visual prediction and perceptual expertise. *International Journal of Psychophysiology*, 83(2), 156-163.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03), 181-204.

Coffman, B.A., Hunter, M.A., Jones, A.P., Saxon, H.A., Kolodjeski, K., Lockmiller, B., Khan, O., Collar, T., Stephen, J.M., & Clark, V.P. (2013). Using independent components analysis (ICA) to remove artifacts associated with transcranial direct current stimulation (tDCS) from electroencephalography (EEG) data: A comparison of ICA algorithms. *Brain Stimulation*, 7(2).

Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT Press.

Cohen, M. X., & Donner, T. H. (2013). Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior. *Journal of Neurophysiology*, 110(12), 2752-2763.

Cole, R. A., & Perfetti, C. A. (1980). Listening for mispronunciations in a children's story: The use of context by children and adults. *Journal of Verbal Learning and Verbal Behavior*, 19(3), 297-315.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6(1), 84-107.

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(3), 658-668.

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, 125(2), 159.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671-684.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of experimental Psychology: General*, 104(3), 268.

Curran, T. (1999). The electrophysiology of incidental and intentional retrieval: ERP old/new effects in lexical decision and recognition memory. *Neuropsychologia*, 37(7), 771-785.

Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition*, 28(6), 923-938.

Curran, T. (2004). Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia*, 42, 1088-1106.

Curran, T., & Dien, J. (2003). Differentiating amodal familiarity from modality-specific memory processes: An ERP study. *Psychophysiology*, 40(6), 979-988.

Curran, T., & Hancock, J. (2007). The FN400 indexes familiarity-based recognition of faces. *Neuroimage*, 36(2), 464-471.

Curran, T., Schacter, D. L., Johnson, M. K., & Spinks, R. (2001). Brain potentials reflect behavioral differences in true and false recognition. *Journal of Cognitive Neuroscience*, 13(2), 201-216.

Curran, T., Tanaka, J. W., & Weiskopf, D. M. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cognitive, Affective, & Behavioral Neuroscience*, 2(1), 1-18.

de Groot, A. M., Thomassen, A. J., & Hudson, P. T. (1982). Associative facilitation of word recognition as measured from a neutral prime. *Memory & Cognition*, 10(4), 358-370.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58, 17–22.

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 of London*, 369(1634).

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

DeLong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, 48(9), 1203-1207.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117-1121.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG sources are dipolar. *PloS One*, 7(2), e30135.

den Heyer, K. (1985). On the nature of the proportion effect in semantic priming. *Acta Psychologica*, 60(1), 25-38.

den Heyer, K., Briand, K., & Dannenbring, G. L. (1983). Strategic factors in a lexical-decision task: Evidence for automatic and attention-driven processes. *Memory & Cognition*, 11(4), 374-381.

Donchin, E., & Cohen, L. (1967). Averaged evoked potentials and intramodality selective attention. *Electroencephalography and Clinical Neurophysiology*, 22(6), 537-546.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11(3), 357-374.

Duarte, A., Ranganath, C., Winward, L., Hayward, D., & Knight, R. T. (2004). Dissociable neural correlates for familiarity and recollection during the encoding and retrieval of pictures. *Cognitive Brain Research*, 18(3), 255-272.

Duff, M. C., & Brown-Schmidt, S. (2012). The hippocampus and the flexible use and processing of language. *Frontiers in Human Neuroscience,* 6, 69.

Duffy, S. A., Henderson, J. M., & Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 791.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, 14(5), 456-467.

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biological Psychology*, 14(1), 1-52.

Düzel, E., Yonelinas, A. P., Mangun, G. R., Heinze, H. J., & Tulving, E. (1997). Event-related brain potential correlates of two states of conscious awareness in memory. *Proceedings of the National Academy of Sciences of the United States of America*, 94(11), 5973-5978.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20(6), 641-655.

Fallon, M., Trehub, S. E., & Schneider, B. A. (2002). Children's use of semantic cues in degraded listening environments. *The Journal of the Acoustical Society of America*, 111(5), 2242-2249.

Fay, S., Isingrini, M., & Clarys, D. (2005). Effects of depth-of-processing and ageing on word-stem and word-fragment implicit memory tasks: Test of the lexical-processing hypothesis. *European Journal of Cognitive Psychology*, 17(6), 785-802.

Fay, S., Pouthas, V., Ragot, R., & Isingrini, M. (2005). Neural correlates of word-stem priming. *Neuroreport*, 16(11), 1169-1173.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491-505.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of memory and Language*, 41(4), 469-495.

Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, 42(2), 133-141.

Federmeier, K. D., McLennan, D. B., Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, 39(2), 133-146.

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.

Fell, J., Dietl, T., Grunwald, T., Kurthen, M., Klaver, P., Trautner, P., Schaller, C., Elger, C.E., & Fernández, G. (2004). Neural bases of cognitive ERPs: more than phase reset. *Journal of Cognitive Neuroscience*, 16(9), 1595-1604.

Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP 'old/new' effects: memory strength and decisional factor(s). *Neuropsychologia*, 40(13), 2288-2304.

Fischer-Baum, S., Dickson, D. S., & Federmeier, K. D. (2014). Frequency and regularity effects in reading are task dependent: Evidence from ERPs. *Language, Cognition and Neuroscience*, 29(10), 1342-1355.

Fischler, I. (1977). Associative facilitation without expectancy in a lexical decision task. *Journal of Experimental Psychology: Human Perception and Performance*, 3(1), 18.

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), 1-20.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, 20(4), 400-409.

Forster, K.I. (1981). Priming and the effects of sentence and lexical contexts on naming time: evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology*, 33A, 465–495.

Freunberger, D., & Roehm, D. (2016). Semantic prediction in language comprehension: evidence from brain potentials. *Language, Cognition and Neuroscience*, 31(9), 1193-1205.

Freunberger, D., & Roehm, D. (2017). The costs of being certain: Brain potential evidence for linguistic preactivation in sentence processing. *Psychophysiology*, 54(6), 824-832.

Friederici, A. D., Steinhauer, K., Mecklinger, A., & Meyer, M. (1998). Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology*, 47(3), 193-221.

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.

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(5), 862.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, 38(7), 833-848.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": an electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, 8(2), 89-106.

Geng, H., Qi, Y., Li, Y., Fan, S., Wu, Y., & Zhu, Y. (2007). Neurophysiological correlates of memory illusion in both encoding and retrieval phases. *Brain Research*, 1136, 154-168.

Geraci, L., Hamilton, M., & Guillory, J. J. (2015). Age effects in implicit memory: The role of response competition induced by relative word frequency. *Experimental Aging Research*, 41(5), 496-509.

Gerrig, R. J., & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes*, 26(2-3), 67-86.

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

Gilbert, D. T., & Wilson, T. D. (2009). Why the brain talks to itself: Sources of error in emotional prediction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1521), 1335-1341.

Glanzer, M., & Adams, J.K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, 13, 8-20.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(1), 5.

Glanzer, M., & Bowles, N. (1976). Analysis of the word frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 21-31.

Goetz, E. T., Anderson, R. C., & Schallert, D. L. (1981). The representation of sentences in memory. *Journal of Verbal Learning and Verbal Behavior*, 20(4), 369-385.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior*, 23(5), 553-568.

Graf, P., & Williams, D. (1987). Completion norms for 40 three-letter word stems. *Behavior Research Methods, Instruments, & Computers*, 19(5), 422-445.

Guillaume, F., & Etienne, Y. (2015). Target-context unitization effect on the familiarity-related FN400: A face recognition exclusion task. *International Journal of Psychophysiology*, 95(3), 345-354.

Guillaume, F., & Tiberghien, G. (2013). Impact of intention on the ERP correlates of face recognition. *Brain and Cognition*, 81(1), 73-81.

Gunter, T. C., Jackson, J. L., & Mulder, G. (1992). An Electrophysiological Study of Semantic Processing in Young and Middle-Aged Academics. *Psychophysiology*, 29(1), 38-54.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304(5669), 438-441.

Hagoort, P., & van Berkum, J. (2007). Beyond the Sentence Given. *Philosophical Transactions: Biological Sciences*, 801-811.

Hald, L. A., Bastiaansen, M. C., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, 96(1), 90-105.

Hanslmayr, S., Pastötter, B., Bäuml, K. H., Gruber, S., Wimber, M., & Klimesch, W. (2008). The electrophysiological dynamics of interference during the Stroop task. *Journal of Cognitive Neuroscience*, 20(2), 215-225.

Hartman, M., & Hasher, L. (1991). Aging and suppression: memory for previously relevant information. *Psychology and Aging*, 6(4), 587-594.

Hasher, L., Quig, M. B., & May, C. P. (1997). Inhibitory control over no-longer-relevant information: Adult age differences. *Memory & Cognition*, 25(3), 286-295.

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, 115(5), 1090-1103.

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision and action* (pp. 1–58). Hove: Psychology Press.

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, 124(1), 62.

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences*, 95(3), 781-787.

Hillyard, S. A., & Kutas, M. (1983). Electrophysiology of cognitive processing. *Annual Review of Psychology*, 34(1), 33-61.

Holcomb, P. J., Grainger, J., & O'rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, 14(6), 938-950.

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(4), 2088-2095.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, 31(1), 19-31.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151-171.

Ibáñez, A., López, V., & Cornejo, C. (2006). ERPs and contextual semantic discrimination: degrees of congruence in wakefulness and sleep. *Brain and Language*, 98(3), 264-275.

Ikkai, A., McCollough, A. W., & Vogel, E. K. (2010). Contralateral delay activity provides a neural measure of the number of representations in visual working memory. *Journal of Neurophysiology*, 103(4), 1963-1968.

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.

Jordan, T. R., & Thomas, S. M. (2002). In search of perceptual influences of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 34.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, 2(4), 647-670.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133-156.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42(3), 627-633.

Kielar, A., Meltzer, J. A., Moreno, S., Alain, C., & Bialystok, E. (2014). Oscillatory responses to semantic and syntactic violations. *Journal of Cognitive Neuroscience*, 26(12), 2840-2862.

Kleiman, G. M. (1980). Sentence frame contexts and lexical decisions: Sentence-acceptability and word-relatedness effects. *Memory & Cognition*, 8(4), 336-344.

Kleinman, D., Runnqvist, E., & Ferreira, V. S. (2015). Single-word predictions of upcoming language during comprehension: evidence from the cumulative semantic interference task. *Cognitive Psychology*, 79, 68-101.

Klonek, F., Tamm, S., Hofmann, M. J., & Jacobs, A. M. (2009). Does familiarity or conflict account for performance in the word-stem completion task? Evidence from behavioural and event-related-potential data. *Psychological Research*, 73(6), 871-882.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, 95(1), 95-127.

Kompus, K., Eichele, T., Hugdahl, K., & Nyberg, L. (2011). Multimodal imaging of incidental retrieval: the low route to memory. *Journal of Cognitive Neuroscience*, 23(4), 947-960.

Konkel, A., & Cohen, N. J. (2009). Relational memory and the hippocampus: representations and methods. *Frontiers in Neuroscience*, 3, 23.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32-59.

Kurczek, J., Brown-Schmidt, S., & Duff, M. (2013). Hippocampal contributions to language: Evidence of referential processing deficits in amnesia. *Journal of Experimental Psychology: General*, 142(4), 1346.

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

Kutas, M., DeLong, K. A., & 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, pp. 190–207. Oxford University Press.

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

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, 62, 621.

Kutas, M., & Hillyard, S. A. (1980a). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1980b). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, 11(2), 99-116.

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

Kutas, M., & Hillyard, S. A. (1989). An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience*, 1(1), 38-49.

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

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 Effects of Prediction from Association in Single-word Contexts. *Journal of Cognitive Neuroscience*, 25(3), 484-502.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de) constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920-933.

Leutheuser, H., Gabsteiger, F., Hebenstreit, F., Reis, P., Lochmann, M., & Eskofier, B. (2013). Comparison of the AMICA and the InfoMax algorithm for the reduction of electromyogenic artifacts in EEG data. *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6804-6807.

Leynes, P. A., & Phillips, M. C. (2008). Event-related potential (ERP) evidence for varied recollection during source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 741.

Lorsbach, T. C., Wilson, S., & Reimer, J. F. (1996). Memory for relevant and irrelevant information: Evidence for deficient inhibitory processes in language/learning disabled children. *Contemporary Educational Psychology*, 21(4), 447-466.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, 383(6601), 616-618.

Luke, S. G., & Christianson, K. (2012). Semantic predictability eliminates the transposed-letter effect. *Memory & Cognition*, 40(4), 628.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, 88, 22-60.

Maess, B., Mamashli, F., Obleser, J., Helle, L., & Friederici, A. D. (2016). Prediction signatures in the brain: Semantic pre-activation during language comprehension. *Frontiers in Human Neuroscience*, 10.

Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalography and Clinical Neurophysiology*, 86(4), 283-293.

Mandler, G. (1980). Recognizing: the judgment of previous occurrence. *Psychological Review*, 87, 252–271.

Mandler, G. (1991). Your face looks familiar but I can't remember your name: A review of dual process theory. In E. William, E. Hockley, & E. S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 207–225). Hillsdale, NJ: Erlbaum.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology: Human Perception and Performance*, 17(4), 1057.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8(1), 1-71.

Martin, S., Dressaire, D., Guerdoux, E., Trouillet, R., Brouillet, D., Brouillet, T., & Maury, P. (2009). French word-stem completion norms by participants from 30 to 93 years old. *European Review of Applied Psychology*, 59(1), 25-36.

Masson, M. E. (1986). Comprehension of rapidly presented sentences: The mind is quicker than the eye. *Journal of Memory and Language*, 25(5), 588-604.

McCallum, W. C., Farmer, S. F., & Pocock, P. V. (1984). The effects of physical and semantic incongruites on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology*, 59(6), 477-488.

McClelland, J. L., & O'Regan, J. K. (1981). Expectations increase the benefit derived from parafoveal visual information in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 7(3), 634.

McDonald, S. A., & Shillcock, R. C. (2003a). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, 14(6), 648-652.

McDonald, S. A., & Shillcock, R. C. (2003b). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, 43(16), 1735-1751.

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

Migo, E. M., Roper, A., Montaldi, D., & Mayes, A. R. (2010). British English norms for the spontaneous completion of three-letter word stems. *Behavior Research Methods*, 42(2), 470-473.

Moreno, E. M., Casado, P., & Martín-Loeches, M. (2016). Tell me sweet little lies: An event-related potentials study on the processing of social lies. *Cognitive, Affective, & Behavioral Neuroscience*, 16(4), 616-625.

Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1), 92.

Moss, H. E., & Marslen-Wilson, W. D. (1993). Access to word meanings during spoken language comprehension: effects of sentential semantic context. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 19(6), 1254.

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, 26(2-3), 131-157.

Narayanan, N. S., Cavanagh, J. F., Frank, M. J., & Laubach, M. (2013). Common medial frontal mechanisms of adaptive control in humans and rodents. *Nature Neuroscience*, 16(12), 1888-1895.

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

Neely, J. H., & Keefe, D. E. (1989). Semantic context effects on visual word processing: A hybrid prospective-retrospective processing theory. *Psychology of Learning and Motivation*, 24, 207-248.

Nessler, D., Mecklinger, A., & Penney, T. B. (2001). Event related brain potentials and illusory memories: the effects of differential encoding. *Cognitive Brain Research*, 10(3), 283-301.

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(1), 75-92.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus--norepinephrine system. *Psychological Bulletin*, 131(4), 510.

Nigbur, R., Cohen, M. X., Ridderinkhof, K. R., & Stürmer, B. (2012). Theta dynamics reveal domain-specific control over stimulus and response conflict. *Journal of Cognitive Neuroscience*, 24(5), 1264-1274.

Norris, D. (2006). The Bayesian reader: explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, 113(2), 327.

Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA.

Olichney, J. M., Van Petten, C., Paller, K. A., Salmon, D. P., Iragui, V. J., & Kutas, M. (2000). Word repetition in amnesia: Electrophysiological measures of impaired and spared memory. *Brain*, 123, 1948.

Otten, M., & Van Berkum, J. J. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, 45(6), 464-496.

Paller, K. A. (1990). Recall and stem-completion priming have different electrophysiological correlates and are modified differentially by directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(6), 1021.

Paller, K. A. (2001). Neurocognitive foundations of human memory. In D. L. Medin (Ed.), *The Psychology of Learning and Motivation* (Vol. 40, pp. 121-145). San Diego: Academic Press.

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(4), 375-392.

Palmer, J.A., Kreutz-Delgado, K., Makeig, S. (in prep). AMICA: An Adaptive Mixture of Independent Component Analyzers with Shared Components.

Payne, B. R., Lee, C. L., & Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. *Psychophysiology*, 52(11), 1456-1469.

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

Penolazzi, B., Hauk, O., & Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology*, 74(3), 374-388.

Pfurtscheller, G. (1977). Graphical display and statistical evaluation of event-related desynchronization (ERD). *Electroencephalography and Clinical Neurophysiology*, 43(5), 757-760.

Phillips, J. M., Vinck, M., Everling, S., & Womelsdorf, T. (2014). A long-range fronto-parietal 5- to 10-Hz network predicts" top-down" controlled guidance in a task-switch paradigm. *Cerebral Cortex*, 24(8), 1996-2008.

Pijnacker, J., Geurts, B., Van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2010). Exceptions and anomalies: An ERP study on context sensitivity in autism. *Neuropsychologia*, 48(10), 2940-2951.

Polich, J. (1986). Attention, probability, and task demands as determinants of P300 latency from auditory stimuli. *Electroencephalography and Clinical Neurophysiology*, 63(3), 251-259.

Polich, J. (1987). Task difficulty, probability, and inter-stimulus interval as determinants of P300 from auditory stimuli. *Electroencephalography and Clinical Neurophysiology*, 68(4), 311-320.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116(1), 129.

Posner, M. I., & Snyder, C. (1975). Attention and cognitive control. *In R. L. Solso (Ed.), Information processing and cognition: The Loyola symposium*. Hillsdale. NJ: Erlbaum.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109(2), 160.

Pritchard, W. S. (1981). Psychophysiology of P300. *Psychological Bulletin*, 89(3), 506.

Pulvermüller, F., Lutzenberger, W., & Preissl, H. (1999). Nouns and verbs in the intact brain: evidence from event-related potentials and high-frequency cortical responses. *Cerebral Cortex*, 9(5), 497-506.

Rabbitt, P., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *The Quarterly Journal of Experimental Psychology*, 29(4), 727-743.

Ranganath, C. (2010). Binding items and contexts the cognitive neuroscience of episodic memory. *Current Directions in Psychological Science*, 19(3), 131-137.

Ranganath, C., Cohen, M. X., & Brozinsky, C. J. (2005). Working memory maintenance contributes to long-term memory formation: neural and behavioral evidence. *Journal of Cognitive Neuroscience*, 17(7), 994-1010.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, 7(1), 65-81.

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

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist*, 45(9), 1043.

Roediger, H.L., III, & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 803–814.

Roediger, H. L., Weldon, M. S., Stadler, M. L., & Riegler, G. L. (1992). Direct comparison of two implicit memory tests: Word fragment and word stem completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(6), 1251.

Roll, M., Horne, M., & Lindgren, M. (2010). Word accents and morphology—ERPs of Swedish word processing. *Brain Research*, 1330, 114-123.

Roll, M., Söderström, P., & Horne, M. (2013). Word-stem tones cue suffixes in the brain. *Brain Research*, 1520, 116-120.

Roll, M., Söderström, P., Mannfolk, P., Shtyrov, Y., Johansson, M., van Westen, D., & Horne, M. (2015). Word tones cueing morphosyntactic structure: Neuroanatomical substrates and activation time-course assessed by EEG and fMRI. *Brain and Language*, 150, 14-21.

Rommers, J., Dickson, D. S., Norton, J. J., Wlotko, E. W., & Federmeier, K. D. (2016). Alpha and theta band dynamics related to sentential constraint and word expectancy. *Language, Cognition and Neuroscience*, 1-14.

Rommers, J., & Federmeier, K. D. (2016). Downstream repetition effects reveal impoverished representations for predictable words. *Poster presented at the Society for Psychophysiological Research (SPR), Sept 21-25, Minneapolis, USA*.

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, 51(3), 437-447.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory & Cognition*, 18(4), 367-379.

Rugg, M. D. (1995). ERP studies of memory. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of Mind: Event-related brain potentials and cognition* (pp. 132-170). Oxford: Oxford Science Publications.

Rugg, M. D., Allan, K., & Birch, C. S. (2000). Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. *Journal of Cognitive Neuroscience*, 12(4), 664-678.

Rugg, M. D., Cox, C. J., Doyle, M. C., & Wells, T. (1995). Event-related potentials and the recollection of low and high frequency words. *Neuropsychologia*, 33(4), 471-484.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, 11(6), 251-257.

Rugg, M. D., Doyle, M. C., & Melan, C. (1993). An event-related potential study of the effects of within-and across-modality word repetition. *Language and Cognitive Processes*, 8(4), 357-377.

Rugg, M. D., Doyle, M. C., & Wells, T. (1995). Word and nonword repetition within-and across-modality: An event-related potential study. *Journal of Cognitive Neuroscience*, 7(2), 209-227.

Rugg, M., Mark, R., Walla, P., Schloerscheidt, A., Birch, C., & Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature*, 392(6676), 595-598.

Rugg, M. D., Walla, P., Schloerscheidt, A. M., Fletcher, P. C., Frith, C. D., & Dolan, R. J. (1998). Neural correlates of depth of processing effects on recollection: evidence from brain potentials and positron emission tomography. *Experimental Brain Research*, 123(1-2), 18-23.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2(6), 110-114.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience*, 8(9), 657-661.

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience*, 26(8), 1736-1747.

Schuberth, R. E., & Eimas, P. D. (1977). Effects of context on the classification of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance*, 3(1), 27.

Schuberth, R. E., Spoehr, K. T., & Lane, D. M. (1981). Effects of stimulus and contextual information on the lexical decision process. *Memory & Cognition*, 9(1), 68-77.

Schvaneveldt, R. W., Meyer, D. E., & Becker, C. A. (1976). Lexical ambiguity, semantic context, and visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 2(2), 243.

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(2), 344.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, 24(2), 232-252.

Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly*, 114-126.

Senkfor, A. J., & Van Petten, C. (1998). Who said what? An event-related potential investigation of source and item memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(4), 1005.

Shaw, R. J. (1997). Unprimed stem completion is only moderately predicted by word frequency and length. *Behavior Research Methods, Instruments, & Computers*, 29(3), 401-424.

Shohamy, D., & Turk-Browne, N. B. (2013). Mechanisms for widespread hippocampal involvement in cognition. *Journal of Experimental Psychology: General*, 142(4), 1159.

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(1), 88.

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two Neurocognitive Mechanisms of Semantic Integration during the Comprehension of Visual Real-world Events. *Journal of Cognitive Neuroscience*, 20(11), 1-21.

Smith, N., & Levy, R. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In *Proceedings of the Cognitive Science Society* (Vol. 33, No. 33).

Söderström, P., Horne, M., Frid, J., & Roll, M. (2016). Pre-Activation Negativity (PrAN) in Brain Potentials to Unfolding Words. *Frontiers in Human Neuroscience*, 10, 512-512.

Soler, M. J., Dasí, C., Ruiz, J. C., & Cervera, T. (2017). Spanish norms for completion of restricted length word stems. *Behavior Research Methods*, 49(5), 1920-1929.

Staub, A., & Clifton Jr., C. (2006). Syntactic Prediction in Language Comprehension: Evidence From Either… or. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 32(2), 425.

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General*, 112(1), 1.

Steffens, M. C., & Mecklenbräuker, S. (2007). False memories: Phenomena, theories, and implications. *Zeitschrift Für Psychologie/Journal of Psychology*, 215(1), 12-24.

Stenberg, G., Hellman, J., Johansson, M., & Rosén, I. (2009). Familiarity or conceptual priming: Event-related potentials in name recognition. *Journal of Cognitive Neuroscience*, 21(3), 447-460.

Stenberg, G., Lindgren, M., Johansson, M., Olsson, A., & Rosén, I. (2000). Semantic processing without conscious identification: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(4), 973.

Strijkers, K., Costa, A., & Thierry, G. (2009). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, 20(4), 912-928.

Stróżak, P., Abedzadeh, D., & Curran, T. (2016). Separating the FN400 and N400 potentials across recognition memory experiments. *Brain Research*, 1635, 41-60.

Sussman, R. S., & Sedivy, J. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, 18(2), 143-163.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, 150(3700), 1187-1188.

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, 68(4), 297-314.

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, 12(1), 43-47.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632.

Taylor, W. L. (1953). "Cloze procedure": a new tool for measuring readability. *Journalism Quarterly*, 30(4), 415.

Thompson-Schill, S. L., Kurtz, K. J., & Gabrieli, J. D. (1998). Effects of semantic and associative relatedness on automatic priming. *Journal of Memory and Language*, 38(4), 440-458.

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.

Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1266.

Tsivilis, D., Otten, L. J., & Rugg, M. D. (2001). Context effects on the neural correlates of recognition memory: an electrophysiological study. *Neuron*, 31(3), 497-505.

Tulving, E., & Gold, C. (1963). Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *Journal of Experimental Psychology*, 66(4), 319.

Tweedy, J. R., Lapinski, R. H., & Schvaneveldt, R. W. (1977). Semantic-context effects on word recognition: Influence of varying the proportion of items presented in an appropriate context. *Memory & Cognition*, 5(1), 84-89.

Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception & Psychophysics*, 34(5), 409-420.

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

Van de Vijver, I., Ridderinkhof, K. R., & Cohen, M. X. (2011). Frontal oscillatory dynamics predict feedback learning and action adjustment. *Journal of Cognitive Neuroscience*, 23(12), 4106-4121.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, 18(4), 380-393.

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open-and closed-class words. *Memory & Cognition*, 19(1), 95-112.

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

Verfaellie, M., & Cermak, L. S. (1999). Perceptual fluency as a cue for recognition judgments in amnesia. *Neuropsychology*, 13(2), 198.

Vilberg, K. L., Moosavi, R. F., & Rugg, M. D. (2006). The relationship between electrophysiological correlates of recollection and amount of information retrieved. *Brain Research*, 1122(1), 161-170.

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, 37(2), 190-203.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, 438(7067), 500-503.

Von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld. *Psychologische Forschung*, 18(1), 299-342.

Voss, J. L., & Federmeier, K. D. (2011). FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing. *Psychophysiology*, 48(4), 532-546.

Voss, J. L., Schendan, H. E., & Paller, K. A. (2010). Finding meaning in novel geometric shapes influences electrophysiological correlates of repetition and dissociates perceptual and conceptual priming. *NeuroImage*, 49(3), 2879–89.

Wang, W., Li, B., Gao, C., Xiao, X., & Guo, C. (2014). Electrophysiological correlates associated with contributions of perceptual and conceptual fluency to familiarity. *Frontiers in Human Neuroscience*, 9, 321-321.

Wang, W. C., Ranganath, C., & Yonelinas, A. P. (2014). Activity reductions in perirhinal cortex predict conceptual priming and familiarity-based recognition. *Neuropsychologia*, 52, 19-26.

Wang, W. C., & Yonelinas, A. P. (2012). Familiarity is related to conceptual implicit memory: An examination of individual differences. *Psychonomic Bulletin & Review*, 19(6), 1154-1164.

Wang, L., Zhu, Z., & Bastiaansen, M. (2012). Integration or predictability? A further specification of the functional role of gamma oscillations in language comprehension. *Frontiers in Psychology*, 3, 187.

West, R. F., & Stanovich, K. E. (1978). Automatic contextual facilitation in readers of three ages. *Child Development*, 717-727.

Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, 346, 165-168.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender Agreement in Spanish Sentence Reading. *Journal of Cognitive Neuroscience*, 16(7), 1272-1288.

Wilding, E. L. (2000). In what way does the parietal ERP old/new effect index recollection?. *International Journal of Psychophysiology*, 35(1), 81-87.

Wilding, E. L., & Rugg, M. D. (1996). An event-related potential study of recognition memory with and without retrieval of source. *Brain*, 119(3), 889-905.

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language*, 101(3), 278-289.

Wills, A. J., Lavric, A., Croft, G. S., & Hodgson, T. L. (2007). Predictive learning, prediction errors, and attention: Evidence from event-related potentials and eye tracking. *Journal of Cognitive Neuroscience*, 19(5), 843-854.

Wlotko, E. W., & Federmeier, K. D. (2007). Finding the Right Word: Hemispheric Asymmetries in the Use of Sentence Context Information. *Neuropsychologia*, 45(13), 3001-3014.

Wlotko, E. W., & Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology*, 49(6), 770-785.

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, 68, 20-32.

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(4), 975.

Wlotko, E.W., Zeitlin, M., & Kuperberg, G.R. (2015). An event-related brain potential investigation of multi-level probabilistic expectations in sentence comprehension. *Poster presented at the 22nd Annual Meeting for the Cognitive Neuroscience*.

Wlotko, E.W., Zeitlin, M., Riley, S., Cunha-Lima, M.L., & Kuperberg, G.R. (2016). The brain dissociates between different levels of prediction during language comprehension. *Poster presented at the 8th Annual Society for the Neurobiology of Language Conference*.

Wolk, D. A., Schacter, D. L., Berman, A. R., Holcomb, P. J., Daffner, K. R., & Budson, A. E. (2004). An electrophysiological investigation of the relationship between conceptual fluency and familiarity. *Neuroscience Letters*, 369(2), 150-155.

Wolk, D. A., Schacter, D. L., Lygizos, M., Sen, N. M., Holcomb, P. J., Daffner, K. R., & Budson, A. E. (2006). ERP correlates of recognition memory: Effects of retention interval and false alarms. *Brain Research*, 1096(1), 148-162.

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, 11(18), R729-R732.

Woodruff, C. C., Hayama, H. R., & Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Research*, 1100(1), 125-135.

Woodward, S. H., Ford, J. M., & Hammett, S. C. (1993). N4 to spoken sentences in young and older subjects. *Electroencephalography and Clinical Neurophysiology*, 87(5), 306-320.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46(3), 441-517.

Yu, S. S., & Rugg, M. D. (2010). Dissociation of the electrophysiological correlates of familiarity strength and item repetition. *Brain Research*, 1320, 74–84.

Yun, H., Mauner, G., Roland, D., & Koenig, J. P. (2012). The Effect of Semantic Similarity is a Function of Contextual Constraint. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*.