ABSTRACT

Title of Dissertation:	EEG EFFECTS OF EVENT MODELS IN STORY COMPREHENSION
	Benjamin Bogart Rickles, [Doctor of Philosophy], 2023
Dissertation directed by:	Associate Professor Donald J. Bolger, Department of Human Development and Quantitative Methodology

Cognitive models can offer deep insights into how stories are comprehended. Models which follow event segmentation theory (EST) focus on the processing of brief episodes or events within a narrative and the boundaries between events. To test the brain mechanisms proposed by EST to occur at the event boundaries we looked at electroencephalographs (EEG) recorded from 49 participants as they were tasked with both listening to and recalling 9 blocks of ~ 6 minute-long audio clips in one of three conditions: single ordered stories, unrelated events from unrelated stories, or single stories in scrambled order. All stimuli were designed to contain event boundaries spaced at semi-regular intervals. Accuracy during an inference recognition task administered after each block was highest in the single ordered stories condition. Analysis 1 examined the effects of event boundary vs. local semantic context on evoked negativities (N400) related to lexical processing of each word. Effects of condition suggest that narrative structure affected lexical processing, more so than event-level structure and sentence-level semantic context. Analysis 2 Examined changes in alpha (8.5-12.5 Hz) and theta (4-8 Hz) band power of the EEG induced by the onset of the event boundary. Boundary-induced changes in both frequencies were recorded, in all conditions. The largest increases were recorded during the ordered stories over large portions of the scalp. How these findings relate to cognitive mechanisms suggested by event segmentation theory is discussed.

EEG EFFECTS OF EVENT MODELS IN STORY COMPREHENSION

by

Benjamin Bogart Rickles

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctorate in Philosophy
2023

Advisory Committee:
Associate Professor Donald J. Bolger, Chair
Associate Professor Ellen Lau
Assistant Professor Rachel Romeo
Senior Research Scientist Joseph Dien
Professor Charles Perfetti

Dedication

To Eddie, brother to me.

Acknowledgements

First and foremost, I would like to acknowledge my advisor, DJ Bolger, whose willingness to give me freedom in my research was afforded by the breadth of his knowledge. Beyond providing me with the tools and guidance to complete this project, his zeal to achieve a more equitable society by spreading the gospel of language science to instructors was motivating to be around. I must acknowledge the help of my wife, Kira, who helped me edit this and who made me excited to finish this thing so we could continue to enjoy life together. I also would like to thank my committee, Joseph Dien, Ellen Lau, Rachel Romeo, and Charles Perfetti who were each patient with me throughout the process and who have taken the time to provide me with helpful advice. I would also like to thank my business partner, Kristina Christmon, a true leader, whose bravery I have witnessed on several occasions and who gave me a chance to branch out of my field and get into the *fields*. I would also like to thank my former roommates at 8819 Patricia Ct. whose curiosity and camaraderie lead them on one occasion to listen to and attempt to recall scrambled stories.

Going way back, I must also acknowledge the members of the LRDC at the University of Pittsburgh, including my first labmates and colleagues. I want to thank Tessa Warren whose courses introduced me to questions about how the brain does language, and Charles Perfetti who has motivated me to do my own research as a research assistant in his lab. I also need to thank Gwen Frishkoff who trained me in electroencephalography and who first took me as her student. I need to thank Naomi

Feldman and Jo Shoemaker for introducing me to vector space models and Christian Brodbeck for helping me find the forced aligner.

From there I would have to thank each person who helped support me along the way in addition to the financial, intellectual, and emotional support of my parents, family, and friends as well as fellow grad students at GSU, U of O, and UMD. I would like to thank Daniel Finkel who gave me the clue to start focusing on event segmentation theory more seriously. I would like to thank Kevin Dunbar whose attitude towards the use of metaphors in conversation aligns with my own and whose discussions about schemata were unforgettable. Conversations with Jessica and Matt Turner, Regina Calloway, Joe Stafura, Jeremy Purcell, Juan-Camillo Barbosa, Edgar Perez ether lead me to the ideas in this thesis or helped me to sharpen them. I would also like to thank Chris Tennsmeyer at Adobe for hooking me up with a summer job and Pam Komarek and NACS faculty and students at UMD who provided me with a stipend a few years ago and a fellowship last fall while I was able to pilot some of these stimuli and write the proposal for the thesis. I would also like to thank Penny Gomes who helped find the room for the defense.

I would be remiss if I did not thank my research assistants, Jaylen Worthy and Carlos Koudouovoh who were with me the whole way on this project. Lastly, I must thank all of the UMD students who participated in this study.

List of Tables

List of Figures

List of Abbreviations

DFB: distance from boundary

ESPM: event segmentation process models

EST: Event Segmentation Theory

SO: Semantic Overlap

LSP: Lexico-semantic processing

EEG: Electroencephalography

fMRI: functional magnetic resonance imaging

ERP: Event related Potential

BOLD: blood oxygenation level-dependent signal

ANOVA: analysis of variance

WM: Working memory

LTM: Long-term memory

MTL: Medial temporal lobe

MTG: middle temporal gyrus

FFT: Fast Fourier transform

TF:Time-frequency

VSM: Vector Space Models

GloVe: Global Vectors

Chapter 1: Introduction

And everything we believe comes from a story we've been told.

– Heather McGhee (2022. p xxiii)

1.1 The story of comprehending stories

Comprehending stories is essential to making meaning of our lives and for successfully communicating with others. Comprehension of stories and other genres of narrative relies on creating mental representations or *situation models* of the actions described in the text, movie, play, epic poem, etc. which are stored in memory (Van Dijk & Kintsch, 1983, Johnson-Laird, 1983). Thus comprehension refers to the act of creating situation models. Various cognitive theories have been offered to explain how larger situation models, such as those for stories, are structured (Bartlett, 1932, Thorndyke, 1977, Mandler & Johnson, 1977, Rumelhart, 1977, Radvansky & Zacks, 2017, Cohn, 2020). As structures, situation models for stories are not indivisible blobs, but gestalten of a series of smaller representations. One set of theories stemming from the recently established field of event cognition (Radvansky & Zachs, 2016) propose that the smaller structures which comprise situation models are smaller models representing events. The operational definitions of *events* emerged from behaviors of experimental subjects who were asked to apply their intuitive notions of human activities while watching images depicted in videos, or else to recall what they heard described in narratives (Newtson & Enquist, 1979, Newtson et al., 1987). People's ability to demarcate boundaries between activities and to recall details clustered around certain activities is explained by the existence of mental representations of activities or event models. The theory of *event* *models* was thus built upon perception and memory of human activities. Stories as a genre of narrative comprise series of descriptions of human activities, and as such are most easily translatable into events (compared to, but not excluding biology textbooks). In this way the goal of story comprehension in comparison to comprehension of other genres can be most clearly seen as the successful creation of a series of event models.

This thesis is dedicated to understanding how the brain parses these event representations during story comprehension. If the serial creation of event models can be tracked through neurophysiological methods such as electroencephalography (EEG) then the event model can be treated as a unit of comprehension. Such a unit would provide some reference with which to consider the multitude of cognitive operations required for creating the situation model structure of a longer discourse, like that of a story, and may explain some of the extralinguistic effects seen in studies of discourse beyond the isolated sentence (Ferstyl et al., 2007, Perfetti & Helder, 2022). With the experiment and analyses discussed in this thesis, we can begin to test specific processes in relation to the event model. However, achieving this perspective requires that we must overcome various impediments to studying longer, *naturalistic*, or *extended* discourse such as stories (see Sassenhagen, 2018, for a discussion). We look to overcome these difficulties through careful balancing and control, within the confines of what is possible, while maintaining the entertainment factor of our stimuli. How we achieved this can be seen in the methods section of Chapter 3.

1. 2 Event segmentation process models

According to event segmentation theory and the process models of comprehension it inspired (Kurby & Zacks, 2008, Radvansky & Zacks, 2011, Richmond, Gold, & Zacks, 2017, Bauer & Varga, 2017), while experiencing a story, comprehension starts somewhat freshly, or *from scratch* with each new event (Kurby & Zacks, 2008). I refer to these process models of what the brain is doing during comprehension as *event model theories,* or ESPM. As such, ESPM make a number of hypotheses about the mechanisms which guide comprehension, each with respect to the place where one event model should begin and another should end, the *event boundary*. While these mechanisms provide explanations for a number of behaviors observed while people comprehend stories, and an intriguing path for future research, precisely how they are substantiated in the brain remains unclear.

That event model creation starts *from scratch* could have multiple interpretations within event model theories, given the multitude of processes assumed to occur during comprehension. To say something begins *from scratch* usually implies that there is no extant process to utilize. However, when successfully switching sub-events within a story or narrative, it is not as if the entire narrative representation or situation model preceding the event model is lost. It is more likely that even if event models are made from scratch, certain ingredients are already removed from the shelves and are sitting on the kitchen counter. We know that some processes are preserved between event models since people with amnesia have normal recall ability for narratives if cued immediately afterwards (Baddeley & Wilson, 2002). We also know that event models have access to long-term memory, for example, the event models created while watching Fast and Furious 9 were in some way informed by event models formed while watching Fast and Furious 8. We also often experience continuity when a detail refers back to something just seen in the previous event. For example, in *Taledega Nights: the balled of Ricky Bobby* (McKay et al., 2006), an opening segment of a modern stockcar race is preceded by an event wherein of a father gave his son advice about racing before driving his car fast down the street, fading to black. Although the setting has shifted, the immediate relevance of the racetrack to the context of the previous event (mention of *racecar*, racing advice, car speeding down the street) would not be possible without some lingering semantic content from the previous event. Surely, the idea of starting context from scratch is hyperbolic. However the number of behavioral and neural phenomena surrounding events and even boundaries requires we understand what processes are affected, when this occurs, and to which degree. It is therefor most likely that even if event models are made from scratch, certain ingredients are already removed from the shelves and on the counter.

1.3 Brain measurement of ESPM hypotheses

In addition to these conscious references to specific items in memory we also expect that unconscious processes are available to guide comprehension between stimuli at shorter lengths of time. It soon becomes less clear however, whether the same memory systems are responsible for allowing event models to consciously reference information from other event models across longer times for example, across adjacent event models, or across the first and last events of a story, novel, play (see Ericcson & Kintsch, 1995, for discussion). For example, is brief activation of context able to support unconscious processing across the gap between event models? ESPM would suggest that it is not.

Must information which is within the event model remain active? Leveraging the predictions of processes in ESPM, we can make hypotheses, such as *yes, they are.* We can also look to ESPM to suggest hypotheses about long-term memory during comprehension which they predict occurs after a new event boundary is detected. We can then test those hypotheses with neurophysiological measurements.

1.4 Lexico-semantic processing

It has long been suspected that the semantic coherence of the words in a story should decrease after boundaries between episodes, with less-related word meanings reflecting the new location, activities, goals, or characters in the new episode (Gernsbacher, 1990). ESPMs suggest that detection of lexical or semantic incongruity may act as a boundary detection mechanism, triggering the brain to initiate the creation of a new event model (Kurby & Zacks, 2008). Since words that are either semantically (ball-bounce) or lexically (snow- cone) related may facilitate unconscious processing of proximal words both in isolation (Neely, 1977, McNamara, 1999), and in sentences (Duffy, Henderson, & Morris, 1989), we would suspect that this facilitation would no longer be available after the boundary. This could be for two reasons, both that the context shifted and because the event boundary has removed the activated context. For these reasons, even if relevant context such as related words had been experienced and activated (which is expectedly less likely across the event boundary), the boundary itself would void this activation.

For the measurement of unconscious processing from briefly-activated context we have event-related voltage potentials or ERPs. ERPs are deflections in the EEG recorded in response to a stimulus event such as the onset or offset of a word and can be informative about a range of contextual factors on word processing. In particular, the N400 is an ERP which is often used as a proxy for the unconscious contextual facilitation of word processing by its context and is ontologically characterized as a negative dip in the voltage over central-posterior scalp reaching its minima between 200-600 milliseconds after the onset of a word in vision or audition. Ambiguity about whether the N400 best reflects ease of access to the lexical or semantic representation of a word, causes us to refer to it as a measure of *lexico-semantic processing* or LSP (Kutas & Federmeier, 2011). The context which may help or encumber LSP at the word level may come from a variety of levels of discourse, as reflected by the N400, either through world knowledge, (Hagoort, et al, 2004) message level and word-level semantics (Otten & Van Berkum, 2007, Ettinger et al., 2016).

Another ERP related to word processing in context is the P600, which is characterized as a positive peak in voltage located over central-posterior midline scalp between 600-1000 milliseconds after a word is presented. The amplitude of the P600 has been shown to increase with a range of contextual manipulations, and in conjunction with the N400, has been through to reflect processes related to inference (Burkhardt, 2006), or other memory processes which may benefit the creation of a coherent mental model (Kim & Osterhout, 2005). Alternatively, the P600 may also reflect surprisal in a similar way to the P3b, an ERP which is a response to infrequent but not necessarily incorrect variation (Coulson, King, and Kutas, 1998, Sassenhaggen, Schlesewsky & Bornkessel-Schlesewsky, 2014, c.f. Frisch, et al., 2003). While we do not examine the P600 in this analysis, we do look at the N400 in chapter 3.

1.5 Alpha EEG power and Working memory

ESPM predict that information in the event model becomes less accessible after the event boundary. Evidence for this loss of accessibility, comes primarily from early experiments showing an alteration of memory at a boundary in the situation model for narratives (Gernsbacher, 1985, Bransford & Franks 1971, Dooling & Lachman, 1968, Rink & Bower, 2000), as well as first-person virtual experiences (Pettijohn & Radvansky 2016). While ESPM suggest that the contents of working memory are less accessible following an event boundary, we would expect decreases in measures of active working memory, such as those processes suggested in the episodic buffer suggested by Baddeley (2000).

While we do not necessarily have a direct brain measure of working memory, various uses of EEG frequency measurements over time, or *time-frequency* measurements, have been used to get a sense of the different brain responses during periods in which memoranda are successfully maintained in memory for brief periods of a few seconds, or longer. One consistent brain measurement which increases linearly with working memory load is power in the alpha band, which refers to the absolute value of amplitude of oscillations in the frequency range of 8-12 Hz (Klimesche, 2012).

The increase in alpha power over perceptual cortex in congruence with the domain of task stimuli (Thut et al, 2006) can be interpreted in multiple ways. During memory maintenance alpha might be related to blocking out external stimuli as some suggest (Jensen & Mazaheri, 2010, Smallwood, 2011), or in the executive engagement of task relevant knowledge, as others suggest (Klimesche, 2012). Some have taken multiple approaches to interpreting alpha power depending on location, with alpha power over frontal electrodes to signify loss of consciousness awareness, while alpha over posterior areas alpha power was detectable during successful use of working memory (Boudewyn & Carter, 2018). Evidence that event boundaries increase conscious attention to outside stimuli (Eisenberg et al., 2016, Faber et al., 2018) would suggest by any and all accounts that alpha over task-relevant perceptual areas such as visual cortex, or perceptual cortex, should decrease at the boundary. By the Klimesche (2012) account, decreased access to contents of working memory would also suggest a decreased executive control of knowledge, and decreased alpha. Thus, both accounts of the functionality of alpha power would cause us to expect to see greater alpha power in the middle of the event model, and less following a boundary. We examine alpha EEG power invoked by the event boundary in chapter 4.

1.6 Theta EEG power and Long-Term memory Encoding

ESPM predict that information in the event model becomes transferred to long-term memory after the boundary (Kurby et al., 2008, Richmond et al., 2017, Radvansky, et al., 2017, Bauer et al., 2017). Measures of long-term memory storage have been tested with paradigms which expose participants to test items during a *study* phase, and then test whether participants remember specific, individual items during a later, test phase. The so-called *subsequent memory effects* (SME) are meant to estimate proper encoding of memory by looking back to the study phases for items successfully remembered. SME for words (Klimesch et al., 1997) pictures (Khader et al., 2010), and discourse (Sato & Mizahura, 2018) are seen in increased theta EEG power over frontal midline (Klimesch, et al., 1997), parietal (Khader et al., 2010), and temporal (Sato et al., 2018) electrodes during the study phase. Theta activity is in the range of 3-7 Hz, has been linked with activity of the hippocampus and cortex (Buzsaki, 1996, Lega, Jacobs, & Kahana, 2012) presumably to coordinate cortical sources of perception (Hassabis et al., 2007, Hsie & Ranganath, 2014). Increasingly relevant to this thesis are studies using eletrocorticography (eCOG) which have shown single cells within the hippocampus to respond to event boundaries during encoding (Zheng et al., 2022, Yoo, Umbach, & Lega, 2022). This fits with the predictions of ESPM as to when SMEs for event models should occur. We would hope that these effects would be detectable with EEG as well. However, the exact locations of where the effects event boundary effects will be detected is difficult to predict a priori. From Sato et al. (2018) we might expect the effects to be left lateralized. We examine theta EEG power invoked by the event boundary in chapter 4.

1.8 Implications

Whether this lingering knowledge which passes or does not pass between event models is in the domain of working memory, schematic knowledge in long-term memory, lower level lexical or semantic activation, or any combination is important to understand, not just for elaborating the theory but for helping to understand the emergence of higher-level deficits of comprehension from both clinical and educational perspectives.

The study of how comprehension is organized by events might also come to offer an analogue of how people organize their behavior in terms of everyday actions and goals. Evidence that the ability to parse continuous actions into separate events is a skill which slowly grows during development (Baldwin, et al., 2001) and decreases with age

(Zacks, et al., 2006, Cannizzaro & Cohelo, 2013), lends credence to the idea that studying event models may lead to insights in our study of cognition across the lifespan. Evidence that instructing people to identify event boundaries during comprehension leads to enhancement of memory for stories (Thompson & Radvansky, 2014, Flores et al., 2018), has led some to promote their use in educational and clinical interventions to aid deficits in comprehension (Richmond, et al., 2017, Cannizzaro & Coelho, 2013). If these ESPM are going to be put to good use, understanding the cognitive mechanisms which underlie how event models constrain comprehension of stories is a worthwhile endeavor.

Studying story comprehension is made difficult by the subjective nature of comprehension which adds randomness to empirical experimentation and decreases statistical power (Sassenhagen, 2019). To date, I could only find one study which explicitly examined EEG effects of the event boundary, and this was for participants watching videos (Sharp, 2010). In this thesis we plan to overcome the hurdles of quantifying EEG effects during comprehension with a design centered on monitoring the event boundary while participants listen to stories in a few experimental conditions. This method will allow us to control the stimuli enough to evaluate some of these mechanisms provided by ESPM, specifically those related to how contextual facilitation of word processing, perception, and memory storage occur at the event boundary, while people listen to events. The experiment is designed to detect deviations in each of the three EEG measurements discussed above with respect to the event boundary at the level of the sentence. While it should be noted that the interpretations of the EEG findings may differ according to the specific functional interpretations linking brain activity to cognitive models, boundary effects should help elaborate the hypotheses ESPM make about comprehension of stories and more.

1.9 Specific Aims

1. Analysis 1: Do lexico-semantic effects of context vary with event model creation? From the studies detailed above and in the previous section, two factors driving lexico-semantic processing (LSP) appear: one in which the N400 is driven by semantic relatedness of nearby words, and another in which the N400 is driven by the distance from the event boundary (DFB):
   a. While other Event-Related Potentials (such as the P600) may be affected by context, we will focus on whether the DFB interferes with LSP via the N400.
      i. DFB will predict Recognition accuracy (not in Scrambled Stories)
   b. Semantic overlap (SO):
      i. SO (as measured by a computational model) will predict mean N400, beyond the relationship between SO and DFB

2. Analysis 2a: Does the event model predict EEG correlates of attention during story comprehension?

   As Event Segmentation Theory (EST) predicts that perceptual attention is expected to increase after crossing an event boundary, we expect that changes in EEG TF measures of perceptual attention (temporal alpha waves) will show the least power at this point.

a. Distance from Boundary (DFB)

i. Alpha EEG (8-12 Hz) power over large parts of scalp will decrease after the event boundary (not in Scrambled stories)

3. (Analysis 2b): Does the event model predict EEG correlates of memory encoding during story comprehension?

a. Analysis 2b) Does the event model predict EEG correlates of episodic memory encoding during story comprehension?

i. Theta EEG (3-7 Hz) power over frontal midline or left temporal sites will increase after the event boundary (not in Scrambled stories)

Chapter 2: Background

2.1 Defining Events

The study of action perception, like comprehension, has been concerned with events as units of memory. Researchers in action perception have defined events as wherever an outside observer perceives goal-directed actions to begin and end (Lichtenstein & Brewer, 1980, Newtson, 1987). These theories specify that when the goal of a character is recognized, the actions which lead to its completion are grouped in memory as a whole. Combining event perception and situation models for comprehension gave rise to the event-indexing model (Zwaan, Radvansky, 1998), which stipulated that comprehenders organize their situation models according to events. Elaborations of this model lead to event segmentation theory (Zacks et al., 2007), which promoted the view of comprehension as the processing of contiguous *event models*: individual situation models for goal-directed actions within a larger narrative, such as a story. Event segmentation theory of Zacks et al, (2007) was elaborated with the creation of event model theories of comprehension (Kurby & Zacks, 2008), which we call ESPM. These ESPM include the event horizon model (Radvansky & Zacks, 2017), the ERISS model (Bauer & Varga, 2017), as well as an earlier, compatible theory, the structure building framework of Gernsbacher, (1990). According to ESPM, story comprehension centers on creating event models in series, one at-a-time as a story is comprehended. This pattern of event model creation in ESPM offers a detailed and parsimonious explanation of how memory and perception operate during comprehension. This also makes for specific hypotheses about the processes that are active at any given time during comprehension.

Zacks and Tversky (2001) define an event as “a segment of time at a given location that is conceived by an observer as having a beginning and an end” (p 17). In stories, these changes in location and time and character’s goals (as well as other variables see Zwaan & Radvansky, 1998) usually coincide with the changes of the goals of a protagonist as they usually travel to a new location to complete a new goal which usually takes place over a single period of time. In ESPM the processing of a sentence (n) depends on whether n is an elaboration of the previous event from sentence (n-1), or whether sentence n describes a new event. The boundaries between events are thus important markers for change in the story and story processing.

There may be many factors which contribute to noticing a boundary in text (see Appendix E i for discussion). The precursor to ESPM, the event-index model (Zwaan, Langston, & Graesser, 1995), posits that violations of assumptions about time, space, intentionality, causality, and protagonist are interpreted as boundaries between episodes, or events. Consider the following sentences from Zwaan (1996):

(1) (a) The professor started analyzing the data. An hour later, her phone rang.
(b) The professor started analyzing the data. At that moment, her phone rang.

Semantic cues about time in (2), (e.g., “An hour later”, vs “At that moment”) are thought to indicate that these actions occur within either separate, or the same episode. The test of this assumption has been done with multiple studies which asked people to indicate when they thought boundaries between meaningful actions occurred in a text. While a variety of changes in a text indicate boundaries between events, (Zwaan &

Radvansky) temporal changes are among the most salient (Zwaan 1996, Speer et al., 2005, Zacks et al., 2009). Speer and Zacks, (2005) and Ezzyat and Davachi, (2011) directly manipulated texts by adding the phrase “an hour later” and “a while later” between events in order to increase the ability for subjects to recognize event boundaries.

2.2 Event schemas

The comprehension models proposed by ESPM are not unlike previous attempts to explain the aberrant behavior of patients with frontal lobe damage. Grafman, et al., (1993), borrowed from structural theories of story comprehension such as those of Rumelhart (1977) to hypothesize structures which guided behavior in terms of hierarchically organized schemas or scripts, whereby behaviors were grouped into overall goals (e.g. going to restaurant), and subgoals (e.g., paying for check).

Knowing when to separate events may also rely on knowledge about goal directed activity (Lichtenstein, 1980). For example, *folding clothes* is a goal with a beginning and an end, as is *mailing an envelope*. Such knowledge about how events play out is referred to as an *event schema*. In ESPM, sensing a change to a new event invokes a new event schema (Kurby & Zacks 2008). Event schemas have been theorized to factor into the interpretation of a protagonist’s goal-directed actions by making what comes next more predictable (Brewer & Lichtenstein, 1980, Richmond, Gold, & Zacks, 2017) (see Appendix E i for exhaustive discussion).

2.3 Event Boundaries

Despite Zacks et al., (2001) vague definition of events, individuals have the ability to identify event boundaries within a story somewhat consistently (Newtson & Engquist, 1976, Bower, Black, & Turner, 1979, Speer & Zacks, 2005, Speer, Zacks & Reynolds, 2007). One method for eliciting the reliable placement of event boundaries is to ask subjects reading a story to mark in time where "one meaningful action ends and another begins" (Newtson, et al., 1976, Zacks, et al., 2009, Ezzyat & Davachi, 2011) (see Appendix E ii for discussion).

The assumption that participants place consistent event boundaries is supported by brain research using functional magnetic resonance imaging (fMRI). With fMRI, brain activity has been seen to reliably track human-derived event boundaries when the people in the scanner were not instructed to attend to them. Fluctuations in the hemodynamic response of a multitude of locations in primary and association cortex coincide with the boundary of an event (Zacks et al., 2001, Ezzyat & Davachi, 2011, Ben-Yakov, Eshel, & Dudail, 2013, Baldassano et al,. 2017) (see Appendix E vii 3 for discussion). This evidence gives credence to the notion that event boundaries are a meaningful structural unit of story segmentation. The brain evidence suggests that cognitive processes which may be involved in comprehension not only track the onsets and offsets of events as they occur, but also relate to how event models are formed between boundaries as a story progresses.

2.4 Situation models

ESPM make specific predictions about how cognitive representations of information from the story are created. In cognitive models of comprehension, the situation model is "the cognitive representation of the events, actions, persons, and in general the situation, a text is about" (Van Dijk & Kintsch, 1983 p 11). In the case of a story comprehension, the text is the story. Thus the situation model forms the basis of knowledge which we keep in memory while comprehending a story (Van Dijk & Kintch, 1983). The situation model is supposed to be the representation of all of the events in the story whether it is read, heard, or watched in a movie, play or real life (Johnson-Laird, 1983, Zwaan & Radvansky, 2018). In this way the situation model during story comprehension is invoked by the sentences and discourse of the story. The situation model is thought to be embellished by life experience such that comprehenders construct their situation model for a story with details which may not be explicitly mentioned in the text, and some of which are unique to each person (Van Dijk & Kintsch, 1983, Zwaan, 2016). Despite the idiosyncrasies in phenomenal details between how two comprehenders construct a situation model in response to the phrase "He was carjacked by a child", evidence using brain imaging techniques to decode information in the brain supports the idea that there is at least some similarity between the content of situation models (Wehbe et al., 2014, Kurby & Zacks, 2013, see Appendix E v for discussion).

Not only is the situation model the subjective experience of events occurring during comprehension, the situation model is also what must be remembered after comprehension (Van Dijk & Kintsch, 1983). Individuals tend to immediately forget perceptual elements or *surface forms* of stories (exact words, grammar, or pictures), while retaining other information about what happened (Bransford & Franks, 1971, Gernsbacher, 1985, Kintsch, et al., 1990). Situation models are thought to represent what happens in a story in what is commonly referred to as the *gist.* Early studies of comprehension showed that the gist remains in memory long enough to be remembered years later (Bartlett, 1995). A domain of long-term memory (LTM) which stores detailed information from first-hand experience with temporal and locational fidelity is called *episodic memory* (Tulving, 2002, Hassabis et al., 2007). Episodic memory allows recall of episodes from personal experiences (Tulving, 2002). While a character's experience in a story is not our own personal experience, our memory of being exposed to the events in the story is. This is seen in the inability of patients with anterograde amnesia to recount neither novel personal details nor details from novel stories after a short delay (Baddeley & Wilson, 2002, Zou et al., 2020). While other forms of memory are required to comprehend a story (see MacDonald & Christiansen, 2002), it is the long-term retention allowed by episodic memory which make it a prime candidate for storing the situation model (Erikkson & Kintsch, 1995).

2.5 Story schemas

It has long-been suggested that situation models for stories were structured in some way (Rumelhart, 1977). Theories of story comprehension since Aristotle have suggested that stories are segmented into events regarding the goals of the protagonist (Mar, 2004). For example, a story typically involves an overarching goal of a protagonist, and a series of attempts to complete that goal (Rumelhart, 1977). From a structural perspective, these goals and subgoals constitute events which can be organized within a larger structure (Kintsch, Mandel, & Kozminsky, 1977, Cohn, 2020). Organization of the situation model is necessary given the huge amount of information relayed and comprehended in stories.

Story schemas have been theorized to explain how events in a story are related or structured together in memory, positing unique functions derived from preconceived knowledge for how stories typically unfold (Rumelhart, 1977, Mandler & Johnson, 1977) (see Appendix E iii for discussion). The mental function of story schemas is to create slots which remain empty until comprehenders fill them in (Rumelhart, 1980). The reliance on story schemas during comprehension has been able to explain behavior in the ability to reorganize scrambled stories into their proper order (Kintsch, Mandler, et al., 1977), to summarize stories into most-relevant details (Rumelhart, 1977), and to remember some events in a story over others (Thorndyke, 1977, Mandler & Johnson, 1977). Other structural theories have suggested situation models for stories occur through causal links, or goal-states, connecting each action in a story, without relying on preconceived story schema, but instead on logical reasoning and knowledge about events in the world (Omanson, 1980, Trabasso & Sperry, 1985, Black & Bower, 1980) (see Appendix E iii for discussion of other structural theories).

While structural theories may prove able to predict memory for a story, the ability for these theories to explain the momentary processing of a person comprehending a story are limited (though see Cohn, 2020). As has been suggested by Thorndyke and Yekovich (1980), many of the effects of story structure on LTM memory could be explained offline, that is, after the initial encoding of the events, or after the story is over. As all structural theories discuss relationships between actions, understanding how actions within them are processed is a step in understanding how a situation model for a story is structured. One example of this is Thompson et al. (2014), who found that while causal connections between clauses of a story predicted later recognition, the existence of event boundaries also played a role. This emphasizes the necessity to understand how events are processed before moving on to studying stories.

2.6 Event Models

Theories of memory have been forced to explain how memory, despite its limits, is able to simultaneously handle the situation model while guiding moment-to-moment behavior during comprehension (Baddeley & Wilson, 2002, Ericsson & Kintsch, 1995, Schacter & Addis, 2007). ESPM have tried to explain this process by assuming a situation model is segmented into event models. According to ESPM, event models are segments of situation models which represent a meaningful action occurring at the same location as described within discourse such as a story (Zacks et al., 2007). ESPM further assume that the information within the event model is privileged above information in the situation model in its ability to guide cognition *online*, during comprehension (Richmond, et al., 2017, Bauer, et al., 2017). The creation of separate event models has been thought to be beneficial for memory (Thompson & Radvansky, 2014, Flores et al., 2018, Pettijohn & Radvansky, 2016) (see Appendix E vi & vii1 for discussion).

As we discussed above, event boundaries in the medium (movie, book, podcast, etc), whether imposed by semantic cues or event schemas, are thought to bias comprehension by inducing one event model to end, and another to begin. Thus, despite the fact that event models, like situation models, are personal creations of the comprehender and as such contain admixtures of information collected over time from the story along with information from the comprehender's life experience and reasoning skills (Zwaan & Radvansky, 2018, Anderson et al.,1983), the timing of event model creation is thought to occur *as consistently* across healthy subjects as would be their ability to identify boundaries (though see Zacks et al., 2006, Zahaa, et al., 2003).

2.6 Mechanics of Event Model Theories

ESPM have relied on a synthesis of experimental findings. According to a model put forward by Kurby and Zacks, (2008) and elaborated by Richmond, et al. (2017), Bauer et al., (2017), and Radvansky et al., (2017), event model creation was imagined to occur in the following way (figure 1). In this model, while an event is being described, the current event model provides prediction of what will happen next. The deviations between the incoming sensory information and the prediction based on the current event model are monitored. Once an event boundary occurs in the discourse, the differences between the incoming sensory information and the current event model increase, causing prediction error. This error signals a "global reset" is used in the brain to trigger a change in event models (Kurby et al., 2007, p 74). The old event model is stored in memory, and the new event model begins to form. Sensory prediction resumes once a new event model is created.

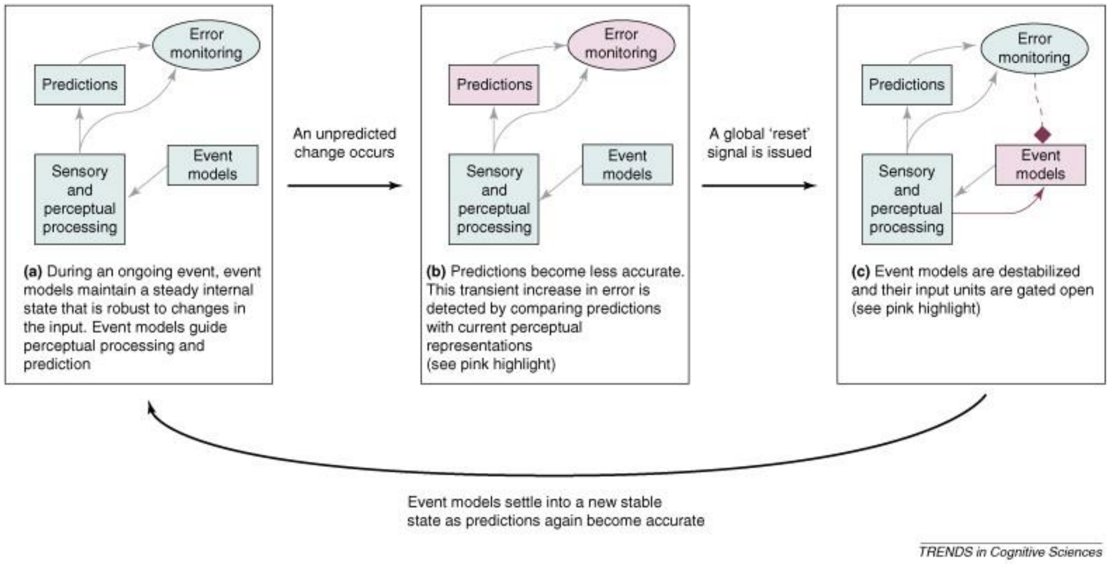


Figure 1. Mechanism for event model processing, Kurby and Zacks (2008).

This mechanism is assumed to be analogous for written and aural discourse. One repeated finding among researchers in event segmentation while people read stories is the slowing of reading speeds following an event (Zwaan, Magliano, & Graesser, 1995, Kelter et al., 2004, Speer et al., 2005). This slowing is explained by ESPM as owing to either updating of the past event model, or increased difficulty in processing new information, or both. Explanations for this slowing range from a lack of resources due to changing event models, as well as increased conflict due to integration of the previous model (Speer, et al., 2005, Pettijohn et al., 2016). Both of these processes could be responsible, and thus overlapping in time in the initial stages of event model building. In Appendix E (see Appendix E vii), we detail research related to how each process might operate according to the above model (figure 1) proposed by ESPM (Kurby & Zacks, 2008, Richmond et al., 2017, Bauer et al., 2017, Radvansky & Zacks, 2017).

2.7 Psychophysiological measurement of event models

2.7.1 Limitations of fMRI

Taking ESPM into account, integration of information from each sentence into the situation model for the story does not occur at a uniform rate following encounters with each word, nor the completion of each sentence. Instead integration occurs somewhat in parallel at two levels, the event model and the situation model. The event model must be updated within the duration that the event is displayed, before being interrupted at the event boundary, at which point a new event model begins to be formed. At this point old events are simultaneously being stored in the situation model for the rest of the narrative. Much of the previous brain research supporting mechanistic hypotheses of event model theories has focused on the brain processes that occur during segmentation, that is, effects seen in the brain once the comprehender encounters boundaries between events. Functional magnetic resonance imaging (fMRI), however is limited in the number of measurements it can make within a short period of time. This makes it hard to clearly depict and separate overlapping neural processes occurring within the hypothesized timespan of segmentation, which, given the high speed of neuronal communication, can be presumed to occur on the order of milliseconds. As a result, fMRI studies can detect which brain areas might be active during event model creation but are less able to properly attribute the neural contributions of each parallel processes, for example, those related to boundary detection, event-model wrap-up, storage, and new model creation (cf. Ezzyat et al., 2011).

2.7.2 Importance of EEG

In the analyses laid out here we will rely on electroencephalography (EEG), which has long been used to measure contextual facilitation of word processing (Lau, Phillps, & Poeppel, 2008, Kutas & Federmeier, 2011), as well as transient processes of perception and memory (Kintsch, 1999, 2012). By providing a recording of brain activity that is in high fidelity to the actual neuro-cognitive processes which are active during comprehension, we hope to use EEG to show whether the timing of the effects of crossing an event boundary while people listen to stories are either in accordance with the hypotheses of ESPM or whether they can be explained by other theories.

As unimodal accounts for processing in various domains (both visual and linguistic), ESPM make broad claims about cognition during comprehension of both video and verbal discourse (for more, see Appendix E vii). For example, ESPM suggest that changes to memory and attention should both be prevalent at the same time at the event boundary (Zacks et al., 2007, Bauer, et al., 2017). Psycholinguistic research on comprehension using EEG might shed light on these mechanisms in story comprehension. EEG offers a higher temporal resolution than fMRI and thus can better separate the activity of linguistic processes which occur in parallel such as those which occur within a sentence. Using EEG it is possible to see how perceptual, semantic, and grammatical processes overlap to affect the processing of each word (Hagoort, 2005, 2016).

Some dynamic brain function is visible in oscillatory activity coming from the cell membrane which fluctuate at frequencies within the order of milliseconds (Buzsaki, 2006). While some of these oscillations are correlated with blood oxygenation levels detectable with fMRI (Sheeringa, et al., 2011), these oscillations can be studied directly with magneto-electrical measurement techniques such as magnetoencephalography (MEG) and EEG. A long history of these oscillations, beginning with the invention of EEG by Hans Berger in the 1920s, have led to a range of hypotheses regarding their functional significance during a range of tasks including language comprehension (Berger, 1929, Bastiaansen et al., 2010). In this way, EEG allows us to contribute to the ongoing discussion about the function of these oscillations during comprehension, as well as draw on the brain frequency literature in order to test hypotheses set forth by ESPM.

2.7.3 Limitations of EEG

It is most common to study comprehension with EEG via context effects; by observing the effects which a context exerts on the cognitive processing of specific *target*s (e.g words, images). Rigid methodological requirements of EEG and behavioral experiments, (tightly-controlled manipulations comprising many repeat trials) could exhaust the experimental subject when target stimuli consist of longer materials such as stories.

Furthermore, classical EEG techniques such as event-related potentials (ERPs) look for causal relationships between a stimulus and fluctuations in brain waves; they are typically able to detect brain processes only when researchers can predict precisely when they should begin, obscuring processes that a subject might generate internally at unpredictable times (Luck, 2014). The subjectivity of the situation model which contains mental representations that are not explicitly mentioned in the text discourages examination of evoked potentials related to the situation model in long or naturalistic stimuli. Despite this, many attempts have been made to do so (St. George, Mannes, & Hoffman, 1994, Sato & Mizuhara, 2018, Boudewyn & Carter, 2018). Other EEG methods do allow the data from free comprehension to be analyzed over longer timescales (see King & Dehaene, 2014, Broadbeck, Pressaco, & Simon, 2018, Sassenhaggen, 2019, Fyshe, 2020, see Appendix E v for discussion), but have not been used to look at event boundaries.

Both heartbeat and muscle activity can show up in the EEG recording, obscuring brain activity. Eye movement from macroscopic eye blinks to more subtle micro-saccadic movements can have a detrimental effect on interpreting EEG activity by adding noise to the signal (see Dimigen, et al., 2011 for review). In cases such as with reading, this noise can have a periodic nature as the eyes move, causing saccade activity to be confused with oscillatory brain activity (Yuval-Greenberg et al., 2008).

Regardless of reasons why, the effort to apply EEG methods to the study of event boundaries (c.f. Sharp, 2010) or event model integration has not been widely adopted yet. So far, only one study has looked at the EEG effects of the boundary (Sharp, 2010). EST predicts that meaningful changes occur at and within event boundaries, thus offering a means with which to apply EEG methods to study comprehension of larger materials such as naturalistic texts (Zacks et al., 2007).

Chapter 3: Lexical Semantic Priming Across the Event Boundary

3.1 Overview

3.1.1 Research Question

For Analysis 1, we wish to answer the first research question: **whether lexico-semantic effects of context vary with event model creation**. Behavioral studies of reading with respect to events have shown slowed reading times at event boundaries (Rinck & Bower, 2000, Speer & Zacks, 2005, Magliano, Zwaan, Graesser, 1995), but it is still unclear if slowed reading at the event boundary is due to interference from encoding the old event, or from lack of facilitation from context, assuming the reader has cleared their mind in preparation for a new event model, or both (for further discussion see Appendix E vii1). ESPM hypothesize that changes to events make information from previous text (and thus the situation models) less available for subsequent processing. We want to know whether that includes lexico-semantic information from sentences before the boundary. If true, such a mechanism would change the way we understand how different forms of memory interact to uphold comprehension, and add mechanisms of ESPM specifically for linguistic stimuli. Below are examples of how lexico-semantic processing changes in discourse which we expect to be implicated in our study.

Alternatively, other models suggest that some information remains accessible at all times equally throughout the story, e.g., information about goals and states of a main character (Albrecht & O'Brien, 1993, Trabasso & Sperry, 1995) (for further discussion see Appendix E x). Other theories such as the minimalist hypotheses suggest that the amount of information held actively to guide comprehension in the moment is limited to what is needed for comprehension of the next sentence (McKoon & Ratcliff, 1992). In our study, in addition to examining the effects of the boundary, we compare multiple conditions of events in stories with unrelated events, to see whether the increased contextual information outside of the boundary affects lexico-semantic processing.

3.1.2 Priming, lexico-semantic processing, and the event boundary

The context in which a word appears has long been shown to facilitate processing of individual words, affecting the speed at which people access their pronunciations and recognize them (Myers & Schvanaveldt, 1979, see McNamara, 2005 for a review). Schvaneveldt & Meyers (1979) showed that children were faster to read pairs of words when they were semantically related (e.g., Doctor – Nurse) than when the words were unrelated (e.g., Doctor – Bread). Faster reading times and faster recognition of words comes from information that is active in the brain to subsequently facilitate processing of what comes next. Facilitated processing from context is called *priming* (Neely, 1977). In the case of words being primed this is thought to occur when context leads to transient increases in lexical or semantic networks, which bias word processing (Collins & Loftus, 1975). As such the cognitive domain giving rise to priming word recognition can be referred to as lexico-semantic processing (LSP). When a sentence context highly biases the appearance of one word, for example "tea" in "He liked lemon and sugar in his ____ ", LS-P for "tea" would be primed more so than "coffee" or "sock" (Hillyard & Kutas, 1984). If event models are accompanied by active lexico-semantic networks, then we would expect that when the event model changes, say at the event boundary, this activity would no longer be able to prime LS-P. The slowing of reading seen by Speer et al., 2005, and others, (Zwaan, 1996, Magliano, Zwaan, Graesser, 1995), could be owed to other processes, such as memory processes, which simultaneously occur when a new event begins. To parse whether this is the case without disturbing people as they comprehend, we can rely on EEG.

3.1.3 Lexico-semantic processing and the N400

The neuro-cognitive processes underlying LSP during comprehension have been measured using EEG most commonly using a method called event-related potential technique (ERP). Using this technique, it is possible to see averaged voltage potentials that index the degree of LSP while they read or listen to texts. One such voltage potential is the N400, which is a negative deflection seen 200-600 ms after the onset of a visual word, and is maximal over centro-parietal scalp locations (see Lau et al, 2008 for a review). The N400 is larger (more negative) in response to words which are less probable to occur with the context. For example, in word pairs, the N400 is smaller in response to "nurse" when preceded by "doctor" (Neely, 1991). In sentence contexts, the N400 is larger in response to "coffee" than "tea" when either word appears at the end of a sentence "He liked lemon and sugar in his ____" (Kutas & Hillyard, 1984). World knowledge and semantic relatedness of nearby words is thought to combine to support LSP as reflected in the N400 (Hagoort, 2005, 2013).

3.1.4 Semantic Vectors

Semantic relatedness of words within a sentence can be estimated using semantic vectors (Landauer & Dumais, 1997, Milikov, et al., 2013, Pennington, Socher, &

Manning, 2014). Semantic vectors are outputs of vector space models which take in a corpus of text excerpts, and derive a list or *vector* of numbers for each word. The ability of these vectors to capture the full semantic depth of what is provided by a sentence (e.g., a situation model) is probably limited (Perfetti, 1998, for discussion please see Appendix E viii). However, as a measure of semantic relatedness, they still have been observed to impact sentence processing. Ettinger et al., (2016) used semantic vectors to derive a value for the relatedness between the context and target words within them. The experimenters saw that semantic relatedness using semantic vectors was able to predict N400 amplitude in a variety of sentence stimuli. In a somewhat exploratory study of comprehension, Sato and Mizuhara, (2018) used semantic vectors to compare recall elicited after reading a text to portions of the actual text. They found that the semantic vector similarities predicted certain brain activity, which they attributed to memory encoding. While this finding might be cautiously interpreted, it demonstrates that semantic vectors can be used in meaningful ways to study discourse.

3.1.5 The N400, discourse, and events

Studies looking at the N400 with respect to the event boundary are limited. In my search I could find only one set of experiments which measured the effects of the event boundary using ERP (Sharp, 2010). Those experiments were part of a dissertation, and measured ERP effects of the event boundary while subjects watched single scene, continuous movies depicting an actor doing various household chores (e.g., fertilizing a houseplant, making a bed). The averaged waveform at the boundary appeared to resemble the N400, as a negative deflection over posterior scalp, occurring within 600 ms of the onset of a new event. While this was in response to film, it is harder to directly infer whether a similar effect would occur in response to LSP of words during auditorily or visually-presented stories.

However, there is a good amount of research studying the N400 in response to words in discourse which might allow such inferences to be made. Sections of text larger than a sentence are referred to as discourse, as such an event or a story context would be referred to as *discourse context*. Theories of comprehension posit that contextual effects on the N400, and thus LS-P, from both semantic relatedness of nearby words as well as from discourse context (Hagoort, 2005). A great example of how discourse context constrains LSP, is a study by St. George, Mannes, and Hoffman (1994) in which context was heightened by the inclusion of a title accompanying an otherwise ambiguous paragraph. Averaging each word of the paragraph in the titled and non-titled conditions revealed that the average N400 was greatly reduced by the inclusion of a title. While their stimuli were disorganized lists within a paragraph, it demonstrated that LSP can be persistently affected by context across sentences.

The N400 also tracks semantic overlap between words across the sentence boundary. For example, Stafura & Perfetti (2014) looked at whether LSP would be enhanced for target words across the break between two sentences when words in the previous sentence were semantically related. They did see effects in the N400 suggesting that semantic relatedness aids LSP across the sentence boundary. This was also seen in Ditman & Kuperberg, (2007), who saw effects of larger discourse context on the N400 (see Appendix E for discussion). While they explained their results in terms of causal reasoning, this finding suggests that event schemas exert an influence on word processing.

The effects of discourse context are not always seen to affect the N400. In another study of contextual facilitation, Burkhardt, (2007), failed to find any effect of the discourse context on the N400 response to critical words. An example of their stimuli is provided below:

(7) (a) Yesterday, a Ph.D student was **shot/killed/found dead** downtown

(b) The press reported that the pistol was from army stocks.

In this case, the authors saw no difference among all three antecedent sentences (7a) in the N400 response to target word (pistol in 7b). Interestingly, the contexts in Burkhardt (2007) were not only gruesome but crossed the event boundary. This is the opposite of the events in Stafura et al., (2015), in which sentences were described as happening within the same event. Thus, this leaves open the possibility that event boundary, cued by explicit or implied changes in time or place, between sentences, is having an effect in prior studies of discourse.

In the one ERP study examining the event boundary, Sharp (2010) looked at brain effects of unconscious, semantic processing while people watched single scene, continuous movies depicting an actor doing various household chores (e.g., fertilizing a houseplant, making a bed). In that study one ERP resembling the N400 was more active in response to the boundary. If that is in fact what was being measured by Sharp (2010), it would suggest that semantic processing is more effortful once a new event model has begun.

3.2 Hypotheses

3.2.1 Lexico-semantic Processing at the Event Boundary

From the studies detailed above and in the previous section, two factors driving LSP appear: one in which the N400 is driven by semantic relatedness of nearby words, and another in which the N400 is driven by the discourse context. If the ability of discourse context to affect LSP decreases either momentarily or permanently because of the event boundary, we expect to see increased N400 amplitudes for words in sentences following the boundary, than words in sentences happening right before. As a control, we will measure semantic relatedness of the previous sentence to offer a baseline of how sentences might facilitate one another if events were not driving the N400.

To study the effect of semantic relatedness owing to the situation model of the story, we contrast events in a story (Story Events) with a condition in which events are isolated and unrelated (Unrelated Events). This should allow us to remove semantic relatedness from the previous event more easily. In both cases we expect the boundary between events to decrease LSP. Whether this will be driven more by semantic relatedness between words in events, or truly is an effect of the event boundary remains to be seen (for detailed hypotheses see Appendices F & G).

3.3 Methods

3.3.1 Participants

57 subjects participated in the experiment in exchange for course credit. Of these, data from 49 was included in Analysis 1. Two were not included for inability to follow directions during recording. Six others were removed for technical difficulties during recording (connectivity issues or needing to end session early).

3.3.2 Materials and Procedure

3.3.2.1 Folktales.

Five folktales, and one short story were chosen based on their length and the hope that the plots would be unfamiliar to the participant population. Three folktales were chosen from compilations of folktales available via Project Gutenberg (http://projectgutenberg.com), and one folktale was taken from http://indigenouspeople.net. The folktales included the *Seven Ravens*, a German folktale (https://www.gutenberg.org/files/2591/2591-h/2591-h.htm); *The Good Thunder*, a Japanese folktale (https://www.gutenberg.org/files/35853/35853-h/35853-h.htm), *The Dam*, a South African folktale (https://www.gutenberg.org/files/38339/38339-h/38339-h.htm), and *Grandfather and his Two Grandsons*, a folktale attributed to the Ute tribe. (http://www.indigenouspeople.net/twogrand.htm). Another folktale *The story of the Donkey*, was taken from materials used by Trabasso and Sperry et al., (1985). In addition, we adapted a short story, The Summer of the Beautiful White Horse by William Saroyan (1940) which was used in an experiment by Huettner, Rosenthal, and Hynd (1989).

Ordered Stories

In the Ordered Stories condition, folktales were rewritten in order to control for sentence length. In order to increase the probability that participants would notice event boundaries, sentences were altered to include both temporal or location changes meant to indicate change in event. To ensure that event boundaries would be recognizable to listeners, a separate group of participants (n=7) read each altered folktale. This separate group of participants chosen from among friends and family was asked to mark which sentence they "thought a significant action began". Agreement among raters was over (75%) indicating reliable identification of event boundaries. From their responses, the number of events per story was 11, except for two stories (The Seven Ravens and The Dam) which both contained 13 events. Sentences were then omitted or added to create events that were between 3 and 7 sentences long. Mean sentences per event in every condition (Unrelated Events, Scrambled Stories, Ordered Stories) was 4.26 (s.d. 4.1).

Unrelated Events

We created a separate set of 33 unrelated vignettes in order to allow us to contrast event model creation in absence of a situation model of a story. The Unrelated Events condition was created with 33 vignettes sampled from a range of other folktales These folktales came from a range of traditions including Russian, Indian, Arabian, First Nations, Central and South American, German, as well as some written by the author BR. The original events in the Unrelated Events condition were rewritten with similar attention to sentence length and sentences per event, as with the other two conditions. To ensure that they were maximally unrelated, care was taken to change character names to avoid co-reference across events. Unrelated events in the Unrelated Events condition were then randomly assigned to 3 blocks of 11 events each. The effect was that no events in these conditions would reference any other events in the experiment.

Scrambled stories.

In order to avoid event model creation using adjacent sentences while still allowing for a high level of semantic relatedness between words, we created a condition in which stories were arranged with no adjacent sentences presented in sequential order. This was the Scrambled Stories condition. Scrambled Stories were each created from the same sentences as the corresponding Ordered Stories, such that there existed a scrambled version of each ordered story. The scrambling was pseudo random, so that no sentences appeared in the same position as in the corresponding ordered story, and no two adjacent sentences appeared in linear order. In order to control for boundaries, sentences containing event boundaries in the ordered story were placed, out of consecutive order, but spaced every 3-7 sentences as with the other conditions. In this way it was possible to see the effect of event boundary sentences, without allowing linear construction of event models. Each of the 6 stories was scrambled only once for 6 Scrambled Stories stimuli.

3.3.2.3 Stimulus Conditions Overview.

Each folktale/unrelated event was rewritten so that sentence lengths were between 9-13 words in length. Mean sentence length was 10 words (s.d. 0.02). Mean number of sentences per condition was 48.6 (s.d., 3.6). Mean words per story was 496 (s.d. 44.8) (See Table, Appendix A). In all conditions, care was taken to avoid pronouns without explicitly highlighting their referents in the same sentence. This was done to limit the impact of referential coherence between conditions; to make it possible for comprehenders to attribute the appropriate actions to characters in the Scrambled Stories condition. Word frequency for each word in the stories was obtained using lexical frequency counts from the Corpus of Contemporary American English (Davies, 2008). Word frequency and length of words for each story was compared to each other using analysis of variance (ANOVA). Some words were changed in order to increase or reduce the length or frequency of a particular story.

During the experiment, sentences within each block were each played with a 500 ms interstimulus interval.

3.3.2.4 Stimulus Recording.

Sentences in all conditions were recorded individually in a soundproof booth using Audacity software (http://audacityteam.org/). The speaker was a 34 year-old male with a regional accent from the Northeast United States, who calls a sandwich made with a cylindrical roll a *hoagie* and pronounces *merry*, *Mary*, and *marry* differently. During recording care was taken to annunciate each word. After recording, individual audio tracks were normalized to correct for volume differences and trimmed so that each sentence track began with the onset of the first word and ended with the offset of the last word. Mean sentence length was 4.9 seconds (s.d., 0.7).

3.3.2.5 Semantic Overlap (SO).

Semantic overlap for each word was determined using a context model based on that of (Ettinger et al., 2016). The context model calculated the context vector as well as the distance between the context vector and the vector of a particular content word, using a dictionary of vectors from pretrained vector space model (VSM) (GloVe; Pennington et al., 2014) obtained from https://github.com/stanfordnlp/GloVe. The pretrained vector model was trained on a corpus of 5 billion word tokens gleaned from Wikipedia, which resulted in a vocabulary of 400 thousand words.

To derive an SO value for each content word, the story stimuli were first tokenized and stripped of function words and punctuation using the Gensim package (Rehurek & Sojka, 2011) in python. A full list of function words is provided in Appendix A. SO for each word was derived by creating a context vector specific to each content word, and then finding the distance (cosign angle) between each context vector (CVn) and the vector for that word. The context vector for each target word (word *n*) consisted of the sum of the vectors of the 4 content words determined to be most-highly-related to the target word (based on cosine angle), sampled from all of the 10 content words preceding the target word in context, based on cosign angle. Each word was then weighted based on distance (1-10), such that the vector for the n-10 word was multiplied by 0.1 and the vector for the n-1 word was multiplied by 1.

The resulting context vectors created by this process provide estimated semantic content of the context preceding each word (see Appendix E xi). To calculate SO, we then calculated the cosine angle of each context vector and its corresponding word vector using the Gensim package. This should also be equivalent to the product of the context vector for the word n and that of the context vector, divided by the product of their euclidean norms. This relationship can be plotted such that for each word, there is a value corresponding to the estimated semantic overlap between each content word and the context preceding it.

3.3.2.6 Distance from boundary (DFB)

Each word of each sentence was given a distance from boundary based on the distance of that sentence from each boundary. Words from the sentence containing the boundary were given a “0” to indicate that they came from a sentence in which the boundary occurred. All words in the sentence after were given a “1”, and so on, until the next event occurred. The result is roughly 33 sentences per condition marked as 1, 33 marked as 2, etc. (see Appendix A).

3.3.3 Procedure

After EEG cap was applied (See EEG recording, below), subjects were positioned roughly 18 inches from two computer speakers. The volume level was adjusted to a reasonable level. Before the task began, subjects were given instructions about the task. They were told that the task would involve listening to *audio passages* each 5-7 minutes long and that those passages would contain either folktales in regular order, scrambled randomly sentence-by-sentence, or would contain short paragraphs from unrelated stories. In each condition, subjects were told to attend to the passages as best they could and were encouraged to remember each by explaining that recall would be asked of them immediately after each passage was played. When subjects asked for clarification about whether they were required to recall the scrambled stories in proper order (i.e., to unscramble them), the experimenter reiterated simply to recall as many details from each passage as possible.

A short practice example of a ten-sentence-long passage in regular order was presented followed by an example of the recall and comprehension task before the main blocks began. Stimuli appeared by block, such that one block contained one passage from one of the three conditions, as well as a change for recall, and then 20 True/False inference recognition questions. After that block, subjects were allowed to rest before beginning another block presented in another passage from another condition. Following the practice, the order of the nine blocks was as follows: [OS, UE, SS, OS, UE, SS, OS, UE, SS]. Presentation of conditions always began with an Ordered Story (OS) block, followed by an Unrelated Events (UE) block, followed by a Scrambled Stories (SS) block, before repeating this pattern. Each type of condition was shown 3 times, for a total of 9 blocks, lasting roughly 7-10 minutes each depending on the comprehension tests (recall and recognition) which were self-paced.

Each subject saw only one story in each condition and did not see the same story stimuli used in any other block. All of the Unrelated Events stimuli came from separate stories, not used in the other conditions. Separate stimulus lists were created to ensure that subjects did not see the same inference questions, and to add a correction for effects of block order.

3.3 Behavioral measures

3.3.4.1 *Recall.* After each block, subjects were asked to recall the story, verbally, to the best of their ability. A recording of their responses was created using a digital audio recorder. All responses were transcribed using online software (Trint.com). No further scoring of recall responses was made for this analysis. Scoring of their responses will be made based on a rubric for each set of sentences from each block. Each rubric will contain the total set of propositions in the block, as well as the order of each. An example is given below (table B1, Appendix B). Points will be given for each idea that was present in the block and in the recall response (Keven et al., 2018, Poulsen et al., 1979).

3.3.4.2 Inference Recognition

After recall was elicited for each block, subjects completed an inference recognition task consisting of 20 sentence probes. Participants were asked to read each sentence in isolation and respond whether the sentence described a situation that occurred in the block they just listened to. Participants were given unlimited time to respond to each probe. Stimulus materials were created based on paraphrasing specific sentences with specific care made not to repeat unique words. For example, in the story *The Beautiful White Horse*, a sentence describes how the neighbor informed someone that his horse was missing and that he could no longer use his horse-drawn wagon. A corresponding *True* probe for that sentence was "Neighbor had no other horses", while a corresponding *False* probe was "The Neighbor had plenty of other horses". If answering an inference question required attending multiple sentences, as was sometimes the case, the last necessary sentence in linear order was considered the corresponding one. In this way each probe could be referenced to a specific sentence, and thus a value for distance from boundary (DFB). Examples of each are given in table C1 (Appendix C). A total of 20 trials were presented in each block, with 10 of each probe type (Paraphrase, Incorrect Paraphrase).

3.4 EEG recording.

During EEG recording subjects were seated in a partially sound-proofed, dimly-lit room in front of an LCD monitor. Subjects were briefly taught about muscle and movement artifacts in the EEG data and encouraged to relax muscles of the head and face in order to reduce these artifacts during recording. The effects of closing the eyes on posterior alpha EEG activity were demonstrated and it was explained that while blinking was allowed, that the eyes must be kept open for the duration of the audio passages. One subject reported that they did not follow these instructions and their data were removed from the data set. The EEG data were recorded from 32 Ag/Cl lined electrodes fitted into a neoprene cap (Neuroelectrics, Barcelona, Spain) placed in locations corresponding to the 10-10 system nomenclature (Chatrian et al., 1988): P7, P4, Cz, Pz, P3, P8, O1, O2, T8, F8, C4, F4, Fp2, Fz, C3, F3, Fp1, F7, Oz, PO4, FC6, FC2, AF4, CP6, CP2, CP1, CP5, FC1, FC5, AF3, T7, PO3. A clip electrode on the earlobe served as both ground and reference during recording. EEG data were digitized at 500Hz. Data and event triggers were sent from the analogue-to-digital converter, located on the cap, to the computer via a cable, and recorded using NIC2 recording software (Neuroelectrics, Barcelona, Spain). During recording, impedances were all kept below 10 kOhms. Participants were seated in a chair with their heads roughly 30 inches away from the Dell LCD desktop monitor. Before recording, participants were coached on how to minimize muscular artifacts during recording.

3.5 Word Onsets

Onsets of each word were determined using Montreal forced aligner on each sentence (McAuliffe et al., 2017, https://montreal-forced-aligner.readthedocs.io/en/latest/index.html). To validate onsets, onsets for one story were manually measured. Mean difference between manual onset detection and the forced aligner was roughly 20 ms, which seemed acceptable.

3.6 EEG preprocessing

EEG data were preprocessed using matlab. After recording, data from each subject were filtered using a .03 to 45 Hz bandpass FIR filter (EEGLAB, Delorme & Makeig, 2004). Independent components related to ocular artifacts were identified using automatic Multiple Artifact Rejection Algorithm (MARA) (https://github.com/irenne/MARA , Winkler et al., 2016). After ICA component rejection, bad channels were identified as the thee channels with the highest frequency power within the range of 12-30 Hz. These were then interpolated using data from surrounding electrodes. After creating epochs relative to word onsets (-250 before to 1,500 ms after), any epochs containing values outside of the range of -100 to 100 microvolts on any channel were removed from the analysis. Data were then re-referenced to the average reference and baseline corrected from -250 ms to 0 ms before word onset.

3.7 N400 Measurement

Location of the N400 was determined by visual inspection of the grand average of electrodes (n=49). A cluster of electrodes was chosen to cover the topography of the prominent negative peak within the stereotypical time range (200-600 ms) and location (central-posterior scalp) (Lau et al., 2008). The central-posterior cluster included electrodes over Pz, Cz, CP1, and CP2 locations according to the international 10-10 system. An adaptive mean procedure was applied to the mean voltage of the central-posterior cluster to determine the value of the N400 for each word. The adaptive mean procedure hones in on the peak within a time-window (in this case most negative), and forms a mean based on the average values surrounding that peak. After viewing the individual subject averages for the central-posterior cluster, the time window to search for the negative peak was 150-750 ms. Means of 40 ms centered in time around the peak value within this window were used as our measure of the N400. Sentences from the first event in each condition were discarded due to the possibility of unique discourse processes dealing with beginning a story (Van den Broek & Helder, 2017). Each condition (Story Events, Scrambled Story, Unrelated Events) is yielded 30 events, each comprised of 3-7 sentences.

3.8 Statistical analysis

3.8.1 Mediation

To test the effects of crossing the event boundary on the availability of lexical-semantic context, we created mediation models for each of the three conditions (Ordered Story, Scrambled Story, and Unrelated Events) using the mediation package (Tingley et al., 2014) built for R software (R Team, 2013 http:/www.R-project.org). In each model, distance from boundary (DFB) was treated as the independent variable, with N400 amplitude treated as the dependent variables and semantic overlap (SO) treated as the mediator. DFB (DFB=0-4) was treated as continuous independent variables.

3.8.2 Inference Recognition Accuracy 5 x 3 ANOVA

In addition to the mediation analysis, behavioral results from the inference recognition test were submitted to a within-subjects ANOVA with DFB and Condition as within-subjects factors. Tukey's tests were used for post-hoc comparisons.

3.8.3 Block Order

Order of stimulus block was entered as an independent variable in a linear model predicting inference recognition accuracy separately for each condition.

3.4 Results

After measurement of the N400 for each word any values outside of three standard deviations from the mean (-5.35 μV ± 14.47) were removed from the analysis. The same outlier identification procedure was used for Semantic Overlap (SO) where we rejected any values above or below .31 cosine similarity ± .51. Distance from Boundary (DFB) above 5 were excluded from analysis, due to low numbers of words in this category (< 25 per condition).

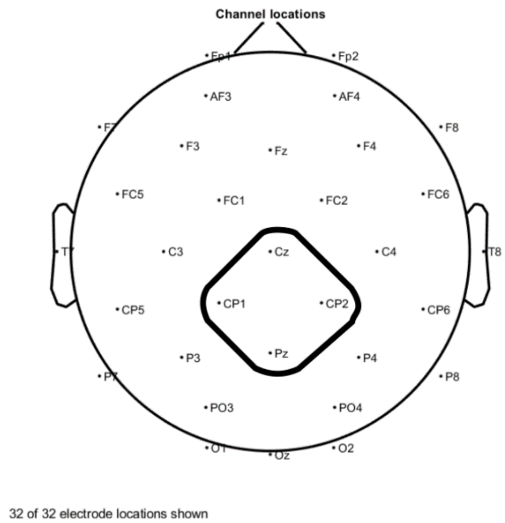


Figure 2. N400 region of interest. Topographic locations showing electrode locations for the posterior cluster used to measure the N400.

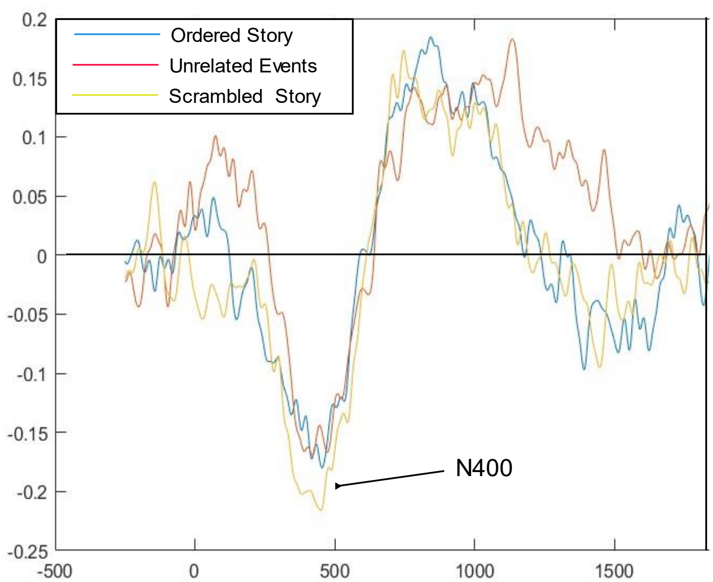


Figure 3. N400 at Pz averaged by condition.

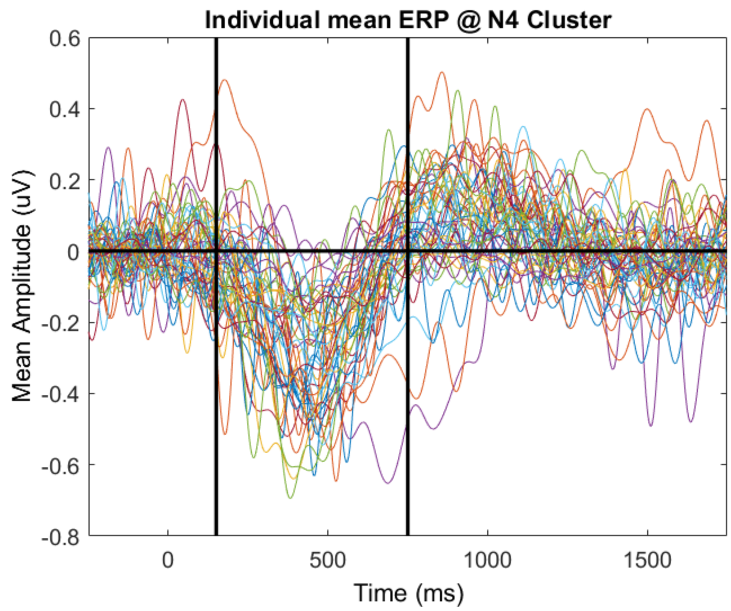


Figure 4. Mean N400 averaged by subject. Black vertical lines indicate the time window chosen for the adaptive mean. Data in figure are depicted with a lowpass 10 Hz lowpass filter.

Condition	DFB	mean	sd	n
SE	0	-5.15	4.11	7885
SE	1	-5.18	4.09	7633
SE	2	-5.11	4.10	7685
SE	3	-5.18	4.09	5355
SE	4	-5.13	4.08	2735
SS	0	-5.09	4.03	7987
SS	1	-5.06	4.05	7620
SS	2	-5.08	4.02	7502
SS	3	-4.98	3.94	5327
SS	4	-4.95	4.03	3096
UE	0	-4.91	3.96	7469
UE	1	-4.98	4.04	8348
UE	2	-4.86	4.01	7653
UE	3	-4.95	4.05	6377
UE	4	-4.81	4.02	2359

Table 1. Mean N400 for each condition and DFB.

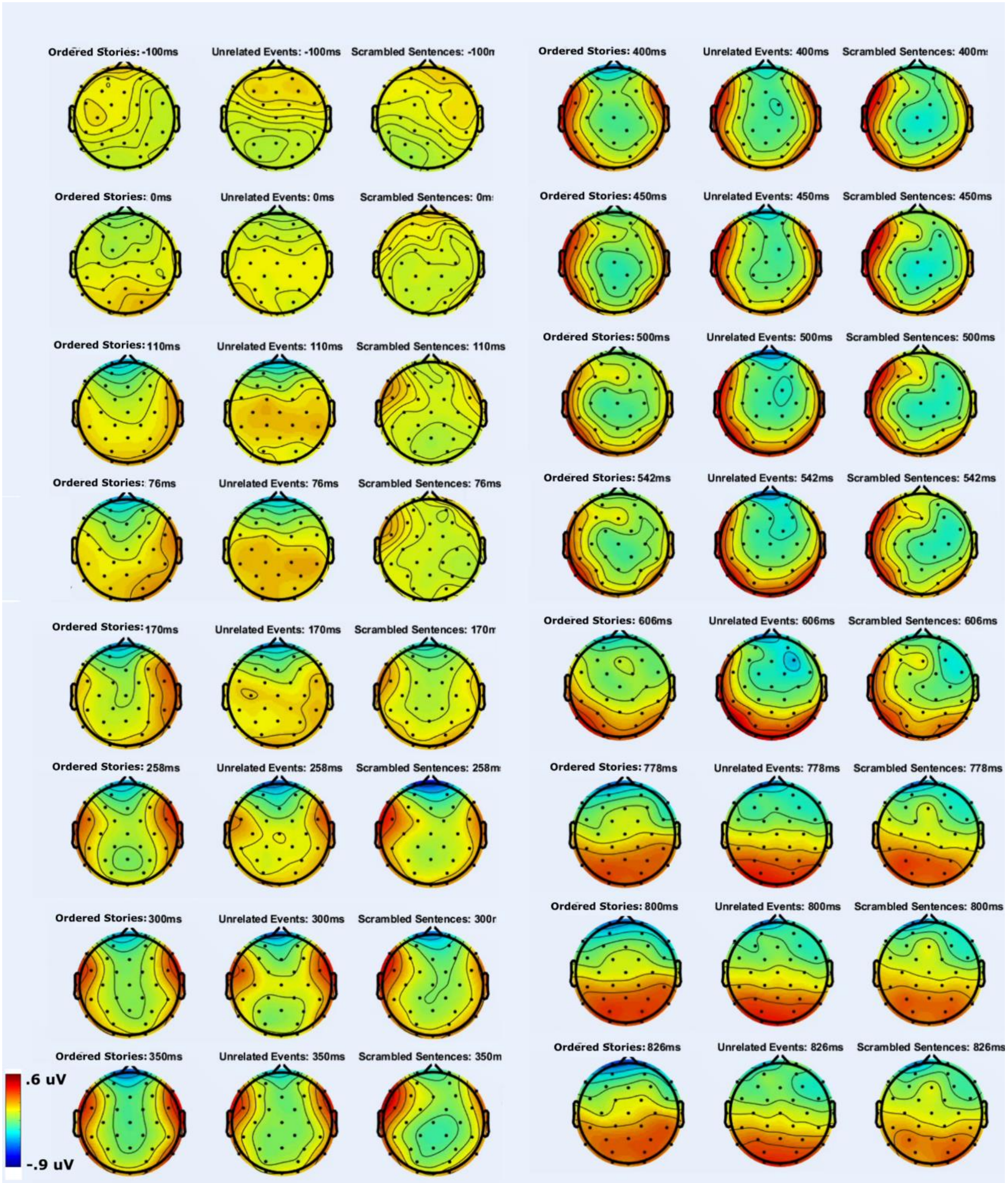


Figure 5. Topographic Maps across time. Topographic maps of voltages time-locked to onset of content words in each of the three conditions (Ordered Stories, Unrelated events, Scrambled Stories).

3.4.1 Behavioral analysis

Recall was recorded and transcribed, but not analyzed here. Mean accuracy for the inference recognition task is show in table X.

3.4.1.1 Recall.

Recalls were transcribed using online functionality provided by Trint.com (http://www.trint.com). No further analysis has been done but it is possible to provide anecdotal evidence that the number of details recalled from the Ordered Stories greatly exceeded those from the other two conditions. The worst performance was seen in the Unrelated events conditions where some participants could recall only one or two details from 11 events. Often participants were able to recall only 2 or 3 details from each Unrelated Events condition block. Participants almost exclusively attempted to recall details in the Scrambled Stories condition in order despite lack of instruction to do so. In terms of details recalled, this condition was much more successful than the Unrelated Events condition.

3.4.1.2 DFB effects of inference accuracy.

The 2x5way ANOVA for inference recognition test failed to find any significant effects of DFB ($F(4,8216)= 0.068$, $p = 0.99$). While graphs of DFB x Condition suggest that participants were most accurate for inferences informed by sentences which appeared as the third sentence from the boundary (DFB=2) in the two Story conditions, and more accurate for inferences informed by sentences at DFB=4 for the Unrelated Condition, these results are not significant.

3.4.1.3 Condition effects of Inference Accuracy

The 2x5way ANOVA for inference recognition test detected significant effects of condition (F(2,8216)= 0.98.28, p = 0.99). Separate post-hoc tests show that participants were more accurate in Ordered Stories condition than in either the Scrambled Stories (difference=0.15, 95% CI=[0.12, 0.18], p<.001) or Unrelated Events conditions (difference=0.13, 95% CI=[0.10, 0.16], p<.001). There was no difference in mean accuracy between Scrambled Stories and Unrelated Events (difference=0.01, 95% CI=[-0.04, 0.01], p<.001).

3.4.1.4 Inference accuracy Condition x DFB effects

The 2x5way ANOVA for inference recognition test accuracy detected a significant interaction of Condition and DFB (F(8,8216)=3.46, p=.001). Post-hoc tests (table 3) show an interesting pattern where accuracy was greater for Ordered Stories vs Scrambled Stories at each DFB, and greater for Ordered Stories than Unrelated Events at all DFB except the last (difference=0.09, 095% CI=[.209,0.040], p=.58). Post-hoc tests show no differences within conditions at any DFB.

a.

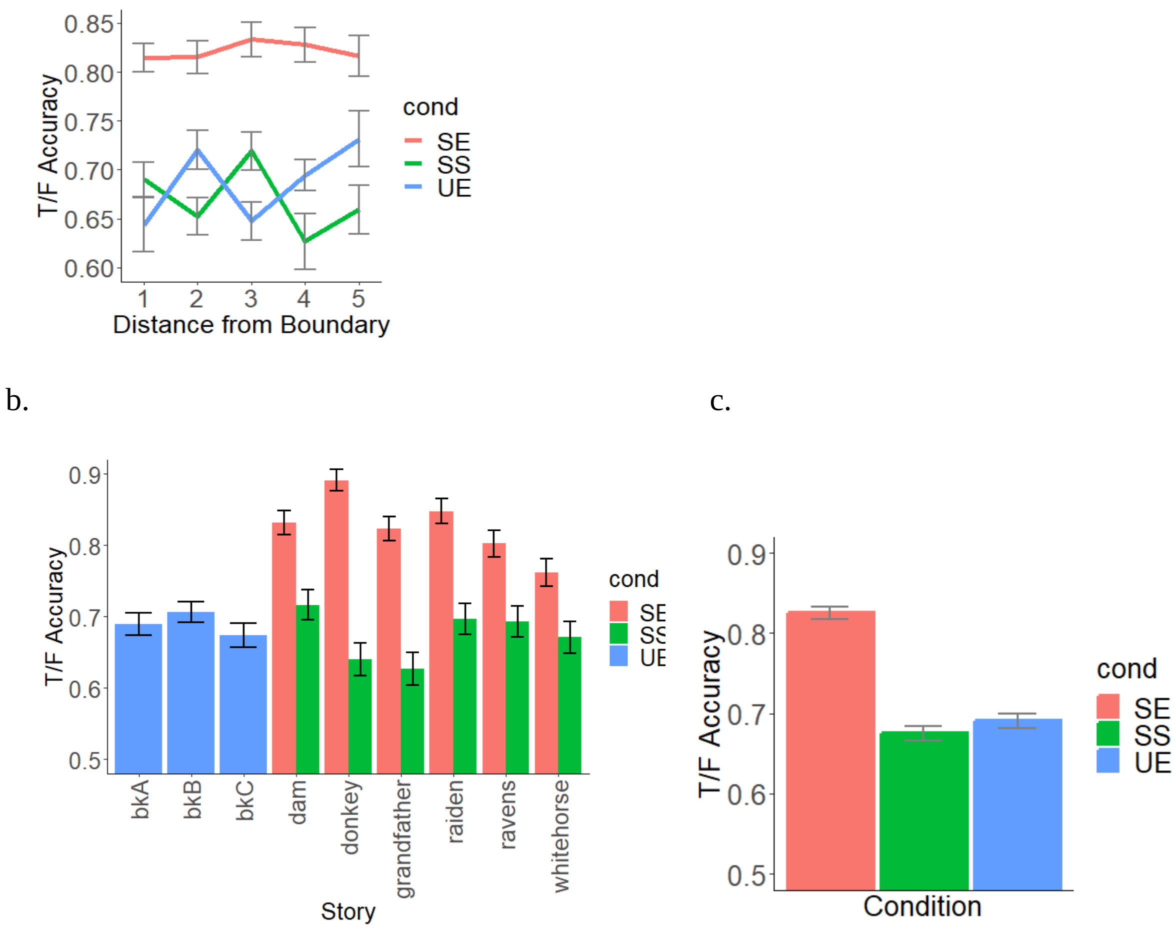


Figure 6 . Inference recognition accuracy. Inference recognition accuracy is displayed in mean percentage correct. (a) Accuracy relative to the distance from boundary of each event (DFB). (b) Accuracy displayed by block with story type shown (colors: green =Scrambled Stories, red=Ordered Stories, blue=Unrelated Events). (c) Overall accuracy across stories.

condition	n	Mean acc	se	sd
OS	2758	0.825	0.007	0.340
SS	2746	0.674	0.009	0.469
UE	2727	0.690	0.009	0.462

Table 2. Accuracy for responses to the true/false inference recognition test averaged across condition.

Condition Comparison	DFB	Mean difference	Lower 95% CI	Upper 95% CI	p-value
UE-OS	0	0.171	0.273	0.069	0.000***
UE-OS	1	0.095	0.187	0.003	0.036*
UE-OS	2	0.186	0.279	0.093	0.000***
UE-OS	3	0.133	0.220	0.047	0.000***
UE-OS	4	0.085	0.209	0.040	0.579
UE-SS	0	0.047	0.150	0.056	0.970
UE-SS	1	0.068	0.021	0.157	0.374
UE-SS	2	0.072	0.160	0.017	0.274
UE-SS	3	0.068	0.035	0.170	0.631
UE-SS	4	0.072	0.052	0.196	0.814
SS-SE	0	0.124	0.204	0.045	0.000***
SS-SE	1	0.163	0.249	0.077	0.000***
SS-SE	2	0.114	0.210	0.018	0.005**
SS-SE	3	0.201	0.314	0.089	0.000***
SS-SE	4	0.157	0.268	0.046	0.000***

Table 3. Accuracy Post-Hoc tests. Post-hoc (Tukey's test) for accuracy to the true/false inference recognition test averaged across distance from boundary (DFB) by condition. Significant mean differences between conditions are denoted with "*". "*":p <.05; "**":p<.01, "***":p<.001.

3.4.1.4 Order effects of inference accuracy

Block order significantly affected accuracy in the Order Story blocks (b=-0.007, R^2 = 0.002, $F(1, 2756) = 5.85$, $p = 0.016$), and Scrambled Story blocks (b=-0.007, R^2 = 0.001, $F(1, 2744) = 4.44$, $p = 0.035$), but not in the Unrelated Events blocks (the Order Story (b=-0.002, R^2 = 0.001, $F(1, 2725) = 0.42$, $p = 0.52$). This means that for both Story conditions, participants accuracy on the inference recognition task suffered 0.007 of a percentage with each proximal block. While intriguing, this appears to be so miniscule as to be meaningless.

3.4.2 N400 Analysis

3.4.2.1 Multiple regression

SO predicting N400: SO predicting N400 values were centered for the linear models. None of the linear models predicting N400 from SO had significant F values, indicating that they were each a poor fit. This was true for Ordered Stories, R^2 = <.0001, $F(1, 31291) = 0.011$, $p = 0.92$; Scrambled Stories, b= 0.021, R^2 = <.0001, $F(1, 31530) = 0.58$, $p = 0.45$; and Unrelated Events, R^2 = <.0001, $F(1, 32204) = 1.28$, $p = 0.26$.

DFB predicting SO: Linear models predicting SO from ranked levels of DFB linear model were significant. This was true in the Ordered Stories, b=-0.44, R^2 = <.003, $F(1, 31291) = 99.9$, p <.0001; Scrambled Stories, b=-0.51. R^2 = 0.004, $F(1, 31530) = 141.9$, p <.0001; and Unrelated Events, b=-0.71. R^2 = 0.01, $F(1, 32204) = 322.8$, p <.0001.

DFB predicting N400. None of the linear models predicting N400 from DFB had significant F values, indicating that they were each a poor fit. This was true for Ordered

Stories, R^2 = <.0001, F(1, 31291) = 0.011, p = 0.92; Scrambled Stories, b= 0.021, R^2 = <.0001, F(1, 31530) = 0.58, p = 0.45; and Unrelated Events, R^2 = <.0001, F(1, 32204) = 1.28 , p = 0.26.

SO and DFB predicting N400: None of the multiple regression linear predicting N400 from SO and DFB were significant. This was true for Ordered Stories, R^2 = <.0001, F(2, 31290) = 0.013, p = 0.99; Scrambled Stories, R^2 = .0002, F(2, 31529) = 2.30, p = 0.1; and Unrelated Events, R^2 = <.0001, F(2, 32203) = 0.94 , p = 0.39. However, despite the models not being significant, there one condition where DFB was significant in predicting the N400 the Scrambled Stories condition (b=0.007, p=0.045).

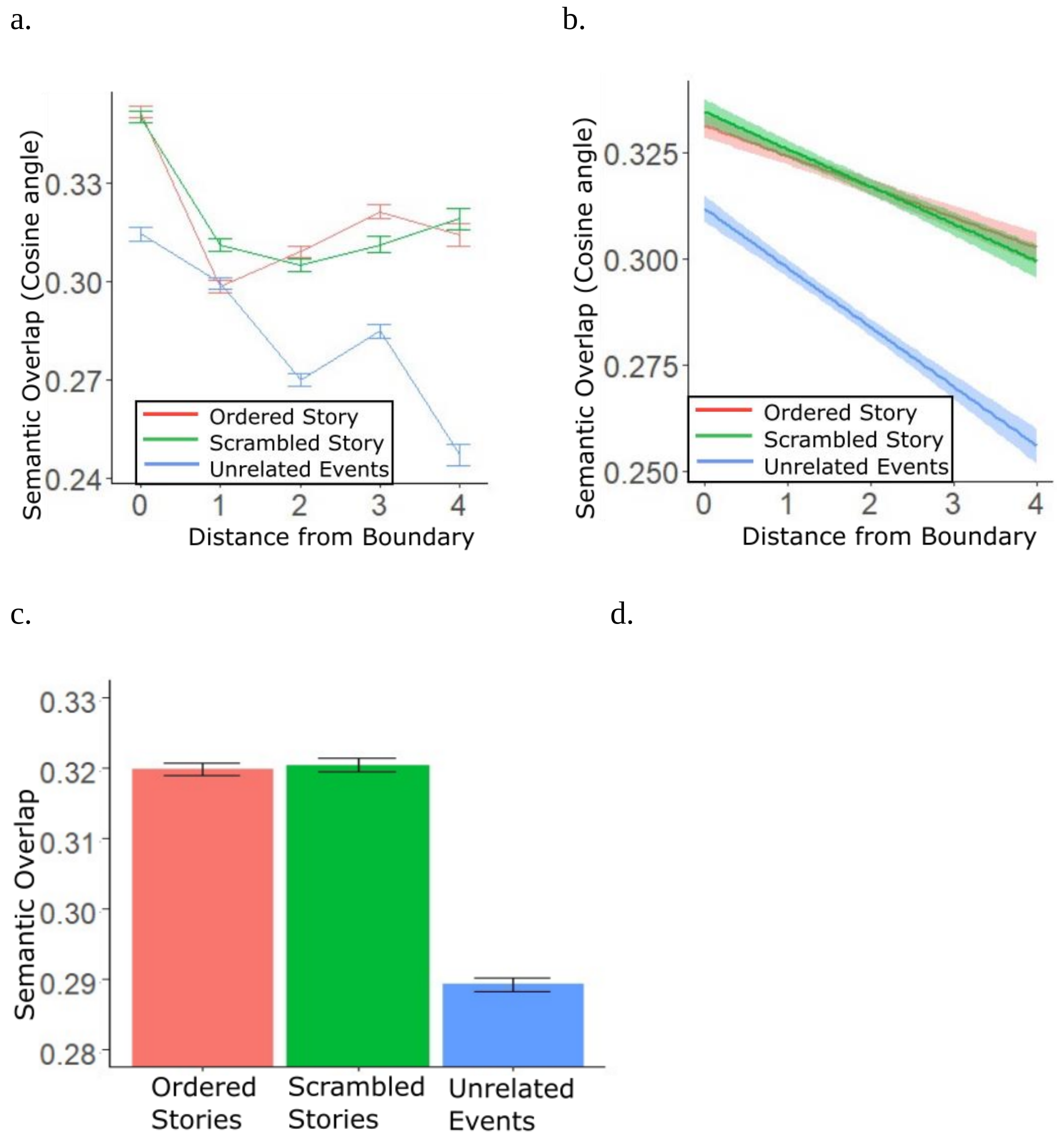
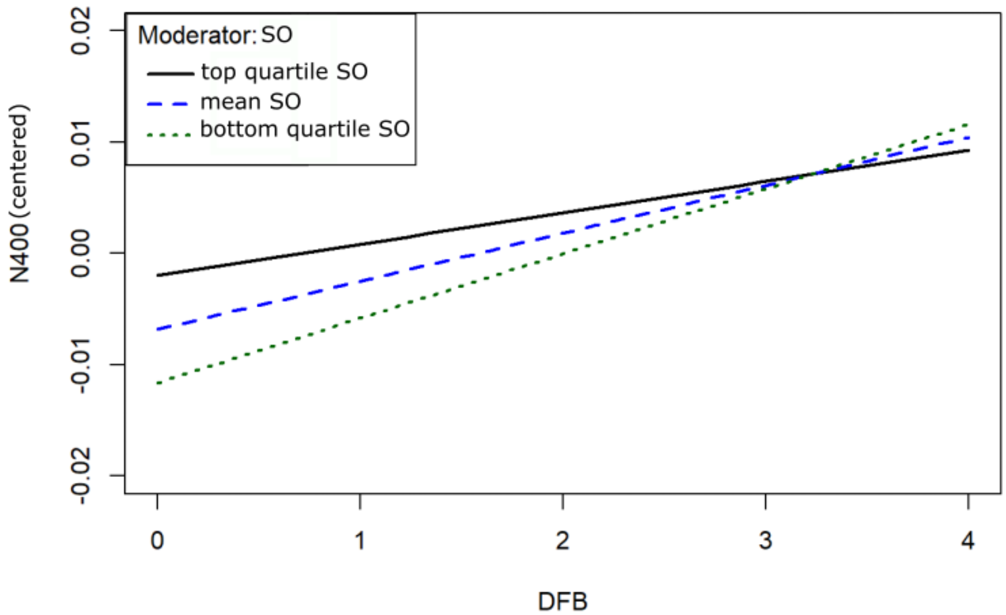


Figure 7. Semantic overlap and distance from boundary. Relationship between semantic overlap (cosign angle between each word vector and its individually assigned, weighted context vector). (a) Mean values with standard error bars. (b) linear function depicting relationship between semantic overlap (SO) and distance from boundary (DFB). (c) SO averaged across condition.

3.4.1.6 Mediation for Scrambled Stories

Graphing the results shows that for the Scrambled Stories, the N400 became less negative with increased DFB (see Figure 8a). The average estimated causal mediation effect was not significant (0.009, 95% CI[-0.0009,0.00], p=0.38). This is the combined effect of DFB on the SO (-0.009, p<.001) and DFB on the N400 (0.008, p=.05). The estimated average indirect effect of DFB on the N400 though SO in the Scrambled Sentences condition was marginally significant (0.0009, 95% CI[-0.009,0.02], p=.05). There was a marginally significant total effect of DFB on the N400 (0.008, 95% CI=[-0.0003,0.02], p=0.06).

a.



b.

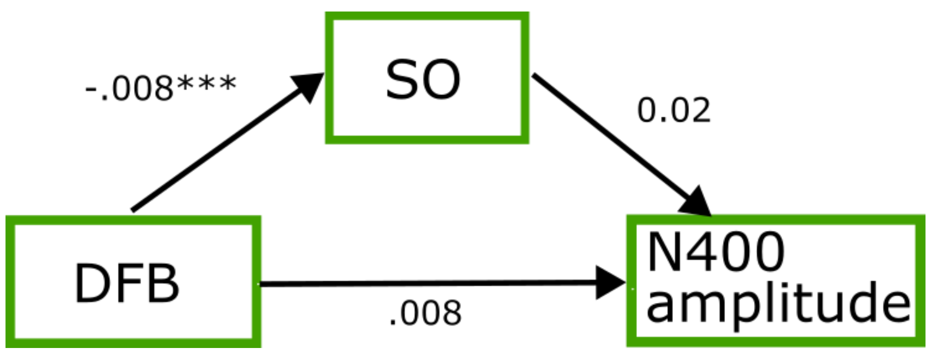


Figure 8. Scrambled Story Specific Effects. Specific effects of distance from boundary (DFB), Semantic Overlap (SO), and N400 amplitude. (a) N400 plotted as a function of DFB at three levels of SO, mean (blue dashed), bottom quartile (solid black) and upper quartile (dotted green). The parallel lines show a decrease with increasing SO. (b) The beta values and significance levels depicting the mediation effects of SO on the relationship between DFB and the N400. The Total effect of DFB on N400 was marginally significant , at p=.058. “***”=p<.001.

3.4.1.7 Condition, distance from boundary, and semantic overlap effects

In an exploratory manner we decided to look at condition effects, DFB, and their interaction on SO. In a 3X2 ANOVA comparing mean SO across conditions, with post-hoc mean comparisons examined using Tukey's tests. Results of the ANOVA are shown in table 4. From the results of the ANOVA we saw a significant difference between Conditions (F(2, 95016) = 351.7 , p<.001), DFB (F(4,95016)=255.1.p<.001) and the interaction of DFB and Condition (F(8,95016)=33.1,p<.001).

Post-hoc tests of the Condition effects show that SO in Unrelated Events differed from that in both Ordered Stories (difference=0.03, 95% CI=[-0.03,-0.03], p<.001) and Scrambled Stories (difference=0.03, 95% CI=[-0.03,-0.03], p<.001) but that the SO in Ordered stories did not differ from the Scrambled (difference=0.0, 95% CI=[0.0,0.0]p=.93) (see figure 7a).

Post-hoc test of the interaction are shown in Table 3. By condition, we see that in the Unrelated Events, the SO is different in all of the DFB (p>.01). In the Ordered Story, all DFB are different except for DFB 4 vs 2 (p=0.99) and DFB 4 vs 3 (p=.90). In the Scrambled stories, the best way to characterize the difference is that DFB=0 differed from all other DFB (viz., 1,2,3,4) (all p<.005) and DFB=4 differed from DFB=2 (difference=0.01, p=.01). No other differences in DFB were detected for the Scrambled condition (al p>0.6).

condition comparison	DFB comparison	diff	lower	upper	p
Ordered Story-Ordered Story	1-0	0.05	0.06	0.04	0.00**
Ordered Story-Ordered Story	2-0	0.04	0.05	0.03	0.00**
Ordered Story-Ordered Story	3-0	0.03	0.04	0.02	0.00**
Ordered Story-Ordered Story	4-0	0.04	0.05	0.02	0.00**
Ordered Story-Ordered Story	2-1	0.01	0.00	0.02	0.01*
Ordered Story-Ordered Story	3-1	0.02	0.01	0.03	0.00**
Ordered Story-Ordered Story	4-1	0.02	0.00	0.03	0.00**
Ordered Story-Ordered Story	3-2	0.01	0.00	0.02	0.00**
Ordered Story-Ordered Story	4-2	0.01	0.01	0.02	0.99
Ordered Story-Ordered Story	4-3	0.01	0.02	0.01	0.90
Scrambled-Scrambled	1-0	0.04	0.05	0.03	0.00**
Scrambled-Scrambled	2-0	0.05	0.05	0.04	0.00**
Scrambled-Scrambled	3-0	0.04	0.05	0.03	0.00**
Scrambled-Scrambled	4-0	0.03	0.04	0.02	0.00**
Scrambled-Scrambled	2-1	0.01	0.02	0.00	0.62
Scrambled-Scrambled	3-1	0.00	0.01	0.01	1.00
Scrambled-Scrambled	4-1	0.01	0.00	0.02	0.62
Scrambled-Scrambled	3-2	0.01	0.00	0.02	0.72
Scrambled-Scrambled	4-2	0.01	0.00	0.03	0.01
Scrambled-Scrambled	4-3	0.01	0.00	0.02	0.75
Unrelated-Unrelated	1-0	0.02	0.02	0.01	0.00**
Unrelated-Unrelated	2-0	0.04	0.05	0.04	0.00**
Unrelated-Unrelated	3-0	0.03	0.04	0.02	0.00**
Unrelated-Unrelated	4-0	0.07	0.08	0.05	0.00**
Unrelated-Unrelated	2-1	0.03	0.04	0.02	0.00**
Unrelated-Unrelated	3-1	0.01	0.02	0.01	0.00**
Unrelated-Unrelated	4-1	0.05	0.07	0.04	0.00**
Unrelated-Unrelated	3-2	0.01	0.01	0.02	0.00**
Unrelated-Unrelated	4-2	0.02	0.04	0.01	0.00**
Unrelated-Unrelated	4-3	0.04	0.05	0.02	0.00**

Table 4. Post-hoc tests for semantic overlap by distance from boundary. Post-hoc tests (Tukey's) showing difference in SO at various levels of DFB (my sentence) within conditions. Significant within-condition differences show that SO varied most within the Unrelated events. SO varied least within Scrambled Stories, as expected, however in all conditions, first sentences (DFB=0), differed with later sentences. This points to lexico-semantic artifacts at the boundary.

	Df	SumSq	MeanSq	F	p
DFB	4	28.6	7.138	255.1	0
condition	2	19.7	9.841	351.72	0
DFB:condition	8	7.4	0.925	33.06	0
Residuals	95016	2658.6	0.028		

Residual standard error: 0.1673 on 95016 degrees of freedom
Multiple R-squared: 0.0205, Adjusted R-squared: 0.02035
F-statistic: 142 on 14 and 95016 DF, p-value: < 2.2e-16

Table 5. Semantic Overlap differences. Table showing results for the ANOVA looking at effects of distance from boundary (DFB) and Condition on (SO).

3.4.1.8 Condition and N400 effects

In an exploratory manner we decided to look at condition effects by DFB using a 3(Condition)x5(DFB) analysis of variance ANOVA with post-hoc comparisons using Tukey's tests. Results are shown in table 4. From the ANOVA we saw a significant difference between Conditions ($F(2, 95016) = 26.8$, $p<.001$). From the post-hoc tests we see that the mean N400 for each condition was significantly different from the others (table 7), with Ordered Stories the most negative, Unrelated Events least negative, and Scrambled Stories in between (table 7, figure 9).

	Df	SumSq	Mean Sq	F	p
DFB	4	4	0.895	1.255	0.285
condition	2	38	19.116	26.805	0
DFB:condition	8	5	0.591	0.829	0.577
Residuals	95016	67761	0.713		

Residual standard error: 0.84 on 95016 degrees of freedom
Multiple R-squared: 0.0007, Adjusted R-squared: 0.0006
F-statistic: 4.66 on 14 and 95016 DF, p-value: <.0001

Table 6. N400 Difference. Table showing results for the ANOVA looking at effects of distance from boundary (DFB) and Condition on N400 amplitude.

	difference	lower	upper	p
SS-OS	0.02	0.01	0.04	0.00
UE-OS	0.05	0.03	0.06	0.00
UE-SS	0.03	0.01	0.04	0.00

Table 7. Mean N400 differences by condition (microvolts). Tukey Post-hoc test. Difference = mean difference. SS=Scrambled Stories. OS=Ordered Stories, UE=Unrelated Events. p-value is shown rounded to the second decimal place. All p values <0.001.

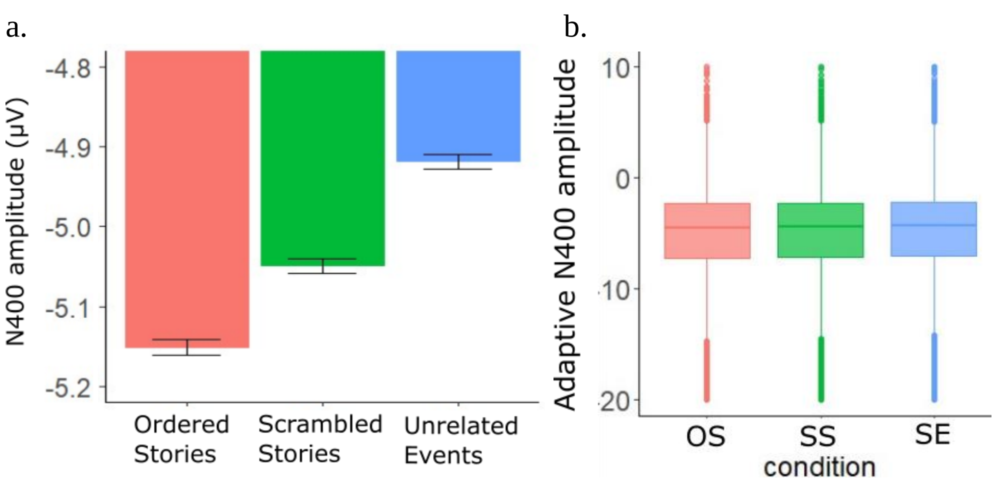


Figure 9. (a) Mean N400 by condition. (b) Distribution of adaptive N400 amplitudes.

3.5 Discussion

Accuracy in the inference recognition task was higher in the Ordered Stories condition compared with both of the other conditions, and the other two did not differ from each other. The reasons for the decrease in inference recognition in the other two conditions is likely due to lack of robust event models in these conditions. In the Scrambled Story condition, it is possible that inferences were not made because they require being created online, within the event model (in active working memory), rather than offline, after being integrated into the situation model for the entire narrative (in latent episodic memory). This is speculative and could be explored further if we find the inference recognition scores of some of the participants who were able to recreate the scrambled stories. The ability to reorder scrambled stories has been documented (Kintsch, Mandler, & Kozminsky,1977), and presumably is made possible by reliance on schemata. It might be the case that memory consolidation is required to make inferences from scrambled stories that were put into situation models piecemeal via schemata rather than by event models. However, it still appears that participants were able to correctly guess above chance for the Scrambled Stories allowing the possibility that some inference was possible.

3.5.0 Event-wise garden path

For the Unrelated stories, creation of the event models would be expected to occur, however it was obvious from the recall data that participants suffered massive interference of event model encoding during this condition. Although it is not examined here in detail, the anecdotal first impressions from having been present for nearly all of the recalls would be that recall in the Ordered Stories conditions was better than in the Scrambled Stories, which were far better than the Unrelated Events. In the Unrelated events condition, participants were often able to recall details from only 2 or 3 out of 11. What events were recalled were often predicted by recency and primacy effects. This was often to the participants' own bewilderment since many would have previously recalled a majority of the events in the Ordered Story experienced on the previous block. When interviewed about why the Unrelated events were so difficult to remember, one particularly introspective participant said:

> *It takes a few seconds to figure out that the event has changed and the change is causing the previous one you heard to [be forgotten]…It is like having a flow, or a stream of consciousness and someone says something and you get irritated that they said something and just forget everything.*

The introspective account elicited from that participant is coherent with the timeline of cognitive mechanisms hypothesized by event model theories, particularly the onset of encoding the event model into long-term memory. This is scheduled to take place upon realization that an event boundary had occurred. If a boundary had not been noticed, the effect could be that information from two, unrelated events had accidentally been integrated into one event model. During comprehension of linearly presented discourse, blending consecutive event models would be less likely to create an infelicitous event model; fusion might only result in overloading WM, or confusing of a minor detail such as location of where an event took place; and it might go unnoticed.

However, in non-linear narratives it follows that the stakes are higher: blending consecutive events into one model would result in something entirely different from what the narrator intended. This would become evident in proximal sentences where coherence would be more likely to break down. This is similar to the *garden path* effect which describes misattribution errors during processing of tricky sentences such as “The old man the boat”.

During reading this might result in a regressive eye movement to initiate a reread. During aural narratives for lack of time and resources, it might be beneficial for comprehenders to dissolve what binds the tainted event model in memory (“popping the event model bubble”), rather than to attempt to extricate propositions from the false chimera of the two, or else to consciously believe a lie. This might also speak to a difficulty in parsing event models into pieces.

It might be the case that the event boundaries were less recognizable in the Unrelated Event condition, and that this caused the realization that the event had changed to come as a surprise. It is possible that surprise disrupts episodic encoding or “wrap-up” and induce some sort of amnesia. The precise timing of this surprise-induced amnesia might be interesting to explore, as would varying the degree of noticeability of the event boundaries in non-linear narratives. For example, if external cues were used to correctly cue the event boundary we might expect by this account that behavioral measures of comprehension (recall, recognition) in the Unrelated Events condition would greatly increase. This would be similar to the beneficial effect of increasing boundary visibility on recall seen by (Flores et al., 2018).

Before discussing the ERP data, the linear effects of DFB on SO found in each condition were unexpected. Looking at the figures (figure 7a), it would appear that the effect could have been driven by the first sentence in each event (DFB=0). The large discrepancy between the SO at DFB 0 and the sentences at other DFB cause the negative linear effects to be considered an artifact of this discrepancy, rather than a true linear effect. If we were to remove the first sentence, we would likely see no significant negative linear effect at least in the ordered and scrambled story conditions, though possibly still in the Unrelated Events condition. The cause of such an effect could be due to the wording of the first sentence of each event which often invoked the main characters or else common characters such as “father”, “woman”, “brother” or common locations (“road”, “castle”) or common verbs (e.g., “walked”, “travelled”) which could have biased the SO model to score these words and thus these sentences highly. Another possibility is that high coherence in the boundary sentence was due to high intra-sentence coherence within the beginning and end of boundary sentences (DFB=0).

3.5.1 Semantic Overlap as Semantic Coherence

The measure of SO detected changes across the event which were in the opposite direction of what was expected. The pattern seen across all conditions was that the first sentence of each event had a higher degree of semantic overlap than the rest of the sentences in the event. This would be unexpected given an event-centric view (Gernsbacher, 1990, Kurby and Zachs, 2008) who suggest that it is break from continuity with the previous context which triggers a coherence monitor to detect incoherence and signal an event boundary. If the model we have used for semantic overlap is measuring anything close to semantic coherence, then this suggests that such a mechanism in the mind would not become active until the second sentence of a new event, since the first sentences have the highest semantic overlap with what came before them. In our model, it appeared that the second sentence of each condition was the one which incurred the largest decrease in semantic overlap. If, in later analyses human raters show a delay in noticing event boundaries until after the second sentence, this would support the idea that decreased lexico-semantic coherence is being monitored to trigger event model creation and wrap-up. Otherwise, if large intra-sentence coherence is to blame for the large SO of the boundary sentences, then it might be the opposite: coherence monitoring detects high intra-sentence coherence to trigger a boundary. This is completely speculative and would require more computational modelling.

3.5.2 Condition SO effects.

The SO in the Unrelated condition was the lowest of the three. While the Story Conditions were not different. This was not surprising given that the DFB=0 in a story condition is more likely to contain higher overlap with the previous sentence from a previous event (DFB=-1) due to the continuity of characters, actions, and locations than in unrelated events condition, which did not share characters or overlap (unless by accident).

The decrease in SO in the Unrelated events condition is surprising, however, given that the intra-event sentences should have been more semantically related to one another than sentences spanning events. That is, sentence DFB=1 should contain more semantic overlap than DFB=0 by virtue of containing some words from DFB=0 in the SO model. This was not the case with the stimuli used here. To explain why this might not be the case, might require a look at narrative structure. Borrowing terms from Neil Cohn's narrative grammar (2013). In explaining the structure of a four-panel comic, Cohn invokes the concept of narrative *pace* or the speed at which the story is told, jumping from exposition (Cohn's *Orientation* and *Establishment*) to eventually a *Peak* or "height of narrative tension" (Cohn 2013 p 73), and a *Release* or denouement. Pace could be a factor, especially in the Unrelated Events, where sentences within events are more likely to progress from establishment to "rising action" in sentences DFB=0 and DFB=1, and finally a peak in DFB=3. Such pace requires information to be transmitted with less redundancy and coherence in the lexical semantic content as new objects and characters are introduced quickly from sentence to sentence. This could be an artifact of the 3-6 sentence pace.

In the Story conditions, apart from SO differences between DFB=0 and the other DFB, the pattern is more expected. The SO did not differ as much within DFB1-4 in the Scrambled Stories, as it did within DFB1-4 in Unrelated Stories. From this, it appears that randomizing sentences in the Scrambled Story was mildly successful in reducing systematicity in semantic relationships between sentences within events.

3.5.3 Condition SO and N400 effects

The effects of condition on the N400 did not follow the effects of condition on SO. The N400 was most negative in the condition with the highest SO, the Ordered Story Condition. This runs counter to the findings of others such as St. George et al., (1994), who found that decreased coherence in the discourse caused general increases in the amplitude of the N400 across content words. Our results show the opposite relationship between the N400 if we consider each experiment block as one unit of discourse. In our study, the N400 is greatest (most negative) in the Ordered Stories condition. In this condition integration of each word into a model of the narrative should be easiest.

This also conflicts with experiments by Otten and Van Berkum, (2008), who saw increases in the N400 from both the message (event model) level as well as local context of neighboring words. We see neither.

This finding is aligned with the results of a study by Otten and Van Berkum (2007) which modulated message-level constraint and semantic overlap of target words embedded in two-sentence passages. In such scenarios where message-level constraint was weak, the semantic overlap was both weaker and more left lateralized. Similar to our study, Otten et al. (2007, 2008) did not use nonsensical or implausible words in their low-semantic overlap condition. The strength of the N400 in the unrelated condition appears positively related to message-level constraint. In our study, the message level constraint in the Unrelated Events condition was weaker compared to the Story Events condition due to the lack of accumulated background information. While we do not have an overt manipulation of semantic relatedness, given the lack of an attempt to overtly create highly-related contexts, we assume that the mean relatedness in our stimuli (mean=0.31, s.d.=0.17) was lower than what Otten et al., used for their related target words, and more comparable to their semantically less-related target words. Therefore, the constraint-based modulation of the N400 is explanation for our condition-wise results.

3.5.4 Mediation analysis by condition

The marginally significant effect of DFB on the N400 in the Scrambled Sentences Condition was explainable by differences in DFB reflected in both SO and partially on its own (Figure 8). The findings from the analysis of the mediation of the effect of the DFB by SO on the N400 for the Scrambled Stories condition are surprising given that this condition was intended to serve as a control condition for the effects of DFB. The presence of a marginally significant effect of DFB on the N400 causes us to think that some process governing LSP was present during the scrambled condition and was able to track the event boundary.

One explanation is that ordering of the blocks lead to entrainment, or learning of some form. The ordering of blocks was such that Scrambled Stories always came after Ordered Stories and Unrelated Events. This could have resulted in a bias in detecting boundary sentences. The diminution of N400 amplitude (more positive) with increased DFB is what we would have expected in the other two conditions which *did* contain coherent events. Anecdotal evidence from the recall of scrambled stories leads us to believe that in some instances, participants were forming event models from adjacent sentences in the Scrambled Stories. The erroneous fusing of consecutive sentences during the recall of scrambled stories likely occurred for multiple participants on multiple occasions. It is possible that this was due to another strategy of event model creation. The size of the linear effect with the effectiveness of the strategy will be something to investigate in future analyses.

3.6 Conclusion

While the hypothesized effects of Semantic Overlap (SO) and distance from the event boundary (DFB) on the N400 were not supported (expect for the Scrambled Stories condition), the behavioral analysis and analysis of condition effects on the N400 revealed a number of surprising findings. First, we speculate that decreased inference recognition in the Scrambled Stories and Unrelated Events conditions was due to schematic processing and event model disruption, respectively; both resulting in the inability to create event models. Pending a more formal analysis of recall measures, the discrepancy between inference recognition and recall in the Unrelated Events condition could be due to the timing of when event boundaries were noticed relative to when event models were actively being created. The event boundaries in this condition were less likely to contain changes in time or location and might have gone unnoticed. A replication with the same stimuli, this time adding clear signals to when an event has changed would be evidence that lack of awareness of the boundary is what was to blame for the poor performance in this condition. It may be the case that comprehenders often miss boundaries, but this would go undetected in linear narratives with temporally and locationally contiguous events. For non-linear narratives, boundary recognition must overcome this expectation to allow proper event model encoding.

Decreases in SO with DFB were likely artifactual and due to drastically greater SO in the first sentence post-boundary compared to all other sentences. Reasons for decreasing SO from the beginning to the end could be due to boundary-sentences having generic words which were highly related to any context. Alternatively, the drop off in SO with DFB was due to narrative pace. It might be wise to consider the narrative pace in relation to constraint when designing any multi-sentence stimuli. For example, some narrative sequences might contain a *release* sentence after a *peak* sentence and other sequences might end at the peak sentence. The existence of a release sentence might cue an upcoming event boundary. If subjects are sensitive to the narrative arc when considering when an event boundary would occur, the peak-final sentences could have also contributed to the difficulties in recognizing event boundaries in this experiment. In light of this consideration, theories of text processing in natural conditions should consider how event-level constraint narrative pace. Condition differences in the N400 are best explained as the effects of heightened message level constraint in the Story conditions similar to that seen by Otten et al,. 2007. This can be seen as an effect of narrative structure on the N400. Further comparisons of the N400 at various levels of SO are warranted and further refinement of the location of the effect to ensure that location differences were not great across conditions. It could be valuable to further examine the predictive power of other computational models of lexico-semantic processing or on other ERPs, such as the P600 or earlier frontal negativities (see Perfetti & Frishkoff, 2008).

Finally, the DFB effects on the N400 were partially explained by SO in the Scrambled stories condition. For Scrambled Stories the largest effect of SO was in the beginning of the event, becoming increasingly weak as the sentences within the event continued. The main implication from this is that event boundaries were exerting some effect on the processing of events in the Scrambled Sentences condition. Whether such a tendency is increased by the recognition of the event boundary, or by entrainment to the pace or timing of the events from the other two conditions, or both, remains to be explored. Anecdotally, relatively intact recall in the scrambled stories opens the possibility that event models were being created with alternative strategies. These strategies could include admixtures of contiguous and non-contiguous actions as a mnemonic strategy to aid in later memory. Future research might consider replicating event-level entrainment. It may also be interesting to explore the degree to which the linear DFB-N400 relationship corresponds with the effectiveness of comprehension for scrambled stories.

Chapter 4: Time-frequency Measurements of Attention and Episodic Memory at the Event Boundary During Story Comprehension

4.1 Research Questions

Analysis 2 is designed to answer the following two research questions: (1) Does the event model predict EEG correlates of attention during story comprehension? (2) Does the event model predict EEG correlates of memory encoding during story comprehension.

4.2 Overview

Event segmentation process models (ESPM) focus on the structural grouping of actions contained in narratives into separate events. The implication is that there is lots of neural action shortly after a comprehender passes a boundary between events. Once this boundary has been detected, ESPM hypothesize a cascade of mechanisms required to create new mental representations for the upcoming event (event model), while integrating the previous event model into the larger situation model of the narrative as a whole (Zacks et al., 2007). Existing cognitive neuroscience experiments involving participants reading and watching narratives has lead ESPM to make predictions about how and when cognitive operations such as attention, working memory (WM) and episodic memory are activated to handle the demands of such hypothesized models (Kurby et al., 2008, Radvansky & Zacks, 2017). More specifically ESPM predict that people will be maximally attentive at the beginning of events; and that event models integrate into the situation model of the entire narrative in episodic memory upon the realization that some event described in the narrative has changed. The behavioral phenomenon whereby readers slow down after an event boundary can thus be explained by effortful semantic, perceptual processing, lingering effects of storage from the previous event, or all three (Speer et al., 2005, Radvansky & Zacks, 2011) (for discussion, see Appendix E vii). Other methods could be employed to further understand these mechanisms which handle narrative structure, to reconcile them with neuro-behavioral evidence supporting semantic integration (Mason & Just, 2004, Ferstyl, 2007).

Electrophysiological techniques like EEG can be used to separate concomitant neural processes such as we would expect at the event boundary. Frequencies of EEG activity have been associated with general cognitive mechanisms of attention and episodic memory, specifically alpha (8-12 Hz), and theta (3-7 Hz) band power. In addition, recent electrocorticographic (eCoG) recordings have shown hippocampal cells firing at the theta frequency (4-8 Hz) while comprehenders view video narratives, with specific populations of cells to event boundaries and others responding to periods within events (Zheng et al., 2022). We propose studying neurological effects of the event boundary during narrative comprehension using time-frequency analysis of EEG. Because ESPM theorize drastic changes to attention and memory after the event boundary (see Appendix E vii for discussion), EEG frequency power in these various bands (as measured at the surface of the scalp) can be expected to change within and across the event boundary. Such findings of systematic change in these frequency bands could reconcile other cognitive theories of meaning representation during narrative comprehension which focus on inter- and intra-sentential inference but are less-event centric (see Perfetti & Frishkoff, 2008 for a review). The experimental data is the same as was recorded for Analysis 1, however, the methods for analysis are different and aimed at different cognitive mechanisms (attention, WM, and episodic memory vs. LSP) and at a larger timescale (sentence vs. word). Below is a summary of research findings relating alpha power and theta power to attention and memory during comprehension, followed by hypotheses for what we expect to observe for each band at the event boundary.

4.3 Hypotheses:

4.3.1 Alpha Frequency Power and Comprehension

Alpha frequencies (8-12 Hz) consistently vary with the perceptual demands of a task (Klimesch, 2012). Klimesch (see Klimesch, 2012) and others (Scheeringa et al., 2008, Jensen & Mazaheri, 2010, Meeuwissen et al., 2011) have suggested that cortical alpha power tracks task demands by suppressing perceptual cortex when it is irrelevant to the task, reducing activity in those areas and thus reducing the ability for the brain to respond to certain irrelevant inputs. For example, in simultaneous fMRI and EEG recordings, Scheeringa et al., (2008) found increased alpha was associated with decreased blood flow in perceptual cortex.

Such a mechanism could be advantageous when alpha waves appear in cortex that is unnecessary for processing task stimuli (Foxe, Simpson, & Ahlfors, 1998) or sensory input during an interval during which people are trying to retain information in memory (Klimesch, et al., 1997, Meeuwissen, et al., 2011, see Appendix E xv for discussion). However, this *blocking* hypothesis may only tell part of the story. Alpha increases have also been seen when sentences were processed that required an additional semantic judgment in addition to reading (Rohm et al., 2001). For example, when subjects were asked to offer a superordinate category for a word in a sentence (e.g., “box” in “A rabbit is in the box, hiding." is a "container"), alpha power increased in comparison to when subjects were asked to read without making this judgment (Rohm, et al., 2001). This latter finding could have been indexing memory processing required by the task. These latter findings suggest that alpha may also index executive engagement of cortical areas in the service of applying semantic knowledge. This is in line with the idea that increased alpha activity over perceptual areas is beneficial to performance in tasks requiring the preservation of internal working memory processes, or what Klimesch has called *semantic orientation*. From the semantic orientation account, we would expect alpha power in brain areas necessary for semantics (temporal lobes) to increase between event boundaries for maintenance of event models. Where this would appear on the scalp is difficult to predict given the low quality of EEG source localization.

A third alpha effect, that can impact comprehension is frontal alpha power. Large-scale alpha power increases over frontal electrodes index low awareness, and as such are seen across the cortex during periods in which subjects report zoning out or *mind wandering* during comprehension, presumably devoting perceptual resources to their own internal thoughts away from the outside world (Smallwood, 2011, Boudewyn & Carter, 2018, Compton, Geeringer, & Wild, 2019) (see Appendix E xvi for discussion). Mind wandering is considered to be a state of low perception (Smallwood et al., 2008, Smallwood, 2011). Boudewyn and Carter (2018), relied on mind wandering to attribute increases in alpha power, particularly over frontal electrodes, to "missed" or poorly perceived words or sentences during comprehension of narratives. In line with the mind-wandering account, we would expect increased alpha over large portions of the scalp to correspond with decreased awareness during story comprehension.

4.3.2 EEG Alpha Power and the Event Boundary

From the blocking account, we would expect decreased alpha in perceptual areas (temporal electrodes in an auditory task) after an event boundary is crossed to the extent that we see heightened perception of external stimuli (Eisenberg et al,. 2016). This is hypothesized in ESPM to build up new event models from scratch, requiring decreased top-down control (Kurby et al., 208, Zacks, et al., 2011, Bauer et al., 2017, Richmond et al., 2017). In accordance with ESPM, we expect the boundary to force executive control shift semantic knowledge or some other process which encumbers access to recently seen information. From the semantic orientation perspective this would also suggest a decrease in alpha power over task-relevant areas after the boundary, reflecting this shift. Given mind-wandering is found to be least likely to occur after a new event begins (Faber et al., 2018), we expect the event boundary to incur heightened awareness of the task, decreased mind wandering or zoning out, and thus decreased alpha waves over frontal areas (Boudewyn & Carter, 2018).

Alpha power between the event should, by all account be the opposite. From the mind wandering account (Smallwood, 2011), between boundaries, we would expect mind wandering to become more likely, and bottom-up perception to decrease in proportion to the amount of executive attention required by event model building. To the extent that event model creation utilizes syntactic and semantic knowledge, which have both been shown to occur alongside increased alpha activity (Rohm et al., 2001, Sheeringa et al., 2009), we would expect increased EEG alpha power across the event as the event model builds. From the blocking perspective, alpha would increase over perceptual cortex not relevant to the task. Combining these three interpretations we hypothesize that EEG alpha power will increase over most electrodes on the scalp during the event, while immediately after the boundary, alpha power will decrease in all areas corresponding to an increase in bottom-up perception and renewed awareness of the task (see figure F1 in Appendix F). The location of these effects could have theoretical implications for both ESPM and each of the functional interpretations of alpha EEG power more generally.

4.3.3 Theta EEG power and Comprehension

In line with the idea activity in the hippocampus fluctuates in the theta frequency (3-7 Hz) (Miller, 1989, Busczaki 2002), human EEG theta recorded at the scalp is thought to originate from cortical areas that (at least indirectly) coordinate their activity with the hippocampus in some way. Theta rhythms in neurons located in the hippocampus and the cortex have been implicated in episodic memory formation (Buzsaki, 1998, Lega et al., 2012, Yoo et al., 2012, see Hsieh & Ranganath, 2014 for a review). This makes theta rhythms especially interesting to the study of comprehension and memory for event models, as episodic memory has been implicated as the memory domain in which situation models and event models are stored (Ericsson & Kintsch, 1995, Zuo et al., 2020). More recently, theta rhythms in the hippocampus have been linked to event model processing (Zheng, et al., 2022, Yoo et al, 2022), further justifying their candidacy as a means of testing ESPM.

While Tulving (2002) warned that successful recognition of an item as having been seen before does not necessitate that a subject used their episodic memory to do so, episodic memory has nonetheless been implicated as the type of memory involved in recognition paradigms. Increased theta EEG power has been recorded while subjects viewed stimuli that they later remembered having seen (Klimesch, et al., 1997, Khader, et al., 2010, see Hsieh & Ranganath, 2014, Nyhus & Curran, 2010, Herweg, Solomon, & Kahana, 2020 for a review). These so-called, *subsequent memory effects* (SMEs) in the theta band were originally shown for memoranda that included stimuli such as word pairs or lists of words (theta: Klimesch et al, 1997) and pictures (Khader et al., 2010).

Theta measurements have been seen to correspond with comprehension of more complex, linguistic stimuli such ass sentences (Bastiaansen, Magyari, & Hagoort, 2010) and informative passages (Sato & Mizuhara, 2018). In a post-hoc analysis, Bastiaansen, et al., (2010), saw theta EEG power increase during sentence comprehension across the sentence until the point when an error was detected. As the stimuli were not related discourse, but isolated sentences, Bastiaansen et al., (2010) interpreted theta EEG as an index of either LSP, or else syntactic processing.

A more direct connection between EEG theta power and discourse was found by Sato et al., (2018) who measured EEG while subjects read naturalistic texts (Sato & Mizuhara, 2018). Sato et al. (2018), had subjects read articles about scientific topics. Later the experimenters asked subjects to recall all that they could about each article. In a complex behavioral analysis, they used semantic vectors derived from a vector space model (VSM), word2vec (Mikilov, et al., 2013), in order to estimate semantic overlap (SO) between the responses and sections of the text. The results allowed them to detect sections of the articles which maximally overlapped with the responses, presumably indicating parts of the text for which memory formation was strongest. When correlating the EEG frequency data from multiple bands with semantic overlap of recall responses, theta EEG power over the left hemisphere was found to have the only significant correlation. The correlation was positive, suggesting that higher theta EEG activity over the left hemisphere during encoding was associated with better SME. As episodic memory is hypothesized to be the means in which event details are remembered (Ericsson & Kintsch, 1995), this suggests that EEG theta power during comprehension implicates theta power in episodic memory encoding for text. One caveat, however, is that the EEG theta activity seen by Sato et al., (2018) did not occur over frontal midline electrodes. It is therefore possible that it is an entirely different phenomenon, not related to SME experiments listed above (see Appendix E xix for discussion). Sato et al., (2018) interpreted their results as relating to the operation of the default mode network during comprehension.

4.3.4 EEG Theta Power and the Event Boundary

To the extent that theta power is related to encoding information in episodic memory we expect theta power will be implicated in the formation of event models. To the extent that SME effects will correspond with increases in frontal midline, or temporal theta EEG power, we expect theta power to increase at points vital to LTM encoding. Since ESPM predict updating of the situation model with event model information to occur after the event boundary is discovered, (i.e., when the event model is concluded), we would expect frontal midline and or temporal theta to increase at the point when this is noticed. Unlike video stimuli, however, the grammatical differences across sentences which signify a boundary and individual differences in inference making may vary the point of boundary recognition. Thus, we expect that memory updating to occur in either the first or second sentence of a new event after the boundary (Appendix G, figure G1). If it is the case that theta EEG activity is related to updating the situation model incrementally, sentence by sentence, we may expect increases in theta EEG power after each sentence in situations when the sentences of a story are scrambled (Scrambled Story condition), and thus each sentence represents an event boundary (for reasons why we might not see theta increases, see Appendix E xix).

4.4 Methods

4.4.1 Participants

Data from 49 subjects was analyzed. After inspection of the data, 3 subjects were removed due to excessive 13-30 hz power, on over half of the trials, or from having partially recorded sessions due to technical error. One subject was removed for having extreme values. Data from forty-five participants were retained in the final analysis.

4.4.2 EEG preprocessing

Participants, materials, procedure, and EEG recording, independent components analysis for artifact removal, filtering, the protocol for channel removal and interpolation, and re-referencing were identical to what was performed in Analysis 1.

4.4.3 Epochs

EEG data was segmented into epochs of 8.5 seconds beginning 1.5 seconds before the onset of each sentence and lasting until 7 seconds after the onset of each sentence. Epochs which contained 12-30 Hz activity beyond a threshold were considered to contain muscle artifacts and were removed from the analysis. Mean number of epochs removed was .25 (s.d., 7.9). One subject was removed from the analysis due to having an excessively low number of trials in all 3 conditions (< 3 s.d. from the mean).

4.4.4 Time-Frequency analysis.

Frequency analysis was performed using the complex wavelet transform described by Cohen (2014, 2019). This method was chosen due to its computational efficiency. In this procedure, a family of Morlet wavelets are tailored to the specific frequencies of interest. EEG recordings for each channel and epoch are then subjected to a fast Fourier transformation, and the resulting transformation is convolved with each of the wavelets. The inverse of this dot-product results in a section of the original input, bandpass filtered at the frequency corresponding to each wavelet, equivalent to the narrowband-bandpass signal. Squaring this results in frequency power at that specific band. Frequency was measured in .5 Hz increments from .5 to 30 Hz. For each increment, 1-second long (500 sample) wavelets were created (7 cycles). Measures of frequency power at each frequency within alpha and theta ranges were averaged together to create mean band power within the alpha and theta ranges for each sentence epoch, and participant. The theta range was chosen to be 4-8 Hz, after the frequency range of hippocampal neurons found by Zheng et al. (2022) to be sensitive to both event boundaries and the middle of events. The range of frequencies classified as alpha was 8.5-12.5 Hz based on the analysis by Maurer et al (2015).

4.4.4.1 Baseline correction.

After averaging within each frequency band (alpha and theta), each epoch for each frequency band was baseline corrected to the mean of the epoch preceding the event boundary (DFB=-1). This was not possible for epochs of the first event of each block (no epoch nor event preceded them) and were discarded from the analysis. The result is that if some of the power values are negative it is because they were less than the baseline segment.

4.4.5 Electrodes

Electrode clusters were chosen based on a mixture of the hypotheses as well as visually inspecting topographical maps of the grand averaged data, across each sentence organized by DFB and condition. We were interested in effects over posterior temporal, as well as frontal scalp locations. The end result of this hypothesis-guided visual inspection was 6 clusters which covered most of the frontal and posterior scalp. These electrode clusters included Left Frontal (FP1, AF3, F3, F7, FC1, FC5), Right Frontal (FP2,AF4,F4,F8, FC2, FC5), Midline frontal (Fz), Left Posterior(P3, CP5, PO3, P7,O1), Right Posterior (P4, CP6, PO4, P8,O2), and Midline Posterior (Pz).

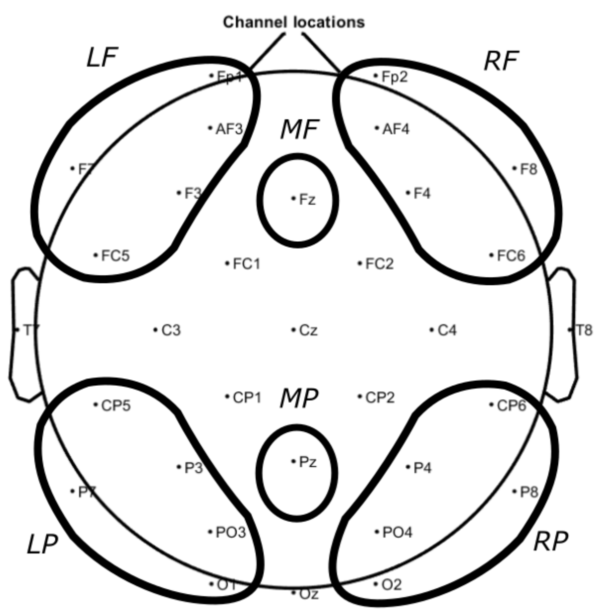


Figure 10. Electrode groups for time-frequency analysis. Topographic layout of scalp showing location of electrode groups for analysis 2 (arial view, nose pointed toward top of page for reference).

4.4.6 Averaging

For each epoch and subject, TF values for each channel cluster were then averaged within the range of 0-4.5 seconds. In order to avoid including data from the next sentence in epochs with sentences that were shorter than average, this range corresponded to slightly length less average length of each sentence (4.9 seconds, s.d., 0.69 seconds). After this, any epochs containing cluster values over 3 standard deviations from the mean power across clusters within any band were removed from the analysis. This procedure resulted in the removal of an average of (4) epochs per subject per condition. After this procedure, three subjects were removed for having fewer trials than 3 times the standard deviation from the mean number of trials. After this, the average number of trials removed in the final sample was 2 (s.d.=4). The end result was an average of 128 (s.d.=8.9) epochs per subject for Ordered Sentences, 129 (s.d.=9.1) for Scrambled Stories, and 123 (s.d.=4.4) for the Unrelated Events condition. TF data for each cluster and frequency range were stored with unique identifiers for each sentence, event, block, subject, DFB, and condition.

4.4.7 Statistical Analysis

Statistics were performed using R Studio software (). Separate 5-way, within subjects analysis of variance (ANOVA) was performed for each frequency band of interest (Alpha) and (Theta) predicting changes in frequency power with 3 levels of condition (Ordered Stories, Scrambled Stories, Unrelated Events), 5 levels of DFB (0,1,2,3,4), 3 levels of Laterality (Left,Midline,Right), 2 levels of Frontality (Frontal, Posterior). The results are summarized below. Greenhouse-Geisser correction was applied to all p-values. Tukey's post-hoc tests were performed to explain significant factor effects.

4.5 Results

4.5.1 Alpha ANOVA

The results of the ANOVA predicting Alpha power are shown below in table table 7. There were significant effects of both Condition ($F(2, 102973)=5.5$, $p=.004$), and DFB ($F(4, 102973)=100.41$, $p<.001$), as well as a significant interaction of Condition and DFB ($F(2,102973)=3.78$, $p=.023$).

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
laterality	2	0.75	0.37	0.15	0.858
frontality	1	1.62	1.62	0.66	0.416
condition	2	26.89	13.45	5.50	0.004**
DFB	1	245.68	245.68	100.41	0.000***
laterality:frontality	2	1.68	0.84	0.34	0.709
laterality:condition	4	2.37	0.59	0.24	0.915
frontality:condition	2	6.70	3.35	1.37	0.254
laterality:DFB	2	0.90	0.45	0.18	0.832
frontality:DFB	1	6.43	6.43	2.63	0.105
condition:DFB	2	18.49	9.25	3.78	0.023*
laterality:frontality:condition	4	2.46	0.62	0.25	0.909
laterality:frontality:DFB	2	5.69	2.84	1.16	0.313
laterality:condition:DFB	4	1.12	0.28	0.11	0.977
frontality:condition:DFB	2	0.38	0.19	0.08	0.924
laterality:frontality:condition:DFB	4	4.60	1.15	0.47	0.758
Residuals	102973	251946	2.45		

Residual standard error: 1.564 on 102973 degrees of freedom
Multiple R-squared: 0.001291, Adjusted R-squared: 0.0009519
F-statistic: 3.804 on 35 and 102973 DF, p-value: 2.347e-13

Table 8. Alpha ANOVA table. Results of the ANOVA showing mean differences for alpha (8.5-12.5 Hz) power (narrowband amplitude, squared) by Condition and Distance from boundary (DFB) as well as the interaction (DFB x Condition). Effects of Laterality (Left/Midline/Right) and Frontality (Anterior/Posterior) were insignificant.

4.5.2 Alpha and Condition pairwise comparisons

Although describing the effect of a single variable from a significant interaction can be misleading, we suggest that it might be informative for other EEG research to compare the effects of Condition here. Tukey’s post-hoc tests showed significant differences between mean alpha power in Unrelated Events and both Scrambled Stor y (diff=.03 95% CI [.06 .001], p=.046) and Ordered Story (diff=.04, 95% CI [.07 .01], p=.005) conditions. There were no significant differences in alpha power between Scrambled and Ordered Story conditions (diff=.01, 95% CI [.04 .02]. p=.73).

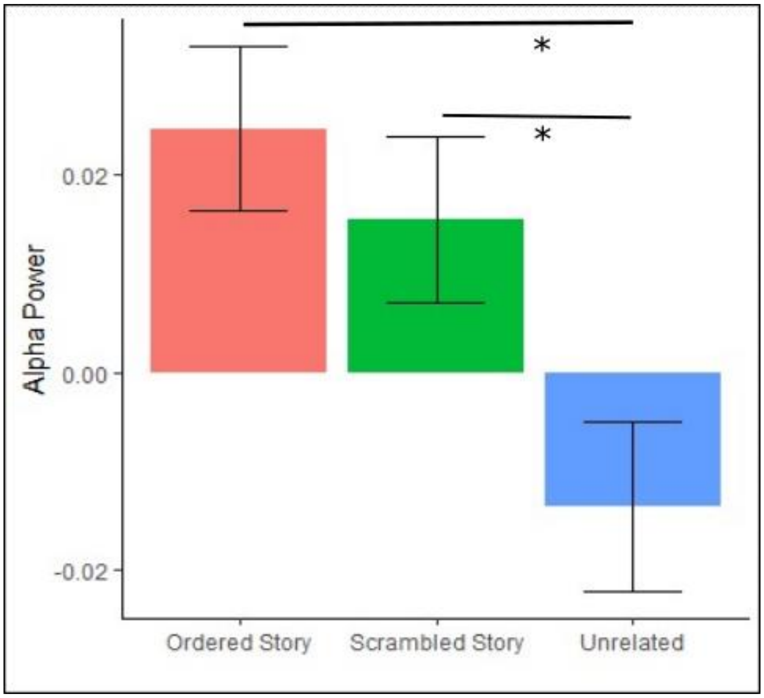


Figure 11. Mean alpha by Condition. Mean alpha (8.5-12.5 Hz) across condition. Error bars depict mean standard error. Significant effects seen in post-hoc tests are labelled "*"=p<.05.

4.5.3 Alpha and DFB pairwise comparisons

Although describing the effect of a single variable from a significant interaction can be misleading, we suggest that it might be informative for other EEG experiments to compare the effects of DFB here. Tukey's post-hoc tests comparing mean alpha power at each DFB showed that alpha power at DFB=2 was greater than at DFB=1 (difference=.061, 95% CI [0.023,0.099], p<.001), and alpha power DFB=4 was greater than at DFB=3 (difference=.086, 95% CI[0.033,0.139], p<.001). There was no significant increase in DFB1 vs DFB0 (difference=0.007 [0.031,0.046], p=0.99), and DFB3 vs DFB2 (difference=0.014, 95%CI [0.03, 0.06], p=0.889).

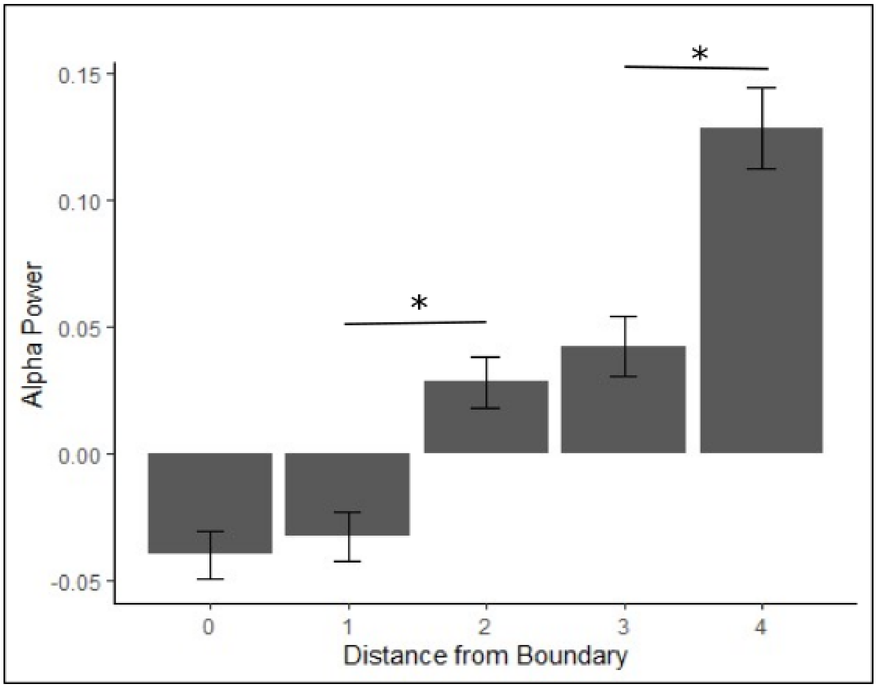


Figure 12. Mean alpha by distance from boundary. Mean alpha (8.5-12.5 Hz) across distance from boundary (DFB). Error bars depict mean standard error. Significant effects seen in post-hoc tests are labelled “*”=p<.05. See table 9 for more detail.

4.5.4 Alpha and DFB x Condition pairwise comparisons

Tukey’s post-hoc test for each Condition at each DFB revealed that the Unrelated Events invoked significantly less alpha power than the Scrambled Stories at DFB=1 (difference=.101, 95% CI [.018, 0.183], p=.003),).Likewise The Unrelated Events invoked less alpha EEG power than the Ordered Stories at DFB=1, however that difference was only marginally significant (difference=.081, 95% CI [0.002,0.164], p=.061). Alpha power averaged across all locations was not significantly different between Ordered and Scrambled Story conditions at any DFB (all $p > 0.95$).

Across DFB in the Unrelated Events condition, alpha activity was significantly greater than DFB=0 at all DFB except DFB=1 (DFB=2,DFB=3,DFB=4), and alpha power at each DFB after DFB=1 (DFB=2, DFB=3, DFB=4) was significantly greater than DFB=1. DFB=3 and DFB=4 in the Unrelated Events condition were not significantly greater than DFB=2 (see table 9).

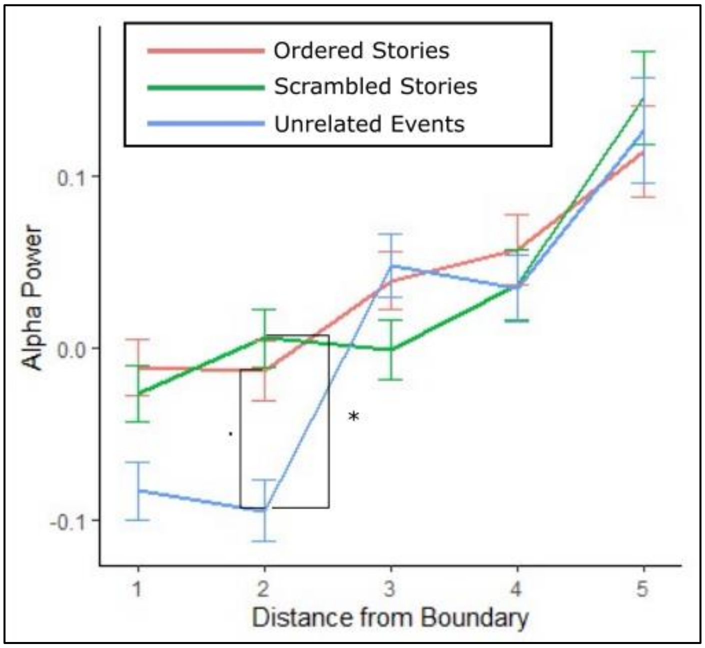
Across DFB in the Scrambled Stories condition, alpha activity was significantly greater in DFB=4 than in each DFB=0, DFB=1, and DFB=2 (see table).

Across DFB in the Ordered Stories condition, alpha activity was significantly greater in DFB=4 than in each DFB=0, DFB=1 (see figure table 9).

Condition	DFB com-pari-son	diff	lower 95% CI	upper 95% CI	p-value
Unrelated Events	1 - 0	-0.012	-0.096	0.072	1.00
Unrelated Events	2 - 0	0.131	0.047	0.215	0.000**
Unrelated Events	2 - 1	0.142	0.058	0.226	0.000**
Unrelated Events	3 - 0	0.117	0.029	0.206	0.001
Unrelated Events	3 - 1	0.129	0.041	0.217	0.000**
Unrelated Events	3 - 2	-0.013	-0.102	0.075	1.000
Unrelated Events	4 - 0	0.209	0.094	0.324	0.000**
Unrelated Events	4 - 1	0.221	0.106	0.336	0.000**
Unrelated Events	4 - 2	0.078	-0.037	0.193	0.584
Unrelated Events	4 - 3	0.092	-0.026	0.210	0.348
Scrambled Stories	1 - 0	0.033	-0.049	0.114	0.991
Scrambled Stories	2 - 0	0.026	-0.056	0.108	0.999
Scrambled Stories	2 - 1	-0.007	-0.088	0.075	1.000
Scrambled Stories	3 - 0	0.063	-0.026	0.153	0.518
Scrambled Stories	3 - 1	0.031	-0.058	0.120	0.998
Scrambled Stories	3 - 2	0.037	-0.052	0.127	0.986
Scrambled Stories	4 - 0	0.172	0.065	0.279	0.000**
Scrambled Stories	4 - 1	0.139	0.033	0.246	0.001**
Scrambled Stories	4 - 2	0.146	0.039	0.253	0.000**
Scrambled Stories	4 - 3	0.109	-0.004	0.222	0.072
Ordered Stories	1 - 0	-0.002	-0.084	0.080	1.000
Ordered Stories	2 - 0	0.051	-0.032	0.133	0.743
Ordered Stories	2 - 1	0.052	-0.029	0.134	0.687
Ordered Stories	3 - 0	0.068	-0.022	0.158	0.386
Ordered Stories	3 - 1	0.070	-0.019	0.159	0.335
Ordered Stories	3 - 2	0.018	-0.072	0.107	1.000
Ordered Stories	4 - 0	0.125	0.019	0.232	0.006**
Ordered Stories	4 - 1	0.127	0.021	0.234	0.005**
Ordered Stories	4 - 2	0.075	-0.032	0.181	0.527
Ordered Stories	4 - 3	0.057	-0.055	0.170	0.925

Table 9. Post-hoc alpha mean difference by distance from boundary and condition. Results of post-hoc tests of the mean difference in alpha (8.5-12.5 Hz) power (narrowband amplitude, squared) across electrode groups (LF, MF, RF, LP, MP, RP), within block conditions comparing various distance from boundary (DFB=1-DFB=0, etc.).

a.



b.

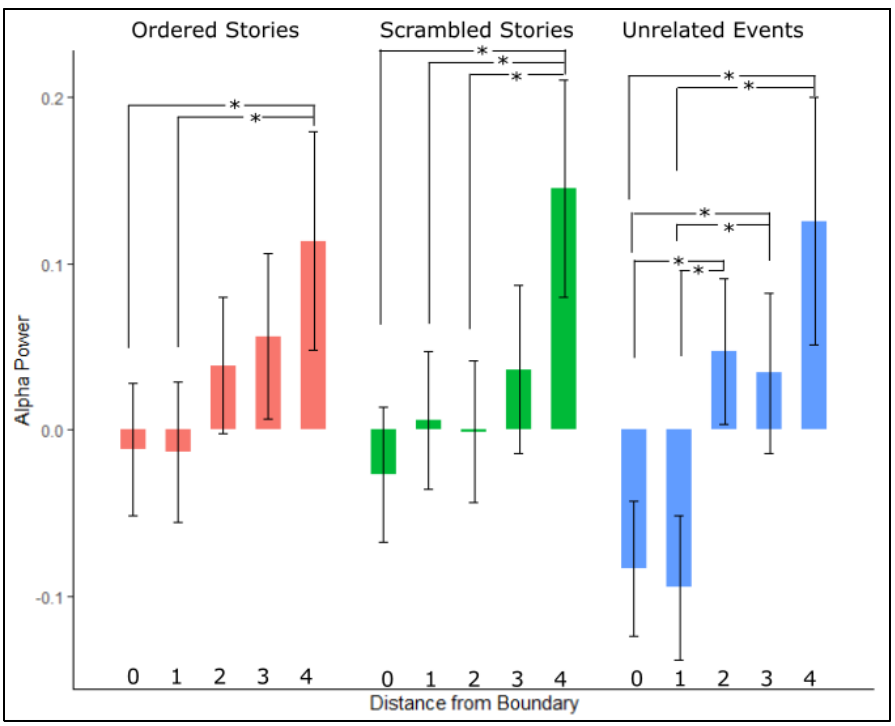


Figure 13. Alpha (8.5-12.5 Hz) power (narrowband amplitude, squared) by condition and distance from boundary (DFB), across all electrode groups (LF, MF, RF, LP, MP, RP).

(a) and (b) depict same data in different arrangement for ease of viewing across and within condition effects respectively.

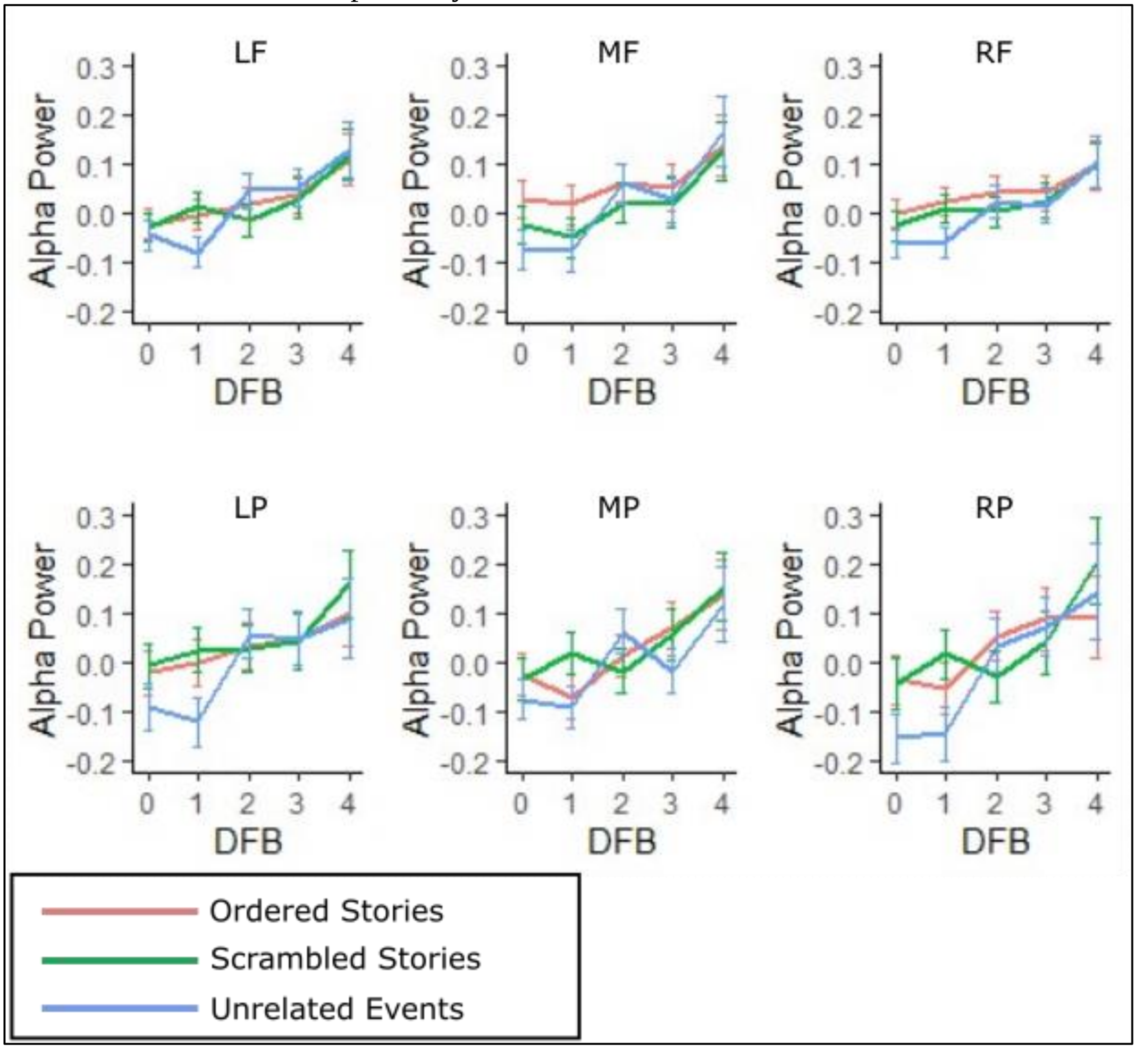


Figure 14. Alpha (8.5-12.5 Hz) power (narrowband amplitude, squared) by condition and distance from boundary (DFB), at each electrode group (LF, MF, RF, LP, MP, RP).

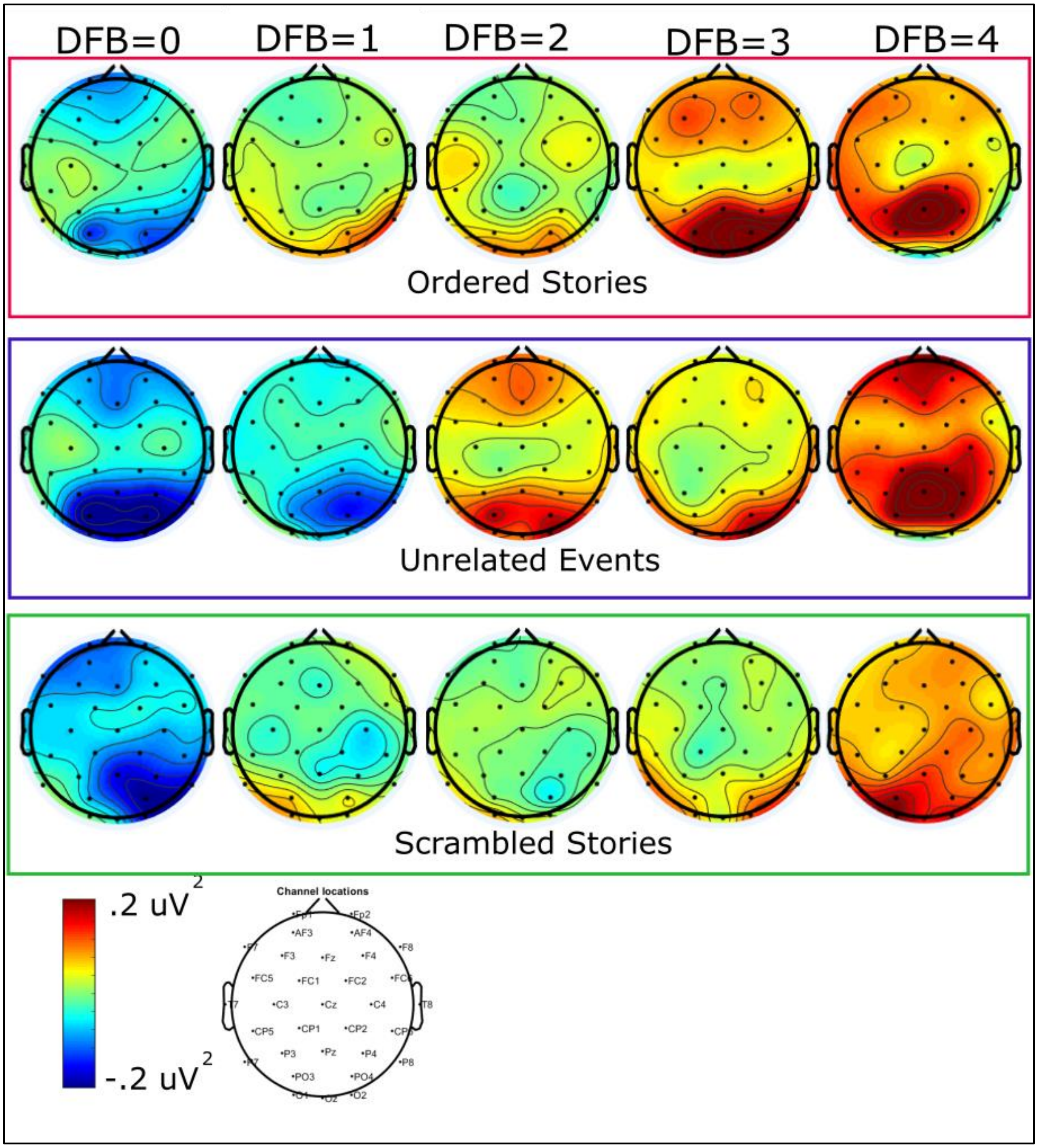


Figure 15. Topographic plots of alpha power over time. Topographic plots showing the alpha (8.5-12.5Hz) power (narrowband amplitude, squared) averaged over sentence (0-4.5 seconds) for each of the distances from the boundary (DFB=0-4), for the separate conditions.

4.5.5 Theta ANOVA

The results of the ANOVA predicting Theta power are shown below in table 9. There were significant effects of both Condition (F(2, 102973)=6.47, p=.002), and DFB (F(1, 102973)=65.03, p<.001), as well as a significant interaction of Condition and DFB (F(2,102973)=4.09, p=.017).

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
laterality	2	0.021	0.011	0.046	0.955
frontality	1	0.005	0.005	0.022	0.883
condition	2	3.021	1.510	6.517	0.001**
DFB	1	15.073	15.073	65.036	0.000***
laterality:frontality	2	0.402	0.201	0.868	0.420
laterality:condition	4	0.463	0.116	0.499	0.736
frontality:condition	2	0.129	0.064	0.278	0.758
laterality:DFB	2	0.061	0.031	0.132	0.877
frontality:DFB	1	0.088	0.088	0.378	0.539
condition:DFB	2	1.895	0.948	4.089	0.017*
laterality:frontality:condition	4	0.823	0.206	0.888	0.470
laterality:frontality:DFB	2	0.097	0.048	0.209	0.811
laterality:condition:DFB	4	0.436	0.109	0.471	0.757
frontality:condition:DFB	2	0.021	0.010	0.045	0.956
laterality:frontality:condition:DFB	4	0.894	0.224	0.965	0.425
Residuals	102973	23865.9	0.232		

Residual standard error: 0.4814 on 102973 degrees of freedom
Multiple R-squared: 0.0009807, Adjusted R-squared: 0.0006412
F-statistic: 2.888 on 35 and 102973 DF, p-value: 2.498e-08

Table 10. Theta ANOVA table. Results for testing difference in mean theta (4-8 Hz) power (narrowband amplitude, squared) by Condition and Distance from boundary (DFB) as well as the interaction (DFB x Condition). Effects of Laterality (Left/Midline/Right) and Frontality (Anterior/Posterior) were insignificant.

4.5.6 Theta and DFB pairwise comparisons

Although describing the effect of a single variable from a significant interaction can be misleading, we suggest that it might be informative for other EEG research to compare the effects of DFB here. Tukey's post-hoc tests comparing mean alpha power at each DFB showed that theta power at DFB=2 was greater that at DFB=1 (difference=.019, 95% CI [0.007,0.031], p<.001), and theta power DFB=4 was greater than at DFB=3 (difference=.019, 95% CI [0.003,0.035], p=.012),). There was no significant increase in theta power between DFB1 vs DFB0 (difference=0.007, 95% CI [0.005,0.018], p=0.54), DFB3 vs DFB2 (difference=0.004 [0.017,0.009], p=0.92), and DFB4 vs DFB2 (difference=0.015, 95%CI [0.0, 0.031], p=0.06). This is in line with the analysis of the means in figure 15 which show that maximum peaks in theta across DFB appear at DFB=2 and DFB=4.

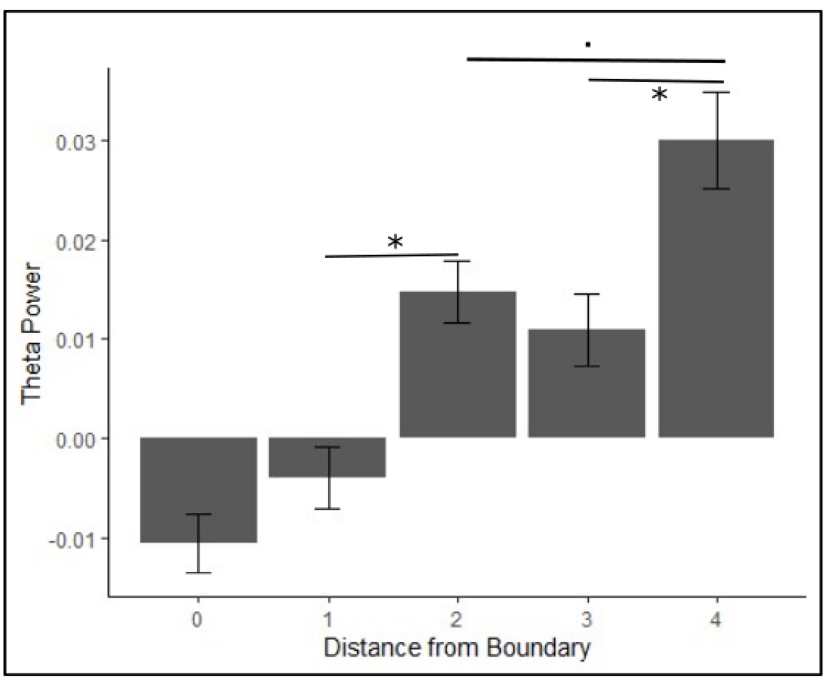


Figure 16. Mean theta (4-8 Hz) power (narrowband amplitude, squared) across distance from boundary (DFB). Error bars depict mean standard error. Significant effects seen in post-hoc tests are labelled "*"=p<.05. See table 10 for more detail.

4.5.7 Theta and Condition pairwise comparisons

Although describing the effect of a single variable from a significant interaction can be misleading, we suggest that it might be informative for other EEG research to compare the effects of Condition here.

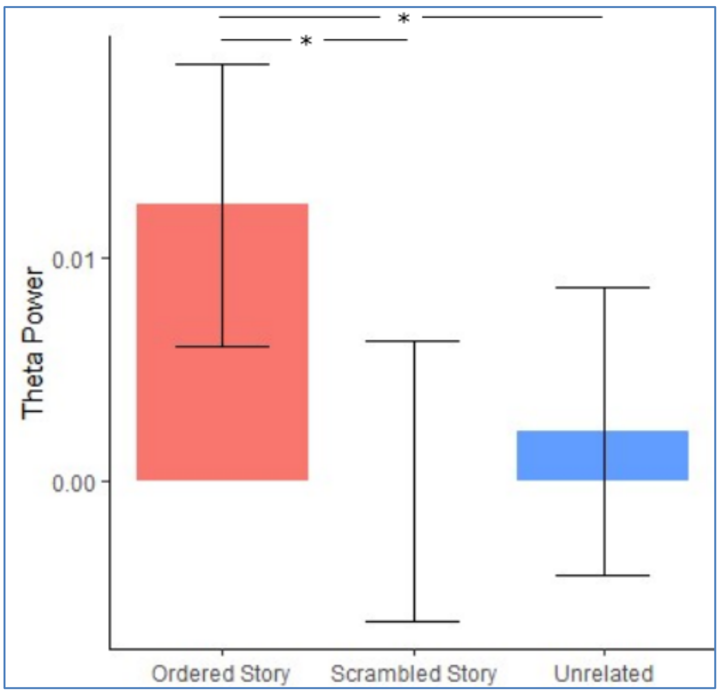


Figure 17. Mean theta (4-8 Hz) power (narrowband amplitude, squared) across distance from boundary condition. Error bars depict mean standard error. Significant effects seen in post-hoc tests are labelled “*”=p<.05. See table 11 for more detail.

4.5.6 Theta and Condition x DFB pairwise comparisons

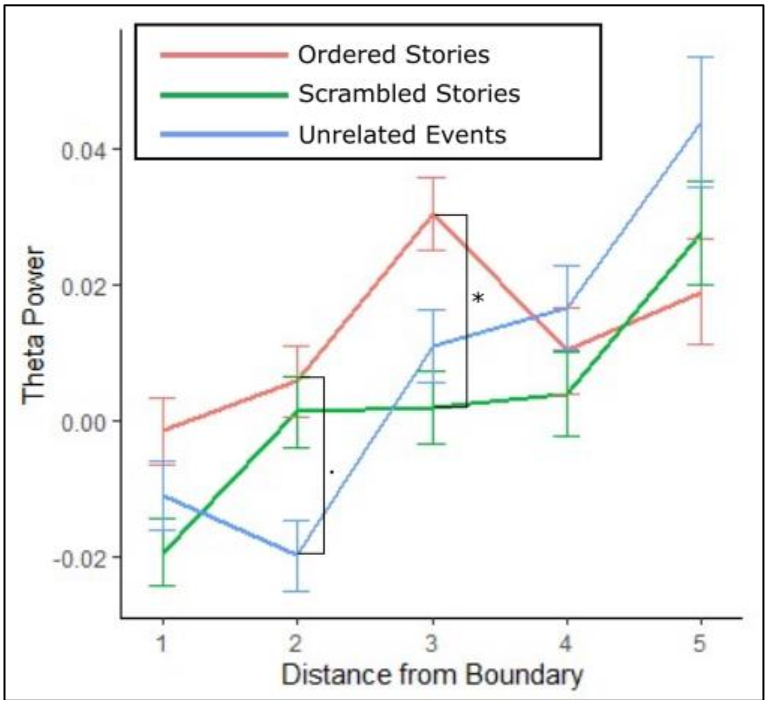
Analysis of Tukey's tests for each Condition at each DFB revealed higher theta power in the Unrelated Events than the Ordered Stories condition at DFB=1 (difference= -0.03, 95% CI[-0.052, -0.001], p=.031). Mean theta power was not significantly different between Unrelated and Scrambled Story conditions at any DFB (all >0.1). Mean theta power was significantly greater in the Ordered Story than in the Scrambled Story condition only at DFB=2 (difference=.0283, 95% CI[0.003,0.053], p=.027).

Across DFB in the Unrelated Events condition Tukey's tests show that theta power was significantly greater than DFB=1 at all DFB except DFB=1 (DFB=2,DFB=3,DFB=4), and theta power was significantly greater than DFB=1 at each DFB after DFB=1 (DFB=2, DFB=3, DFB=4). No other comparisons were significant (see table 11).

Across DFB in the Scrambled Stories condition Tukey's tests show that theta activity was significantly greater in DFB=4 than in DFB=0 (difference=0.047, 95% CI = [0.014,0.08], p<0.000). This was the only significant comparison for DFB in this condition (see table 11).

Across DFB in the Ordered Stories condition Tukey's tests show that theta activity was significantly greater in DFB=2 than in DFB=0 (difference=0.032, 95% CI = [0.006,0.057], p=0.002), and that this was the only significant difference between DFB in this condition (see table 11).

a



b

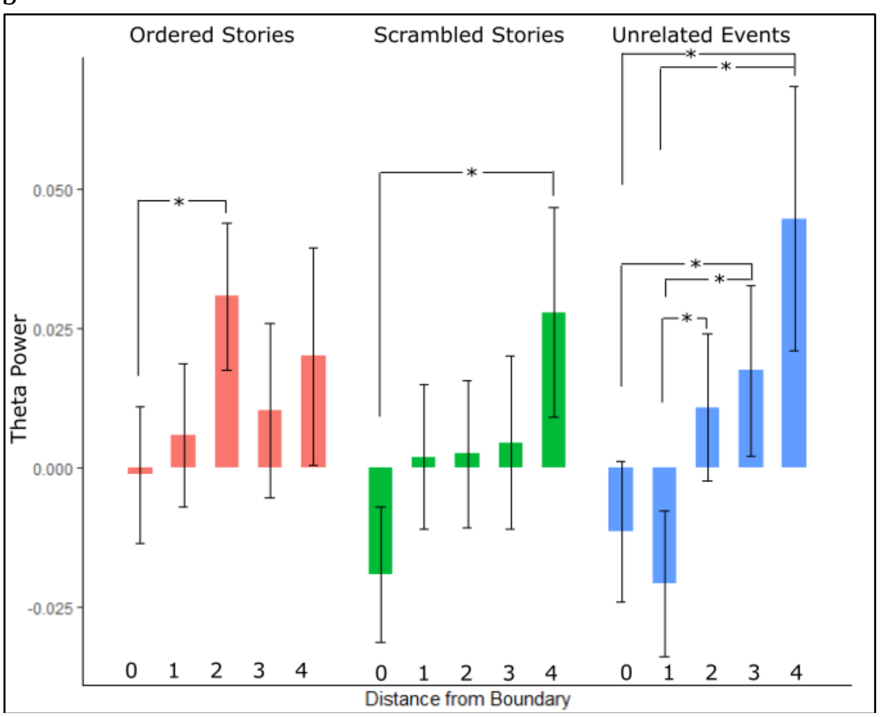


Figure 18. Theta (4-8 Hz) power (narrowband amplitude, squared) by condition and distance from boundary (DFB), across all electrode groups (LF, MF, RF, LP, MP, RP). (a) and (b) depict same data in different arrangement for ease of viewing across and within condition effects respectively.

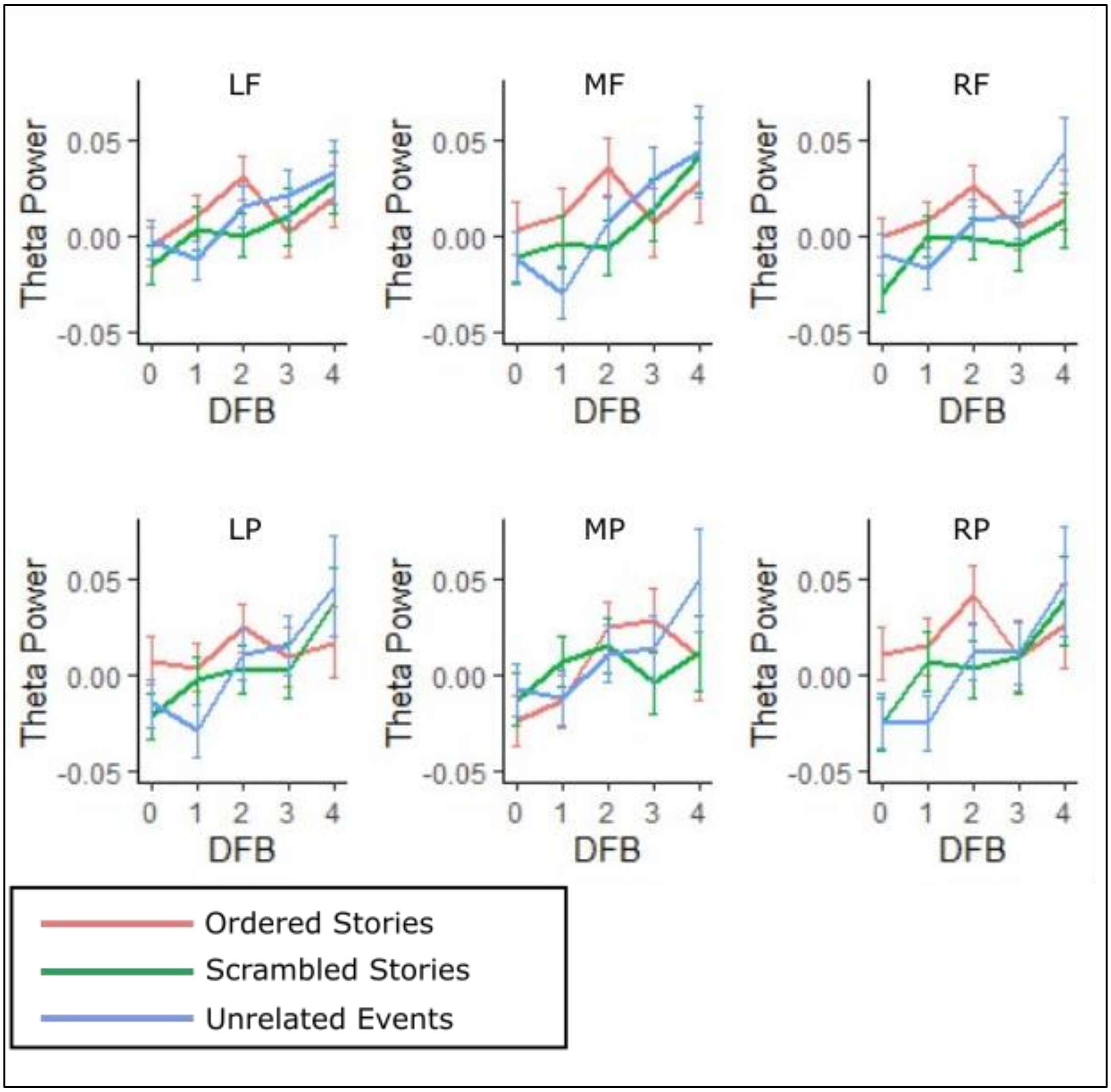


Figure 19. Theta (4-8 Hz) power (narrowband amplitude, squared) by condition and distance from boundary (DFB), at each electrode group (LF, MF, RF, LP, MP, RP).

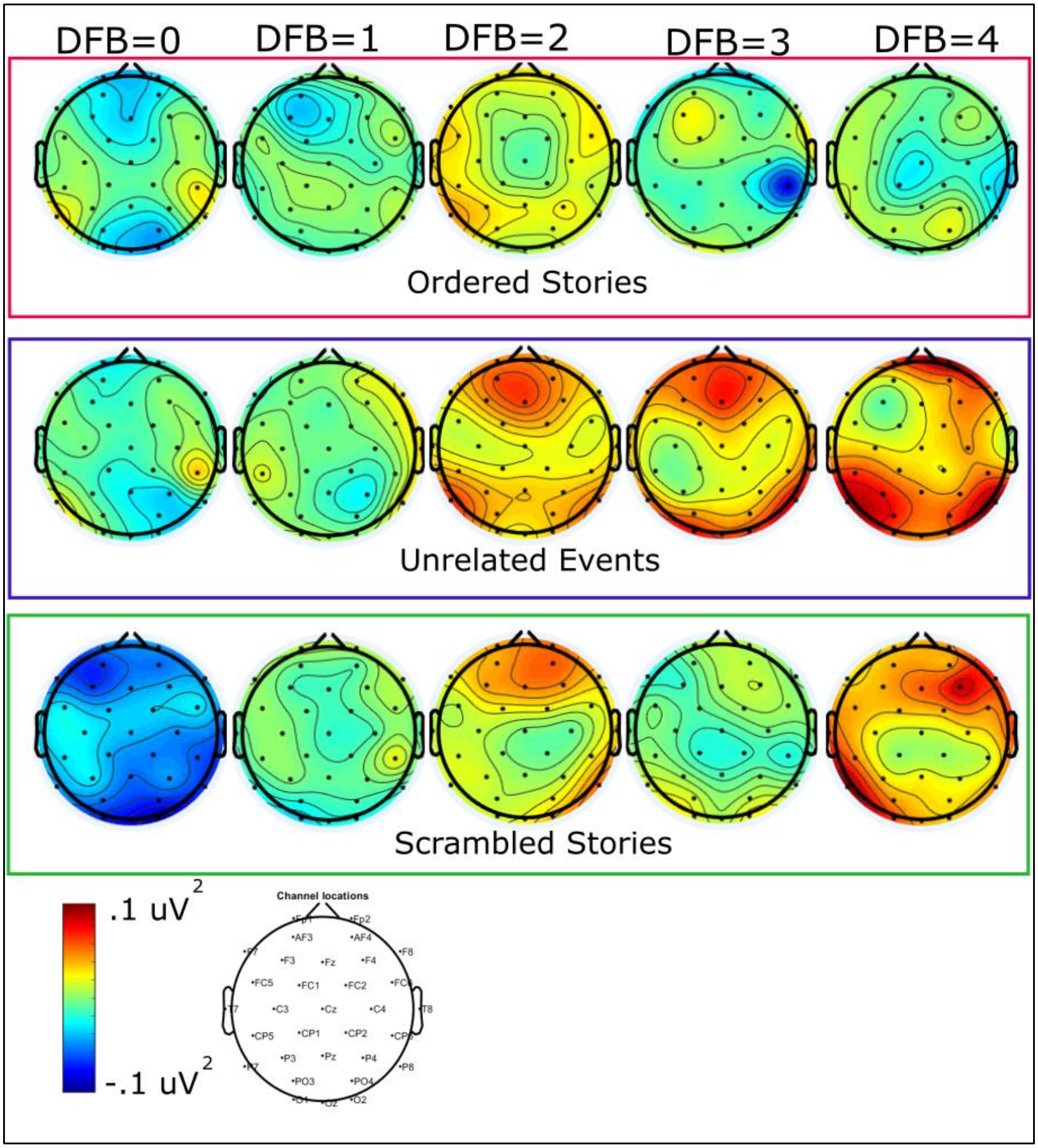


Figure 20. Theta power across the event. Topographic plots showing the theta (4-8Hz) power (narrowband amplitude, squared) averaged over sentence (0-4.5 seconds) for each of the distances from the boundary (DFB=0-4), for the separate conditions.

Condition	DFB comparison	Mean diff	lower 95% CI	upper 95% CI	p-value
Unrelated Events	1-0	0.009	-0.035	0.016	0.997
Unrelated Events	2-0	0.022	-0.004	0.048	0.185
Unrelated Events	3-0	0.029	0.002	0.056	0.025*
Unrelated Events	4-0	0.056	0.021	0.092	0.000***
Unrelated Events	2-1	0.032	0.006	0.058	0.003**
Unrelated Events	3-1	0.038	0.011	0.065	0.000***
Unrelated Events	4-1	0.066	0.030	0.101	0.000***
Unrelated Events	3-2	0.007	-0.021	0.034	1.000
Unrelated Events	4-2	0.034	-0.001	0.069	0.077
Unrelated Events	4-3	0.027	-0.009	0.064	0.405
Scrambled Stories	1-0	0.021	-0.004	0.046	0.216
Scrambled Stories	2-0	0.022	-0.004	0.047	0.187
Scrambled Stories	3-0	0.024	-0.004	0.051	0.186
Scrambled Stories	4-0	0.047	0.014	0.080	0.000***
Scrambled Stories	2-1	0.001	-0.025	0.026	1.000
Scrambled Stories	3-1	0.003	-0.025	0.030	1.000
Scrambled Stories	4-1	0.026	-0.007	0.059	0.316
Scrambled Stories	3-2	0.002	-0.025	0.030	1.000
Scrambled Stories	4-2	0.025	-0.007	0.058	0.356
Scrambled Stories	4-3	0.023	-0.011	0.058	0.600
Ordered Stories	1-0	0.007	-0.018	0.032	1.000
Ordered Stories	2-0	0.032	0.007	0.057	0.002**
Ordered Stories	3-0	0.012	-0.016	0.039	0.985
Ordered Stories	4-0	0.021	-0.012	0.054	0.673
Ordered Stories	2-1	0.025	0.000	0.050	0.055
Ordered Stories	3-1	0.005	-0.023	0.032	1.000
Ordered Stories	4-1	0.014	-0.019	0.047	0.981
Ordered Stories	3-2	-0.020	-0.048	0.007	0.428
Ordered Stories	4-2	-0.011	-0.044	0.022	0.999
Ordered Stories	4-3	0.010	-0.025	0.044	1.000

Table 11. Post-hoc theta mean difference by Distance and Condition Pairwise. Results of Tukey’s test for theta (4-8 Hz) power (narrowband amplitude, squared) at various levels of distance from boundary within conditions.

4.6 Discussion

4.6.1 Alpha power: Condition and distance from boundary

The pattern across all conditions was a general increase in alpha across the event boundary. In the two conditions which were privileged to allow event model creation (Ordered Story, and Unrelated Events), these increases began after a slight decrease in alpha power across the initial two sentences. Differences in alpha power between conditions were seen following the event boundary (DFB=1), but not immediately after (DFB=0). This difference appeared as significant reductions in the Unrelated Events relative the Story events in the second sentence of the new event (though this difference was only marginally significant in the Ordered story condition). This direction of the alpha difference (a decrease relative to the Scrambled Sentence condition), was expected following an event boundary, given the hypothesized effects of the event boundary. While a significant effect of Laterality * Posteriority * Condition * DFB was not found, visual inspection of the specific electrode clusters (Figure 15) appears to indicate that the largest decrease in alpha across the event boundary occurred in the Unrelated Events condition at DFB=1over posterior-temporal areas and more so on the right than on the left.

What was surprising was the lack of a significant decrease in alpha power in the Ordered Stories relative to the Scrambled Stories, which we would have expected if the event boundary were to trigger increased awareness in this condition (Kurby et al., 2008). While the temporal and locational cues used in our creation of event boundaries in the Story conditions were effective in our small normed group, it is possible that these boundaries were less detected in the story condition. Detection of the boundaries was not tested in the Unrelated Events stimuli, since agreement would be assumed to be 100% between subjects surveyed offline. It is thus likely that people online, missed the event boundaries in the Ordered Story more than we would expect.

Another interesting finding was the similar alpha increases in the Ordered and Scrambled Story conditions. This could be due to entrainment to the rhythms of the event models in other conditions. Further evidence supporting this view would be behavioral evidence demonstrating mnemonic grouping of Scrambled Sentences based on their presentation during the experiment (as opposed to their linear position in the story schema). Since behavioral evidence reported on the same experiment in Chapter 3 shows significantly higher accuracy in the inference recognition task for the Ordered Stories than for the Scrambled Stories, we can say that the patterns are not perfectly correlated. It is still possible that either situation or event model building did proceed in a rhythmic way, however poorly, in the Scrambled Story condition.

4.6.2 Alpha and mind wandering.

We expected the Unrelated Events and Scrambled Stories conditions would have elicited the most mind wandering or, more extremely, complete lack of awareness of the task or, *zoning out* (Smallwood, 2011), compared with the Ordered Stories. Based on accuracy in behavioral tests (see results in Chapter 3) not to mention the general sense we received from participants random self-reports between blocks we would have expected that alpha over frontal areas followed the patterns reported in Boudewyn and Carter (2018) who found increased alpha over frontal areas in periods during comprehension just preceding mind-wandering survey probes. Surprisingly, however we saw higher alpha overall in the condition in which people performed the best in both inference recognition, and (anecdotally) in recall (Ordered Stories). Further adding to the surprise is that alpha power appears greatest in the Ordered Stories at the frontal midline electrode group (see figure 15). This supports ideas of shared mechanisms between narrative comprehension and mind wandering (Sato & Mizahura, 2018, Mar, 2004).

4.6.3 Theta power: Condition and distance from boundary

The pattern of theta power across conditions mirrors the pattern of inference recognition accuracy. Ordered Stories invoked the highest theta EEG power and the highest accuracy on True/False inference recognition (see discussion in Chapter 3). This would support the findings of Sato et al., (2018), who found that left frontal theta during comprehension was correlated with recall. Visual analysis of scalp distribution shows that theta in our study is also prominent over left frontal scalp (see figure 21). One caveat to this support of Sato et al. (2018) is that that alpha activity seemed to follow the pattern of anecdotal recall (pending formal analysis, see description in Chapter 3). This is extremely intriguing and warrants further analysis.

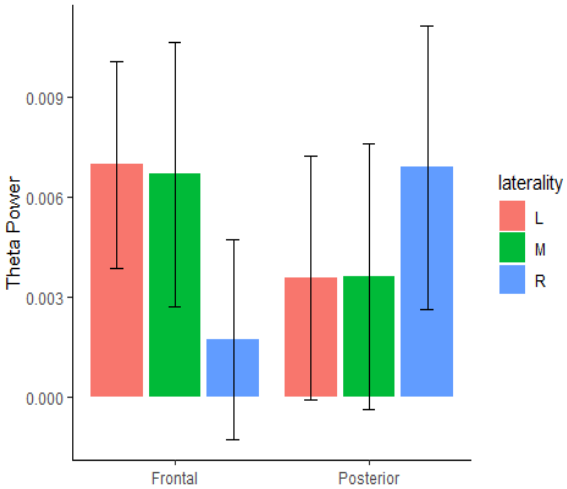


Figure 21. Scalp distribution of theta power. Distribution of Theta (4-8 Hz) power (narrowband amplitude, squared) scalp-wise, across electrode groups by Laterality (left/midline/right) x Frontality (Frontal/Posterior)

4.6.4 Theta power: Latency of effects

Theta power peaked in the Ordered Story condition during the third sentence following the boundary (DFB=2). The timing of this would put the peak within 9-13.5 seconds of an event change. The peaking trajectory of theta power in the Ordered Story condition supports our hypothesis, that theta power would increase for brief intervals following the crossing of an event boundary in conditions where event models were possible (Ordered Stories and Unrelated Events). While speculative, this is not too far from the time course of would track the cortical and hippocampal (Dubrow et al., 2016) BOLD signal after the boundary was seen.

2.6.3 Limitations and future directions

The time course of the effect within events is only approximated here, due to averaging. Temporal variance may be expected due to differences in the onset of information within a sentence that might trigger a boundary. For example, boundary sentences that begin with adverbial phrases might cue comprehenders to a boundary more quickly than other cues (see Gernsbacher, 1985). A strategy for increasing the temporal accuracy of brain activity related to event boundary onset would be to obtain normative data about when people judge a boundary to occur in time.

One limitation of our analysis could be the lack of an adaptive epoch size for the sentences. We sampled from the first 4.5 seconds of each sentence. Condition differences in sentence length were minimal, but significant (5.21 vs 4.75 seconds), with Unrelated Events being 455 ms longer than the Story Sentences. The variability in sentence length means that it is possible that some end-of-sentence effects in longer sentences were not picked up, or that epochs included onsets of other sentences in when sentences were below 4 seconds long (.5 second ISI). It is important to remember here that each frequency band was baseline corrected to the mean of that frequency band across the epoch preceding each particular event boundary (DFB=-1). Whether this significantly affected the baseline in the present analysis should be tested.

Density of electrode coverage was a limiting factor in our analysis. Increasing electrode coverage might have allowed us to better-separate patterns of alpha activity occurring in occipital (V1) vs. auditory sensory cortices (Heshyl’s gyrus). In addition, cluster-based data-driven measures could have been employed to detect regions of interest in an unbiased manner. Adaptive measures could have been employed to find maxima and minima of EEG power rather than averaging over large windows such as the sentence. Statistics such as generalized additive mixed models, or Bayesian statistics could be implemented to test for patterns which the present methods did not have the power to detect.

Choice of frequency bands could have been tailored to the specifics of each subject, as is the recommendation of Wolfgang Klimesch (see Klimesch, 1999). The ranges chosen for theta (4-8 Hz) and alpha (8.5-12.5 Hz) may have missed other sub bands of theta which have been reported using other methods, such as better oscillation detection (BOSC) (Caplan, et al., 2001). These could be explored in our dataset using other data-driven methods such as PCA to home in on frequency bands of interest.

In the future, analysis of Time-Frequency effects in different frequency bands such as beta (13-21 Hz) which have been shown to vary with grammatical processing (see Weiss & Muller, 2012, for a review) could show effects event boundary, as would more advanced time-frequency methods some of which are more sensitive to neural activity within the bands or less sensitive to artifacts. Measurements of power in a frequency band reduce the polarity (negativity and positivity relative to 0) of the energy in that band. Theories of sequence encoding attributes different functional significance to the negative and positive polarity of hippocampal and cortical theta rhythms (Herweg et al., 2020). Whether theta activity in the Ordered Stories condition was detecting a neural process specific to storing event models from WM in long-term memory, or some other process (e.g., event-model building) cannot be clarified by this analysis. Future studies might uncover to what extent conscious awareness of the boundary corresponds with the changes in EEG rhythms.

4.7 Conclusion

This is the first EEG study we are aware of that examines EEG activity specifically with respect to event boundaries in linguistic narratives. While EEG power in both alpha and theta bands accompanied successful comprehension of stories, a quantitative brain-behavior test is necessary to further examine the connection between these brain frequencies, neural mechanisms, and overt behavior.

Chapter 5: General Discussion

5.1 Summary of Findings

5.1.1 N400 results

From the N400 analysis in Chapter 3, the N400 seen was not predicted by our measure of semantic overlap. This was highly surprising given the well-established finding that the N400 decreases in amplitude when the local preceding context is semantically related vs when it is unrelated (Lau, Phillips, & Poeppel, 2008, Otten & Van Berkum, 2008, Kutas & Federmeier, 2012). This is also counter to the idea that lack of coherence, that is, lack of ability to integrate a word within the context, due to, for example, lack of background knowledge, should lead to increases in the N400 as well (St. George et al., 1994).

One explanation for the difficulty comparing these N400 results to results from other studies is that we used naturalistic texts, as opposed to short, highly-controlled materials, or *textoids* (Graesser, Millis, & Zwaan, 1997). While we did alter the stories, there was little control over the semantic relatedness of individual words, nor did our words of interest fit into context-target word pairings. It is possible that our results disagree with the literature on the N400 to the extent that previous studies utilizing textoids instigated unnatural processing strategies.

Another explanation comes from the perspective of message-level constraint effects (Federmeier & Kutas, 1999, Otten & Van Petten, 2007), as well as block-level effects of relatedness on the N400 (Lau, Holcomb, & Kuperberg, 2013). In our study, the N400 in response to content words was more negative in the Ordered Stories condition which also had the highest relatedness. This is also the condition in which constraint would seem higher. Findings of increased N400 negativity with increased constraint were found by Otten et al., (2007). Block level effects of relatedness were seen by Lau et al., (2013) to affect the N400 amplitude as well, interacting with relatedness of context and target words. While the single-word contexts used by Lau et al. (2013) make comparison to the current study difficult, block-wise effects seen in our study may be explained by a similar mechanism; the differential reliance on context from which to make predictions about upcoming words in various blocks. A predictive mechanism sensitive to coherence could explain the condition effects seen here: block-level coherence had more of an effect on LSP than the event structure (i.e., the event boundary).

5.1.2 Time-frequency results

The TF findings of alpha and theta power fluctuations at the event boundary follow a pattern of decrease or maintenance after the boundary, followed by increases within the event. We interpret these fluctuations as evidence for event model creation, maintenance, and encoding, corroborating fMRI (Speer et al., 2007) and eCog (Zheng et al., 2022) evidence of event processing in narrative, and supporting claims made by ESPM.

5.1.3 Learning effects

In all three analyses we saw effects of event boundaries in the Scrambled Sentences condition. This condition was used as a control to limit the ability to create an event model. Boundary sentences were randomized but placed at predictable intervals (every 3-7 sentence). Between these each scrambled story could be said to have contained 11-13 pseudo-events which each contained random arrangements of non-linearly ordered sentences, all from a single story. Also relevant is the order effect, whereby each Scrambled Story was presented after exposure to both one Ordered Story and one Unrelated Events passage. Explanations for effects of DFB seen in the Scrambled Story condition are (1) that there was a cascade triggered by the boundary sentence, and (2) that there was some sort of learning effect due to ordering. Both (1) and (2) are not mutually exclusive. The idea that story schema and event schema interact is a tenet of event cognition (Radvansky & Zacks, 2016). We can formulate a speculative mechanism using schema theory for what these effects represent here: it is possible that comprehenders in the Scrambled Story condition learned to utilize the boundary sentence as a placeholder for event schemas. If participants were able to understand that the boundary sentence was always the chronologically-first sentence of an event, the boundary sentences in the Scrambled Story should have cued the creation of an event model, stored offline, in long-term memory whose constituent details would then be sought out, as the comprehender furnished each event model in a non-linear manner. If such a mechanism is what our EEG effects index, then the condition differences we saw are the differences between mechanisms furnishing event models linearly and non-linearly.

5.2 Limitations and future directions

Further study of the N400 in relation to SO needs to be done to determine the relationship in each condition. The lack of significant linear models means that linear models were not the best way to measure the N400 in terms of SO. The stimuli in this study offer contexts that are much more rich than those used in studies of isolated sentences, and thus offer a unique way to measure how situation and event model driving LSP.

Future studies should also explore other language-related ERPs such as the P600 (Bornkessel-Schlesewsky & Schlesewsky, 2008). Figures 3 and 5 appear to show a posterior positivity occurring after the N400 in roughly the same area. It is likely a P600 component and it appears more positive in the Unrelated Events condition. While it is purely speculation at this point, it is possible that this positivity is due to some mechanism utilized by the comprehender to correct the event model in cases of difficult, but not impossible repair as would be most encountered in the Unrelated events condition.

Another limitation of the present study was the use of a single region of interest for each subject, based on grand averages. Other means of individual regions of interest may be a more precise means with which to measure the N400 for each subject. Another limitation was the use of the adaptive mean which, in comparison to the grand average waveforms, appear to have included large amounts of noise. In this instance, the use of adaptive means was justified by the presence of variability in the calculation of the onsets of the auditory words by the forced aligner and audio quality.

Future studies should compare our results directly to findings from fMRI. Despite encouraging connections between BOLD and EEG (Scheeringa et al., 2008) connecting the effects in the EEG to effects with other brain imaging techniques such as fMRI is difficult to do in an unbiased way. It is possible that boundary-evoked alpha and theta rhythms measured in our analysis are related in some way with BOLD activations in hippocampus (Dubrow et al., 2016), posterior parietal cortex (Speer et al., 2007, Yarkoni et al., 2011). Unfortunately, localization of EEG sources with only 32 electrodes is difficult. More sensitive equipment and methods would increase our ability to detect other boundary effects in the EEG data. From the patterns of EEG power over the time course of events, we can see some neural effects of distance from the event boundary. Increased density of coverage might better allow us to localize the sources of posterior EEG activity to either the parietal or occipital lobes, which would greatly influence the interpretation.

There is much work to be done here to improve upon the methods here. Future work should test whether these boundaries are correlated with comprehension. If so, it may be that boundary evoked fluctuations in EEG activity would be more indicative of successful encoding than raw power sampled from other points in the text. This would be the case according to event segmentation theory. This work should also embolden other researchers to use EEG for studying comprehension of natural texts.

The present analysis is limited by the strict interpretation of alpha power as an index of WM and theta power as LTM. Many studies have tied theta to WM as well (Hsieh & Ranganath, 2014). Current models of memory have difficulty explaining phenomena surrounding comprehension, as some have pointed out (Ericcson & Kintsch, 1995, Baddeley et al., 2002). A model of memory which better distinguishes the formation, degradation, and retention of episodic memory, as well as the transformation of episodic traces via consolidation is required to better understand comprehension of narrative. One hopeful area of research would be testing the analogues between cortico-hippocampal interaction during ambulatory movement and story comprehension. The intuitive connection between the location updating effect (Pettijohn et al., 2016) and ESPM for narrative comprehension is limited by being indoors. More elaborate 3D environments and wearable electrode arrays similar to those used by (Liang, Starrett, &

Eckstrom, 2018) would open doors for improved functional interpretations of the EEG rhythms studied here (Liang, Starrett, & Eckstrom, 2018).

Lastly, the translation of these findings into advice to writers or storytellers is limited by the story genre. At this moment it would be said that novels and longer narratives certainly differ in structure from the folktale genre used here. How they differ remains interesting to speculate on. The relationship between the size of events, event models and situation models and their memorability will only be better understood with increasingly realistic ESPM.

5.3 Conclusion

The goal of this thesis was to examine a cognitive approach to comprehension using EEG. The more distant or *global* hope is that someone will eventually use insights here to make comprehension easier and better. What can be added beyond the wisdom of writers, directors, avid videogame critics, and narratologists is the possibility that we can test the outcomes of comprehension in a way that not-only does not burden the comprehender but also adds a new perspective. While the cognitive perspective is not new (see Mar, 2004), the utility of EEG to provide this perspective on comprehension (how comprehension *is going* at various points in a story or overall) is what could be improved. Thus, the novelty of the work presented here is to inspect the hypotheses of EST regarding how memory encoding occurs during narrative comprehension.

It is important to be reminded that the various formats in which we convey information to one another are technologies, whether it be stories, jokes, expressions of love, mantras, etc. Like any tool, they have changed and been improved over the years, and will continue to do so into the future. The user experience and versatility of a story will eventually be improved as well. Other tools will be created to allow usability for people with different needs. The cognitive perspective of narrative comprehension holds the most hope expanding practices to guide those who may find themselves lost in a story but still desire to follow along.

References

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of experimental psychology: Learning, memory, and cognition*, *19*(5), 1061.

Andersen, S., & Slator, B. M. (1990). Requiem for a theory: the ‘story grammar’ story. *Journal of Experimental & Theoretical Artificial Intelligence*, *2*(3), 253-275.

Anderson, A., Garrod, S. C., & Sanford, A. J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text. *Quarterly Journal of Experimental Psychology*, *35*(3), 427-440.

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. Trends in cognitive sciences, 4(11), 417-423.

Baddeley, A., & Wilson, B. A. (2002). Prose recall and amnesia: Implications for the structure of working memory. *Neuropsychologia*, *40*(10), 1737-1743.

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, *95*(3), 709-721.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child development*, *72*(3), 708-717.

Bartlett, F. C., (1995). *Remembering: A study in experimental and social psychology*. Cambridge university press.

Bastiaansen, M., Magyari, L., & Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of cognitive neuroscience*, *22*(7), 1333-1347.

Bastiaansen, M. C., Van Berkum, J. J., & Hagoort, P. (2002). Event-related theta power increases in the human EEG during online sentence processing. *Neuroscience letters*, *323*(1), 13-16.

Bauer, P. J., & Varga, N. L. (2017). Similarity and deviation in event segmentation and memory integration: Commentary on Richmond, Gold, & Zacks. *Journal of Applied Research in Memory and Cognition*, *6*(2), 124-128.

Ben-Yakov, A., Eshel, N., & Dudai, Y. (2013). Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. *Journal of Experimental Psychology: General*, *142*(4), 1255.

Berger, H. (1929). Über das elektroenkephalogramm des menschen. *Archiv für psychiatrie und nervenkrankheiten*, *87*(1), 527-570.

Black, J. B., & Bower, G. H. (1980). Story understanding as problem-solving. *Poetics*, *9*(1-3), 223-250.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain research reviews*, 59(1), 55-73.

Boudewyn, M. A., & Carter, C. S. (2018). I must have missed that: Alpha-band oscillations track attention to spoken language. *Neuropsychologia*, *117*, 148-155.

Boudewyn, M. A., Long, D. L., Traxler, M. J., Lesh, T. A., Dave, S., Mangun, G. R., Carter, C., & Swaab, T. Y. (2015). Sensitivity to referential ambiguity in discourse: the role of attention, working memory, and verbal ability. *Journal of cognitive neuroscience*, *27*(12), 2309-2323.

Bowman, A. D., Griffis, J. C., Visscher, K. M., Dobbins, A. C., Gawne, T. J., DiFrancesco, M. W., & Szaflarski, J. P. (2017). Relationship between alpha rhythm and the default mode network: an EEG-fMRI study. *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, *34*(6), 527.

Braboszcz, C., & Delorme, A. (2011). Lost in thoughts: neural markers of low alertness during mind wandering. Neuroimage, 54(4), 3040-3047.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive psychology*, *2*(4), 331-350.

Brewer, W. F., & Lichtenstein, E. H. (1980). Event schemas, story schemas, and story grammars. *Center for the Study of Reading Technical Report; no. 197*.

Brodbeck, C., Presacco, A., & Simon, J. Z. (2018). Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. *NeuroImage*, *172*, 162-174.

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *Neuroreport*, *18*(17), 1851-1854.

Buzsáki, G. (1996). The hippocampo-neocortical dialogue. *Cerebral cortex*, *6*(2), 81-92.

Buzsáki, G. (2006). *Rhythms of the brain.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195301069.001.0001

Cannizzaro, M. S., & Coelho, C. A. (2013). Analysis of narrative discourse structure as an ecologically relevant measure of executive function in adults. *Journal of psycholinguistic research*, *42*(6), 527-549.

Caplan, J. B., Madsen, J. R., Raghavachari, S., & Kahana, M. J. (2001). Distinct patterns of brain oscillations underlie two basic parameters of human maze learning. *Journal of neurophysiology*, 86(1), 368-380

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

Cohn, N. (2013). *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images*. A&C Black.

Cohn, N. (2020). Your brain on comics: A cognitive model of visual narrative comprehension. Topics in Cognitive Science, 12 (1), 352–386.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407.

Compton, R. J., Gearinger, D., & Wild, H. (2019). The wandering mind oscillates: EEG alpha power is enhanced during moments of mind-wandering. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(5), 1184-1191.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13(1), 21-58.

Davachi, L., & Wagner, A. D. (2002). Hippocampal contributions to episodic encoding: insights from relational and item-based learning. *Journal of neurophysiology*, *88*(2), 982-990.

Davies, Mark. (2008-) Word frequency data from The Corpus of Contemporary American English (COCA). Data available online at https://www.wordfrequency.info.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: analyses and review. *Journal of experimental psychology: General*, *140*(4), 552.

Ditman, T., & Kuperberg, G. R. (2007). The time course of building discourse coherence in schizophrenia: an ERP investigation. *Psychophysiology*, *44*(6), 991-1001.

Dowty, D. R. (1986). The effects of aspectual class on the temporal structure of discourse: semantics or pragmatics?. *Linguistics and philosophy*, 37-61.

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.

Ettinger, A., Feldman, N., Resnik, P., & Phillips, C. (2016). Modeling N400 amplitude using vector space models of word representation. In *CogSci*.

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory?. *Psychological science*, *22*(2), 243-252.

Eisenberg, M. L., & Zacks, J. M. (2016). Ambient and focal visual processing of naturalistic activity. *Journal of Vision*, *16*(2), 5-5.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological review*, *102*(2), 211.

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.

Ferstl, E.C. (2007). The functional neuroanatomy of text comprehension: What's the story so far? In F. Schmalhofer & C.A. Perfetti (Eds.), Higher level language processes in the brain: Inference and comprehension processes (pp. 53–102). Mahwah, NJ: Erlbaum.

Frisch, S., Kotz, S. A., Von Cramon, D. Y., & Friederici, A. D. (2003). Why the P600 is not just a P300: The role of the basal ganglia. Clinical Neurophysiology, 114(2), 336-340.

Flores, S., Bailey, H. R., Eisenberg, M. L., & Zacks, J. M. (2017). Event segmentation improves event memory up to one month later. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(8), 1183.

Foxe, J. J., Simpson, G. V., & Ahlfors, S. P. (1998). Parieto-occipital ~10Hz activity reflects anticipatory state of visual attention mechanisms. *Neuroreport*, *9*(17), 3929-3933.

Fyshe, A. (2020). Studying language in context using the temporal generalization method. *Philosophical Transactions of the Royal Society B*, *375*(1791), 20180531.

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive psychology*, *17*(3), 324-363.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual review of psychology*, 48(1), 163-189.

Grafman, J., Sirigu, A., Spector, L., & Hendler, J. (1993). Damage to the prefrontal cortex leads to decomposition of structured event complexes. *The Journal of Head Trauma Rehabilitation*.

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. (2005). On Broca, brain, and binding: a new framework. *Trends in cognitive sciences*, *9*(9), 416-423.

Hagoort, P. (2016). MUC (Memory, Unification, Control): A model on the neurobiology of language beyond single word processing. In *Neurobiology of language* (pp. 339-347). Academic Press.

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences, 104*(5), 1726-1731.

Herweg, N. A., Solomon, E. A., & Kahana, M. J. (2020). Theta oscillations in human memory. *Trends in cognitive sciences, 24*(3), 208-227.

Hsieh, L. T., & Ranganath, C. (2014). Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. *Neuroimage, 85*, 721-729.

Huettner, M. I., Rosenthal, B. L., & Hynd, G. W. (1989). Regional cerebral blood flow (rCBF) in normal readers: bilateral activation with narrative text. *Archives of Clinical Neuropsychology*, *4*(1), 71-78.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Frontiers in human neuroscience*, 186.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (No. 6). Harvard University Press.

Kelter, S., Kaup, B., & Claus, B. (2004). Representing a described sequence of events: a dynamic view of narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 451.

Keven, N., Kurczek, J., Rosenbaum, R. S., & Craver, C. F. (2018). Narrative construction is intact in episodic amnesia. *Neuropsychologia*, *110*, 104-112.

Khader, P. H., Jost, K., Ranganath, C., & Rösler, F. (2010). Theta and alpha oscillations during working-memory maintenance predict successful long-term memory encoding. *Neuroscience letters*, *468*(3), 339-343.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. Journal of memory and language, 52(2), 205-225.

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. Trends in cognitive sciences, 18(4), 203-210.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological review*, *95*(2), 163.

Kintsch, W., Mandel, T. S., & Kozminsky, E. (1977). Summarizing scrambled stories. *Memory & Cognition*, *5*(5), 547-552.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and language*, *29*(2), 133-159.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, *29*(2-3), 169-195.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in cognitive sciences*, *16*(12), 606-617.

Klimesch, W., Doppelmayr, M., Schimke, H., & Ripper, B. (1997). Theta synchronization and alpha desynchronization in a memory task. *Psychophysiology*, *34*(2), 169-176.

Klimesch, W., Doppelmayr, M., Schwaiger, J., Auinger, P., & Winkler, T. H. (1999). 'Paradoxical' alpha synchronization in a memory task. *Cognitive Brain Research*, *7*(4), 493-501.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in cognitive sciences*, *12*(2), 72-79.

Kurby, C. A., & Zacks, J. M. (2012). Starting from scratch and building brick by brick in comprehension. *Memory & cognition*, *40*(5), 812-826.

Kurby, C. A., & Zacks, J. M. (2013). The activation of modality-specific representations during discourse processing. *Brain and language*, *126*(3), 338-349.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621-647.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161-163.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, *104*(2), 211.

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

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.

Lega, B. C., Jacobs, J., & Kahana, M. (2012). Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus*, 22(4), 748-761.

Liang, M., Starrett, M. J., & Ekstrom, A. D. (2018). Dissociation of frontal-midline delta-theta and posterior alpha oscillations: A mobile EEG study. *Psychophysiology*, 55(9), e13090.

Lichtenstein, E. H., & Brewer, W. F. (1980). Memory for goal-directed events. *Cognitive Psychology*, *12*(3), 412-445.

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

Nyhus, E., & Curran, T. (2010). Functional role of gamma and theta oscillations in episodic memory. *Neuroscience & Biobehavioral Reviews*, *34*(7), 1023-1035.

Magliano, J., Kopp, K., McNerney, M. W., Radvansky, G. A., & Zacks, J. M. (2012). Aging and perceived event structure as a function of modality. *Aging, Neuropsychology, and Cognition*, *19*(1-2), 264-282.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive psychology*, *9*(1), 111-151.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996*).* *Psychological Review*. 109(1), 35–54.

Mar, R. A. (2004). The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia,* 42(10), 1414-1434.

Mason, R.A. & Just, M.A. (2004). How the brain processes causal inferences in text. *Psychological Science*, 15(1), 1–7.

Maurer, U., Brem, S., Liechti, M., Maurizio, S., Michels, L., & Brandeis, D. (2015). Frontal midline theta reflects individual task performance in a working memory task. Brain topography, 28, 127-134.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017, August). Montreal forced aligner: Trainable text-speech alignment using kaldi. In *Interspeech* (Vol. 2017, pp. 498-502).

McGhee, H. C. (2021). The sum of us: what racism costs everyone and how we can prosper together (First edition.). One World.

McKay, A., Miller, J. (., Apatow, J., Ferrell, W., Bibb, L., Baron Cohen, S., Reilly, J. C., Cole, G., & Duncan, M. C. (2006). *Talladega nights: the ballad of Ricky Bobby.*

Widescreen unrated edition. Culver City, California, Sony Pictures Home Entertainment.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*(3), 440–466. https://doi.org/10.1037/0033-295X.99.3.440

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Psychology Press.

Meeuwissen, E. B., Takashima, A., Fernández, G., & Jensen, O. (2011). Increase in posterior alpha activity during rehearsal predicts successful long-term memory formation of word sequences. *Human brain mapping*, *32*(12), 2045-2053.

Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv1301.3781.

Miller, R. (1989). Cortico-hippocampal interplay: Self-organizing phase-locked loops for indexing memory. *Psychobiology*, *17*(2), 115-128.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current opinion in neurobiology*, *7*(2), 217-227.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of experimental psychology: general*, *106*(3), 226.

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, 12(5), 436-450.

Newtson, D., Hairfield, J., Bloomingdale, J., & Cutino, S. (1987). The structure of action and interaction. *Social Cognition*, 5(3), 191-237.

Omanson, R. C. (1982). An analysis of narratives: Identifying central, supportive, and distracting content. *Discourse processes*, *5*(3-4), 195-224.

Otten, M., & Van Berkum, J. J. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research*, *1153*, 166-177.

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532-1543.

Perfetti, C.A. (1998). *The limits of co-occurrence: Tools and theories in language research*, Discourse Processes, 25(2-3), 363-377, DOI: 10.1080/01638539809545033

Perfetti, C. A., & Frishkoff, G. A. (2008). The neural bases of text and discourse processing. *Handbook of the neuroscience of language*, *2*, 165-174.

Perfetti, C., & Helder, A. (2022). Progress in reading science: Word identification, comprehension, and universal perspectives. *The Science of Reading: A Handbook* (Eds Snowling, M.J. and Hulme, C.E. Blackwell Publishing (5-35).

Pettijohn, K. A., & Radvansky, G. A. (2016). Walking through doorways causes forgetting: Event structure or updating disruption?. *Quarterly Journal of Experimental Psychology*, *69*(11), 2119-2129.

Poulsen, D., Kintsch, E., Kintsch, W., & Premack, D. (1979). Children's comprehension and memory for stories. *Journal of experimental child psychology*, *28*(3), 379-403.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, 98*(2), 676-682.

Radvansky, G. A., Tamplin, A. K., Armendarez, J., & Thompson, A. N. (2014). Different kinds of causality in event cognition. *Discourse Processes*, *51*(7), 601-618.

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current opinion in behavioral sciences*, *17*, 133-140.

Rajagovindan, R., Ding, M., 2010. From prestimulus alpha oscillation to visual-evoked response: an inverted-U function and its attentional modulation. J. Cogn. Neurosci. 23, 1379–1394.

Rehurek, R., & Sojka, P. (2011). Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic, 3*(2).

Richmond, L. L., Gold, D. A., & Zacks, J. M. (2017). Event perception: Translations and applications. *Journal of Applied Research in Memory and Cognition*, *6*(2), 111-120.

Rinck, M., & Bower, G. H. (2000). Temporal and spatial distance in situation models. *Memory & Cognition*, *28*(8), 1310-1320.

Röhm, D., Klimesch, W., Haider, H., & Doppelmayr, M. (2001). The role of theta and alpha oscillations for language comprehension in the human electroencephalogram. *Neuroscience letters*, *310*(2-3), 137-140.

Rumelhart, D. E., (1977). Understanding and summarizing brief stories. In Laberge, D., & Samuels, J. (Eds.) Basic processes in reading: Perception and comprehension (pp. 265-303), Hillsdale. *NJ Erlbaum*.

Rumelhart, D. E.(1980) Schemata: The Building Blocks of Cognition. In Spiro, R., Bruce, B. & Brewer, W. (Eds.), *Theoretical issues in reading comprehension : perspectives from cognitive psychology, linguistics, artificial intelligence, and education* (pp.33-58). Hillsdale, N.J: L. Erlbaum Associates.

Saroyan W. & Freeman D. (1940). My Name is Aram. Harcourt Brace.

Sassenhagen, J. (2019). How to analyse electrophysiological responses to naturalistic language with time-resolved multiple regression. *Language, Cognition and Neuroscience, 34*(4), 474-490.

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, 137, 29-39.

Sato, N., & Mizuhara, H. (2018). Successful encoding during natural reading is associated with fixation-related potentials and large-scale network deactivation. *eneuro*, *5*(5).

Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., Gruber, W.R., Birbaumer, N., 2005. A shift of visual spatial attention is selectively associated with human EEG alpha activity. Eur. J. Neurosci. 22, 2917–2926.

Schacter, D. L., & Addis, D. R. (2007). The ghosts of past and future. *Nature*, *445*(7123), 27-27.

Scheeringa, R., Petersson, K. M., Oostenveld, R., Norris, D. G., Hagoort, P., & Bastiaansen, M. C. (2009). Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. *Neuroimage*, *44*(3), 1224-1238.

Scheeringa, R., Fries, P., Petersson, K. M., Oostenveld, R., Grothe, I., Norris, D. G., Hagoort, P., & Bastiaansen, M. C. (2011). Neuronal dynamics underlying high- and low-frequency EEG oscillations contribute independently to the human BOLD signal. *Neuron*, *69*(3), 572-583.

Sharp, R. (2010). Electrophysiological correlates of event segmentation : how does the human mind process ongoing activity? Doctoral dissertation, University of Edinburgh.

Smallwood, J., Beach, E., Schooler, J. W., & Handy, T. C. (2008). Going AWOL in the brain: Mind wandering reduces cortical analysis of external events. *Journal of cognitive neuroscience*, *20*(3), 458-469.

Smallwood, J. (2011). Mind-wandering while reading: Attentional decoupling, mindless reading and the cascade model of inattention. *Language and Linguistics Compass*, *5*(2), 63-77.

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of memory and language*, 53(1), 125-140.

Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological science*, 18(5), 449-455.

Stafura, J. Z., & Perfetti, C. A. (2014). Word-to-text integration: Message level and lexical level influences in ERPs. *Neuropsychologia*, *64*, 41-53.

St. George, M. S., & Mannes, S. Hoffman, J.E. (1994). Global semantic expectancy and language comprehension. *Journal of cognitive neuroscience, 6*(1), 70-83.

Swallow, K. M., Barch, D. M., Head, D., Maley, C. J., Holder, D., & Zacks, J. M. (2011). Changes in events alter how people remember recent information. *Journal of cognitive neuroscience, 23*(5), 1052-1064.

Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive psychology*, *9*(1), 77-110.

Thorndyke, P. W., & Yekovich, F. R. (1980). A critique of schema-based theories of human story memory. *Poetics*, *9*(1-3), 23-49.

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). α-Band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *Journal of Neuroscience, 26*(37), 9494-9502.

Tingley D, Yamamoto T, Hirose K, Keele L, Imai K (2014). "mediation: R Package for Causal Mediation Analysis." *Journal of Statistical Software*, **59**(5), 1–38. http://www.jstatsoft.org/v59/i05/.

Tompkins, C. A., Lehman-Blake, M. T., Baumgaertner, A., & Fassbinder, W. (2001). Mechanisms of discourse comprehension impairment after right hemisphere brain damage: Suppression in inferential ambiguity resolution. *Journal of Speech Language and Hearing Research*, *44*(2), 400-415.

Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and language*, *24*(5), 595-611.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual review of psychology, 53*(1), 1-25.

Van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. *The construction of mental representations during reading*, 71-98.

Van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: Passive processes, reader-initiated processes, and evolving mental representations. Discourse Processes, 54(5-6), 360-372.

Van Dijk, T. A., & Kintsch, W. (1983). Strategies of Discourse Comprehension. New York: Academic Press.

Van Dijk, T. A. (1976). Narrative macro-structures. *PTL: A journal for descriptive poetics and theory of literature*, *1*(1976), 547-568.

Varela, F., Lachaux, J. P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2(4), 229-239.

Weiss, S., & Mueller, H. M. (2012). “Too many betas do not spoil the broth”: the role of beta brain oscillations in language processing. *Frontiers in psychology*, *3*, 201.

Wehbe, L., Murphy, B., Talukdar, P., Fyshe, A., Ramdas, A., & Mitchell, T. (2014). Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses. *PloS one*, *9*(11), e112575.

Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, *15*(2), 70-73.

Winkler, I., Haufe, S., & Tangermann, M. (2011). Automatic classification of artifactual ICA-components for artifact removal in EEG signals. *Behavioral and brain functions*, *7*, 1-15

Yoo, H. B., Umbach, G., & Lega, B. (2022). Episodic boundary cells in human medial temporal lobe during the free recall task. *Hippocampus*, 32(7), 481-487.

Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., & Deouell, L. Y. (2008). Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron*, *58*(3), 429-441.

Zacks, J. M., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006). Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychology and aging*, *21*(3), 466.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological bulletin*, *127*(1), 3.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological bulletin*, *133*(2), 273.

Zalla, T., Pradat-Diehl, P., & Sirigu, A. (2003). Perception of action boundaries in patients with frontal lobe damage. *Neuropsychologia*, *41*(12), 1619-1627.

Zalla, T., Joyce, C., Szöke, A., Schürhoff, F., Pillon, B., Komano, O., Perez-Dias, F., Bellivier, F,. Alter, C., Dubois, B., Ruoillon, F., Houde, O., & Leboyer, M. (2004). Executive dysfunctions as potential markers of familial vulnerability to bipolar disorder and schizophrenia. *Psychiatry research*, *121*(3), 207-217.

Zheng, J., Schjetnan, A. G., Yebra, M., Gomes, B. A., Mosher, C. P., Kalia, S. K., ... & Rutishauser, U. (2022). Neurons detect cognitive boundaries to structure episodic memories in humans. *Nature neuroscience*, 25(3), 358-368.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, memory, and cognition*, *22*(5), 1196.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of experimental psychology: Learning, memory, and cognition*, *21*(2), 386.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, *123*(2), 162.

Zuo, X., Honey, C.J., Barense, M.D., Crombie, D., Norman, K.A., Hasson, U., Chen, J. (2020). Temporal integration of narrative information in a hippocampal amnesic patient. *NeuroImage* 213, 116658

Appendix A

Condition	Example sentences	SO	DFB
Scrambled Story	Some time later the brothers spotted another elk which was close.	NA	0
	The man didn't accept any elk meat when the brother offered.	.13	1
	The grandfather killed the man and returned with his body.	.42	2
	Younger brother enjoyed himself, signing songs and eating noisily	.09	3
	Once home, the grandfather gave the older brother specific instructions	.51	0
	The elk was alerted and crossed a meadow and disappeared.	.05	1
	Grandfather and older brother found tracks of the man leading to a lake.	.14	2
	The next day the younger brother emerged from the house.	.23	0
	Neither did the man utter a single word to the younger brother.	.15	1
Unrelated Events	One day, the grandfather sent the brothers out to hunt.	NA	0
	The brothers found hoof prints and began tracking an elk.	.45	1
	Searching, the older brother spotted the elk and took aim.	.65	2
	But the younger brother wanted to shoot and shouted loudly.	.54	3
	The elk was alerted and crossed a meadow and disappeared.	.23	4
	The next day, the animals found Baboon defeated and bloody.	.09	0
	Lion was angry with Baboon for being so easily tricked.	.55	1
	Lion punished Baboon for his inability to capture Jackal.	.65	2
Events in Stories	One day, the grandfather sent the brothers out to hunt.	NA	0
	The brothers found hoof prints and began tracking an elk.	.45	1
	Searching, the older brother spotted the elk and took aim.	.65	2
	But the younger brother wanted to shoot and shouted loudly.	.54	3
	The elk was alerted and crossed a meadow and disappeared.	.23	4
	Some time later the brothers spotted another elk which was close.	.50	0
	The younger brother begged for a chance to shoot it.	.21	1

Table A1. Stimulus Conditions (yellow sentences indicate event boundary likely to occur). SO=Semantic overlap, DFB=Distance from Boundary

Appendix B

Table B1. Rubric for assessing recall

sentence	exact	proposition	order
1	One day, the grandfather sent the brothers out to hunt.	Grandfather sent brothers out	1
		The brothers went out to hunt	2
2	The brothers found hoof prints and began tracking an elk,	The brothers found hoof prints	1
		The brothers began tracking an elk	2

Table B1. Rubric for assessing recall. Propositions in story correspond to actions described by each verb.

Appendix C

Sentence from story	Paraphrase	Incorrect Paraphrase	Distance from Boundary (DFB)
One day, the grandfather sent the brothers out to hunt.	Grandfather asked the brothers to hunt	Grandfather asked the brothers to go fishing.	0
The brothers found hoof prints and began tracking an elk.	The brothers followed the elk's hoofprints	The brothers followed the elk's mating call.	1
Searching, the older brother spotted the elk and took aim.	The older brother found the elk in the woods.	The younger brother found the elk in the woods.	2
But the younger brother wanted to shoot and shouted loudly.	The younger brother was not cautious around the elk.	The younger brother did not want to shoot and remained quiet.	3
The elk was alerted and crossed a meadow and disappeared.	The elk heard the brothers and escaped	The elk didn't hear the brothers and remained still.	4
The next day, the animals found Baboon defeated and bloody.	The animals found out that Baboon had been tricked.	The animals found out that Baboon had been victorious..	0

Table C1. Example recognition probes. Example recognition probes following an Unrelated Events block, alongside corresponding sections of text.

Appendix D

Table D1. Sample stimuli with participant responses

s entID	sentence	s1	s2	s3	s4	s5	s6
1	A grandfather lived in the woods with his two grandsons.						
2	The older brother was often patient with the younger one.						
3	The family's life in the forest was lovely, simple and honest						
4	One day, the grandfather sent the brothers out to hunt.						
5	The brothers found hoof prints and began tracking an elk.						
6	Searching, the older brother spotted the elk and took aim.						
7	But the younger wanted to shoot and so shouted loudly.						
8	The elk was alerted, crossed a meadow and disappeared.						
1 0	Some time later the brothers spotted another elk which was close.						
1 1	The younger brother begged for a chance to shoot it.						
1 2	This time, the older brother allowed the younger to shoot.						
1	The younger brother complained about a tree blocking the						

3	**shot.**						
14	**The elk became heard the complaint and charged at the brothers**						
15	**The older brother quickly shot the elk at the last minute.**						
16	**With the meat packed, older brother prepared to go home.**						
17	**Younger brother begged the older to camp out for the night.**						
18	**The older brother was angry at the request, but agreed.**						
19	**The older brother made a camp and tried to sleep.**						
	Through the night younger brother stayed up eating elk meat.						

Table D1. Sample stimuli with participant responses. Participants were instructed to mark and ‘x’ next to sentences describing a new event in a story.

Appendix E

(i)

Defining Events

As sentences in a story follow a sensible pattern, certain principles have been derived about how people assume each sentence is related to the one before it within a story. For example, a principle of temporal contiguity assumes that neighboring sentences describe actions which happen immediately after each other (Dowty, 1986). For example consider the following sentences:

(2) (a) Bill was walking home one night. He found a bag of money on the side of the road.

Without explicitly stating it, we assume that both sentences in (1) are part of the same event; Bill discovered the bag of money while he was walking home. We also assume that Bill is the referent of "He" in the second sentence such that it was Bill who found the money. However it is not always safe to assume that the actions described by each sentence are related to the ones that came before them.

(ii)

Event boundaries

To get an idea of how subjects might parse a more natural written story into fine-grained events, Speer, Zacks, and Reynolds (2007) asked subjects to read excerpts from a novel about the daily life of a school-aged child, and to mark the smallest units of change which they thought meaningful. According to their participants' responses, finest-grained events were 3 clauses long on average. The clause length refers to discourse time, or a measure of the text.

Assumptions about the ratio of text-to-events is difficult to estimate in other forms of writing. Consider James Joyce's Ulysses, or Lewis Caroll's The Adventures of Alice in Wonderland, in which a variety of events are described in such detail that it might take a much longer time in the discourse (multiple sentences of text) for a character to complete a simple action. As such, eliciting placement of boundaries by surveying human participants appears to be necessary rather than adhering to rules based on an interval of text length.

(iii)

Story schemas and other structures

These story schemas contain labels for actions, such as "CAUSE", and "OUTCOME" (Rumelhart, 1977), or " Beginning" Development", "Outcome", which are positioned in a hierarchical grammar below categories such as "Episode" (Mandler & Johnson, 1977). This is similar to the common literary notions of *rising action,* and *climax,* and *resolution* used for literary analysis, however they are slightly more complex, with sub-episodes separable with them.

Other structural theories suggest that causal links between actions or states allow hierarchies based on their position in the hierarchy of goals and subgoals (Black & Bower, 1980), or that some actions have more causal links attaching them to other actions in the story (Omanson, 1980, Trabasso & Sperry, 1985). These structural theories highlight the importance of life experience when trying to comprehend and have also been shown to predict behavior via recall of stories (Omanson, 1980, Trabasso et al., 1985, Black & Bower, 1980).

(iv)

What it takes to process event schemas

While some theories suggest that event schemas rely on perceptual simulation (Zwaan & Radvansky, 2018), the ability of schemas to affect comprehension may be achievable through associations (Anderson, Garrod & Sanford, 1983). For example, once a sentence invokes the action "At the cinema", an event schema would generate a set of possible actions associated with the goal: standing in line, buying a ticket, avoiding the concession stand, finding the theater. Schemas are thought to help determine when an event will likely end, or whether a sentence refers to a later time (Anderson, et al., 1983). For illustration, consider the example in 3, adapted from an example used by Anderson et al., (1983).

(3) (a) Jenny found the film boring. Ten minutes later she was fast asleep.

(b) Jenny found the film boring. Seven hours later she was fast asleep.

Schematic knowledge of how long a movie takes place allows (3a) to be interpreted as Jenny fell asleep in the theater, while (3b) could be interpreted as Jenny fell asleep somewhere else. The interpretation of (3b) might cause someone to mark an event boundary between sentences in (3b) as opposed to (3a). Invoking a new schema would be a function that utilizes neural resources. Grafman et al., (2013) proposed that such knowledge of events was stored in what they called Structured Event Complexes, which resided in the frontal lobe. The utility of these complexes was seen in their absence, when patients who suffered damage to their prefrontal lobes were no longer able to clearly communicate their daily routine to their supervisors at work (Grafman, et al., 2013). Without being able to group actions as parts of events, the explanation of a daily routine would be highly difficult to produce. Following Grafman et al. (2013), Zalla, Pradat-Diehl, and Sirigu, (2003) tested the ability of prefrontal lobe lesions to detect transitions between actions in videos depicting everyday activities (e.g., a writer at his desk). The patients were able to indicate that little actions were occurring (grabbing pen, writing in notebook, grabbing cigarettes, finding matches, lighting cigarette, grabbing book), but when asked to say when a goal action had changed (e.g., writing in a notebook→smoking a cigarette→reading a book), patients were less accurate compared to non-patient control subjects. Without the ability to enact event schemas due to damage in the prefrontal cortex, the transitions between events were hard to detect.

(v)

Decoding the content of situation models

For example, when reading a well-known about a child wizard, Wehbe et al., (2014) were able to use the functional roles of certain areas of cortex (motor areas, emotional areas, location-tracking areas) and see how activity in these areas changed in congruence with the changes in the story (e.g. actions, visual imagery, location). Different methods lead to similar findings of cortical patterns tracking story meaning by Kurby and Zacks (2013). This supports the idea that despite the uniqueness of each person's situation model, some aspects of situation model creation are consistent across people.

(vi)

Event model benefits to memory

A similar suggestion is made in other model-mapping theories, where new information is integrated or "mapped" onto individuated event models during comprehension (e.g., Gernsbacher, 1990). The creation of separate event models has been thought to be beneficial for memory, based on experiments which demonstrated that memory for stories (Thompson & Radvansky, 2014) increased if an event boundary were imposed either by a phrase denoting an event boundary. Likewise, memory for items in a list was increased by having the subjects of the experiment physically move into a different room (Pettijohn & Radvansky, 2016). Flores et al., (2018) found that attention to boundaries is responsible for increasing memory for movies depicting action sequences. Simply asking people to mark the boundaries while watching a movie increased their recognition of images from the movies as well as their ability to recall what happened. Groups of participants asked to mark segments where events ended and began were able to recall more information for each event both ten minutes and one month later. This all hints at the importance of individuating events with increasing memory for event models.

(vii)

Mechanics of event model theories

From the model in Figure 1 we can extract three predictions about the cognitive processes underpinning event model creation during story comprehension (Kurby & Zacks, 2008, Richmond et al., 2017, Bauer et al., 2017). Research supporting each of these is detailed in separate sections, below.

(1) Information in the event model becomes less accessible after the event boundary.

(2) Perceptual attention increases after the event boundary.

(3) Integration of the situation model occurs after the event boundary.

(vii 1)

(1)Information in the event model becomes less accessible after the event boundary.

ESPM predict Examples of (1), that information in the event model becomes less accessible after the event boundary. This is particularly important in story comprehension, as recognition of words and images can be facilitated by unconscious processing of active context (see McNamara, 2005). If the event boundary alters what information is accessible, this means that recently experienced context before the boundary will not facilitate semantic processing in the same way it did before the boundary. Evidence for this loss of accessibility, comes primarily from early experiments showing an alteration of memory at a boundary in the situation model. For example Gernsbacher (1985) and others (Bransford & Franks 1971, Dooling & Lachman, 1968) showed that memory for the exact images or words in a comic strip or paragraph, faded immediately after a paragraph of text or comic strip concluded, such that people were unable to remember exact images or wording.

Further evidence for (1) comes from behavioral experiments that looked at probe recognition, reading speed, and memory when event boundaries were placed in brief narratives (Rinck & Bower, 2000, Speer et al., 2005, Ezzyat & Davachi, 2011). Rinck and Bower (2000) had subjects study the layout of a fictitious laboratory, and then read a story of a janitor walking through it, room by room. Readers were probed with object names referenced previously mentioned items. They found that probe recognition slowed in relation to a temporal marker in the text “A while later”, rather than by discourse distance (how many sentences back the object had appeared). A similar finding from

Speer et al., (2005), highlighted the effect of the event boundary on reading speed, showing that the effect of slowing down while reading an anaphoric reference was strongly affected by both discourse distance and by the existence of the referent to the anaphor being placed before an event boundary. Kelter et al., (2004) also saw similar findings when observing readers slow down more when an anaphor referenced an event which occurred before a sentence denoting long time shifts (e.g.,"She bakes some cookies to put on the plates") vs. when the intervening sentence denoted a short time shift (e.g., "She puts some cookies on the plates").

The idea that memory is organized based on what is occurring in story time, rather than discourse time, is in line with both theories which suggest that event models contain evidence of embodiment, or simulation (Zwaan, 1996, Zwaan & Radvansky, 1998). This makes the event boundaries similar to the location updating effect, or more colloquially, the "doorway effect" whereby people tend to forget what they were thinking about previously once they enter a new room (Pettijohn & Radvansky, 2016). Without demanding an explanation from embodiment, these effects of the event boundary on what information is available in memory is also consistent with the idea that event models are constrained by knowledge from event schemas (Anderson, et al., 1983), combined with the idea that the current situation model is maintained by general semantic coherence (Gernsbacher, 1990). Breaks in the coherence between one event and another, due to situational changes can come from event knowledge from embodiment, or semantic cues to create a new event model (e.g., "An hour later…", or "Back at home…").

In a more recent study of narrative recall in an amnesic patient, Zuo et al., (2020) had an amnesic patient listen to 2 minute stories that were either intact, or scrambled by concatenating 1 minute sections from separate stories. Probing recall at the end, experimenters noted that recall in the patient for the first half of the story was significantly worse for the scrambled, but not the coherent story. As amnesic patients are generally poorer, but still able to recall short narratives if probed immediately after (Baddeley & Wilson, 2002), this effect of narrative discontinuity on memory in an amnesic patient suggests that certain processes occur at a discontinuity in a story which may segregate active memory. Whether an event boundary in a normal story is strong enough to do this remains to be seen in amnesic patients, but it does suggest that discontinuities, at least strong ones caused by concatenating two separate stories, has the potential to cause changes in active information.

Evidence that brain areas which are damaged in amnesic patients are more active in recognizing information across the boundary comes from experiments by Swallow, et al., 2011. While recording fMRI, they showed people movies, while interrupting the movies to probe recognition of items that appeared or did not appear previously in the films. Event boundaries in the movies were identified using standard procedures. Swallow et al., (2011) found that recognizing images from across event boundaries (from the previous event) increased activity in the parahippocampal gyrus and hippocampus, compared to recognizing an item from within the current event. This suggests that information is stored in a different type of memory before vs. after an event boundary. In line with ESPM, this information from previous event models is less active and thus more effortful to retrieve.

(vii 2)

(2) Perceptual attention increases after the event boundary.

Examples of (2), that perceptual attention increases after the event boundary, come from experiments in which perception was measured in relation to the event boundary as people watched movies and videos (Newtson & Engquist, 1976, Eisenberg & Zacks, 2015, Zacks, Kurby, Eisenberg, & Haroutanian, 2011). Evidence from increased perception comes from both increased ability to notice missing frames of a movie if they occur at the boundary (Newtson & Engquist, 1976), as well as decreased likelihood of distraction after a boundary had been crossed (Faber, Radvansky, & D'Mello, 2018). Evidence that vision changes after the event boundary was found by tracking the gaze of subjects while they viewed videos depicting everyday activities (Eisenberg et al., 2016). The everyday activities included changing a car tire, and hanging up birthday decorations. Boundaries between actions in the movies were coded using a separate group of raters tasked to mark when in the video "one natural and meaningful unit of activity ended and another began" (Eisenberg et al., 2016, p 3).

These boundaries occurred on average every 14-19 seconds. When passively viewing the videos, the lengths of quick eye movements, or *saccades*, slowly decreased before the boundary indicating a change in attention as the gaze was becoming more fixated. At the boundary, the saccade distances returned the baseline levels, suggesting that visual searching related to the task was occurring when the new event started.

Fluctuations in perceptual attention, such as vision and hearing, are thought to occur during states of inattention or *mind wandering*. Mind wandering is thought to represent a state of attention in which perception is decreased as attentional focus is moved to internal, task irrelevant thoughts (Smallwood, 2011). Faber et al., (2018) looked at mind wandering with respect to the event boundary while subjects watched a movie. When probed at intervals of each five seconds, participants watching the movie, *The Red Balloon* (Lamorisse, 1957) were less likely to report that they were thinking about “anything else besides the movie” if an event change had occurred within the previous 15 seconds (Faber, et al., 2018). This loosely fits with the predictions that perception will change at the event boundary, at least as a decrease in mind-wandering, if not as a heightened state relative to baseline levels.

The idea that perception is heightened at the event boundary was partially explained by findings from fMRI that images shown from across an event boundary increase activity in perceptual cortex (Zacks, Kurby, Eisenberg, & Haroutunian, 2011). In a series of experiments they tested the hypothesis that prediction across the event boundary in movies were more difficult to make (Zacks, et al., 2011). Stimuli like those used by Eisenberg, et al., (2016) were used, but the procedure was slightly different. In these experiments viewers were asked to watch movies depicting a series of events related to everyday activities (e.g., washing a car) while brain activity was recorded using fMRI. At points before or after an event boundary, the movie was stopped and viewers were shown an image and asked to predict whether the still frame would appear in the film five seconds later. After an event boundary had been crossed, subjects were less accurate in predicting whether a given frame would occur five seconds later. Supporting the idea that perception is heightened at the event boundary, probe images from new events across the boundary were accompanied by increased responses from perceptual cortex (occipital lobes), in addition to a variety of other regions thought to track event perception. This provides neural evidence that perception may be increased when images appear that are inconsistent with the current event model.

How this corresponds to perception during verbal story comprehension remains to be seen, however it has been noted that EEG activity corresponding to mind-wandering has been reported when people reported task-irrelevant thoughts while listening to a story (Boudewyn & Carter, 2018). This latter finding, in light of the findings of Faber et al., (2018) of decreased mind-wandering after a boundary, it would be interesting to see whether boundaries correspond with the mind-wandering EEG activity found by Boudewyn et al., (2018). While there are differences in modality between the stimuli used by Zack's group (Zacks et al, 2011, Eisenberg et al., 2016, Faber et al., 2018) and Boudewyn et al., (2018), synthesizing these findings we would predict that fluctuations in EEG brain activity associated with perception of the stimulus modality may increase in some relationship to the event boundary.

(vii 3)

(3) Integration of the situation model occurs after the event boundary.

Examples of (3), that event models are updated in LTM after the event boundary, come from behavioral and brain imaging studies of what occurs following an event boundary (Thompson & Radvansky, 2012, Zou et al., 2020, Ben-Yakov, et a., 2013, Baldasano, Chen, Zaboud, Pillow, Hassan, & Norman, 2017, Speer et al., 2007, Ezzyat & Davachi, 2011). Thompson et al., (2012) found that people

Brain imaging studies using fMRI while people watched movies found activity related to the event boundary in the hippocampus, an area thought to be integral for episodic memory formation (Nadel & Moscovitch, 1997, Davachi & Wagner, 2002). In the study of narrative processing in an amnesic patient who lacked a hippocampus (Zou et al., 2020), the patient had more difficulty recalling two sentences when they were not related, while health controls were less affected by the discontinuity. This suggests that the hippocampus might be required for remembering information before a discontinuity in a text, and would be active when such a large disruption to memory following a discontinuity would be encountered.

Brain imaging studies of people watching movies have shown changes in the hippocampus for discontinuities due to event changes. Ben-Yakov et al., (2013) recorded brain activity using fMRI while viewers watched video clips depicting events. Each event was followed either by periods of white noise, or other events. In both cases, Ben-Yakov et al. (2013) found increases in blood flow to the hippocampus occurred in response to the offset of the previous event. Baldassano et al., (2017), followed this finding while participants viewed an episode of the TV series *Sherlock*. Using a novel method for marking the offset of events, they noticed peaks in the hippocampus which corresponded to these offsets, beginning roughly ten seconds before the event boundary and peaking roughly a few seconds (<5 seconds) after it was over.

Ezzyat et al., 2011, and Speer et al., (2007), tested how these predictions of memory encoding at the event boundary correspond with fMRI activity during reading of artificial (experimenter derived) and naturalistic stories (published literature). Ezzyat and Davachi (2011) had subjects read short paragraphs in which the experimenters had added temporal cues to signal readers to create a new event model (e.g., “an hour later”). After reading the whole paragraph, subjects were given cue sentences and asked to provide the sentence which followed. Cued recall performance showed that subjects were less able to recall the proximal sentence based on the cue if the cue came from before an event boundary. Ezzyat et al., (2011), found activity increased in the parahippocampal gyrus, but not the hippocampus, when participants encountered an event boundary. Speer et al., 2007, failed to find activity change in the hippocampus. This could have been due to the lack of clear boundaries between events in their stimuli, excerpts from a fiction novel.

The results of both experiments by Ezzyat et al., (2011) and Speer et al., (2007) showed effects of event boundaries accompanied increased activation of posterior cingulate cortex, and precuneus, both of which are thought to be part of orienting attention to task-directed activity (Raichle, et al., 2001). Whether this is associated with how memory is processed or how perception is changed (2), would be strengthened by other evidence of active memory processing at the boundary such as EEG evidence.

In light of ESPM, event models are thus both “what is happening now” in terms of their constraints on what information is active in memory, how new information is perceived, and “what happened then” in terms of their ability to organize memory for story information. As we argue in the next section, the precise timing of these processes would be better seen with EEG.

(viii)

Explanation of Semantic Vectors

Each vector of numbers represents scores for each word on a set of dimensions. The number of dimensions, and method for deriving scores or weights varies between models. Generally, these semantic scores are derived by optimizing the distance of each word along each dimension depending on how the word co-occurs with other words within a corpus. The window within which two words may be considered to co-occur also depends on the model parameters (Landauer et al., 1997; Mikilov et al., 2013; Pennington, Socher, & Manning, 2014).

These multidimensional vectors for each word allow comparison between words that mimics human performance. This allows these vectors, though lexically derived (from word tokens in context), to be referred to as "semantic" vectors. Semantic vectors allow meaningful operations such as analogy and word similarity by mathematical operations performed on word vectors. For example, by comparing the vector for the word "doctor" to the vector for "nurse", the difference (measured by cosine angle) between the corresponding vectors would be smaller than the difference between the vectors for "doctor" and "fireman", thus mimicking human semantic judgments (Landauer & Dumais, 1997). Semantic vectors have modeled other human semantic phenomena. For example, by adding or subtracting vectors new vectors can be identified which correspond to words in the model. This allows vector models to solve semantic equations, such as man+royal = king, or king-man=queen (Pennington, Socher, & Manning, 2014). Semantic vectors can also be used to find the proper solutions to analogies such as England:London::Brazil: <u>Brasilia</u>) (Pennington et al., 2014).

(ix)

Multiple explanations for slower reading at the event boundary

Explaining the behavioral phenomenon whereby people slow down when reading past an event boundary has been difficult since this could occur for a variety of reasons.

While ESPM hypothesize that changes to events make information from previous text (and thus the situation models) less available for subsequent processing, they also suggest situation model updating occurs at the end of the event (Zacks, et alo., 2007, Richmond, et al., 2017). Thus, under ESPM, slowing reading following the event boundary seen in Rinck et al. (2000) and Speer et al. (2005), could come from either interference from additional processes of situation model updating or decreases in the availability of previous semantic information due to a shift in focus

(x)

Contextual information guiding comprehension in the moment

Experimental evidence for this comes from findings that readers slow down anywhere in a story when information within a sentence is inconsistent with information from previous events. This suggests that certain information, (e.g., knowledge about a character) remains in memory throughout the story. For example, in Albrecht and O'brien (1993) if a character in a story (Bill) was described as physically weak earlier in the story, and later on (a few events later [Bill] was described as performing a physical feat of strength, readers slowed down (Albrecht et al., 1993).

For example, in one experiment reported by Speer et al.,(2005, experiment 2) the researchers were able to replicate the results of Rinck and Bower, (2000) by measuring the time it took to recognize probe words (yes/no) appearing at the end of each sentence of fictional stories. After an event boundary, specifically "An hour later" vs "A moment later" (example 8b, below) subjects were slower to recognize them however the stronger effect was of how many sentences back the probe word had appeared. While the effect of event boundary on recognition time and accuracy partially validated the hypotheses ESPM on the availability of memory, the latter effect of intervening information can be explained by other theories which suggest that information available during comprehension degrades with time.

Further findings from Speer et al., 2005, cast doubt on the ability of ESPM to uniquely explain degraded word processing following an event boundary in text. In experiments 3 and 4, they saw whether referencing a word (e.g., *stream* in 8c) with an anaphor (e.g., *creek* in 8a) was slowed by an intervening event boundary indicating a large time shift (e.g., "an hour later" vs "a moment later").

(8)

(a) She could hear water running and figured there must be a creek nearby

(b) A moment later/an hour later, she was collecting wood for a fire.

(c) Mary heard a noise near the stream/Mary heard a noise in the sky

The results of their measurements of reading time suggest that, while subjects slowed down following the event boundary, they did so equivalently for sentences containing anaphors as well as sentences which did not. This again hints at an effect of the event boundary on processing, but does not support the idea that slowed reading is coming from effects of terminating an old event model.

Another idea which might explain the effects of Speer et al., (2005) is that the distraction from seeing "An hour later" over 10 trials caused some subjects to process the phrase differently in the experiment than they would when reading a typical story.

(xi)

Limitations of Semantic Vectors

For example, approximations from vectors will never fully explain the images or or the emotions a word evokes, much less those evoked by a sentence or discourse (i.e., a situation model). Simple semantic models such as those mentioned above for semantic relations between words (e.g., man+royal=king) might not be able to capture nuance of sentence meaning due to effects like negation, nor discourse. For example, such simple models for sentences would likely fail to catch violations in discourse that readers would quickly find odd.

(9)

(a) Bill made a hot cup of black tea for Sheila.

(b) <u>He sipped the coffee when she told her it was tasty.</u>

(c) <u>She sipped it and told him it was tasty.</u>

In (9), (b) contains semantic and grammatical violations. Reading as a human, we can tell that c is a better continuation from a. A semantic vector for the sum of words in (a) and (b) and a and (c) would likely be quite close. If compared to a, (b) would probably score higher than (c), given that it is lexically more rich. The word "coffee" in (b) is closer to "black" and "tea" in (a) than is the pronoun "it" in (c). As a method for measuring discourse processing, such models may not be perfect.

(xii)

Discourse and the N400: Ditman & Kuperberg 2007

Ditman et al., (2007) had subjects read sets of three sentences, and measured the N400 response to a critical word in the final sentence. Examples are given below, in 4-6.

(1) (a) James was practicing the piano for months.

(b) He won first prize in the competition.

(c) He took the <u>medal</u> with pride.

(2) (a) James was practicing the piano for months.

(b) He played his best in the competition.

(c) He took the <u>medal</u> with pride.

(3) (a) James had never had the measles.

(b) He caught the infection at daycare.

(c) He took the <u>medal</u> with pride.

In healthy subjects, critical words in the final sentence showed equivalent facilitation for target words (e.g., “medal”) in sets 4 and 5, when contrasted with the context in set 6, via a reduced N400 response when the context helped. This meant that despite the lexical-semantic similarity of words from the previous sentence, (e.g., “prize” from 4b), priming was driven more by the concepts conveyed by the context.

(xiii)

Time-Frequency Measurement

Theories of neural communication emphasize the role of oscillatory brain activity (Varela et al., 2001). When the EEG data is transformed from amplitude (voltage) over time to amplitude (energy or power) for various frequency bands, the EEG signal can be seen to be composed of activity at various frequencies, grouped together according to frequency ranges or *bands*. While individual differences exist in the most prominent frequency within a band (See Klimesch, 1999), similarities across people (and animals), permit them to be used as useful classifications. Activity in each band reflects the coordinated activity of different neuronal ensembles, and thus each separate band is thought to have unique functional significance (Buzsaki, 2006). Two prominent rhythms which are visible to the eye when reviewing recorded EEG are alpha (8-12 Hz) and theta (3-7 Hz).

Since frequency, like rate or acceleration, is a measurement of change over time, measuring EEG frequency requires multiple timepoints. As a result, the estimation of frequency power at a specific time point (e.g. 1 millisecond) is slightly imprecise, depending on the measurement technique (Cohen, 2014). Large lengths of time provide precise frequency measurements, but miss the less-stationary, transient brain oscillations. Solutions to this time-frequency trade-off include windowing techniques such as the Welch method (Welch, 1967, for explanation see Appendix E xx). With methods to mark frequency in time it is possible to measure frequency (viz., power, phase) over longer periods of time, such as epochs relevant to the study of comprehension.

(xiv)

Time Frequency and Comprehension

Experimental work has been done to demonstrate the role of oscillatory activity in upholding comprehension processes during reading and listening (Bastiansen et al., 2002, Bastiaansen, Magyari, & Hagoort, 2010, Weiss & Muller, 2012, Boudewyn & Carter, 2018). Various experiments have measured the changes in EEG band power over the course of a sentence (Bastiaansen, Van Berkum, & Hagoort, 2002, Bastiaansen, et al., 2010, Weiss & Muller, 2012). From these experiments, certain bands, such as alpha and theta remain active between words in a sentence, and have been suggested to play various functional roles in comprehension (see Weiss et al., 2012 for information about 13-30 Hz, beta). While the precise basic functional roles of these frequencies remain speculative (e.g., attention vs. memory vs. syntactic vs. semantic), the implication is that these active processes reflect processing in the domains of perception and memory over time during comprehension.

(xv)

Alpha waves and sensory gating

For example, alpha increases in visual cortex preceding an auditory cue (Foxe, Simpson, & Ahlfors, 1998) could block visual activity and thereby free-up resources to make processing of the auditory stimulus more efficient. Various studies have supported the idea that alpha EEG power over visual cortex is advantageous while items are actively being held in memory (Klimesch, Doppelmyer, Shimke, & Ripper, 1997, Meeuwissen, et al., 2011). In a modified Sternberg task, Sheeringa et al., (2009), asked subjects to hold letter strings that were either 3-7 characters long in memory for ten seconds. They found that EEG alpha power during the retention interval increased with the number of letters required to be held in memory. Similar effects of alpha and memory were found by Meeuwissen et al., (2011) who had subjects memorize lists of one to three words, and recall them after a retention interval. They found that alpha activity during the short retention interval predicted whether those participants could recall items held in memory. Klimesch et al.,(1997), found that when words were recalled five minutes after they were asked to be remembered, alpha decreased more when subjects were able to recall them. This last finding suggested that alpha activity was related to the ability of semantic information to cause shifts in attention when recalled.

(xvi)

Alpha waves, mind wandering, and comprehension

Thut et al., (2006) cued participants to expect a briefly presented visual target in either the right or left visual field. They saw decreases in alpha over perceptual cortex (contralateral visual cortex) dedicated to processing the target. Importantly, the degree to which EEG alpha power decreased in the relevant perceptual cortex was proportional to the successful performance of target recognition. This was taken as evidence that decreased alpha power indexes increased perceptual awareness.

Compton et al., (2019) found that when probed for mind wandering after trials of a stroop task, that subjective reports of mind wandering corresponded to increased alpha EEG power over most of the cortex during the trial. Boudewyn and Carter (2015, 2018), relied on this association of alpha with perception to attribute increases in alpha power to "missed" or poorly perceived words or sentences during comprehension. For example, when reading sentences which contained anaphoric referents, subjects' ERP responses to the referents were lower if alpha EEG power over most of the scalp was higher during the presentations of the referent (Boudewyn et al., 2015). Likewise, when subjects read stories, alpha activity was shown to track comprehension (Boudewyn & Carter, 2018). In this latter study, EEG alpha activity over the frontal areas was seen to correspond to lapses in attention as well as decreases in comprehension of events presented in the discourse at those points. In line with these results, we would expect alpha EEG over frontal electrodes to correspond with decreased awareness during processing of stories.

(xvii)

Alpha and theta oscillations

Alpha power, named "alpha" for being discovered first by Berger, (1929), is stationary (resembling a sinusoid which can continue for seconds), and is thought to be generated via communication between ensembles of neurons in the cortical layers (Klimesch, 2017). Activity in the theta range is less stationary appearing as brief sinus patterns in the EEG, and is thought to be generated by ensembles which include cortical and subcortical structures (Miller, 1989, Buzsaki, 1996). While frequency can also be measured in terms of phase relative to a sinusoid we ignore these measurements in this thesis. However, while phase is often used to measure synchrony between electrodes and different frequencies (cross-frequency coupling, or instantaneous phase coupling), *synchrony* also refers to the summated activity of neural generators of a specific frequency generating voltage amplitude, hence why Klimesche and others refer to decreases in power in one band (e.g., alpha) as event-related *desynchronization*, and increases in power as synchronization.

(xviii)

Theta power and subsequent memory

For example, without explicit instruction to remember anything, Klimesch et al., (1997), tested recognition of words subjects had recently seen during a semantic judgment task. The authors found that words which were successfully recognized elicited enhanced theta activity during the semantic task, strongest over frontal midline electrodes.

Khader et al., saw theta activity after studying a picture or letter string was predictive of whether participants would recognize that image later. Like Klimesch et al., (1997), without cuing the subjects to retain the information in long-term memory, Khader et al., (2010) had subjects perform change detection judgments in which subjects were asked to detect whether a stimulus matched another stimulus shown seconds later. After the task was over, subjects were surprised with a recognition task based on the stimuli they had seen during the change detection task. When performance on the recognition task was compared to theta activity while subjects first saw the stimuli, increased theta activity over central and parietal areas was predictive of successful memory. This differs from previous findings of increased theta power over frontal midline electrodes (Klimesch et al., 1997)

(xix)

Reasons why Theta might not increase at the event boundary

There are at least two main reasons why we may not see increased frontal midline nor left hemisphere theta EEG power.

Firstly, theta power may not track formation of episodic memory from text and theta results from Sato et al. (2018), may have been indexing activity in other neural networks related to comprehension such as the default mode network (Raichle, 2001).

Even if theta EEG power indexes episodic encoding, another reason we may not see increased frontal midline nor left hemisphere theta EEG power at the event boundary is that ESPM might be inaccurate in hypothesizing that situation model updating occurs at periodic intervals, i.e., the event boundaries. Other theories might suggest that memory updating occurs at various points, especially for stories, which contain information about specific characters (e.g., the protagonist) which might require updating irrespective of the event boundary (O’Brien & Albrecht, 1993). For this reason we might not see a pattern of increased-EEG theta power at the event boundary when people listen to events which are part of stories, but may see it when events are unrelated to one another and less information (e.g., about protagonists) is available to update.

(xx)

Welch method

The Welch method incrementally slides a shorter window over the length of a longer signal and measures a Fourier Transform within each window. Estimates of the frequency across time are reduced to the intervals at which the smaller window slides. Other forms of bandpass filtering and convolution (e.g., Hilbert transforms and wavelet convolutions) exist as well, and may be less computationally demanding.

Appendix F

Hypotheses #1 In Detail

1. LSP effects of event transition (Distance from event boundary)

    a. In line with behavioral evidence that reading slows around the event boundary, we suspect that LSP will also decrease across the event boundary. We expect this effect to be visible as the increased negative amplitude of the averaged N400 for content words in sentences immediately following the boundary (distance = 0). As the effects of the event boundary decrease, we expect the amplitude of sentences 2 & 3, to become less negative. Event conditions (viz., Story Events, and Unrelated Events).

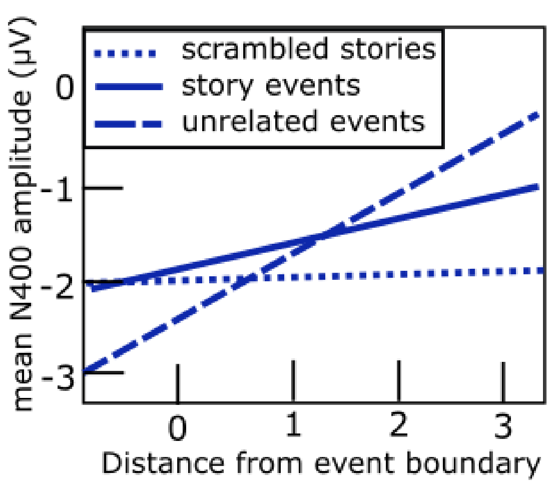


Figure F1. Correlation of N400 amplitude and distance from event boundary. Where effects of transition will be strongest where information preceding the boundary is not available to aid LSP, thus the higher correlation for sentences in the Unrelated Events condition as opposed to Story Events condition.

2. *Semantic overlap and the N400*

   a. The degree to which LSP within events is due to semantic overlap between content words in the the current and previous sentence, will be evident in the N400 averaged for content words in all sentences. We expect this effect to be weaker (less positive correlation) when comprehenders can rely on an event schema. As a result we expert semantic overlap to be more predictive of the N400 amplitude in Scrambled conditions (Related Scrambled condition and the Unrelated Scrambled condition) as opposed to the Event conditions (Unrelated Event and Story Events).

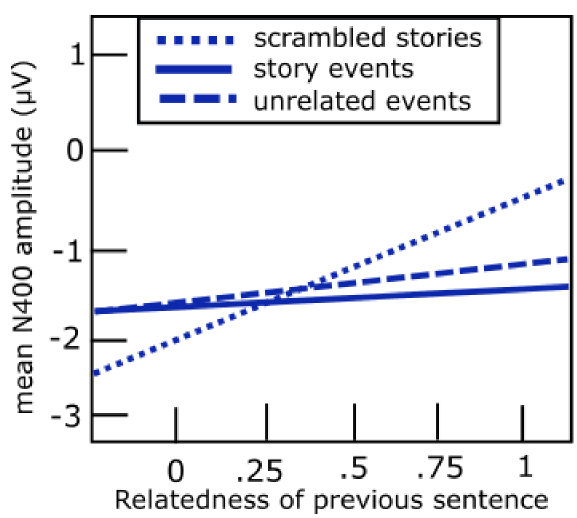


Figure F2. Correlation of N400 amplitude and distance from event boundary. Where effects of transition will be strongest where information preceding the boundary is not available to aid LSP, thus the higher correlation for sentences in the Unrelated Events condition as opposed to Story Events condition.

3. *Behavioral Hypotheses:*
    a. Recall.
        i. If a situation model is formed in the Scrambled Story block using a story schema (as suggested by Kintsch, et al., 1977), this will be seen in equivalent accuracy of recall in the Scrambled Story and Story Events conditions.
        ii. If event models increase comprehension (Flores et al., 2016), recall should be better in the Story Events condition and Unrelated Events than in the Scrambled Story condition.
    b. Recognition
        i. Block Condition:
            1. Accuracy will be higher for Unrelated Events blocks and Story Events blocks than Scrambled Story blocks due to the stronger event models in the former.
            2. This will be true for both Paraphrase and Incorrect Paraphrase conditions (hits and correct rejections).
            3. Reaction time (RT) will be faster for Unrelated Events blocks and Story Events blocks than Scrambled Story blocks due to the effects of the event models.
        ii. Distance from boundary:
            1. Accuracy of recognition will be greater for both items from sentences at the event boundary consistent with boundary

effects seen by Davachi et al. (2011), and Thompson et al. (2016).

2. This will be true for both Paraphrase and Incorrect Paraphrase conditions (hits and correct rejections).
3. Reaction time (RT) for items from the beginning of events will also be faster, consistent with boundary effects seen by Davachi et al. (2011), and Thompson et al. (2016).

Appendix G

Hypothesis 2a and 2b in detail

2a. Alpha and awareness:

EEG: As EST predicts that perceptual attention is expected to increase after crossing an event boundary, we expect that changes in EEG TF measures of perceptual attention (temporal alpha waves) will show the least power at this point. This will allow for comparison of the progress of events (sentence 1-5) which occur as either part of a story (Story Events), or as isolated events, unrelated to the rest of the block context (Unrelated Events). As with N400 hypotheses, we expect linear effects of distance from the boundary to predict alpha power over auditory cortex (temporal electrodes).

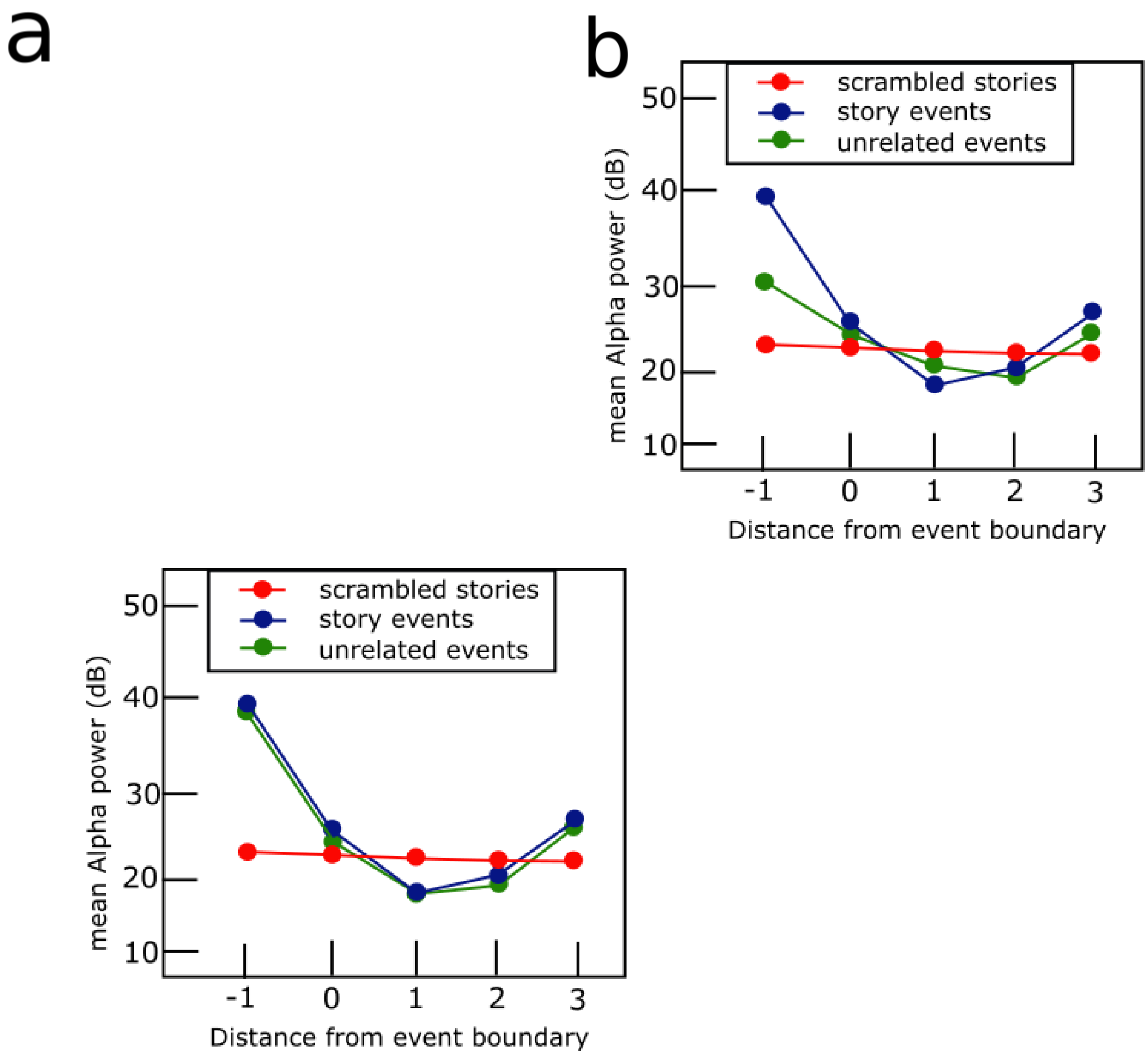


Figure G1. Predictions for alpha power over temporal electrodes following the event boundary. Data will be in decibels (dB) for visual comparison. The arbitrary semi-random event boundary in the Scrambled Stories condition (red) should cause no relationship between the event boundary and alpha power. (a) Prediction that creating a new event model increases perceptual attention (decreased alpha power) following the event boundary. (b) Prediction that changes to perceptual attention following the event boundary should be more distinct for the Story Events condition due to more elaborate event models in this condition.

2b. Theta and episodic encoding of event model:

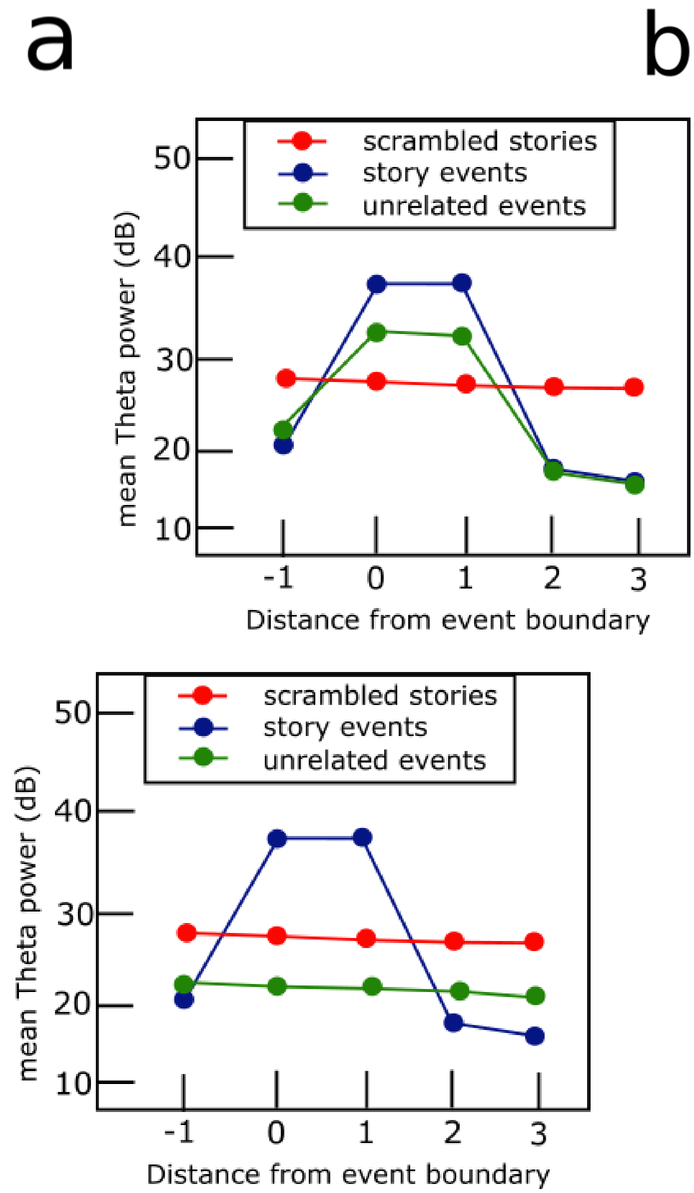


Figure G2. Predictions for theta power over midline frontal electrodes following the event boundary. Data will be in decibels (dB) for visual comparison. (a) Prediction that the event model is encoded in episodic memory following the event boundary (b) Prediction that the event model is integrated with the situation model occurs after the event boundary. In this latter case, unrelated events require less integration due to less information in the situation model.