Title
Causal Inference and Language Comprehension : : Event- Related Potential Investigations

Author
Davenport, Tristan S.

Publication Date
2014

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Causal Inference and Language Comprehension:
Event-Related Potential Investigations

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Cognitive Science

by

Tristan S. Davenport

Committee in charge:

Professor Seana Coulson, Chair
Professor Benjamin Bergen
Professor Jeffrey Elman
Professor Andrew Kehler
Professor Robert Kluender
Professor Marta Kutas

2014

The Dissertation of Tristan S. Davenport is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2014

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Chapter 2, in full, is currently being prepared for submission for publication of the material. Davenport, Tristan S.; Coulson, Seana. The dissertation author was the primary investigator and author of this material.

Chapter 4, in part, is currently being prepared for submission for publication of the material. Davenport, Tristan S.; Coulson, Seana. The dissertation author was the primary investigator and author of this material.

VITA

2001 Bachelor of Arts, Linguistics, University of California, Santa Cruz

2007 Master of Arts, Linguistics, University of California, Santa Cruz

2007-2012 Teaching Assistant, Department of Cognitive Science, University of California, San Diego

2012-2013 Associate Instructor, Department of Cognitive Science, University of California, San Diego

2014 Doctor of Philosophy, Cognitive Science, University of California, San Diego

PUBLICATIONS

Davenport, T.S. & Coulson, S. (2013). Hemispheric asymmetry in interpreting novel literal language: an event-related potential study. *Neuropsychologia*, 51 (5), 907-921.

Coulson, S. & Davenport, T.S. (2011). Cognitive neuroscience of creative language: The poetic and the prosaic. In: Faust, M. (Ed.), *Advances in the neural substrates of language: toward a synthesis of basic science and clinical research*. Oxford, UK: Wiley-Blackwell.

Davenport, T.S. & Coulson S. (2011). Predictability and novelty in literal language comprehension: An ERP study. *Brain Research* 1418, 70-82.

FIELDS OF STUDY

Major Field: Cognitive Science

Studies in Linguistics
Professor Jaye Padgett

Studies in Cognitive Science
Professor Seana Coulson

ABSTRACT OF THE DISSERTATION

Causal Inference and Language Comprehension:
Event-Related Potential Investigations

by

Tristan S. Davenport

Doctor of Philosophy in Cognitive Science

University of California, San Diego, 2014

Professor Seana Coulson, Chair

The most important information conveyed by language is often contained not in the utterance itself, but in the interaction between the utterance and the comprehender's knowledge of the world and the current situation. This dissertation uses psycholinguistic methods to explore the effects of a common type of inference – causal inference – on language comprehension. In 8 experiments, I investigate the effects of causal inference on the neuro-cognitive processes that occur during word processing (Experiments 1-5) and the hemispheric basis of these processing effects (Experiments 6-8).

The goal of Experiments 1-3 was to compare competing theoretical frameworks of language processing with respect to the ordering of "high-level" (causal inferential) and "low-level" (lexical association) context effects on word processing. To that end, participants listened to two-sentence short stories encouraging a causal inference, each followed by a probe word related to some aspect of the context story. ERP results showed that causal information affected word processing earlier than lexical associative information, and that lexical association effects were suppressed in discourse contexts. These results supported dynamic processing theories of the kind inspired by simple recurrent networks.

Experiments 4 and 5 tested the impact of causal relatedness on multiple, semi-redundant discourse cues embedded in sentences. This study investigated whether causal inferences build up over time across several words, or if a full-fledged inference can be activated in response to a single critical word. Results indicated that different participants activated inferences in qualitatively different ways. Some showed evidence of predictive inference, while others showed evidence of a drawn-out inference activation process covering several cues to discourse implausibility. These results reflect individual differences in inference activation that are unrelated to common metrics of processing ability.

Experiments 6-8 tested the hypothesis of a right hemisphere (RH) advantage for activating causal inferences. Results indicated that neither hemisphere had a processing advantage for causal related information, although left hemisphere (LH) experienced facilitated processing for strong lexical associations. This finding suggests causal inference processing is balanced between the two hemispheres, and that causal inference deficits in RH lesion patients are related to a dominant LH tendency to focus on local semantic relationships.

Chapter 1:

Introduction

One of the deep mysteries of human cognition is our ability to transmute physical signals – words on a page or phonation rippling through the air – into meaning. This ability does not depend only on decoding physical words and accessing the meanings that are associated with them. It also depends on inferring meanings that, although not explicitly encoded in the signal, are essential for understanding what a writer or speaker intends to convey. The field of psycholinguistics has focused largely on the sequence of mental operations involved in the decoding aspect of comprehension and the building of formal structures based entirely on the explicit elements of an utterance, for example, the syntactic and propositional semantic structures thought to account for the understanding of a sentence's literal meaning.

The goal of this dissertation is to investigate the other, perhaps larger half of comprehension using the psycholinguistic experimental methods applied to the study of decoding operations. This dissertation focuses on two topics in particular: first, how a successful inference affects the neural signatures of further language processing, and second, the hemispheric basis of these context effects. In order to narrow the scope of inquiry, this dissertation focuses on a specific type of inference, namely those concerning cause-and-effect relationships between the entities and events described in a narrative.

Causal inference is divided into at least two principle types in the literature. *Causal antecedent inferences* occur when comprehenders activate information about the causal prerequisites of a described event. For instance, someone reading, "The dog is having trouble walking with one leg in a cast," might infer that the dog's leg is in a cast because it was broken. This kind of inference is to be distinguished from a *causal consequent inference*, in which the comprehender infers the consequence(s) of an event or state. For instance, someone reading, "The bus ran over the dog's leg," might infer that the dog's leg was broken. A causal antecedent inference may be most likely to occur after a causal coherence gap - a break in the narrative's causal chain (McKoon & Ratcliff, 1992) - or when an event's cause is easily activated based on contextual information (but see also, Graesser, Singer & Trabasso, 1994; Myers & O'Brien, 1998). Although this dissertation primarily concerns antecedent causal inference, causal consequent inferences are touched on in Chapter 3.

Leading theories of discourse processing, including those in the constructivist framework (Kintsch, 1988; Graesser et al., 1994) and those in the memory-based tradition (McKoon & Ratcliff, 1992; Myers & O'Brien, 1998), generally agree that comprehenders readily make causal antecedent inferences if the information to do so is available. In fact, such inferences are routinely made in order to restore causal coherence in a discourse (Singer, Halldorson, Lear & Andrusiak, 1992). For example, reading a short narrative like, "Phil was at the bar drinking tequila all evening. He felt sick the next morning," comprehenders will normally activate information about the intermediate causal steps (e.g., *The tequila gave Phil a hangover*), while reading the second sentence. This inference serves to restore causal coherence, the sense of logical continuity in a narrative's chain of events. Causal consequent inferences, however, are generally agreed to be more difficult to construct (Graesser & Bertus, 1998) and in many cases are not constructed at all (McKoon & Ratcliff, 1986). Causal consequent inferences are thought to be most likely in highly constraining situations (that is, when the results of an event are obvious) and when "the information needed to make the inference is readily available in memory" (McKoon & Ratcliff, 1992).

Causal inferences have been quite well studied in terms of their effects on text comprehension. Successful causal inferences result in reduced reading times for later sentences (Graesser & Bertus, 1998; St. George, Mannes & Hoffman, 1997; Stewart, Pickering & Sanford, 2000) as well as improved memory for the contents of the text (Keenan, Baillet & Brown, 1984; Myers, Shinjo & Duffy, 1987). However, in a great deal of this research, the temporal granularity of measurement has been rather coarse. Sentence-by-sentence or line-by-line self-paced reading is common (Keenan et al., 1984; Myers et al., 1987; Klin, 1995; Graesser & Bertus, 1998). These are excellent tools for understanding whether or not certain types of inferences have been made, but they are poorly suited for understanding other important issues such as the time-courses of inference activation and of the facilitatory effects of a successful inference on word processing.

Causal inference has been shown to facilitate word processing, as indexed by quicker naming (Klin & Myers, 1993; Klin, 1995) and lexical decision performance (Till, Mross & Kintsch, 1988; Millis & Graesser, 1994) on probe words that are presented at key points while participants read or listen to a narrative. By varying the timing at which a probe word is presented with respect to a key point in the context, researchers can investigate how long it takes for an inference to begin facilitating word processing (Till et al., 1988; Millis & Graesser, 1994), but this method cannot shed any light on *which* aspects of word processing are affected by the inference. In order to investigate that question, which is central to this dissertation, a more temporally sensitive measurement technology is needed.

Causal Inference and Event-Related Brain Potentials

In order to take detailed measurements of the mental processes involved in comprehending a word in context, and the impact of causal inference on those processes, seven of the eight experiments described in the dissertation are conducted with the event-related potential (ERP) technique. In ERP experiments of the type described in this dissertation, the electroencephalogram (EEG) is recorded from sensors placed on the scalp as participants read or listen to linguistic stimuli one word at a time. Epochs of EEG activity timelocked to the onset of a critical word (for example, a word that is either consistent or inconsistent with an inference available from context) can be averaged together within each condition. The brain activity unrelated to the word is as likely to be a positive voltage fluctuation as a negative one, and therefore averages to zero if the number of items is sufficient. However, the brain activity that bears a consistent temporal relationship to the word remains, and the resulting record of this average brain activity that is timelocked to a particular type of stimulus is an event-related potential (ERP).

Because electricity propagates through flesh at a speed that is indistinguishable – for our purposes – from the speed of light, and because it can be sampled hundreds of times per second, the ERP technique is exquisitely sensitive to the timing of neurocognitive processes. One aspect of the ERP signal that is particularly useful for investigations of language phenomena is the N400 component (Kutas & Hillyard 1980;

1984). The N400 is a negative-going waveform peaking about 400ms after the onset of any meaningful stimulus. The N400 indexes the effortfulness of semantic activation processes (Kutas & Federmeier, 2011), and its amplitude is inversely proportional to the predictability of a stimulus given its context (Kutas & Hillyard, 1984; DeLong, Urbach & Kutas, 2005). This predictability and therefore the amplitude of the N400 can be modulated by discourse-level factors (Van Berkum, 2009; Van Berkum, Brown, Zwitserlood, Kooijman & Hagoort, 2005), including consistency with information provided in previous sentences (Van Berkum, Hagoort & Brown, 2003; Federmeier & Kutas, 1999) and with general world knowledge (Hagoort, Hald, Bastiaansen & Petersson, 2004).

With its fine-grained temporal sensitivity to brain activity, the ERP technique has much to offer for investigations of the effects of inference on word processing. Nevertheless, relatively few electrophysiological studies of causal inference have been published to date. In the first such study (St. George et al., 1997), subjects read four-sentence paragraphs in which the final sentence fell into one of four categories: *bridging inference*, explicitly stating a causal inference that was necessary to comprehend the paragraph; *elaborative inference*, information that could be guessed from the paragraph but was not necessary to maintain coherence; *lexical priming*, in which the final sentence had stronger lexical relations to the context but its propositional content could not be inferred; and *no inference*, in which the final sentence was neither discourse coherent in relation to its context nor lexically related. The authors found that participants' brain response to the words in the final sentences differed depending on their reading span score (Daneman & Carpenter, 1980). Low-span subjects appeared to make bridging (but not elaborative) inferences and also to make use of lexical relations. Both effects were evident in reduced N400s (Kutas & Hillyard, 1980, 1984). By contrast, high-span subjects showed reduced N400s to both the bridging and elaborative conditions, but not to the lexical priming condition. Although this study does suggest that semantic processing is facilitated by causal coherence, the effect of causal inference on word processing was obfuscated by an analysis strategy that involved averaging ERPs across all the words within the target sentence.

Yang and colleagues (2007) investigated coherence relations across sentence boundaries. Subjects read sentence pairs that varied in the degree of relation between the situation described in the first sentence and the first lexical word in the second sentence. Of particular interest was the *inference* condition: *...the plane hit turbulence. The* ***spilled*** *wine stained....* The fact that significant N400 reduction did not occur in the *inference* condition, but only when the first sentence ended with a lexical associate of *spilled,* was taken as evidence that only paraphrase and coreference relations, but not inferences, could arise entirely through memory activation processes.

Burkhardt (2007) came to a similar conclusion when she found that the degree of relatedness between a description of an event and a word denoting the cause of that event did not affect the semantic activation processes indexed by N400 amplitude. Burkhardt found only P600 effects on the critical word (e.g., “pistol”) in a story like, “A PhD student was found dead/killed/shot downtown. The press reported that the pistol was…” She suggested that the enhanced P600 in the less related conditions (following “dead” or “killed") indexed the cost of adding new information to the discourse representation in working memory (cf. Brouwer et al., 2012), an operation that was not as difficult when the new information ("pistol") was implied by a previous word ("shot").

However, as Kuperberg et al. (2010) point out, Burkhardt's (2007) and Yang et al.'s study (2007) cannot illuminate what role, if any, the situation model plays in inference generation independently of lexical associative influence. In response, Kuperberg et al. (2010) conducted a similar study in which participants read short vignettes that varied in the causal coherence between the final sentence and the previous two. ERPs to a critical word in the final sentence showed that increasing causal coherence reduced N400 amplitude in a graded fashion. However, there was no P600 effect. Crucially, lexical associative relationships between the critical words and their contexts were matched across conditions using Latent Semantic Analysis (LSA) (Landauer, Foltz & Laham, 1998), so the N400 reduction observed in the more causally coherent conditions could not be attributed to purely lexical relationships.

The ERP studies discussed above are focused on investigating which stage(s) of linguistic processing are impacted by causal inference, as revealed by the nature and timing of the ERP effects on critical words. Of particular interest is the difference between the processing facilitation due to a causal inference and that due to a purportedly "lower-level" process such as lexical association. The extant findings, however, are currently ambiguous. While some studies find earlier effects of lexical association than of causal inference (Yang et al., 2007; Burkhardt, 2007), others found that the effect of lexical association began at a similar latency to that of causal relatedness (St. George et al., 1997) or that causal relatedness affected the N400 (Kuperberg et al., 2007), which is also sensitive to lexical association (Kutas & Hillyard, 1989). The timing and nature of linguistic context effects are in turn relevant to ongoing debates about which of many competing pyscholinguistic models best account for the nature of the real-time language processing.

Causal inference effects on word processing, and their theoretical consequences, are investigated in chapters 2 and 3 of this dissertation. Chapter 2 presents three ERP experiments that investigate the relationship between causal and lexical associative influences on word processing, using narratives that involve a causal coherence gap ("A stack of papers was sitting next to the open window. Moments later, the pages were fluttering across the yard.") as a prime for probe words that bear different types of relationship to their context. In Chapter 3, I examine the impact of causal relatedness on words embedded in the narratives themselves. Chapter 3 describes one ERP experiment and one self-paced reading experiment that examine the impact of multiple, semi-redundant cues to an event that is either plausible or implausible in light of its causal coherence with setting established in the previous sentence.

Neural Basis of Causal Inference

Although language processing has traditionally been viewed as a predominantly left hemisphere (LH) phenomenon, there is evidence that the right hemisphere (RH) plays a critical role in pragmatic processes such as causal inference. Most of this evidence comes from patients with damage to the right hemisphere. Although they are usually able to understand straightforward, literal language, such patients often have a great deal of difficulty with language use involving pragmatics or social cognition. Studies of RHD patients have revealed impairment of sarcasm comprehension (Giora, Zaidel, Soroker, Batori & Kasher, 2000), emotional prosody (Bowers, Coslett, Bauer, Speedie & Heilman, 1987), jokes (Brownell, Michel, Powelson & Gardner, 1983), indirect requests (Kaplan, Brownell, Jacobs & Gardner, 1990), inferences that require theory of mind (Siegal, Carrington & Radel, 1996; Happé, Brownell & Winner, 1999), as well as causal inference (Beeman, 1993; Tompkins et al, 2008; Lehman-Blake & Tompkins, 2001; Tompkins, Fassbinder, Lehman-Blake, Baumgaertner & Jayaram, 2004).

At the same time, there is substantial evidence from functional imaging for bilateral activation and maintenance of causal inferences in neurologically normal comprehenders (Kuperberg, Lakshmanan, Caplan & Holcomb, 2006; Mason & Just, 2004; Prat, Mason & Just, 2011). This contrasts with behavioral studies using the hemifield presentation method suggesting a right hemisphere advantage for activating causal information (Beeman, Bowden & Gernsbacher, 2000; Coulson & Wu, 2005; Coulson & Williams, 2005). The hemifield presentation method is a technique whereby a visual stimulus (in this case a word) is presented outside of the fovea to the left or right of fixation. Information about a stimulus presented in right visual field is transmitted to LH first (rvf/LH presentation), and vice versa (lvf/RH). Although inter-hemispheric transfer of information begins very rapidly (Banich, 2002; Saron & Davidson, 1989), ERP and response time measures show a processing advantage for the hemisphere contralateral to stimulus presentation for a second or more after presentation (see e.g., Davenport & Coulson, 2013).

Chapter 4 of this dissertation presents the first ever experimental studies to combine the temporal precision of ERP measures with the hemifield presentation method in order to investigate the use of causal information in word processing. These three experiments (Experiments 6-8 in this dissertation) test the hypothesis that the right hemisphere plays a primary role in activating the causal information, using the materials developed for the experiments described in Chapter 2. Although Experiment 6 indicates a processing advantage for causal related words in right hemisphere, Experiments 7 and 8 demonstrate that was an artifact of confounding lexical associative information. In the final reckoning, no evidence is found to support a right hemisphere advantage for activating causal inferences.

References

Altmann, G., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. Cognitive Science, 33(4), 583-609.

Banich, M. (2002). The divided visual field technique in laterality and interhemispheric integration. Experimental Methods in Neuropsychology 21, 47–63.

Beeman, M. (1993). Semantic processing in the right hemisphere may contribute to drawing inferences from discourse. Brain and language, 44(1), 80-120.

Beeman, M. Bowden, E. & Gernsbacher, M. (2000). Right and Left Hemisphere Cooperation for Drawing Predictive and Coherence Inferences during Normal Story Comprehension. Brain Lang. 71, 310-336.

Bowers, D., Coslett, H., Bauer, R. M., Speedie, L. J., & Heilman, K. M. (1987). Comprehension of emotional prosody following unilateral hemispheric lesions: processing defect versus distraction defect. Neuropsychologia, 25(2), 317-328.

Brouwer, H., Fitz, H. & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. Brain Research 1446, 127-143.

Brownell, H., Michel, D., Powelson, J. & Gardner, H. (1983). Surprise But Not Coherence: Sensitivity to Verbal Humor in Right-Hemisphere Patients. Brain Lang 18, 20-27.

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

Coulson, S. & Williams, R. (2005). Hemispheric asymmetries and joke comprehension. Neuropsychologia 43, 128-141.

Coulson, S. & Wu, Y. (2005). Right hemisphere activation of joke-related information: An event-related potential study. J. Cogn. Neurosci. 17(3), 494–506.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of verbal learning and verbal behavior, 19(4), 450-466.

Davenport, T. & Coulson, S. (2013, in press). Hemispheric asymmetry in interpreting novel literal language an event-related potential study. Neuropsychologia.

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

Elman, J. L. (1990). Finding structure in time. Cognitive science, 14(2), 179-211.

Federmeier, K. & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. Journal of Memory and Language 41, 469–495

Giora, R., Zaidel, E., Soroker, N., Batori, G., Kasher, A., (2000). Differential effects of right- and left hemisphere damage on understanding sarcasm and metaphor. Metaphor Symb. 15, 63–83.

Graesser, A. C., & Bertus, E. L. (1998). The construction of causal inferences while reading expository texts on science and technology. Scientific Studies of Reading, 2(3), 247-269.

Graesser, A., Singer, M. & Trabasso, T. (1994). Constructing Inferences During Narrative Text Comprehension. Psychological Review 101 (3), 371-395.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. (2004). Integration of Word Meaning and World Knowledge in Language Comprehension. Science, 304, 438-441.

Happé, F., Brownell, H., & Winner, E. (1999). Acquired 'theory of mind' impairments following stroke. Cognition, 70(3), 211-240.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. Psychological review, 87, 329-354.

Kaplan, J., Brownell, H., Jacobs, J. & Gardner, H. (1990). The effects of right hemisphere damage on the pragmatic interpretation of conversational remarks. Brian Lang. 315-333.

Keenan, J., Baillet, S. & Brown, P. (1984). The Effects of Causal Cohesion on Comprehension and Memory. Journal of Verbal Learning and Verbal Behavior 23, 115-126.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review, 95 (2), 163-182.

Klin, C. M. (1995). Causal inferences in reading: From immediate activation to long-term memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(6), 1483.

Klin, C. M., & Myers, J. L. (1993). Reinstatement of causal information during reading. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(3), 554.

Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of discourse: An fMRI study of causal inferencing across sentences. Neuroimage, 33(1), 343-361.

Kuperberg, G., Paczynski, M. & Ditman, T. (2011). Establishing Causal Coherence across Sentences: An ERP Study. J. Cogn. Neuro. 23 (5), 1230-1246.

Kutas, M. & Federmeier, K. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). Annual Review of Psychology 62, 621–47.

Kutas, M. & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205.

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

Landauer, T., Foltz, P. & Laham, D. (1998). Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.

Lehman-Blake, M. T., & Tompkins, C. A. (2001). Predictive inferencing in adults with right hemisphere brain damage. Journal of Speech, Language and Hearing Research, 44(3), 639.

Mason, R. A., & Just, M. A. (2004). How the Brain Processes Causal Inferences in Text A Theoretical Account of Generation and Integration Component Processes Utilizing

Both Cerebral Hemispheres. Psychological Science, 15(1), 1-7.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. Journal of Experimental Psychology: Learning, memory, and cognition, 12(1), 82.

McKoon, G. & Ratcliff, R. (1992). Inference during reading. Psychological Review 99 (3), 440-466.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. Journal of Memory and Language, 38(3), 283-312.

Millis, K. K., & Graesser, A. C. (1994). The time-course of constructing knowledge-based inferences for scientific texts. Journal of Memory and Language, 33(5), 583-599.

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

Myers, J., Shinjo, M. & Duffy, S. (1987). Degree of causal relatedness and memory. Journal of Memory and Language 26, 453-465.

Prat, C. S., Mason, R. A., & Just, M. A. (2011). Individual differences in the neural basis of causal inferencing. Brain and language, 116(1), 1-13.

Saron, C. D., & Davidson, R. J. (1989). Visual evoked potential measures of interhemispheric transfer time in humans. Behavioral Neuroscience, 103(5), 1115.

Siegal, M., Carrington, J., & Radel, M. (1996). Theory of mind and pragmatic understanding following right hemisphere damage. Brain and Language, 53(1), 40-50.

Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992). Validation of causal bridging inferences in discourse understanding. Journal of Memory and Language, 31(4), 507-524.

St. George, M., Mannes, S. & Hoffman, J. (1997). Individual Differences in Inference Generation: An ERP Analysis. J. Cogn. Neuro. 9 (6), 776-787.

Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time-course of the influence of implicit causality information: Focusing versus integration accounts. Journal of Memory and Language, 42(3), 423-443.

Till, R., Mross, E. & Kintsch, W. (1988). Time-course of priming for associate and inference words in a discourse context. Memory & Cognition 16 (4), 283-298.

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.

Van Berkum, J. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In U. Sauerland, & K. Yatsushiro (Eds.), Semantics and pragmatics: From experiment to theory (pp. 276-316). Basingstoke: Palgrave Macmillan.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(3), 443.

Van Berkum, J. J., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. Journal of cognitive neuroscience, 11(6), 657-671.

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

Wlotko, E. & Federmeier, K. (2007). Finding the right word: Hemispheric asymmetries in the use of sentence context information. Neuropsychologia 45, 3001-3014.

Yang, C., Perfetti, C. & Schmalhofer, F. (2007). Event-Related Potential Indicators of Text Integration Across Sentence Boundaries. Journal of Experimental Psychology: Learning, Memory, and Cognition 33 (1), 55–89.

Chapter 2:

Event-related potential investigations of lexical association and causal inference on word processing

Introduction

The characterization of different kinds of context effects on word processing has been a contentious issue in psycholinguistics. Researchers have achieved some degree of consensus on the incremental nature of language processing. However, we are far from having a mechanistic understanding of language processing that satisfactorily explains how different types of contextual information contribute to word processing (e.g., Huang & Gordon, 2011, Grodner, Carbary, Klein & Tanenhaus, 2010; Camblin, Gordon & Swaab, 2007). The relative timings of different context effects therefore constitute an important form of evidence for distinguishing between various theories of the language processing architecture. In broad terms, serial processing models (e.g., Kintsch, 1988, Forster, 1976) posit a rather strict hierarchy of bottom-up processing, predicting that lower-level contextual information, such as lexical association, will impact word processing earlier than more complex forms of information, such as overall sentence context or discourse-level meaning (Forster, 1981). By contrast, dynamic or interactive processing architectures (Tabor & Tanenhaus, 1999; MacDonald, Perlmutter & Seidenberg, 1994; Altmann & Mirkovic, 2009; McRae, Spivey-Knowlton & Tanenhaus, 1998) emphasize the relative strengths of different types of contextual information, as well as their consistency with one another. A strong discourse-level cue to the identity or meaning of an upcoming word (for instance, information necessary to make a predictive inference) may facilitate word processing earlier than a relatively weak lexical associative cue, and vice-versa (Camblin et al., 2007).

Another class of models that lies, as it were, between these two extremes, is memory-based processing models (Myers & O'Brien, 1998; McKoon & Ratcliff, 1992). As in serial models, memory-based models emphasize the importance of bottom-up processing. But, like interactive models, they allow for bottom-up input to activate complex structures important for discourse level processing. For instance, complex propositions may be stored in long-term memory and activated just as quickly as lexical information. Consequently, memory based models explain how certain kinds of inferences occur quickly and automatically, such as linking a pronoun to its only plausible referent in recent context (e.g., "Harvey picked up the hammer, then ***he*** pounded the nail into the wall."). However, not all discourse-level processing occurs automatically; some inferences must be constructed slowly and strategically, such as those triggered when neither the contents of working memory nor readily available semantic knowledge suffices to ensure local coherence of the text. Bridging inferences triggered by coherence gaps in order to restore global coherence are an example of such strategic inferences (e.g., the causal bridging inference required to understand, "Harvey picked up the hammer. A short while later, he was in the hospital getting a splint for his smashed thumb.") (McKoon & Ratcliff, 1992).

Supporting the serial class of models, some experimental results suggest that comprehenders make use of low- versus high- level contextual information along different time courses. A number of studies using behavioral and eye-tracking methods have found that lexical association can modulate earlier effects on processing than discourse-level factors. For instance, while reading a story that invites a causal inference, participants are able to respond to probe words more quickly if the probe is lexically associated with the appropriate meaning of a recent homophone than when the probe word is related to a likely causal inference derived from the larger context (Till, Mross and Kintsch, 1988). Similarly, Kintsch and Mross (1985) found that whereas probes related via lexical association were primed by story contexts (e.g., hurried down to his plane *FLY)*, thematically related probes, (e.g., *GATE*), were not.

Also in keeping with serial models, eye-tracking studies have indicated that lexical factors such as repetition facilitate word reading on earlier processing measures than do discourse level factors, such as instrument inferences (Garrod & Terras, 2000) and quantifier manipulations (Huang & Gordon, 2011). In the latter study, the presence of a repeated lexical item was manipulated, as was the choice of a quantifier to draw attention to either the reference set or to its complement. For example, the context sentence "After the lecture, ***some of the girls*** met with the teacher" facilitates later processing of the reference set "the girls," whereas, "After the lecture, ***only some of the girls*** met with the teacher" puts both the reference set (the girls who met with the teacher) and the complement set (the girls who didn't meet with the teacher) into focus. This focus on two different sets of girls causes a later mention of "the girls" to be referentially ambiguous, potentially slowing processing. Eye movements were recorded on a target sentence that varied in whether it contained a lexical repetition ("The ***girls/boys*** were studying for the test…"). The authors found early facilitation effects of lexical repetition and a later, slowing effect of lexical repetition occurring only when the quantifier ***only***

some was present in the context sentence. This earlier effect of lexical repetition compared to the quantifier manipulation was interpreted in light of the memory-based processing framework (Myers & O'Brien, 1998; Gerrig & O'Brien, 2005; McKoon & Ratcliff, 1992), which emphasizes the importance and ubiquity of bottom-up memory operations in language processing, as well as their precedence over strategic inference construction.

Finally, there have also been findings in which lexical effects begin at latencies similar to or even later than discourse-level and other higher-level context effects. Many such findings come from the event-related potential (ERP) literature. Most of these findings are situated in the rich literature concerning the N400 component, a negative-going waveform focused on parietal scalp sites, which is elicited by any meaningful stimulus (Kutas & Hillyard, 1980; 1984) and is generally believed to index the cost of activating the semantic information associated with a stimulus (see Kutas & Federmeier, 2011 for a review). One crucial feature of the N400 is that its amplitude is reduced by stimulus properties or contextual factors that make the current stimulus easier to process. For example, N400 amplitude is reduced when a word is lexically associated with a recent word (Kutas & Hillyard, 1989; Van Petten, 1993); is predictable in its context, as measured with sentence completion ("cloze") norming (Kutas & Hillyard, 1984); is semantically congruent with its sentence context (Kutas & Hillyard, 1980; Van Petten, 1993); is not predictable in its context but shares semantic features with an expected word (Federmeier & Kutas, 1999); is consistent with relevant world knowledge (Hagoort, Hald, Bastiaansen, & Petersson, 2004); and is semantically congruent with its discourse context (Van Berkum, Zwitserlood, Hagoort & Brown, 2003).

ERP evidence has also been brought to bear on investigations of the distinction between lexical associative influences and sentence- and discourse-level influences on word processing. In one such study, Van Petten (1993) compared effects of lexical association and semantic coherence by manipulating the associational strength of word pairs embedded in sentences that were themselves either semantically coherent ("When the *moon* is full it is hard to see many *stars* or the Milky Way.") or semantically incoherent ("When the *moon* is rusted it is available to buy many *stars* or the Santa Ana."). The effect of lexical association on the second word began at a latency indistinguishable from that of sentence-level context.

Similarly, Camblin and colleagues (2007) compared discourse congruence and lexical association in story comprehension. In stories like the one below, Camblin et al. manipulated discourse congruence by presenting two-sentence discourse contexts that made the final word plausible or implausible, and they manipulated lexical association by varying whether the final word was a lexical associate of its antecedent, while holding its plausibility constant.

Discourse-congruent context: *Lynn had gotten a sunburn at the beach. Nothing she tried would help her dried and irritated skin.*

Discourse-incongruent context: *Lynn's wool sweater was uncomfortable and itchy. She fidgeted as the rough material irritated her skin.*

Final sentence: *Lynn couldn't stop scratching her arms and legs/nose.*

Camblin et al. found that the discourse congruency manipulation affected the ERPs at an earlier latency than the lexical associative manipulation. A number of follow-up experiments using eye-tracking showed that when the discourse context was made uninformative by removing the context sentences altogether, or by scrambling them between stimuli, the effects of lexical association emerged earlier in processing. They inferred from this pattern of findings that reliable discourse-level information can override the influence of lexical associative information on comprehension. In a follow-up study using auditory versions of similar materials, Boudewyn, Gordon, Long, Polse, and Swaab (2012) found that lexical association and discourse congruence effects began at statistically similar latencies.

An important issue raised by these studies is the relative strengths of different types of contextual cues. When Camblin et al.'s materials featured incongruent discourse contexts that were completely anomalous, the discourse congruency effects began much earlier than when they used the subtler discourse congruence manipulation exemplified in the stimuli quoted above. In contrast, Garrod and Terras's (2000) and Huang and Gordon's (2011) experiments used the strongest possible lexical manipulation — repetition priming — and pitted it against relatively subtle discourse-level manipulations. Accordingly, it is not altogether surprising that they found earlier effects of the lexical factor. It may therefore be relative cue strength, rather than the "level" of processing, that determines how quickly different context effects begin to influence word processing. This implies that the language comprehension system has some degree of flexibility in its use of incoming information, including an ability to prioritize more informative aspects of the context.

Relative cue strengths may also account for similar findings in the ERP literature. Yang, Perfetti and Schmalhofer (2007) found that two different types of lexical relationship — repetition (“exploded”) and paraphrasing (“blew up”) — had very early effects on ERPs to the target word ("*explosion*"); however, a predictive inferential relationship ("the bomb hit the ground…. *explosion*") did not significantly affect any ERP component. Interestingly, Burkhardt (2007) found that when a noun was implied with varying degrees of strength to be the instrument of a previously mentioned action ("Yesterday a PhD student was shot/killed/found dead downtown. The press reported that *the pistol* was…"), the strength of that relationship modulated the size of a post-N400 positivity, but did not affect the N400 or any earlier components. Whether the null effect on the N400 occurred because the word "pistol" was not primed in any context, or because it was equally primed in every context (in which case the N400 probably was affected) is unfortunately impossible to tell, because Burkhardt's (2007) data lack an "unrelated" condition for comparison.

The emphasis we have placed on relative cue strengths echoes the main thesis of Van Berkum's (2009) multiple-cause intensified retrieval (MIR) hypothesis, a theory of the N400 which aims to explain why very different contextual manipulations – from lexical association to truth value to speaker identity – all have similar effects on N400 amplitude. Van Berkum appeals to the notion of "readiness" in memory-based processing theories, positing that whatever relevant information is available when a word is encountered will influence its ease or difficulty of activation. Recent neuroscience work has indicated that long-term memory may be structured to optimize short-term prediction (Schacter, Addis & Buckner, 2007). In this view, many inferences that are traditionally ascribed to pragmatics, such as causal inferences and bridging inferences, may "come for free" through the activation of conceptual structure in working memory (Coulson, 2006), which may be used, for instance, to maintain discourse coherence across sentences

(Kuperberg, Paczynski & Ditman, 2011). In this view, the semantics/pragmatics distinction is therefore not strictly relevant to the time course of context effects on word processing. What is more relevant is the expectedness of the word, given the semantic and pragmatic processing that has already been done.

The MIR hypothesis focuses on explaining a class of N400 effects, but it does not address contextual modulations of earlier perceptual components that have also been linked to word-form prediction. One such component is the eLAN or M100 (Dikker, Farmer, Rabagliati & Pylkkänen, 2010; Dikker & Pylkkänen, 2011), which is enhanced when a word-form prediction is violated. Dikker and colleagues attribute this effect to a conflict between visual information about a word and word-form representations active in visual cortex. The opposite effect has been observed on the P2 component, which is larger when a word-form prediction succeeds (Federmeier & Kutas, 2002; Federmeier, Mai & Kutas, 2005; Wlotko & Federmeier, 2007). Federmeier and colleagues attributed this effect to top-down facilitation of visual feature extraction from expected items. A common thread, then, runs through much of the literature on prediction and semantic processing: namely that any type of relevant information can facilitate semantic activation, and that stronger or more specific cues can constrain expectations about word-form, leading to early effects on perceptual and/or attentional processes.

One can therefore ask whether the early and late effects of discourse context on word processing reflect qualitatively different processes or not. On one hand, it is possible that early effects on word processing reflect facilitation of lexical activation, while later discourse-level effects reflect a pragmatic process triggered by the detection of a discourse incongruity. On the other hand, it is possible that strong discourse-level cues (e.g., Camblin et al., 2007) can have early effects on word-processing, but more subtle discourse-level cues (e.g., Huang & Gordon, 2011) have similar effects that are likely to occur later. The latter possibility is highly consistent with dynamic processing models that reject a strict hierarchy of “levels” of processing (e.g., Elman, 1990; 1991; McRae, Spivey-Knowlton & Tanenhaus, 1998; Altmann & Mirkovic, 2009), while the former possibility echoes the claims of the memory-based processing framework (McKoon & Ratcliff, 1992).

The present study examined how a discourse-level factor – causal inference – impacted the real time processing of related words, and to compare the magnitude and timing of these discourse-level congruence effects with those of a well-studied ‘low-level’ factor, namely, lexical association. Accordingly, we designed the present materials to include a causal coherence gap — a feature of discourse long thought to provoke strategic inference construction (McKoon & Ratcliff, 1992) — intended to promote the activation of a causal inference.

For the experiments described in this study, I created 160, two-sentence stories, each containing a causal coherence gap (e.g., "Lucy got off her bike and locked it to a post. The bike was gone when she came out of class."). Associated with each narrative were four probe words in different categories. *Causal related* probes (THIEF) denoted the cause of the event described in the second sentence; *causal unrelated* probes (MOUSE) denoted the cause of an event described in the second sentence of a different narrative stimulus; *lexical related* probes (SCHOOL) were lexical associates of the final word in the narrative but unrelated to the cause of the event; and *unrelated lexical* probes (DISH) were lexical associates of the final word of a different narrative stimulus. Apart from the *causal related* probe category, none of the probes were related to the cause of the event described in the second sentence of their associated narratives.

In Experiment 1, I compare ERP relatedness effects on the lexical and causal probes in order to examine the relative timing of lexical association, a traditional bottom-up factor, and discourse-level information pertaining to the causal inference. The predictions of the three theoretical frameworks of text processing described above were compared in this experiment.

Experiments 2 and 3 expanded on the findings of Experiment 1 and addressed potential confounds. Experiment 2 addressed the possibility that effects of causal relatedness observed in Experiment 1 could have arisen due to lexical priming of the *causal related* probes. In Experiment 3 we explored the role of sentence-wrap up in ERP relatedness effects observed on the causal probes by manipulating the time point at which the *causal related* and *causal unrelated* probes were presented.

Experiment 1

The central question of Experiment 1 is how, if at all, the ERP effect of a causal relatedness manipulation (operationalized with the *causal probes* and *unrelated causal probes*) differ from the ERP effect of a lexical association manipulation (operationalized with the *lexical probes* and *unrelated lexical probes*). Previous ERP research on violations of discourse-level expectations have indicated that the response to words that violate discourse-level expectations is highly similar to the N400 response to semantic incongruities within a sentence (e.g., Kuperberg et al., 2011; Van Berkum, Hagoort &

Brown, 1999) and within word pairs (Kutas & Hillyard, 1989). Based on this literature, we might expect for the effect of causal relatedness to strongly resemble the effect of lexical association.

However, other ERP research has suggested that reasoning and/or inference processes triggered by linguistic stimuli can be reflected in a late frontal negativity (Pijnacker, Guerts, Van Lambalgen, Buitelaar & Hagoort, 2011; Van Berkum, 2009). If that is the case, then the inferential activity encouraged by the causal coherence gap in the stimuli may result in differences in the post-N400 ERP response to the causal probes, but not to the lexical ones.

Methods

Participants

Participants were 16 UC San Diego undergraduates (9 female), who were compensated in course credit and/or payment. All participants gave informed consent to participate. All were right-handed, as assessed by the Edinburgh handedness inventory (Oldfield, 1971). Four participants reported left-handed members of their immediate family. All participants spoke English as their first language, had normal or corrected-to-normal vision and were free of neurological disorders and psychiatric medications. Their ages ranged from 18 to 23, with a mean age of 19.5 years. An additional 3 participants were excluded from analysis due to excessive movement and/or blocking artifacts; one additional participant was released from the study because the EEG recording session was interrupted by an earthquake.

Materials

Experimental stimuli consisted of 160 short spoken narratives consisting of two sentences each, as well as visually presented probe words. In each audio narrative, the first sentence established the scene of the described event. The second sentence described a subsequent event, the direct cause of which was left unstated, thus creating a causal coherence gap between the two sentences. The audio narratives were recorded with Adobe Audition software by a female speaker, reading the sentences at a natural rate of speech and with neutral intonation. The sound files were amplitude-normalized to an average of 70 dB.

Each audio narrative was associated with four visual probe words. The *causal probe* was related to a likely cause of the event described in the second sentence. The *lexical probe* was lexically associated with the second sentence's final word, but unrelated to the event's cause. The lexical probes were chosen from the list of lexical associates to final words given in the USF word association norms (Nelson, McEvoy & Schreiber, 1999). If the final word did not appear in the USF norms, then a causally-unrelated lexical associate was chosen from the list of words generated by the "Near Neighbors" function on the Latent Semantic Analysis website (Landauer, Foltz & Laham, 1998)[1]. The *causal unrelated* and *lexical unrelated probes* served as control conditions. They were unrelated to both the cause of the event and to the sentence's final word, and were drawn from the lists of *causal* and *lexical probes* associated with other stimuli. Four lists were constructed such that each audio narrative appeared with only one probe word per list. A four-cell Latin square design ensured that across lists, all narrative/probe combinations occurred. For example stimuli see **Table 1**.

[1] For all LSA tests reported in this paper, we used the "General reading up to 1st year college" semantic space with 300 factors - the maximum allowable. Test comparisons using 150 factors returned different absolute values, but very similar relative values.

Table 1: Experiment 1 example stimuli.

1. *Lucy got off her bike and locked it a post. The bike was gone when she came out of class.*	
Causal related: *THIEF*	Lexical related: *SCHOOL*
Causal unrelated: *MOUSE*	Lexical unrelated: *DISH*
2. *The cowboy was walking through the high grass. Suddenly, he felt a piercing pain and knew that he had received a poisonous bite on his leg.*	
Causal related: *SNAKE*	Lexical related: *FOOT*
Causal unrelated: OWL	Lexical unrelated: *LUNCH*
3. *Amber was proud that she did her laundry all by herself. But when she took the clothes out, there were white patches all over her favorite shirts.*	
Causal related: *BLEACH*	Lexical related: *SLEEVES*
Causal unrelated: *OIL*	Lexical unrelated: *COUNTER*

The probes were controlled for a variety of word-level statistical factors, as shown in **Table 2.** Causal and lexical probes did not differ significantly in length, number of orthographic neighbors, or orthographic neighbor frequency. Probe words were also roughly matched for concreteness, log written frequency (Kucera & Francis, 1967), and familiarity. In addition, term-to-term LSA comparisons were made to ensure that each lexical probe was more related than its causal counterpart to the associated narrative's final word, as well as to the narrative context as a whole (Landauer et al., 1998).

Table 2: Quantitative features of the probe words. Mean values are reported, with standard deviations in parentheses. Causal relatedness ratings are taken from the norming study described in the text. LSA relatedness values are taken from the "1st Year College Reading" semantic space using 300 dimensions. All other values are taken from the MRC Psycholinguistic Database (Coltheart, 1981). Word frequencies used are those of Kucera and Francis (1967).

	Causal Relatedness	LSA Final Word	LSA Whole Story	Concreteness	Length	Orthographic Neighborhood Size	Mean Orthographic Neighbor Frequency	Familiarity	Log Frequency
Causal Related	4.1 (0.41)	0.13 (0.12)	0.27 (0.11)	561 (68)	4.9 (1.2)	5.6 (5.2)	54 (136)	540 (52)	1.2 (0.7)
Causal Unrelated	1.6 (0.46)	0.08 (0.07)	0.19 (0.1)	561 (68)	4.9 (1.2)	5.6 (5.2)	54 (136)	540 (52)	1.2 (0.7)
Lexical Related	2.0 (0.6)	0.42 (0.18)	0.36 (0.12)	519 (98)	5.1 (1.2)	5.2 (5.3)	64 (308)	563 (42)	1.6 (0.6)
Lexical Unrelated	1.7 (0.49)	0.15 (0.14)	0.3 (0.12)	519 (98)	5.1 (1.2)	5.2 (5.3)	64 (308)	563 (42)	1.6 (0.6)

Finally, a norming study was conducted to test whether students drawn from the same population as the EEG participants considered the causal probes to be likely causes of the events described in the narratives. From the 160 narratives, each with 4 associated probe words, 8 lists were created, each consisting of 80 narratives and one of the 4 possible probe words for each. 145 participants were each directed to one randomly chosen list in the form of an online survey. For each narrative/probe word pair, participants were asked to rate, on a 1-to-5 scale, how likely that word is to be the cause of the situation described. A rating of 1 signified "Definitely not the cause" and a rating of 5 signified "Definitely the cause," with intermediate ratings signifying intermediate degrees of causal likelihood. These norming results are also given in **Table 2.**

Procedure

After filling out an Edinburgh handedness inventory (Oldfield, 1971), a demographic questionnaire and a consent form, participants completed a listening variant of the reading span test (Daneman & Carpenter, 1980) on a computer. After the listening span test, participants were fitted with an EEG cap and seated in a dimly lit, electrically shielded chamber, one meter away from the computer monitor on which stimuli appeared. Audio narratives were played through a pair of external speakers flanking the monitor. Throughout the experiment, the monitor displayed a black field with an orange fixation dot in the center.

Participants were told that they would hear two-sentence-long short stories, and that at the end of each story a word would be flashed in the center of the screen. The participant's task was to listen to the stories and try to understand them, and to pay attention to the words that appeared in the center of the screen. Participants were instructed to fixate on the central dot, to avoid blinking whenever the audio was not playing, and to refrain from body movements. Before beginning the experiment, participants were given two practice trials to get used to inhibiting blinks at the proper times. Throughout the experiment, participants were given feedback on their blinking behavior. The total EEG recording time was about 45 minutes.

The experimental paradigm is represented schematically in **Figure 1.** In each trial, participants fixated on the center of the monitor while listening to the audio narrative stimulus. Each sentence of the narrative lasted between 3 and 8 seconds, and there was a pause of 300 ms between sentences. Following the second sentence of each trial, there was an additional pause of 400 ms before a centrally presented probe word was displayed. This latency was chosen because the duration of the final spoken word ranged from 300-450 ms, producing a total stimulus onset asynchrony (SOA) of 700-850ms. This approximates the SOA of 800ms at which Holcomb & Anderson (1993) found the largest N400 effects of auditory-to-visual cross-modal lexical priming. The visually presented probe word was displayed for 250 ms. The interval between trials was 2000 ms. Participants were offered a short break after every 20 trials.

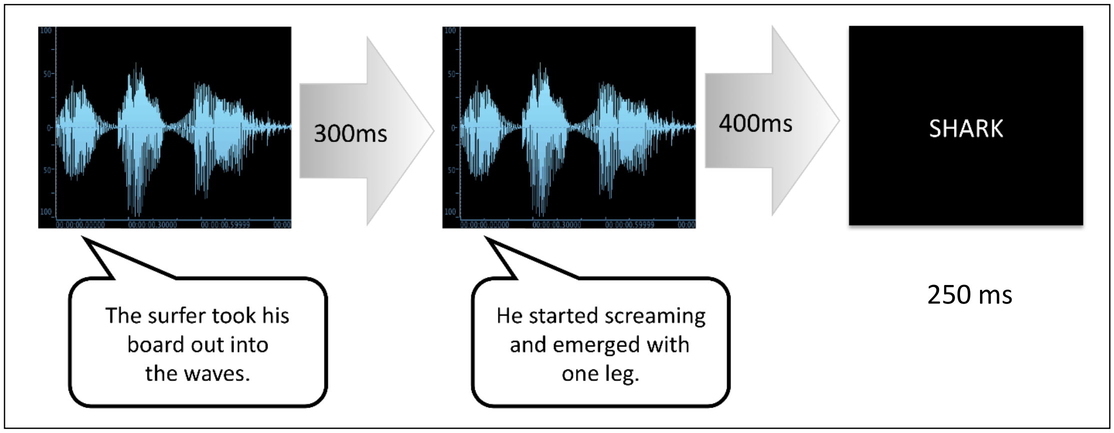

Figure 1: Schematic representation of a *causal related* trial in Experiment 1. Other conditions use the same set of audio narrative primes and timing parameters but a different visual probe word.

EEG Recording

EEG was recorded with 29 tin electrodes in an Electrocap mesh cap, organized in the international 10-20 configuration. Recordings were taken from 8 lateral sites: T5/6, TP 7/8, FT 7/8, F7/8; 10 medial sites: P3/4, CP3/4, C3/4, FC 3/4, and F3/4; 5 midline sites: Pz, CPz, Cz, FCz, and Fz; 3 frontal polar sites: FP1/2 and FPz; and 3 occipital sites: O1/2 and Oz. Three additional electrodes were placed at the outer canthi of the eyes and below the left eye, to record eye movements and blinks. At all sites, electrical impedance was reduced below 5kΩ through gentle abrasion and by applying Electrocap conductive gel. All EEG recorded was referenced on-line to a single electrode on the left mastoid, and re-referenced to an average of the left and right mastoid electrodes before data analysis. The EEG was amplified with half-amplitude cutoffs at 0.1 Hz and 100 Hz using a SA Instrumentation bioelectric amplifier. The data were digitized online at 250 Hz.

Analysis

Only ERPs to the visual probe words were analyzed. ERPs to probes were time-locked to word onset and averaged in a time window spanning 100ms pre-onset to 920 ms post-onset. The period from 100ms pre-onset to stimulus onset served as the baseline. During data analysis, critical EEG epochs were examined and trials containing blocking, drift, and movement artifacts were rejected. If more than 25% of a participant's critical epochs were rejected in this way, that participant was excluded from analysis. As noted above, three participants were rejected for this reason. In the data reported below, 20% of trials were rejected due to artifacts.

Analysis of ERP components focused on three time windows: 150-300ms post-onset (P2 component), 300-500ms (N400 component), and 500-800ms (post-N400 effects). Mean amplitudes were analyzed with a repeated measures ANOVA, using a 2 (relatedness: related vs. unrelated) x 2 (probe type: causal vs. lexical) x 29 (electrode) design. Planned follow-up repeated measures ANOVAs assessed relatedness 2 (relatedness) X 29 (electrode) within each probe type condition. To compensate for violation of the sphericity assumption, the Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied where appropriate.

If a significant or marginal ($p < 0.1$) Relatedness effect, or Relatedness X Electrode interaction, was observed in a particular time window, then a follow-up analysis was conducted in order to examine the scalp distribution of voltage differences. In these cases, we performed a two-tailed *t*-test at each electrode site at each 4-ms time point, testing the null hypothesis that the voltage difference between the *related* and *unrelated* conditions was 0 µV. The *t*-max permutation procedure was used to correct for multiple comparisons, using a family-wise alpha level of 0.05. These permutation tests were carried out with the 'tmaxGND' function using 2000 permutations per time point in the Mass Univariate ERP Toolbox for MATLAB (Groppe, Urbach & Kutas, 2011).

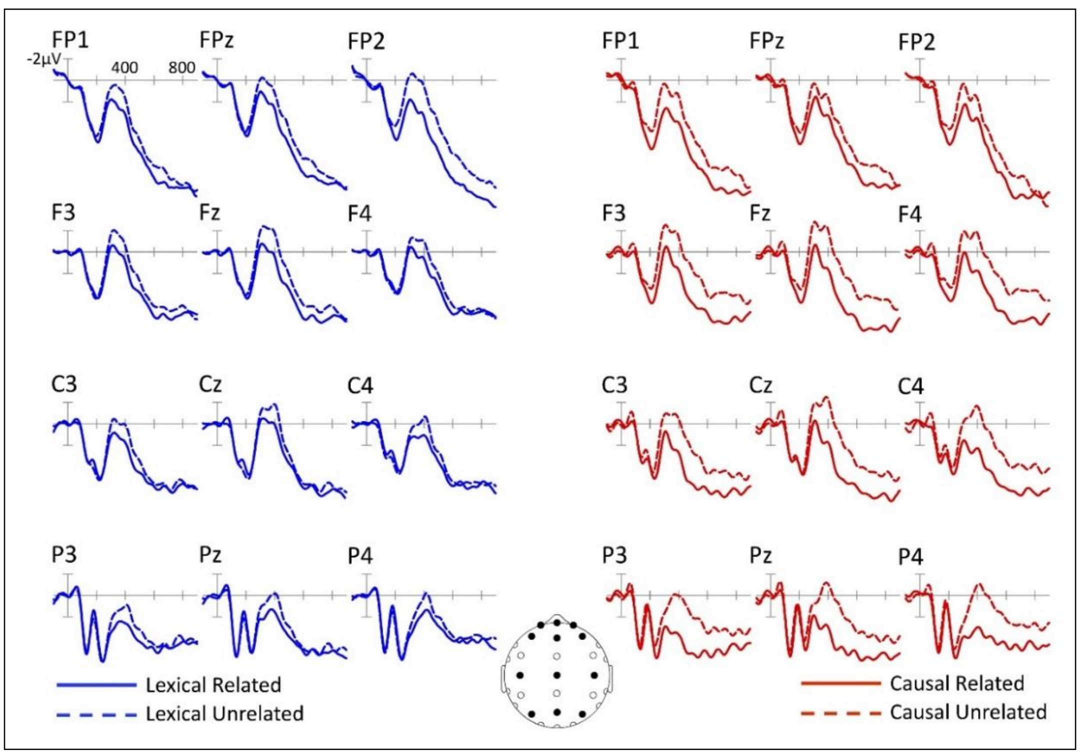

Figure 2: ERP waveforms at 12 electrode sites depicting data from Experiment 1. ERP responses to related items are drawn in solid lines and unrelated items in dashed lines. The lexical relatedness effect is shown in the left column in blue, and the causal relatedness effect is shown on the right column in red. Negative voltage is plotted upward, and the data were low-pass filtered at 10 Hz for presentation purposes.

Results

P2 Component (150-300 ms)

The initial repeated measures ANOVA revealed marginal effects of relatedness ($F[1,15] = 4.5$, $p = 0.052$) and probe type ($F[1,15] = 2.7$, $p = 0.12$). There was a significant interaction between these factors ($F[1,15] = 5.8$, $p = 0.03$), reflecting larger differences between the two causal conditions compared to the two lexical conditions. Planned comparisons within the causal and lexical conditions showed that *causal related* probes elicited a larger P2 than *causal unrelated* probes ($F[1,15] = 8.0$, $p = 0.01$) but that the lexical conditions did not differ ($F[1,15] < 1$).

The *causal* effect was analyzed for topographical distribution using *t*-max permutation tests at each time point as described in section 2.1.5. Significant corrected *t*-tests were observed over frontal sites beginning at 252 ms and continuing to the end of the P2 time window. In all of them, effects reflected more positive ERPs to the *causal related* probes the *causal unrelated* ones. Raster plots showing significant t-tests in all three time windows can be found in Figure 4.

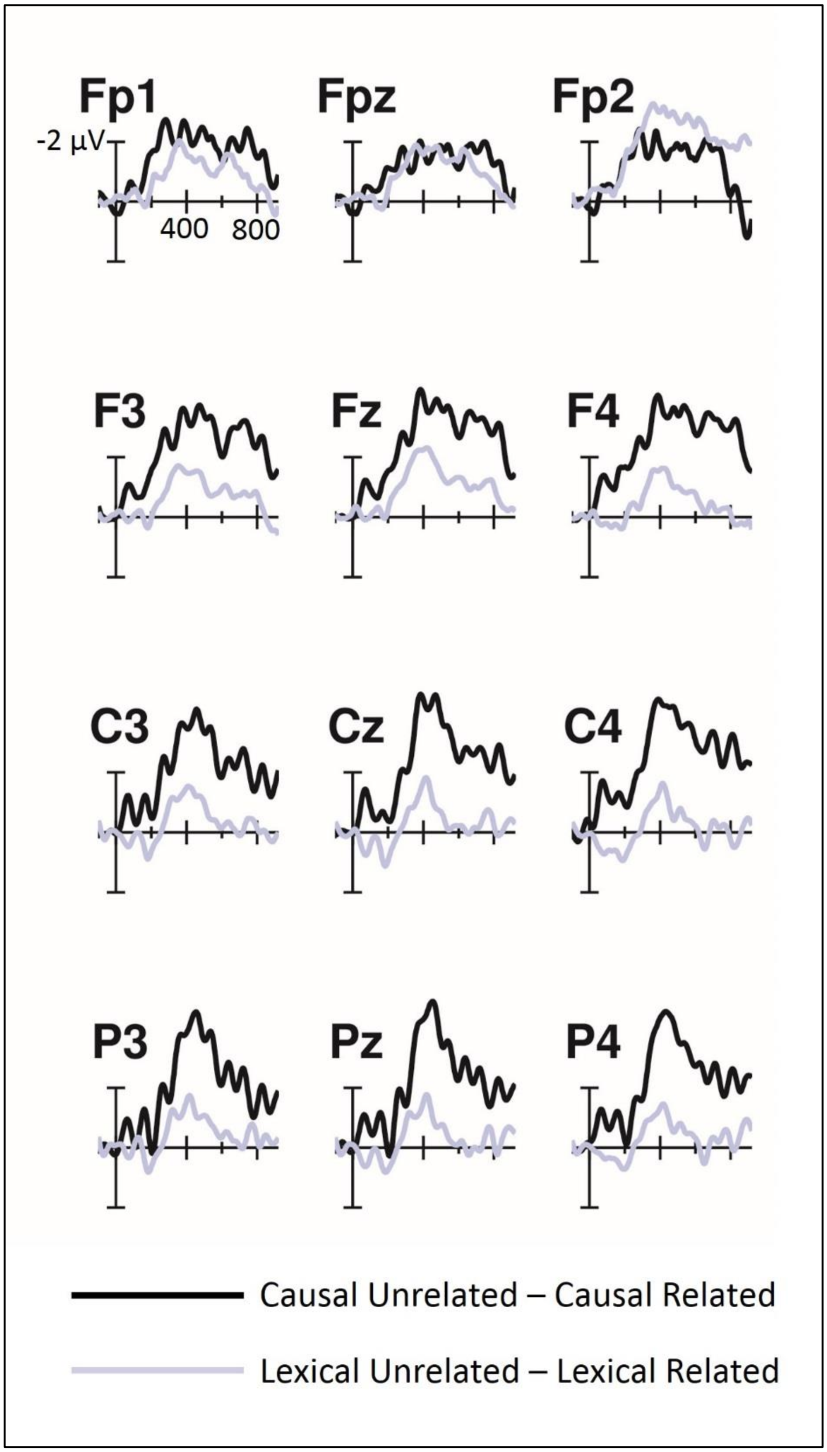

Figure 3: Difference waves showing the relatedness effect (Unrelated – Related) for the *causal* conditions (black line) and the *lexical* conditions (light gray line) in Experiment 1. Negative voltage is plotted upwards.

N400 Component (300-500 ms)

A repeated measures ANOVA revealed a robust main effect of relatedness ($F[1,15] = 61.3$, $p < .0001$) but no main effect of probe type ($F[1,15] < 1$). A significant interaction between these factors ($F[1,15] = 8$, $p = 0.01$) reflected a larger effect of relatedness for causal probes (mean difference = 2.6 μV) than for lexical probes (mean difference = 1.2 μV). Interactions involving the Electrode factor were marginal (Relatedness X Electrode: $F[28,420] = 2.5$, $p = 0.055$; Relatedness X Probe Type X Electrode: $F[28,420] = 2.7$, $p = 0.057$).

Planned pair-wise comparisons within probe type revealed that the effect of relatedness was significant both for the causal probes ($F[1,15] = 38.7$, $p < 0.0001$) and for the lexical probes ($F[1,15] = 11.8$, $p = 0.004$). In the causal, but not the lexical, comparison, there was also a significant relatedness X electrode interaction ($F[28,420] = 4.2$, $p = 0.004$).

The *causal* effect was subsequently analyzed for topographical distribution using *t*-max permutation tests. Significant corrected *t*-tests were observed over central-parietal and parietal sites beginning at 320 ms and expanding to most scalp sites (excluding prefrontal sites) by about 360 ms, a pattern which continued throughout the N400 window. All significant *t*-tests reflected time points at which the *causal related* condition was more positive than the *causal unrelated* condition. The *lexical* relatedness effect was also analyzed for topographical distribution using *t*-max permutation tests. Only a few significant *t*-scores were observed, all on the Fz site.

Post-N400 effect (500-800 ms)

A repeated measures ANOVA revealed a main effect of relatedness ($F[1,15] = 22.6$, $p = 0.0003$) and no main effect of probe type ($F[1,15] < 1$). By contrast, there was an interaction between these factors ($F[1,15] = 7.8$, $p = 0.01$), reflecting larger effects of relatedness on causal probes than on lexical probes.

Planned comparisons within probe type revealed that the *causal unrelated* probes were 2 μV more negative than the *causal related* probes ($F[1,15] = 28.4$, $p = 0.0001$). The relatedness X electrode interaction was significant in the causal comparison ($F[28,420] = 2.9$, $p = 0.03$), apparently indicating a tendency for a right anterior focus. However, ERPs elicited by the lexical related and unrelated probes did not significantly differ (Relatedness: $F[1,15] = 1.3$, $p = 0.26$; Relatedness X Electrode: $F[28,420] = 1.2$ $p = 0.3$).

Once again, the *causal* relatedness effect was analyzed for topographical distribution using *t*-max permutation tests at each time point. Significant corrected *t*-tests (*causal related* more positive than *causal unrelated*) were observed over frontal, central and parietal sites from the beginning of the time window until about 700 ms. This effect was attenuated or even absent during a ~40ms span from 580 ms to about 620 ms, after which it appeared to be largest in right hemisphere and midline electrodes.

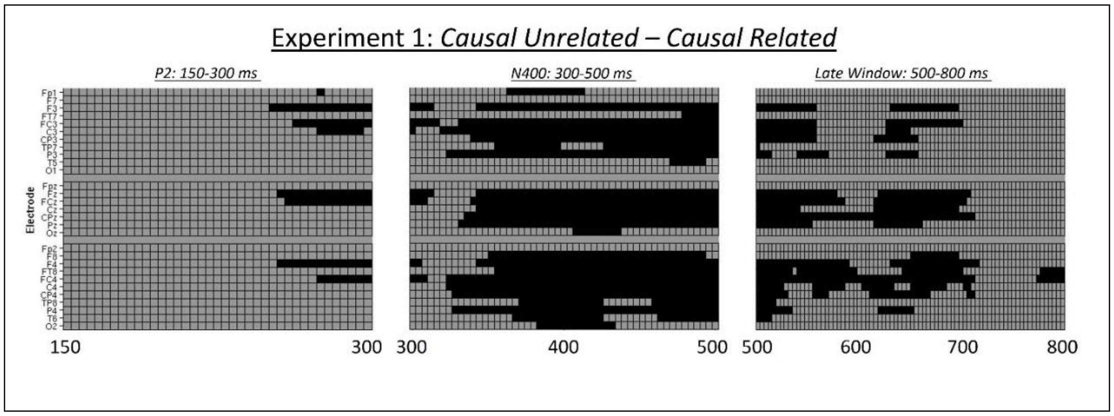

Figure 4: Raster plot of significant 2-tailed *t*-tests against the null hypothesis that the *causal* relatedness effect in Experiment 1 is equal to 0 μV. Black squares represent *causal unrelated* probes eliciting more negative voltage than *causal related* probes at $p \leq 0.05$. *t*-max permutation tests were used to correct for multiple comparisons, with 2000 permutations at each 4-ms time point. Note that in all raster plots, LH electrode sites appear in the top third of each graph, midline sites in the middle third, and RH sites in the bottom third.

Listening Span Scores

Participants' listening span scores were calculated as follows. As in Daneman & Carpenter's (1980) method, a participant was assigned a score for the highest set length at which he/she remembered all of the final words in at least 2 out of the 3 sets. An additional ½ point was scored if the participant remembered all the words in 1 out of 3 sets at a higher level. The maximum score by this method, called the subject's *listening span score*, was 5. Participants were divided into high- and low- listening span (LS) groups using a median split on listening span scores. Both groups of participants showed effects in the same time windows of the same polarity. Although participants in the high-LS group had numerically larger ERP effects in all time windows than participants in the low-LS group, none of these differences reached significance (all p-values > 0.1).

Discussion

The goal of this experiment was to compare the magnitude and timing of discourse- and lexical level information on ERP relatedness effects. In keeping with work by Camblin and colleagues (2007), we found that discourse-level information (the causal relatedness manipulation) had an earlier impact on probe words than did lexical relatedness, as P2 effects were evident on the causal but not lexical probe words. Both sorts of probes elicited effects during the N400 interval, though of smaller magnitude and somewhat more anterior scalp distribution in the lexical conditions. Further, causal relatedness effects continued 500-800 ms after stimulus onset, while lexical effects were no longer evident in the ERPs.

The earlier effect of causal relatedness compared to lexical relatedness matches the predictions made for dynamic models, which can prioritize more informative types of contextual information – in this case causing earlier effects of the causal relatedness manipulation. However, the *causal* relatedness effect also continued later than the *lexical* relatedness effect in form of a sustained negativity that was larger in the *causal unrelated* condition than in the *causal related* condition. This late effect recalls the memory-based framework's prediction that a break in discourse coherence (caused when the probe word fails to fill the causal coherence gap introduced in the narrative context) can trigger an extended, strategic pragmatic process. Indeed, sustained negativities have also been associated with inferential or pragmatic processes triggered by a word inconsistent with a prior inference (Pijnacker et al., 2011; Baggio, van Lambalgen & Hagoort, 2008).

In either case, the results of Experiment 1 are difficult to reconcile straightforwardly with previous findings that lexical-level relations affect processing early on, while discourse-level relations such as relatedness to a causal inference take more time to affect word processing (c.f. Till, Mross, & Kintsch, 1988). One possibility is that the two sorts of probes differed along some dimension that made the causal probes intrinsically easier to process. Perhaps, in spite of our efforts to ensure that the *lexical related* probes were more strongly associated with the story-final words than were the *causal related* probes, the latter were somehow more strongly lexically primed. To test this possibility, in Experiment 2 we removed the discourse-level context from the materials and recorded ERPs as participants read both *lexical related* and *causal related* probe words preceded by only the final word of the associated context story.

Experiment 2

In Experiment 1, we observed larger and earlier ERP effects due to causal relatedness than due to lexical relatedness. This finding appears to contradict some previous results, notably those of Till and colleagues (1988), who found earlier effects of lexical relatedness using a lexical decision task combined with manipulation of stimulus onset asynchrony. One possible explanation for this disparity is that despite our best efforts to ensure that the *lexical related* probe words were more closely related to the story-final words than the *causal related* probes, the *causal* probes nevertheless differed along some dimension that made them intrinsically easier to process in their *related* contexts. Experiment 2 is designed to address that possibility by recording ERPs to the *causal related* and *lexical related* words in absence of their discourse contexts, primed only by the final word of the preceding *related* story.

Methods

Participants

Participants were 8 UC San Diego undergraduates (4 female), who were compensated in course credit. All participants gave informed consent to participate. All were right-handed, as assessed by the Edinburgh handedness inventory (Oldfield, 1971). Three participants reported left-handed members of their immediate family. All participants spoke English as their first language, had normal or corrected-to-normal vision and were free of neurological disorders and psychiatric medications. Their ages ranged from 20 to 30, with a mean age of 22 years.

Materials

Experimental stimuli consisted of 160 spoken words – the final word of each of the context stories used in Experiment 1 – as well as their associated *causal related* and *lexical related* probe words (see Tables 1 and 2). The spoken prime words were recorded with Adobe Audition software by a male speaker. These sound files were amplitude-normalized to an average of 70 dB. Two stimulus lists were constructed, each containing 160 spoken prime words, 80 *causal related* probes and 80 *lexical related* probes, such that all of the prime words appeared with only one probe word each on list 1, and each prime word appeared with its other probe word on list 2.

Experimental Procedure

Experimental procedures were kept as close as possible to those of Experiment 1. After filling out an Edinburgh handedness inventory (Oldfield, 1971), participants were fitted with an EEG cap and seated in a dimly lit, electrically shielded chamber, one meter away from the computer monitor on which stimuli appeared. Audio primes were played through a pair of external speakers flanking the monitor. Throughout the experiment, the monitor displayed a black field with an orange fixation dot in the center.

Participants were told that they would hear single words followed by visually presented words flashed in the center of the screen. The participant's task was simply to pay attention to everything they saw and heard. Participants were instructed to fixate on the central dot, to avoid blinking whenever the audio was not playing, and to refrain from body movements. Before beginning the experiment, participants were given ten practice trials to get used to inhibiting blinks at the proper times. In each of the 160 trials, the participant heard a spoken prime word then, 400 ms after that word ended, saw a printed probe word for 250 ms. The interval between trials was 1000 ms. Throughout the experiment, participants were given feedback on their blinking behavior. The total EEG recording time was about 10 minutes.

EEG Recording

EEG was recorded in a manner identical to that of Experiment 1.

Analysis

Only ERPs to the visual probe words were analyzed. ERPs to probes were time-locked to word onset and averaged in a time window spanning 100ms pre-onset to 920 ms post-onset. The period from 100ms pre-onset to stimulus onset served as the baseline. During data analysis, critical EEG epochs were examined and trials containing blocking, drift, and movement artifacts were rejected. In all, 20% of trials were rejected due to artifacts. Rejection rates did not differ between *lexical related* and *causal related* conditions ($t_7 < 1$, p = *n.s.*).

Analysis of ERP results focused on three time windows: the P2 component (150-300 ms), the N400 component (300-500ms), and the post-N400 window (500-800 ms). Mean amplitudes were analyzed with a repeated measures ANOVA, using a 2 (probe type: causal related vs. lexical related) x 29 (electrode) design.

In order to examine the scalp topographies of the ERP effects of probe type in each time window, *t*-max permutation tests were conducted exactly as in Experiment 1. A two-tailed *t*-test was performed at each electrode site at each 4-ms time point, testing the null hypothesis that the voltage difference between the *causal related* and *lexical related* conditions was 0 μV. The *t*-max permutation procedure was used to correct for multiple comparisons, using a family-wise alpha level of 0.05. Permutation tests were carried out with the 'tmaxGND' function using 2000 permutations per time point in the Mass Univariate ERP Toolbox for MATLAB (Groppe, Urbach & Kutas, 2011). Raster plots showing significant t-tests in the analyzed time window can be found in Figure 7.

Results

At all electrode sites, the *causal related* probes elicited more negative ERPS than the *lexical related* probes, beginning at 300 ms or earlier. The ERPs for Experiment 2 are displayed in Figure 5. A comparison between the *causal related* and *lexical related* conditions in Experiments 1 and 2, is given in Figure 6.

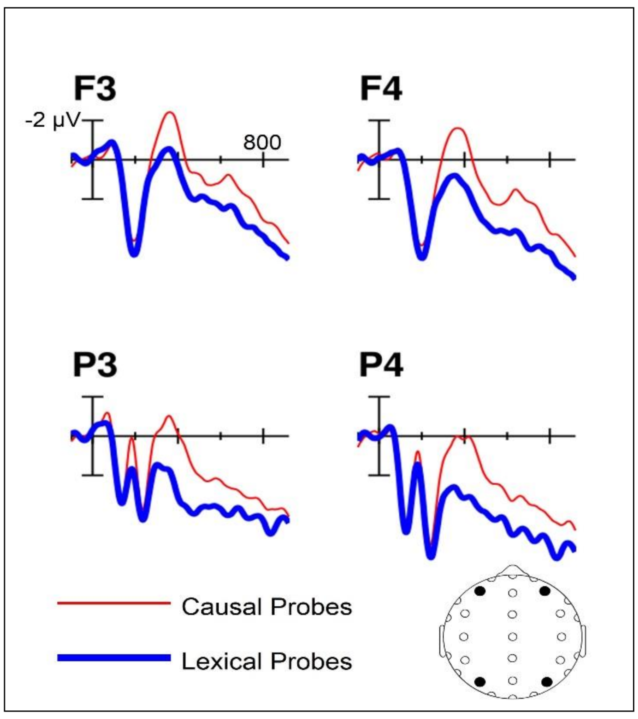

Figure 5: ERP waveforms at 4 electrode sites depicting data from Experiment 2. ERP responses to causal related items are drawn in red dashed lines and lexical related items in blue solid lines. Negative voltage is plotted upward, and the data were low-pass filtered at 10 Hz for presentation purposes.

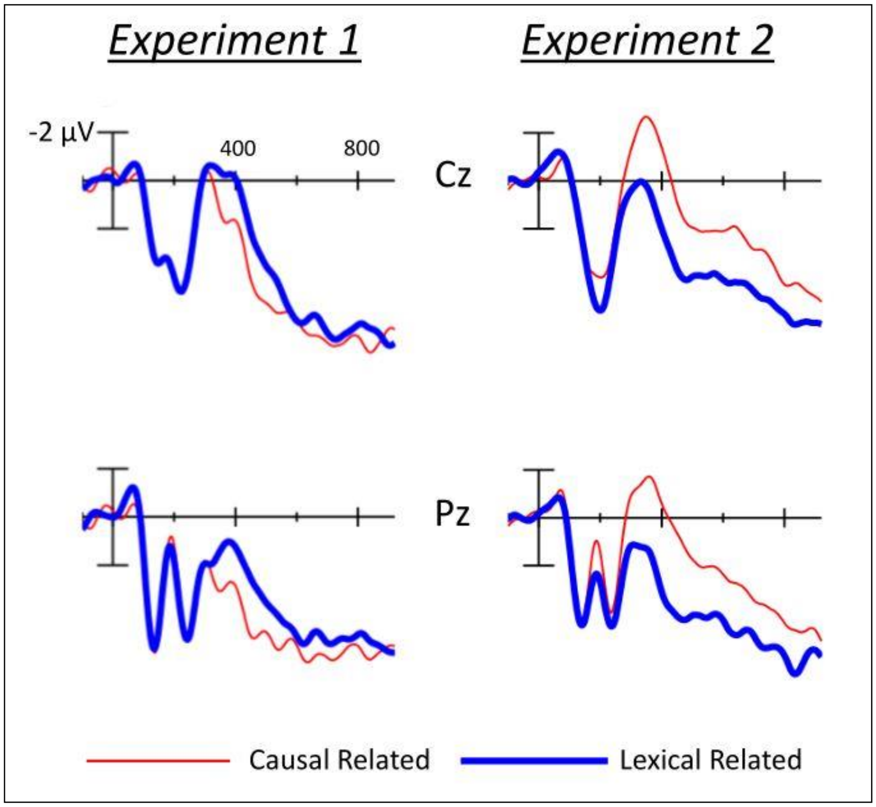

Figure 6: ERP waveforms at 2 electrode sites depicting the *causal related* and *lexical related* probes in Experiment 1 (discourse primes) and Experiment 2 (word primes). ERP responses to causal related items are drawn in red solid lines and lexical related items in black dashed lines. Negative voltage is plotted upward, and the data were low-pass filtered at 10 Hz for presentation purposes.

P2 (150-300 ms)

The *causal related* probes elicited marginally more negative voltage than the *lexical related* probes over the entire scalp ($F[1,7] = 4.4$, $p = 0.07$). This marginal probe type effect was analyzed for topographical distribution using *t*-max permutation tests at each 4-ms time point. No significant *t*-scores were observed.

N400 (300-500 ms)

Averaged across the scalp, the *lexical related* probes elicited N400 some 2.5 µV less negative than the *causal related* probes did (F[1,7] = 19.3, p = 0.003). The significant Probe Type X Electrode interaction indicated that the peak of this effect was to some degree unevenly distributed over the scalp (F[28,196] = 4.4, p = 0.01). The probe type effect in the N400 was analyzed for topographical distribution using *t*-max permutation tests at each time point. Beginning at about 350 ms and continuing until about 450 ms, the *causal related* probes were more negative than the *lexical related* probes, mainly over parietal and temporal electrodes.

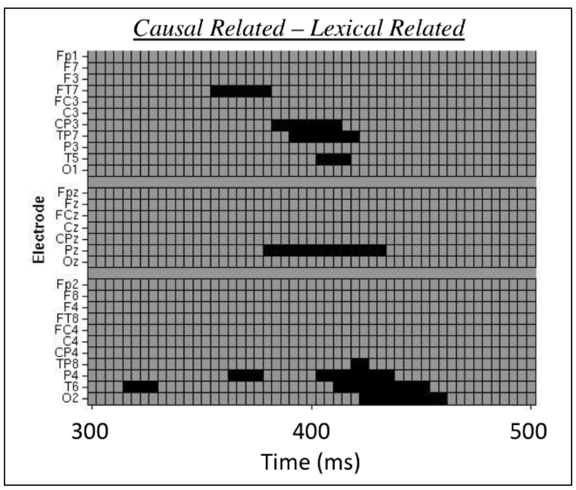

Figure 7: Raster plot of two-tailed *t*-tests against the null hypothesis that the difference between the *causal related* probes and the *lexical related* probes in the N400 window of Experiment 2 is equal to 0 µV. Black squares represent *causal related* probes eliciting more negative voltage than *lexical related* probes at $p \leq 0.05$. *t*-max permutation tests were used to correct for multiple comparisons, with 2000 permutations at each 4-ms time point.

Post-N400 window (500-800 ms)

In addition to the N400 effect, we also observed an effect in our post-N400 window. This effect has the same polarity as the N400 effect, with the *lexical related* probes eliciting more positive voltage than the *causal related* probes ($F[1,7] = 8.5, 0.02$). The Probe Type X Electrode interaction was non-significant. This significant probe type effect was analyzed for topographical distribution using *t*-max permutation tests at each 4-ms time point. No significant *t*-scores were observed.

Discussion

The familiar lexical priming N400 effect was observed, beginning at or shortly before 300 ms and continuing throughout the epoch. Consistent with our norming data, the *lexical related* probes were primed much more strongly than the *causal related* probes. This suggests that the greater facilitation – observed in Experiment 1 for the *causal related* probes – was not due to lexical association with the story-final words.

In addition to the expected N400 lexical priming effect, a broadly distributed late negativity was observed. This finding has previously been reported in ERP lexical priming experiments. (e.g., Lau, Holcomb and Kuperberg, 2013). Lau and colleagues (2013) speculated that this continuing negativity reflected conscious attempts to relate the unassociated target words to the preceding primes.

Experiment 3

Experiment 1 established that manipulations of causal relatedness yielded early effects on the P2 and N400 ERP components, as well as a post-N400 effect that has been observed in some studies of discourse-level processing (Pijnacker et al., 2011; Van Berkum, Hagoort & Brown, 1999), but not others (Kuperberg et al., 2011; Camblin et al., 2007). By contrast, Experiment 1's manipulation of lexical relatedness yielded effects on the N400, but did not modulate the P2 or the post-N400 effect. Earlier effects of discourse- than lexical-level context observed in Experiment 1 were contrary to the findings of a similarly motivated study conducted by Till and colleagues (1988) using lexical decision latencies as their dependent variable. One difference between Till et al. (1988) and the present study, though, was that whereas our inferences about the relative timing of lexical- versus discourse- level contextual factors was based on the divergence of relevant ERP waveforms, Till et al. (1988) relied on differences in priming effects observed at short versus long intervals between the discourse primes and the probes. Accordingly, in Experiment 3 we also varied the timing between the discourse primes and the causal probes in order to examine whether our discourse-level ERP relatedness effects would be attenuated by reducing the interval between the discourse primes and the probes.

The timing paradigm used in Experiments 1 and 2 was based on that used by Holcomb and Anderson (1993), who used auditory primes and visual probe words and observed maximal priming effects on the N400 at a stimulus onset asynchrony of 800 ms. Consequently, we used an inter-stimulus interval of 400 ms, equivalent to an average

SOA of about 800 ms, given a typical story-final word duration of 350-450 ms. The intent of this choice was to maximize the size of the lexical relatedness effect, so that a larger and/or earlier causal relatedness effect would not be observed solely due to an accident of the experimental paradigm. It is possible, however, that some aspects of the causal relatedness effect observed in Experiment 1 depended on processing carried out during the interval between the end of the context story and the onset of the visually presented probe. For example, some or all of the causal relatedness effect may have relied on sentence wrap-up processing (Just & Carpenter, 1980).

If the observed effects of causal relatedness depend on sentence wrap-up processes, interrupting those processes or any other processing occurring during this interval, should selectively attenuate discourse-level congruence effects. To test this possibility, participants in Experiment 3 heard the same context stories as in Experiment 1, with two differences: only the causal relatedness of the probe words was manipulated (not the lexical relatedness), and the inter-stimulus interval was either 400 ms (*delayed* conditions, replicating the *causal* conditions of Experiment 1) or 0 ms (*immediate* conditions). If the early effects of causal relatedness do indeed depend on having ample processing time to prepare for the probe word, then those components of the causal relatedness effect should be eliminated or attenuated in the *immediate* conditions relative to the *delayed* conditions.

Methods

Participants

Participants were 20 UC San Diego undergraduates (12 female) who had not participated in Experiment 1 or in any norming study of these stimuli. They were compensated in course credit and/or payment. All were right-handed, as assessed by the Edinburgh handedness inventory (Oldfield, 1971). Five participants reported left-handed members of their immediate family. All participants spoke English as their first language, had normal or corrected-to-normal vision and were free of neurological disorders and psychiatric medications. Their ages ranged from 18 to 34, with a mean age of 20.4 years. An additional 2 participants were excluded from analysis due to movement and/or blocking artifacts contaminating more than 25% of trials. All participants gave informed consent to participate.

Materials

Experimental materials were identical to those used in Experiment 1, except that only the *causal related* and *causal unrelated probes* were used. This exclusion of the *lexical related* and *lexical unrelated* conditions was necessary to make room in the design for the manipulation of probe word onset latency. The *causal related* and *causal unrelated* probe words were therefore presented in the *delayed* condition 400 ms after the final word of the context story (the same timing paradigm used in Experiments 1 and 2), as well as in the *immediate* condition 0 ms after the final word of the context story. This design resulted in a total of four conditions – *immediate causal related, immediate causal unrelated, delayed causal related,* and *delayed causal unrelated* – with 40 items in each condition. A four-cell Latin Square design ensured that each participant saw each probe word once and that each probe word appeared once in each condition across each group of four participants.

Procedure

After filling out an Edinburgh handedness inventory (Oldfield, 1971), a demographic questionnaire and a consent form, participants completed a listening variant of the reading span test (Daneman & Carpenter, 1980) on a computer. This test was conducted identically to that administered in Experiment 1.

After the listening span test, participants were fitted with an EEG cap and seated in a dimly lit, electrically shielded chamber, one meter away from the computer monitor on which stimuli appeared. Audio narratives were played through a pair of external speakers flanking the monitor. Throughout the experiment, the monitor displayed a black field with an orange fixation dot in the center.

Participants were told that they would hear two-sentence-long short stories, and that at the end of each story a word would be flashed in the center of the screen. The subject's task was to listen to the stories and try to understand them, and to pay attention to the words that appeared in the center of the screen. Participants were instructed to fixate on the central dot, to avoid blinking whenever the audio was not playing, and to refrain from body movements. Before beginning the experiment, participants were given two practice trials to get used to inhibiting blinks at the proper times. Throughout the experiment, participants were given feedback on their blinking behavior. The total EEG recording time was about 45 minutes.

In each trial, participants fixated on the center of the monitor while listening to the audio narrative stimulus. Each sentence of the narrative lasted between 3 and 8 seconds, and there was a pause of 300 ms between sentences. Following the second sentence of each trial, the probe word was presented after 0 ms in the *immediate* conditions, and after 400 ms in the *delayed* conditions. As in Experiments 1 and 2, the probe word was displayed for 250 ms.

EEG Recording

EEG data was recorded in an identical manner to that in Experiments 1 and 2.

Analysis

As in Experiments 1 and 2, ERPs to visual probe words were time-locked to word onset and averaged in a time window spanning 100ms pre-onset to 920 ms post-onset. The period from 100ms pre-onset to stimulus onset served as the baseline. During data analysis, critical EEG epochs were examined and trials containing blocking and eye movement artifacts were rejected. If more than 25% of a participant's critical epochs were rejected in this way, that participant was excluded from analysis. In all, 11% of trials were rejected due to artifacts.

Analysis of ERP components were conducted similarly to those in Experiment 1. Initial analyses included all electrodes and were focused on three time windows: 150-300 ms post-onset (P2 component), 300-500 ms (N400 component), and 500-800 ms (post-N400 effects). Measurements were analyzed with a repeated measures ANOVA, using a 2 (relatedness: related vs. unrelated) x 2 (presentation latency: immediate vs. delayed) x 29 (electrode) design. Planned follow-up comparisons were always conducted within levels of the presentation latency factor using a 2 (relatedness: related vs. unrelated) x 29 (electrode) design. To compensate for violating the sphericity assumption, the Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied where appropriate.

In order to examine the scalp topographies of the delayed and immediate ERP effects of causal relatedness in each time window, *t*-max permutation tests were conducted exactly as in Experiment 1, whenever a Relatedness effect or a Relatedness X Electrode interaction approached significance. In both the delayed and immediate conditions, a two-tailed *t*-test was performed at each electrode site at each 4-ms time point, testing the null hypothesis that the voltage difference between the *related* and *unrelated* conditions was 0 μV. The *t*-max permutation procedure was used to correct for multiple comparisons, using a family-wise alpha level of 0.05. Permutation tests were carried out with the 'tmaxGND' function using 2000 permutations per time point in the Mass Univariate ERP Toolbox for MATLAB (Groppe, Urbach & Kutas, 2011). Raster plots showing significant t-tests in all three analyzed time windows can be found in Figure 10.

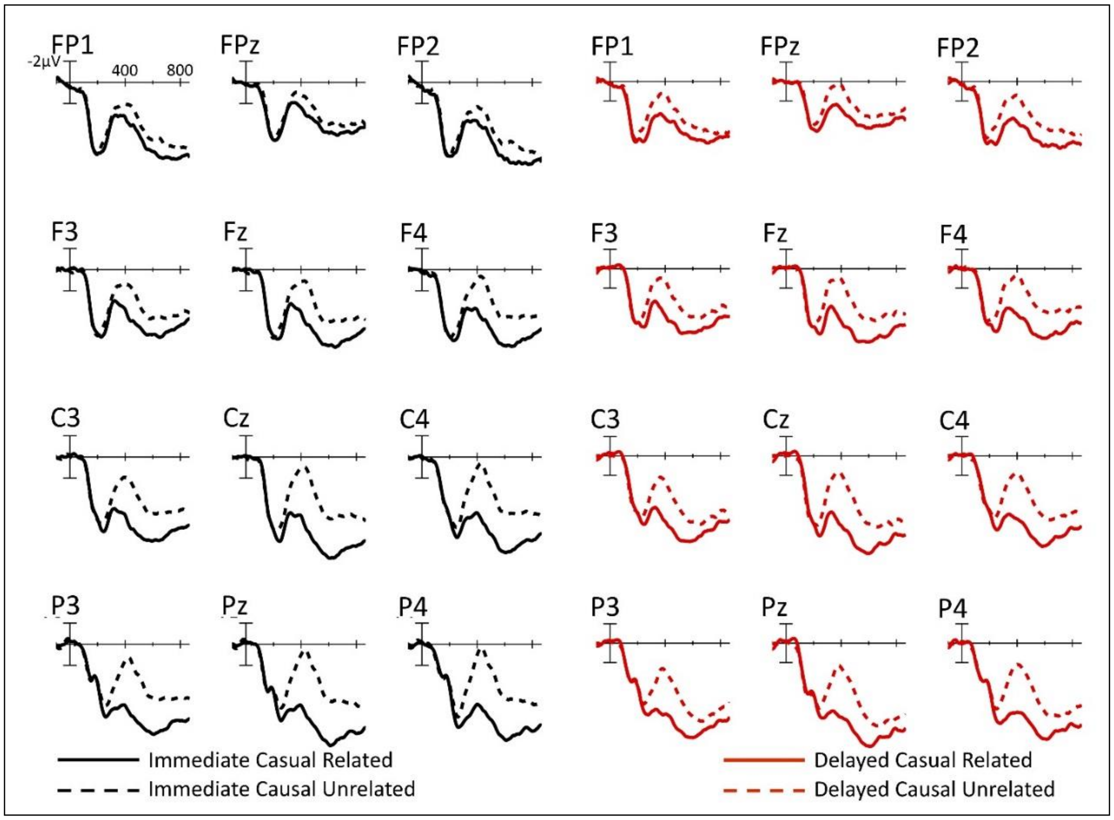

Figure 8: ERP waveforms at 12 electrode sites depicting data from Experiment 2. ERP responses to related items are drawn in solid lines and unrelated items in dashed lines. The immediate causal relatedness effect is shown in the left column in black, and the delayed causal relatedness effect is shown in the right column in red. Negative voltage is plotted upward, and the data were low-pass filtered at 10 Hz for presentation purposes.

Results

P2 Component (150-300 ms)

An initial repeated measures ANOVA revealed a marginal main effect of relatedness ($F[1,19] = 3.6$, $p = 0.07$) on the P2 component, but no main effect of presentation latency on the P2 component, nor interaction between these components significant (all *F*s < 1). Planned follow-up comparisons revealed that in the *delayed* conditions, the *causal related* probes elicited marginally larger P2 responses than the *causal unrelated* probes ($F[1,19] = 3.3$, $p = 0.09$). In the *immediate* conditions, the *causal related* probes also elicited numerically more positive ERPs in the P2 window, but this difference was not significant ($F[1,19] = 2.3$, $p = 0.14$).

The delayed and immediate causal relatedness effects were analyzed for topographical distribution using *t*-max permutation tests at each time point. In the immediate condition, no significant t-scores were observed in the P2 window. However, in the delayed condition, significant corrected *t*-tests were observed at FP2 beginning at 238 ms and then at other anterior sites beginning at 274 ms. Significant voltage differences of the same polarity began at 286 ms at right temporal sites, perhaps representing the onset of the N400 effect. In all significant comparisons, the *delayed causal related* condition elicited more positive voltage than the *delayed causal unrelated* condition.

N400 Component (300-500 ms)

A repeated measures ANOVA using the relatedness and presentation latency factors revealed that *unrelated* probe words elicited larger N400 than *related* probe words, with a mean effect size of 3 μV ($F[1,19] = 56.8$, $p < 0.0001$). The main effect of presentation latency was not significant: *delayed probes* elicited N400 0.6 μV smaller than *immediate* probes ($F[1,19] = 3.4$, $p = 0.08$). The relatedness X presentation latency interaction effect was also non-significant, indicating that N400 relatedness effects were approximately the same size in the *delayed* condition (3.2 μV) and the *immediate* condition (2.9 μV) ($F[1,19] < 1$). Planned follow-up comparisons testing effects of relatedness showed that the N400 elicited by *causal related* probes was smaller than that elicited by *causal unrelated* probes in both the *delayed* ($F[1,19] = 37.6$, $p < 0.0001$; relatedness x electrode: $F[28,532] = 5.6$, $p = 0.002$) and *immediate* conditions ($F[1,19] = 47.4$, $p < 0.0001$; relatedness x electrode: $F[28,532] = 15.6$, $p < 0.0001$).

The delayed and immediate causal relatedness effects were analyzed for topographical distribution using *t*-max permutation tests at each time point. In both timing conditions, the *causal related* condition elicited significantly more positive ERPs than the *causal unrelated* condition. Figure 10 suggests that both N400 effects were somewhat right-lateralized in their early stages, but covered most scalp sites by about 340 ms. Further, the N400 effect was more broadly distributed in the *delayed* than the *immediate* condition.

Post-N400 Effects (500-800 ms)

An initial repeated measures ANOVA revealed that *causal unrelated* probes elicited more negative voltage than *causal related* probes ($F[1,19] = 64.6$, $p < 0.0001$) and that probes in the *immediate* condition elicited more negative voltage overall than those in the *delayed* condition ($F[1,19] = 4.5$, $p = 0.047$). However, the relatedness and presentation latency factors did not interact ($F[1,19] = 1.3$, $p = 0.3$) and the relatedness X latency X electrode interaction was marginal ($F[28,532] = 2.3$, $p = 0.07$). Planned follow-up comparisons were conducted within the *immediate* and *delayed* factors. Within the *immediate* factor, the *unrelated* probe words elicited ERPs an average of 1.7 μV more negative than the *related* probe words ($F[1,19] = 31.8$, $p < 0.0001$). A significant relatedness X electrode interaction reflected the right-central focus of this effect ($F[28,532] = 9.5$, $p < 0.0001$). Within the *delayed* factor, *unrelated* probes elicited post-N400 potentials roughly 1.8 μV more negative than *related* probes ($F[1,19] = 23.8$, $p = 0.0001$). A significant relatedness X electrode interaction effect reflected the right anterior focus of this late difference ($F[28,532] = 4.0$, $p = 0.005$).

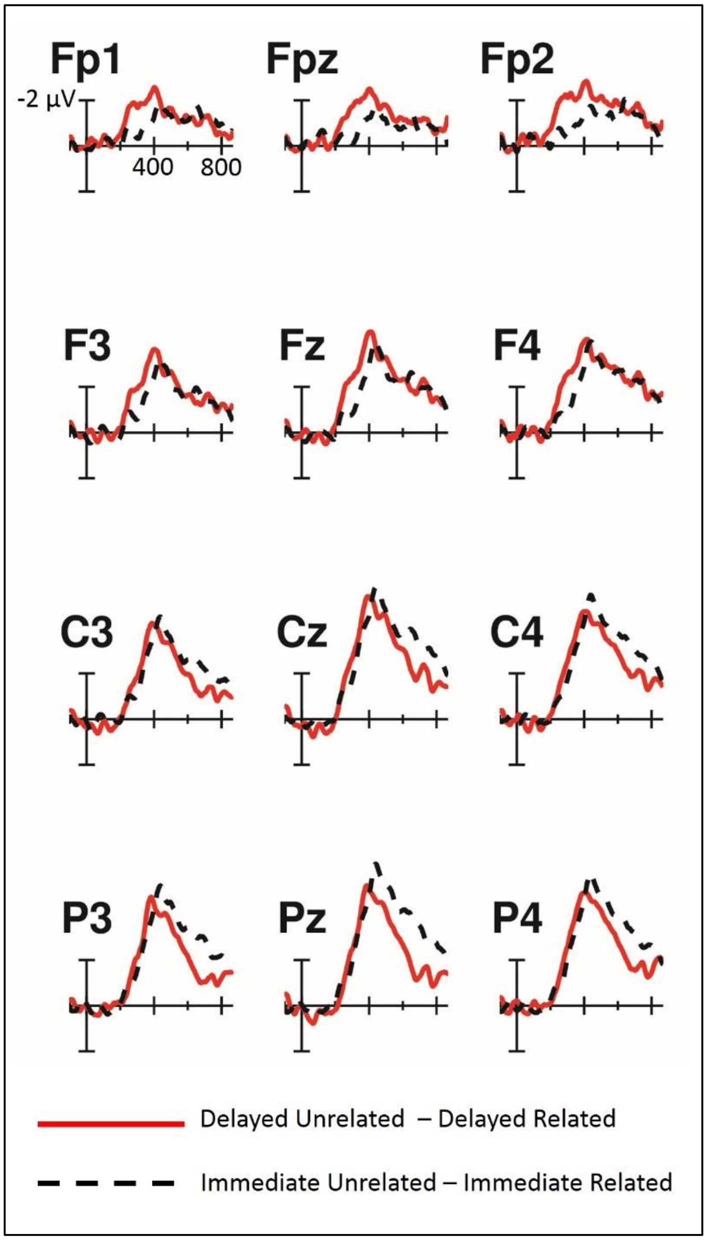

Figure 9: Difference waves showing causal relatedness effects (*causal unrelated – causal related)* in the *delayed* conditions (solid red line) and the *immediate* conditions (dashed black line) of Experiment 3. Negative voltage is plotted upward.

The delayed and immediate causal relatedness effects were analyzed for topographical distribution using *t*-max permutation tests at each time point. Again, in both timing conditions, the *causal related* condition elicited significantly more positive voltage than the *causal unrelated* condition. However, this effect ended much earlier in the delayed condition (at about 630 ms, though continuing until about 730 ms on some right anterior sites) than in the immediate condition (until the end of the epoch on a variety of right hemisphere and midline sites). The sustained negativity thus persisted longer and was distributed more like the N400 effect in the immediate condition as opposed to the delayed.

Post Hoc Comparisons on the Post-N400 Epoch

Observed differences in the topographic profile of the relatedness effects in the immediate and delayed conditions sparked our curiosity as to whether those differences were primarily attributable to ERPs elicited by related or unrelated probes. Consequently, we conducted separate post hoc analyses of the presentation latency effect for each type of probe. A pair of 2 (latency: immediate vs. delayed) x 29 (electrode) repeated measures ANOVA tests showed that while the *causal related* probe words did not differ as a function of latency ($F < 1$), the *unrelated* probe words did ($F[1,19] = 7.6$, $p = 0.01$). A significant Presentation Latency X Electrode interaction reflected the left occipito-parietal focus of this voltage difference $F[28,532] = 12.8$, $p < 0.0001$). See Figure 11 for waveforms and a scalp plot.

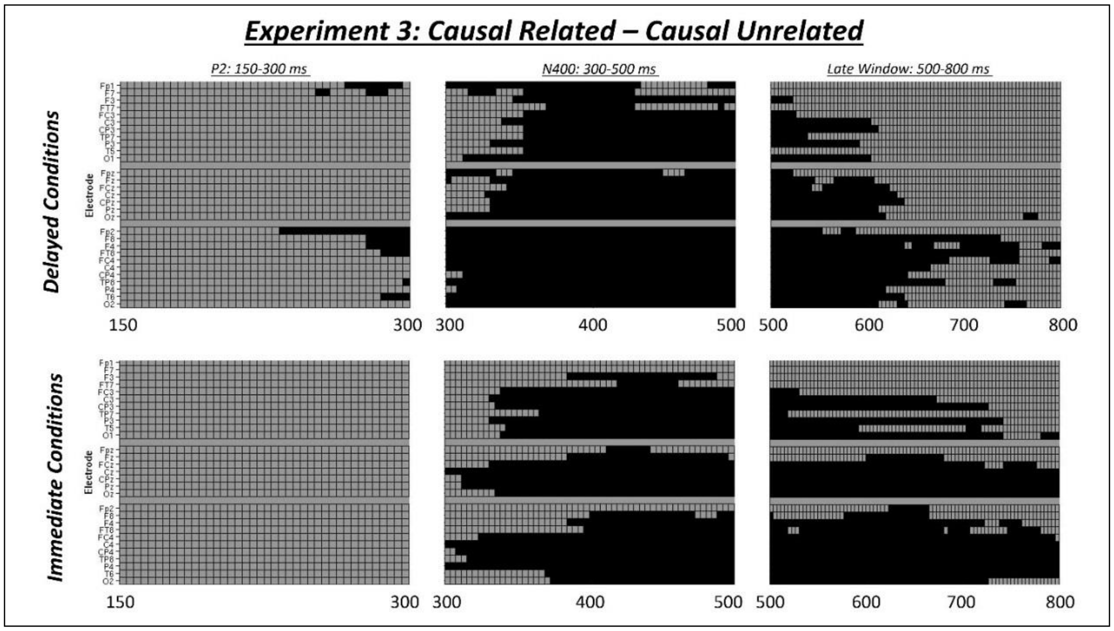

Figure 10: Raster plots of two-tailed *t*-tests against the null hypothesis that the causal relatedness effect is equal to 0 µV. The left column shows each time window in the immediate condition and the right column shows each time window in the delayed conditions. Black squares represent p-values ≤ 0.05. *t*-max permutation tests were used to correct for multiple comparisons, with 2000 permutations at each 4-ms time point.

Listening Span Scores

Participants' listening span scores were calculated using the same method as in Experiment 1. Again, participants were divided into high- and low- listening span (LS) groups using a median split on listening span scores. The high-span group had a mean span of 4.6 (SD: 0.5), and the low-span group had a mean span of 3.1 (SD: 0.9). Both groups of participants showed effects in the same time windows of the same polarity (i.e., *unrelated* condition was more negative), with the exception that there was no significant P2 effect of causal relatedness in the *immediate* conditions. Again, no listening span comparisons reached significance (all *p*-values > 0.1)

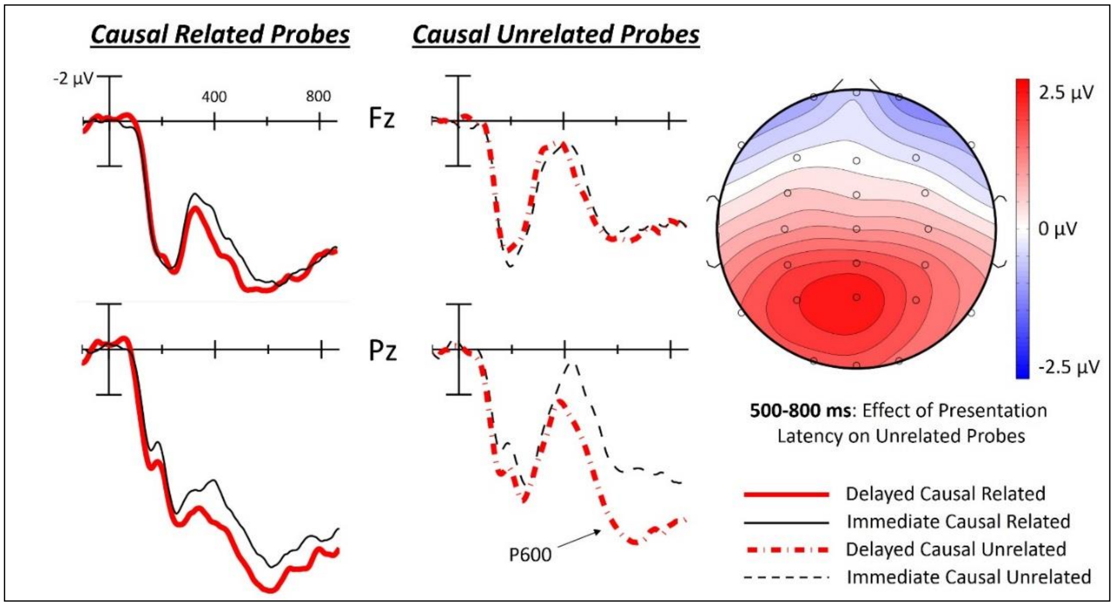

Figure 11: Comparison of the Delayed and Immediate Causal Unrelated conditions in Experiment 3. Negative voltage is plotted upward, and waveforms are low-pass filtered at 10 Hz for presentation purposes. The topographical plot represents scalp distribution of difference wave calculated by subtracting the Immediate Causal Unrelated condition from the Delayed Causal Unrelated condition, using mean amplitude in the 500-800 ms time window.

Discussion

In Experiment 3, we manipulated presentation latency and causal relatedness to investigate the role of sentence-final wrap-up processes on the discourse-level congruence effects observed in Experiment 1. In the *delayed* presentation conditions (identical to the *causal* conditions in Experiment 1), a 400-ms delay after the offset of the story-final word allowed participants time to conduct sentence-final wrap-up before viewing the *causal related* or *causal unrelated* probe word. In the *immediate* conditions, the *causal related* or *causal unrelated* probe word was presented immediately at the offset of the story-final word, interrupting whatever sentence-final wrap-up processes may have been occurring. Although this timing manipulation was quite small, and although it did not impact the processing of any words in the story context (except perhaps for the final word), it had large effects on the ERP response to the visual probe words. In particular, decreasing the inter-stimulus interval reduced or eliminated the P2 effect observed in Experiment 1 in which the *causal related* probes elicited larger amplitude P2 than did *causal unrelated* probes.

Further, while the latency manipulation did not influence the size of the N400 relatedness effect, it did influence the voltage differences *after* the N400, as analysis revealed larger effects in the *immediate* than the *delayed* condition. Post hoc analyses suggested that relatedness effects on the sustained negativity over frontal sites were similar, with differences stemming from ERPs to *causal unrelated* probes over posterior scalp (see Figure 9). *Causal unrelated* probes elicited a P600 only in the *delayed* condition, which reduced the size of the negative-going relatedness effect in the *delayed* relative to the *immediate* condition. Previously linked to word-form predictions that fail at both the semantic and word-form level (Van Petten & Luka, 2012), the P600 in the present study likely reflects participants' recognition of the unrelated probes as unexpected word-forms. As for the P2, the P600 relatedness effect was absent from the immediate condition. Accordingly, these data suggest that elicitation of both ERP effects indexing failed word-form predictions – viz., the P2 and the P600 – required the additional processing available in the delayed condition.

General Discussion

The overall goal of the study was to compare the predictions of three broad classes of processing models with respect to the relative orderings of different types of contextual facilitation effects on word processing. In Experiment 1, discourse-level effects began earlier than lexical-level effects, supporting the flexible parser design advocated by dynamic models. Experiment 2 showed that the results of Experiment 1 could not be attributed to lexical priming effects in the *causal related* condition, and that the size and scalp distribution of the lexical association effect differed depending on pragmatic factors such as the presence/absence of a discourse-level manipulation in the experiment. The latter result also supported the dynamic framework, because the dynamic prioritization of different types of contextual information can be reflected in the engagement of different brain systems to process it. Experiment 3, finally, showed that the early discourse-level effect observed in Experiment 1 was not solely an effect of the probe word or its relationship to its context. It was also influenced by the timing allowed between the end of the context story and the onset of the probe word. Analysis of the ERP effects (explained in more detail below) suggested that this timing manipulation impacted participants' ability to form lexical predictions. If true, this point also favors dynamic models, which are currently the only theoretical framework able to handle lexical prediction. Below, we discuss our results in detail.

N400 effects

In Experiment 1, the manipulation of causal relatedness elicited a large N400 effect, similar to that observed in previous causal inference studies (e.g., Kuperberg et al., 2011). The causal relatedness effect was larger than that of lexical relatedness, a difference due entirely to greater N400 reduction in the *causal related* condition than in the *lexical related* condition, as the two *unrelated* conditions elicited nearly identical N400 responses. This result matches the contemporary theoretical position that by default a word elicits a large N400 response, which can be reduced to varying degrees by supportive context or lexico-semantic association (Kutas & Federmeier, 2011; Van Berkum, 2009). It could be suggested that the difference in N400 relatedness effects in Experiment 1 arose due to the relatively long (400 ms) interval between the offset of the final word in the context story: perhaps the onset of the probe word allowed for causal inferential information to accumulate and for lexical associative information to decay, and with a different timing paradigm the relative effect sizes would be reversed. Given that we chose this timing paradigm to maximize cross-modal lexical priming effects (cf. Holcomb & Anderson, 1993), this is unlikely. In addition, the results of Experiments 2 and 3 suggest that this is not the case. In Experiment 2, the same timing paradigm was used but the discourse context needed to form a causal inference was not available. The lexical relatedness N400 effect size increased relative to Experiment 1, suggesting that Experiment 1's lexical relatedness effect was reduced by the comprehender's focus on the more salient causal relatedness manipulation. In Experiment 3, we reduced the inter-stimulus interval (ISI) between the end of the context and the onset of the probe word to

0 ms. This timing manipulation did not affect the causal relatedness N400 effect (although it did affect other ERP components), indicating that presentation latency was not an important factor for determining N400 relatedness effects (Kutas, 1993).

It is important to note that the difference in N400 elicited by the *causal related* probes between Experiments 1 and 2 contradicts traditional serial processing accounts. Current theories of the N400 identify it with access to lexico-semantic information (Kutas & Federmeier, 2011; Van Berkum, 2009), and serial processing theories denounce discourse-level influence on lexical access (Kintsch, 1988). This serial processing claim is contradicted by the observation that the *causal related* probes elicit a smaller N400 than the *lexical related* probes in Experiment 1, but the opposite is observed in Experiment 2 (see figure 5). This contrast between Experiments 1 and 2 could only occur if the discourse-level factor of causal relatedness reduced N400 responses in the *causal related* condition of Experiment 1, or if the pragmatic effects of the overall experimental structure more greatly facilitated N400 reduction in the *lexical related* condition of Experiment 2, or both. None of those results are compatible with a theory in which lexical access occurs at a “level” of processing unaffected by discourse-level information.

The pattern of N400 results in Experiments 1 and 3 also differs strikingly from the naming time results collected by Till and colleagues (1988), who also manipulated both causal and lexical relatedness in the context, as well as the presentation latency of the probe words to which participants responded. Till and colleagues observed significant priming due to lexical relatedness at shorter latencies than causal relatedness, interpreting this result in favor of serial processing. In Experiment 1, we observed effectively the opposite result: earlier onset of causal relatedness ERP effects. This contrast between our study and Till et al.'s likely arises from the fact that our materials were designed and normed to evoke a single, strong causal inference, while Till and colleagues report that the norming of their context paragraphs elicited a variety of inferences, with the mode response being an inference "of moderate strength." Comparing these studies therefore suggests a relationship between the strength of an inference and the onset latency of its influence on word processing, a position incompatible with serial processing models but a natural fit for dynamic and memory-based processors.

Comparing Experiments 1 and 2 also highlights the suppression of lexical relatedness effects by other factors, undermining traditional ideas that the lexical access process proceeds similarly under different circumstances (Camblin et al., 2007; Boudewyn et al., 2012; 2013). The methodological implication is that using lexical priming paradigms to study lexical activation may lead to a dangerously incomplete picture of the process. Additionally, critiques that experiments involving discourse stimuli inadequately control for lexical factors may be exaggerated, as the influence of lexical factors is greatly diminished in richer contexts (for a review see Van Petten, 1995).

Sustained Negativity Effects

Sustained negativities were observed in all three experiments in the present study, for the causal relatedness manipulations in Experiments 1 and 3, and for the lexical relatedness manipulation in Experiment 2. Similar post-N400 negativities have been observed and discussed with increasing frequency in recent years, as ERP studies of discourse processing have become more popular. For example, Van Berkum and colleagues (1999) in their studies of pronoun interpretation in ambiguous contexts reported a sustained negativity that they referred to as the “nRef” effect that was largest when a pronoun had several possible referents. Sustained negativities have also been observed in studies of logical reasoning (Pijnacker et al., 2011), discourse model revision (Baggio, et al., 2008; but see also Kuperberg et al., 2011), and frame-shifting in jokes (Coulson & Kutas, 2001; Coulson & Wu, 2005). Sustained negativities have typically been interpreted as reflecting operations in working memory (Otten & Van Berkum, 2008; Van Berkum, 2009; Lau et al., 2013). The causal sustained negativity effects observed in the present study fit this picture well, and they likely reflect pragmatic processes triggered by the probe words that do not match the inference made during the context story.

Interestingly, manipulations of lexical relatedness gave rise to a sustained negativity in Experiment 2, in which the probe was preceded only by a single spoken word, but not in Experiment 1, involving discourse contexts. Sustained negativities have been observed in some single-word priming experiments (e.g., Lau et al., 2013; Swaab, Baynes & Knight, 2002), but not in others (e.g., Brown, Hagoort & Chwilla, 2000; Kutas & Hillyard, 1989). In the word priming experiments that elicited sustained negativities, there has generally been some aspect of the experiment that directed the participants’ attention to the semantic content of the words. For instance, Lau and colleagues (2013) instructed participants to respond when they noticed an animal word. Swaab and colleagues (2002) used a semantic relatedness judgment task. However, studies that have used less semantically focused tasks such as lexical decision (Brown et al., 2000) and letter search (Kutas & Hillyard, 1989) have generally not reported sustained negativities. Deeper semantic processing may encourage conscious reflection on prime-probe relatedness, resulting in attempts to construct semantic relationships between primes and unrelated probes. This process is reflected in a sustained negativity. It is possible that the relatively high proportion of *lexical related* probes in Experiment 2 (50%) elicited a similar effect. In Experiment 1, however, participants' attention was captured by the more obvious causal relatedness manipulation.

The sustained negativity fits well with what traditional serial models of processing would term a "late" or "post-lexical" process: it appears to index some kind of higher-level cognitive process indexing conscious attempts at meaning construction (Lau et al., 2013; Van Berkum, 2009). However, the sustained negativity also departs from the classical picture of a post-lexical discourse processing effect in several important ways. First, it can be modulated by differences purely in lexical association, as in Experiment 2. Secondly, it tends to overlap the N400 component, as shown in joke studies (Coulson & Kutas, 2001) and nRef studies (for review see Van Berkum, 2009) and suggested by the broadly distributed negativity observed during Experiment 1's *causal* N400 window (see Figures 2 and 3). These facts indicate that mapping the relative timings of processing effects onto "levels" of representation is likely to lead to an incomplete picture of the underlying processing architecture. The sustained negativity effects observed here fit better with the memory-based processing idea of pragmatic processing triggered by a discourse incongruity (see Huang & Gordon, 2011).

P2 and Predictive Processing Effects

In the present study, P2 effects were observed on the causal relatedness manipulation in Experiment 1 and replicated in the *delayed* conditions of Experiment 3. A P2 effect was not observed in the lexical relatedness manipulation of Experiments 1, nor was it significant in the lexical relatedness manipulation of Experiment 2. These data obviously contradict serial processing models, because they represent a clear case of a discourse-level manipulation affecting processing before a lexical-level manipulation. The P2 has been observed and discussed primarily in the context of visual attention and target detection studies. It is thought to originate from greater activity in orbitofrontal cortex related to detecting task-relevant stimuli and stimulus evaluation (Potts, Liotti, Tucker & Posner, 1996; Potts, 2004). In the field of language processing, the P2 has figured primarily in the work of Federmeier and colleagues. Her account of linguistic P2 effects is that the enhanced P2 responses to visual words occur when expectancy of a particular word form leads to an increased brain response when that word is encountered. Enhanced P2 effects have been observed in high-constraint sentence contexts (Wlotko & Federmeier, 2007) as well as in response to pragmatic manipulations that influenced expectations for particular words (Regel, Coulson & Gunter, 2010).

Because of its association with word-form prediction, the P2 effects in the present study favor dynamic models that incorporate word-form prediction (e.g., Elman, 1990; Altmann & Mirkovic, 2009). Memory-based processing theories have not as yet incorporated recent findings on word-form prediction, and are therefore incompatible with these results. However, a somewhat small extension of the memory-based framework could accommodate pre-N400 effects of word-form prediction. If we suppose that incoming items resonate not only with semantic memory but also with a separate, sensory working memory buffer for pre-activated word forms, then the memory-based processing could be brought into line with the results of contemporary P2 findings in language processing, as well as Dikker and Pylkkänen's (2011) Sensory Hypothesis, which locates a word-form prediction buffer in visual cortex.

Indeed, the P2 effects observed in the present experiment is worth comparing to another early predictive processing effect that has been discussed extensively by Dikker, Pylkkänen, and colleagues: the early MEG effect (the M100 – usually peaking just before 100 ms) originating in visual cortex. Like the P2, the M100 is sensitive to the predictability of a word in its context (Dikker & Pylkkänen, 2011). Also like the P2, it is sensitive specifically to the sensory characteristics of the word-form, rather than to its semantic or syntactic content (although semantic and syntactic information certainly contribute to the inception of a lexical prediction) (Dikker, Rabagliati, Farmer & Pylkkänen, 2010). Research on visual processing in target detection paradigms suggests the N1 and P2 are frequently elicited by the same stimuli, being larger for targets, with the N1 reflecting early visual processing and the P2 evaluative processing relevant to its target status (Potts, 2004). The present experiments had too few stimuli to license measuring N1 effects, but see Figure 2 for hints of a small N1 difference in the *causal* but not *lexical* conditions of Experiment 1.

Not only was the P2 effect absent in Experiment 3's *immediate* conditions, but so was the P600 that overlaid the sustained negativity in the *delayed unrelated* condition. P600s have recently been linked to failed lexical predictions in discourse incongruent situations (Van Petten & Luka, 2012). The absence of both the P2 and the P600 in the *immediate* conditions of Experiment 3, despite no diminution of the N400 effect, indicates that although the same amount of semantic information was available in both timing conditions, lexical prediction only took place in the *delayed condition.* This suggests that at least in some cases, lexical pre-activation takes longer to develop than semantic pre-activation. It is possible that in the present experimental paradigm, the set of operations included in sentence wrap-up include converting sufficiently constrained semantic expectations into word-form predictions. This may be due to the structure of these experiments: the end of the second sentence acts as a cue to expect a visual word, which may trigger the construction of a word form prediction. More generally, it is possible that predictions are generated differently in different types of constraining contexts. In situations like Experiments 1 and 3 of the present study, when the crucial information is provided by a bridging inference, comprehenders may generate word form predictions as part of a larger process of inferring that an explanation of the coherence gap is soon to follow (Rohde, Levy & Kehler, 2011). However, when the information needed to form a lexical prediction can be extracted from the statistical properties of the current sentence, then a word form prediction may be automatically generated intra-sententially (e.g., DeLong, Urbach & Kutas, 2005; Wlotko & Federmeier, 2007; Federmeier, Wlotko, De Ochoa-DeWald & Kutas, 2007).

Conclusion

These experiments show that a discourse-level factor, namely causal inference, can affect all stages of word processing, from visual processing of the word form to semantic processing indexed by the N400 to a variety of integrative and pragmatic processes indexed by the later portion of the wave-form. Huang and Gordon (2011) have suggested that discourse-level factors can affect early stages of word processing via facilitation due to discourse congruence and later stages of processing when discourse incongruity triggers pragmatic processes. Some of our results, in particular the P2 and sustained negativity effects in Experiment 1, appear to match this dichotomy between early facilitation effects and late processes triggered by incongruities. However, Experiment 3 complicates this picture by showing that both early and late effects can depend to some extent on whether comprehenders were able to use discourse-level information to make word-form predictions. The importance of word-form prediction is central to some versions of the dynamic processing framework, in particular models that are trained on the success of their predictions of upcoming items (Elman, 1990) including semantic and word-form information (Altmann & Mirkovic, 2009). Overall, dynamic models that incorporate prediction give the best account of these data because they because they naturally account for semantic and discourse effects on perceptual processing. Such models also fit the recent emphasis in the cognitive neurosciences on predictive processing as a basic aspect of brain function (Bar, 2007; Pickering & Garrod, 2013; Clark, 2012).

References

Altmann, G., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. Cognitive Science, 33(4), 583-609.

Baggio, G., Van Lambalgen, M. & Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. J. Mem. Lang. 59, 36-53.

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

Boudewyn, M. A., Gordon, P. C., Long, D., Polse, L., and Swaab,T.Y. (2012). Does discourse congruence influence spoken language comprehension before lexical association? Evidence from event-related potentials. Lang. Cogn. Process. 27, 698–733.

Boudewyn, M., Long, D. & Swaab, T. (2013). Effects of working memory span on processing of lexical associations and congruence in spoken discourse. Frontiers in Psychology 4 (60), 1-16.

Brown, C. M., Hagoort, P., & Chwilla, D. J. (2000). An event-related brain potential analysis of visual word priming effects. Brain and Language, 72(2), 158-190.

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

Camblin, C., Gordon, P. & Swaab, T. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. Journal of Memory and Language 56, 103–128.

Clark, A. (2012). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci, 1-86.

Coltheart, M. (1981). The MRC psycholinguistic database. The Quarterly Journal of Experimental Psychology, 33(4), 497-505.

Coulson, S. (2006). Constructing meaning. Metaphor Symb. 21 (4), 245-266.

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

Coulson, S., Federmeier, K., Van Petten, C. & Kutas, M. (2005). Right Hemisphere Sensitivity to Word- and Sentence-Level Context: Evidence From Event-Related Brain Potentials. J. Exp. Psych.: Learn. Mem. Cogn. 31 (1), 129-147.

Coulson, S. & Wu, Y. (2005). Right hemisphere activation of joke-related information: An event-related potential study. J. Cogn. Neurosci. 17(3), 494–506.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of verbal learning and verbal behavior, 19(4), 450-466.

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

Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. Psychological Science, 21(5), 629-634.

Dikker, S., & Pylkkanen, L. (2011). Before the N400: Effects of lexical–semantic violations in visual cortex. Brain and language, 118(1), 23-28.

Elman, J. L. (1990). Finding structure in time. Cognitive science, 14(2), 179-211.

Federmeier, K. & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. Journal of Memory and Language 41, 469–495

Federmeier, K. D., & Kutas, M. (2002). Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres. Neuropsychologia, 40(7), 730-747.

Federmeier, K. D., Mai, H., & Kutas, M. (2005). Both sides get the point: Hemispheric sensitivities to sentential constraint. Memory & Cognition, 33(5), 871-886.

Federmeier, K., Wlotko, E., De Ochoa-Dewald, E. & Kutas M. (2007). Multiple effects of sentential constraint on word processing. Brain Research 1146, 75–84

Forster, K. I. Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), New Approaches to language mechanisms. Amsterdam: North-Holland, 1976.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. The Quarterly Journal of Experimental Psychology, 33(4), 465-495.

Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. Journal of Memory and Language, 42(4), 526-544.

Gerrig, R.J., & O'Brien, E.J. (2005). The scope of memory-based processing. Discourse Processes, 39 (2 & 3), 225-242.

Greenhouse, S. & Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, 2. 95-112.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). “Some,” and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. Cognition, 116(1), 42-55.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. Psychophysiology, 48(12), 1711-1725.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. (2004). Integration of Word Meaning and World Knowledge in Language Comprehension. Science, 304, 438-441.

Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. Language and Cognitive Processes, 8(4), 379-411.

Huang, Y. T., & Gordon, P. C. (2011). Distinguishing the time-course of lexical and discourse processes through context, coreference, and quantified expressions. Journal of Experimental Psychology: Learning, Memory, and Cognition, 37(4), 966.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. Psychological review, 87, 329-354.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review, 95 (2), 163-182.

Kintsch, W., & Mross, E. F. (1985). Context effects in word identification. Journal of Memory and Language, 24(3), 336-349.

Kučera, H., & Francis, W. N. (1967).Computational analysis of present day American English. Providence, RI: Brown University Press.

Kuperberg, G., Paczynski, M. & Ditman, T. (2011). Establishing Causal Coherence across Sentences: An ERP Study. J. Cogn. Neuro. 23 (5), 1230-1246.

Kutas, M. & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205.

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

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

Kutas, M. & Federmeier, K. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). Annual Review of Psychology 62, 621–47.

Landauer, T., Foltz, P. & Laham, D. (1998). Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.

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.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. Psychological review, 101(4), 676.

McKoon, G. & Ratcliff, R. (1992). Inference during reading. Psychological Review 99 (3), 440-466.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. Journal of Memory and Language, 38(3), 283-312.

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

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). The University of South Florida word association norms. Tampa, FL, University of South Florida.

Oldfield, R. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

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

Pickering, M. J., & Garrod, S. (2013). Forward models and their implications for production, comprehension, and dialogue. Behavioral and Brain Sciences, 49-64.

Pijnacker, J., Guerts, B., Van Lambalgen, M., Buitelaar, J. & Hagoort, P. (2010). Reasoning with Exceptions: An Event-related Brain Potentials Study. J. Cogn. Neuro. 23 (2), 471-480.

Potts, G. F. (2004). An ERP index of task relevance evaluation of visual stimuli.Brain and cognition, 56(1), 5-13.

Potts, G. F., Liotti, M., Tucker, D. M., & Posner, M. I. (1996). Frontal and inferior temporal cortical activity in visual target detection: Evidence from high spatially sampled event-related potentials. Brain Topography, 9(1), 3-14.

Regel, S., Coulson, S. & Gunter, T. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. Brain Research, 1311, 121-135.

Rohde, H., Levy, R., & Kehler, A. (2011). Anticipating explanations in relative clause processing. Cognition, 118(3), 339-358.

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

Swaab, T. Y., Baynes, K., & Knight, R. T. (2002). Separable effects of priming and imageability on word processing: an ERP study. Cognitive Brain Research, 15(1), 99-103.

Tabor, W., & Tanenhaus, M. K. (1999). Dynamical models of sentence processing. Cognitive Science, 23(4), 491-515.

Till, R., Mross, E. & Kintsch, W. (1988). Time-course of priming for associate and inference words in a discourse context. Memory & Cognition 16 (4), 283-298.

Van Berkum, J. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In U. Sauerland, & K. Yatsushiro (Eds.), Semantics and pragmatics: From experiment to theory (pp. 276-316). Basingstoke: Palgrave Macmillan.

Van Berkum, J. J., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. Journal of cognitive neuroscience, 11(6), 657-671.

Van Berkum, J. J., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. Cognitive Brain Research, 17(3), 701-718.

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. Language & Cognitive Processes 8, 485-531.

Van Petten, C. (1995). Words and sentences: Event-related brain potential measures. Psychophysiology, 32, 511-532.

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

Wlotko, E. & Federmeier, K. (2007). Finding the right word: Hemispheric asymmetries in the use of sentence context information. Neuropsychologia 45, 3001-3014.

Yang, C., Perfetti, C. & Schmalhofer, F. (2007). Event-Related Potential Indicators of Text Integration Across Sentence Boundaries. Journal of Experimental Psychology: Learning, Memory, and Cognition 33 (1), 55–89.

Acknowledgments

Chapter 2, in full, is currently being prepared for submission for publication of the material. Davenport, Tristan S.; Coulson, Seana. The dissertation author was the primary investigator and author of this material.

Chapter 3:

ERP and Self-Paced Reading Investigations of Revising Discourse Models During Reading

Introduction

Many studies have revealed fine-grained information about the time-course of different context effects on individual words, collecting temporally sensitive data like ERPs (e.g., Kuperberg et al., 2011; Chapter 2) or eye movements during reading (e.g., Calvo et al., 2001). These studies are typically designed and analyzed with an eye towards maximizing the chance of finding a significant effect on a single critical word, for example by manipulating the cloze probability or discourse congruity of that word. However, naturally produced sentences, constructed on the fly to induce a cognitive or behavioral change in another person, rarely follow the "single critical word" method of conveying their message. Multiple, partially redundant cues are the norm, with many important words and phrases each making an important but not wholly sufficient contribution to the activation of the speaker's intended message and the suppression of unintended competing messages.

Therefore, although we know a great deal about how individual words are processed as a function of their degree of consilience with the evolving discourse representation, it is still unclear how partially redundant cues to causal relationships are processed in discourse and how they might contribute to the construction of discourse-level meaning. For example, if two separate words in a sentence point to the same overall message but with different degrees of specificity, is the first cue used to activate some general, underspecified event knowledge, while the second cue modifies it in important ways, or do comprehenders use the first cue to predictively activate more specific event knowledge than is necessarily warranted, rendering the second cue useless?

The present study investigated this question with two experiments employing stimuli characterized by discourse-level manipulations of plausibility. In the second sentence of each stimulus (which is word-for-word identical in the *plausible* and *implausible* conditions), a relatively weak cue to the nature of the event being described was presented first, followed by a stronger cue that made clear whether or not the event was causally coherent with the previous sentence. Examples (1a) and (1b) illustrate our experimental manipulation.

1a. *Plausible*: The cowboy was walking through tall grass. Without warning, the cowboy was ***bitten*** by a ***snake***, and he had to call for help.

1b. *Implausible*: The cowboy was driving through tall grass. Without warning, the cowboy was ***bitten*** by a ***snake***, and he had to call for help.

The first critical word (bitten) specified the type of event being described but left important information unspecified, such as whether this event could occur while driving (e.g., a bite from a mosquito), or was unlikely to occur in this environment, like a bite from a snake. The second critical word (snake) made the nature of the situation and consequently the degree of discourse coherence more clear. To address methodological concerns about the conflation of discourse-level semantic influences with lexical associative semantic influences, all critical words were matched between conditions for strength of lexical association using latent semantic analysis (LSA), a computational tool that uses co-occurrence statistics from text corpora to quantify semantic relatedness between pieces of text (Landauer, Foltz & Laham, 1998).

This chapter investigated how multiple, redundant semantic cues contribute to comprehenders' maintenance of a coherent discourse representation in texts like the following, which describe two causally related events: "The cowboy was walking/driving through tall grass. Without warning, the cowboy was ***<u>bitten</u>*** by a ***<u>snake</u>***, and he had to call for help." If sequential discourse cues have independent effects on comprehension, then we would expect relative processing difficulty in the *implausible* condition on both the critical verb ("bitten") and on the critical noun ("snake"). However, if participants fill in a more detailed discourse representation on the basis of incomplete information and easily process later consistent information, then we would expect to observe processing difficulty for the *implausible verb*, but not for the *implausible noun*. Finally, it is possible that the verb, being a weak cue to discourse coherence, will not affect processing differently in the two conditions, and effects of plausibility will only appear on the noun.

Another source of confusion in the literature is that similar experiments conducted with different methods often result in findings that are difficult to reconcile or compare with each other. In order to minimize this problem, the present study comprises two experiments using the same stimuli but different techniques for measuring online processing difficulty. In Experiment 4, using self-paced reading, processing difficulty will be measured with longer reading times, possibly on the critical words but also on so-called 'spillover' words following each critical word. In Experiment 5, ERP data will be collected in a rapid serial visual presentation (RSVP) paradigm. Processing difficulty will be indexed with N400 amplitude, with smaller N400 responses indexing facilitated processing.

The advantage of self-paced reading is that it is relatively natural, because the participant controls the pace of stimulus presentation. However, self-paced reading does not track real time mental processes in response to individual words. Context effects are frequently delayed by a word or two (“spillover effects”), and of RT is a one-dimensional measure of processing difficulty. ERPs provide more detailed, temporally sensitive data, as brain responses to stimuli are conducted through the brain to EEG recording devices at effectively the speed of light. However, the RSVP paradigm imposes a much greater degree of inflexibility on the subject. Using both methods and comparing the results allows us to profit from the benefits and compensate for the deficiencies of both methods.

Experiment 4

Experiment 4 used a single word moving window self-paced reading task in order to investigate the impact of causal relatedness on sentence processing as it unfolds over multiple critical words. Self-paced reading has been widely used in the discourse and inference psycholinguistics literature, but in most cases the unit of interest has been larger than the individual word. For example, investigations concerning which types of inference are activated in different contexts often use self-paced line or sentence reading (e.g., Graesser & Bertus, 1998; Klin, 1995), with hypotheses about, say, predictive causal inference activation being tested based on the reading times of sentences that contain information about a previously available predictive causal inference compared to those that do not. Although measuring sentence reading times is an appropriate method for testing hypotheses about *which* inferences are activated by comprehenders, it is less well-suited for studying the time-course of inference activation, and how inferential activations affect the processing of subsequent words in the text.

One topic in which self-paced single word reading has proven more popular is in the investigations of verbs' implicit causality properties. A phrase like "Phil apologized to Lisa because…" is generally continued with further information about what Phil did that he had to apologize for. However, a phrase with an "object-biasing" verb, like, "Phil praised Lisa because…," is typically continued with information about the praise-worthy thing that Lisa did. These biases are evident both in preferred continuation patterns and also in readers' expectations for upcoming pronouns ('he' is easier to process than 'she' after 'Phil apologized to Lisa because…', and vice-versa for the phrase with praised.). Self-paced reading has proven useful for studying the unfolding in time of this explicit bias, which can be probed by manipulating the compatibility between the main clause containing the biasing verb and a pronoun in the embedded clause. Using self-paced reading, for example, Koornneef and Van Berkum (2006) observed that effects of pronoun congruity emerged in the spillover region immediately following the pronoun, contradicting previous accounts that suggested a slower integration process, based on whole-phrase self-paced reading (Stewart, Pickering & Sanford, 2000).

Self-paced reading is therefore suitable for understanding how context effects unfold over the course of a target sentence. In Experiment 4, the manipulation of plausibility was examined for its effect on reading times over several words in the second sentence of each story, namely the critical verb (e.g., "bitten"), the critical noun ("snake"), and the two-word spillover region following each. In general, we expect to see longer RTs in the *implausible* condition than in the *plausible* condition.

The key question was when these effects would emerge and whether they would be confined to a single region of the sentence. If plausibility effects manifested only on the verb and/or its spillover words, but not on the noun or its spillover words, that would suggest that the initial weak cue to discourse implausibility triggered a relatively comprehensive adjustment of the situation model, such that the continuation of "snake" was equally congruent with the state of the current discourse model in the *plausible* condition and in the *implausible* condition (Van Berkum, 2009). This might occur if comprehenders pre-activate likely biting agents, for example, upon encountering an unexpected cue to a biting event. On the other hand, it is possible that *plausible* words in both regions will be read more quickly than *implausible* words. That result would suggest that updating the discourse model to accommodate implausible information is not a punctate, even predictive event, but is carried out more conservatively, in reaction to each new piece of implausible information.

Finally, a word recognition test of the critical nouns was included at the end of the experiment. Causal inference is known to improve subsequent recall of the items in the conditions that encouraged an inference (e.g., Keenan et al., 1984; Myers et al., 1987). If the manipulation of plausibility facilitates the activation of a sensible causal relationship in the *plausible* condition to a greater degree than in the *implausible* condition, then enhanced memory for *plausible* nouns might be indexed by increased accuracy and/or shorter RTs to those nouns in a recognition test.

Methods

Participants

Participants were 53 UC San Diego undergraduates, who were compensated in course credit. All participants gave informed consent to participate. All participants were native English speakers, free of language or learning deficits, with normal or corrected-to-normal visual acuity.

Materials

80 near-identical pairs of experimental passages were constructed, intended to vary in plausibility due to a difference in causal coherence. In each experimental passage, the first sentence established the scene for an event. The second sentence described that event and always consisted of five parts: (1) a 1-4 word temporal modifier, (2) the subject/patient of the sentence, (3) the main verb in passive form, (4) a noun phrase naming the entity that caused the event, and (5) an adjoined independent clause describing the resolution of the event, always beginning with “and he/she/it was/were/had.” Within each experimental stimulus pair, the first sentence varied between conditions in order to manipulate plausibility, and the second sentence (containing the critical words to which response times (RTs) would be measured) was always the same across the two conditions. Example critical stimuli are shown in **Table 3**.

80 such pairs were used in the experiment, allowing us to construct four stimulus lists as follows: each subject read a total of 40 critical stimuli comprising both conditions. Thus, two stimulus lists were constructed involving stimulus pairs 1-40 and two more stimulus lists involving pairs 41-80. There were two variants of each of those stimulus lists, each containing 20 trials in the *plausible* condition and the other 20 in the *implausible* condition. In addition to the 40 experimental trials per list, each list also included 40 filler items that were plausible, unsurprising, and easy to understand. This resulted in a total of four distinct stimulus lists, together representing each condition/item combination exactly once.

Table 3: Example stimuli for Experiments 4 and 5. The critical noun and critical verb in each stimulus is bolded and underlined. The two spillover words after each critical word are underlined.

Plausible: *Claire was walking with no shoes on the beach.* Implausible: *Claire was walking with thick shoes on the beach.* *A moment later, her toe was **<u>pinched</u>** <u>by</u> <u>a</u> **<u>crab</u>**, <u>and</u> <u>she</u> was crying as she limped back to her car.* Plausible: *Her apartment building was in flames, and the old lady was trapped on the top floor.* Implausible: *Her apartment building was in flames, and the old lady was trapped on the ground floor.* *Before long, the lady was **<u>rescued</u>** <u>with</u> <u>a</u> **<u>ladder</u>**, <u>and</u> <u>she</u> was very thankful.* Plausible: *The steel boat hull was shipped in pieces.* Implausible: *The wooden boat hull was shipped in pieces.* *The next week, the pieces were **<u>fused</u>** <u>by</u> <u>the</u> **<u>blowtorch</u>**, <u>and</u> <u>they</u> were painted the day after that.* Plausible: *The shepherd was tending to his sheep on the hillside*. Implausible: *The farmer was tending to his sheep in the barn.* *That very night, the sheep were **<u>attacked</u>** <u>by</u> <u>a</u> **<u>wolf</u>**, <u>and</u> <u>they</u> were protected by the sheepdog.*

Stimulus Norming

Experimental items were normed for the predictability of the critical words in the second sentence using a cloze task. The cloze task was conducted on a web-based form in two versions: a “verb” version in which the story fragment was complete up until the sentence predicate (e.g., “Amy left her car out on the street during the storm. The next morning, the car was __________”) and a “noun” version in which the story fragment was complete up until the agent noun phrase (e.g., “Amy left her car out on the street during the storm. The next morning, the car was dented by the ___________”). In both cases, participants completed 40 items in the *plausible* condition and 40 in the *implausible* condition, each one from a different stimulus pair. Thus, participants in the cloze study were randomly assigned to either the “verb” task or the “noun” task, and then to one of two lists, so that each item received a roughly equal number of cloze ratings. Overall, 194 participants completed cloze norming. No participant in the cloze task participated in the self-paced reading or ERP experiments. Cloze probabilities can be found in **Table 4**. A two-tailed, unpaired t-test comparing the mean ratings for each item showed that the critical verbs did not differ in cloze probability between the *plausible* and *implausible* conditions ($t_{158} = 1.4$, $p = 0.16$). Among the critical nouns, however, those in the *plausible* condition were rated as significantly more probable than those in the *implausible* condition ($t_{158} = 5.2$, $p < 0.0001$).

Experimental items were also matched for lexical association between those critical words and the entire first sentence using Latent Semantic Analysis (LSA) (Landaueret al., 1998). The comparisons were made with the online LSA tool at http://lsa.colorado.edu, using the “General Reading up to 1st Year College” corpus with 150 factors. LSA ratings for the critical materials are summarized in **Table 2**. Again, two-tailed, unpaired t-tests indicated no difference in lexical association between the *plausible* and *implausible* conditions among the verbs or the nouns (in both cases, t_{158} <

1). Because the same critical words were used in each condition, they were perfectly matched between conditions for all lexical-level factors, such as frequency, concreteness, and length.

Table 4: Quantitative features of the critical words. Mean values are reported, with standard deviations in parentheses.

	LSA: Verbs to 1st Sentence	LSA: Nouns to 1st Sentence	Cloze Probability Verbs	Cloze Probability Nouns
Plausible	0.32 (0.16)	0.33 (0.14)	0.08 (0.15)	0.54 (0.3)
Implausible	0.31 (0.16)	0.32 (0.14)	0.05 (0.13)	0.30 (0.26)

Procedure

After filling out a consent form and an autism quotient questionnaire (Baron-Cohen, Wheelwright, Skinner, Martin & Clubley, 2001), participants completed a listening variant of the reading span test (Daneman & Carpenter, 1980) on a computer. After the listening span test, participants remained in the same testing room for the self-paced reading experiment.

Experiment 4 consisted of two parts. In the first part (hereafter, “the reading experiment”), participants read sentences on a computer screen in a word-by-word non-cumulative moving window self-paced reading paradigm. The experiment was controlled by Linger 2.94 software (Rohde, 2005) on a Windows desktop computer. Reading times were measured with the spacebar of the desktop computer’s keyboard. Each trial began with the two-sentence story blanked out with dash (–) symbols replacing each letter.

Participants tapped the spacebar to display the next word, whereupon the previously displayed word was immediately replaced with dashes. After the spacebar was pressed in response to the last word in the second sentence, there was a short delay before the next trial began. Participants were instructed to read at a natural pace and to pay attention to the stories' content, because there would be a memory test at the end of the experiment. At the beginning of the reading experiment, participants received 10 practice trials followed by a break. They were allowed to take another break after every 20 trials. The reading experiment lasted about 35 minutes in total.

The memory test at the end (hereafter, "the memory experiment") was conducted immediately after the reading experiment. In each trial of the memory experiment, a printed word appeared on the screen. This word was either the critical noun from an experimental trial in the reading experiment (e.g., *blowtorch*) or a lure – a concrete noun that was not included in any word position in any experimental or filler trial in the reading experiment (e.g., *eggplant*). In two blocks of 40 trials each (10 *plausible*, 10 *implausible* and 20 *lures*), participants read these visually presented words and indicated whether each one was familiar from the reading experiment or not with a key press. Participants were instructed to go at a comfortable pace, but to prioritize accuracy over speed. The memory experiment lasted under 5 minutes.

Results and Discussion

Reading Experiment

Before analysis, residual RTs more than 2.5 standard deviations from the mean for each combination of condition and word position were replaced by the cutoff value (~2% of data points). To minimize the effects of word length and participant reading rates, a linear regression was performed on the data for each participant, predicting RT from word length in characters. These regressions were calculated from all words in the experiment, including those in filler trials. The predicted RTs were subtracted from the actual RTs to obtain residual RTs. Repeated measures analyses of variance (ANOVA) were conducted in order to compare residual RTs of *plausible* and *implausible* words at each of six critical word positions: *verb, verb-spillover-1, verb-spillover-2, agent-noun, noun-spillover-1,* and *noun-spillover-2*, as illustrated in example (2) below. In each ANOVA the factor of interest was condition (*plausible* vs. *implausible*), which was within subjects in the F_1 analyses and within items in the F_2 analyses.

(2) Without warning, the cowboy was / bitten / by / a / snake, / and / he / had to...

Data from one item were not included in analysis, due to a coding error that presented both words in the verb spillover region simultaneously. The analyses therefore included data from a total of 79 different items. Results of all ANOVAs were subjected to the Greenhouse-Geisser correction, and only corrected p-values are reported. All the RT data for the reading experiment are summarized in **Table 5**, though only the residual RTs are analyzed.

The analyses of the residual RTs revealed that there was no influence of plausibility on the verb region ($F_1[1,52] = 2.2$, $p = 0.14$; $F_2[1,78] = 1.1$, $p = 0.3$), however *plausible* words were read more quickly in the first verb spillover region ($F_1[1,52] = 8.1$, $p = 0.006$; $F_2[1,78] = 6.6$, $p = 0.01$), though not in the second verb spillover region ($F_1[1,52] = 3.6$, $p = 0.06$; $F_2[1,78] = 1.2$, $p = 0.3$). Residual RTs also did not differ in the noun region ($F_1[1,52] = 1.1$, $p = 0.3$; $F_2[1,78] < 1$), but effects of plausibility again emerged in the following words. The first noun spillover word was read faster in the *plausible* condition ($F_1[1,52] = 9.2$, $p = 0.004$; $F_2[1,78] = 4.7$, $p = 0.03$), as well as the second noun spillover word ($F_1[1,52] = 3.7$, $p = 0.06$; $F_2[1,78] = 18.1$, $p = 0.0001$).

Table 5: Summary of reading experiment results. Mean residual and raw reading times are listed by condition and word position, with standard errors in parentheses. Difference scores were calculated by subtracting the *plausible* RT from the *implausible* RT. For the residual reading times only, results of ANOVAs comparing the two conditions are marked for each word position. A * indicates a significant effect ($p < 0.05$) of plausibility in the subjects analysis, the items analysis, or both. A # indicates at most a marginal effect ($p < 0.1$) of plausibility in one or both analyses.

Condition	Word Position					
	Verb	Spillover 1	Spillover 2	Noun	Spillover 1	Spillover 2
	bitten	by	a	snake,	and	he
Residual RTs						
Plausible	-70 (6)	-11 (3)	-28 (4)	-23 (9)	-12 (5)	-37 (3)
Implausible	-62 (8)	1 (6)	-34 (3)	-28 (8)	1 (6)	-29 (4)
Difference	8	12 *	-6 #	-5	13 *	8 *
Raw RTs						
Plausible	350 (6)	339(4)	324 (4)	376 (10)	355 (6)	317 (4)
Implausible	367 (8)	357 (7)	318 (3)	368 (9)	369 (7)	331 (5)
Difference	17	18	-6	-8	14	14

The results clearly show that *plausible* words were read more quickly than *implausible* words, and that this processing difference was present in both the verb region and in the noun region. This pattern indicates that the effect of causal relatedness on word processing was not resolved at the first cue to discourse incongruity, as an effect only at the verb would have indicated. Rather, effects of causal relatedness occurred in the spillover regions of both critical words, suggesting that an additional discourse updating process was triggered on the critical noun.

It is possible, however, that this pattern of results does not represent a single common comprehension strategy employed by all participants. It might be the case, for example, that some participants experienced plausibility effects primarily on the verb, while others experienced plausibility effects primarily on the noun. If such were the case, then it would be evident in correlations by subject between RT effects in the verb region and RT effects in the noun region. If some participants have large effects in the verb region and small/null effects in the noun region, and vice-versa for other participants, then we would find null or negative correlations between RT differences in those two regions.

Post-hoc tests were conducted to examine this possibility, between RT effects on each of three words in the verb region and those on each of three words in the noun region. All significant correlations found were positive, indicating that participants with large plausibility effects in the verb region tended to have large plausibility in the noun region as well, undermining the possibility of distinct processing strategies within the participant sample. These results are summarized in **Table 6**.

Table 6: Correlation coefficients for correlations between RT effects in the verb region and those in the noun region. Outcomes significant at $p < 0.05$ are marked with a single *. Outcomes that remained significant after the α threshold was adjusted to $p < 0.0056$ with the Bonferroni correction for multiple comparisons are marked with a double **.

	Noun	**Noun Sp1**	**Noun Sp2**
Verb	0.15	0.33*	0.3*
Verb Sp1	0.12	0.02	-0.01
Verb Sp2	0.27*	0.49**	0.38**

Memory Experiment

Responses to words in the memory experiment were also analyzed with repeated measures ANOVAs. Because 4 participants had to leave before completing the memory test, data from 49 subjects were analyzed. Participants differed significantly in both accuracy and RT to the *plausible* and *implausible* words. Participants recognized the *plausible* words more accurately than the *implausible* words (77% vs. 68%) ($F_1[1,48] = 4$, $p = 0.0502$; $F_2[1,39] = 10.1$, $p = 0.003$), but correct responses to the *plausible* words were significantly slower than those to the *implausible* words (2049 ms vs. 1684 ms) ($F_1[1,48] = 7.2$, $p = 0.01$; $F_2[1,39] = 12.4$, $p = 0.001$). Because the *lure* words were qualitatively different from the *plausible* and *implausible* words and demanded a different kind of decision, they were not compared to the other two conditions. Overall, participants were slow but accurate in responding to *lures* (False alarm rate: 12%; mean RT: 2140 ms).

Greater accuracy for *plausible* nouns suggests that causally related words were encoded more successfully than words in the implausible condition, for which a causal relationship may have been unavailable during the reading experiment. This result replicates past findings of enhanced memory for items that were originally read in conditions that allowed for a successful causal inference (e.g., Myers et al., 1987). However, the speed-accuracy observed here is unexpected, as stronger causal relatedness generally speeds, rather than slows, memory test responses (e.g., van den Broek & Lorch, 1993). This discrepancy between our finding and those reported in the literature may be due to the fact that we used an old/new recognition paradigm for single words rather than the more typical cued recall test for entire sentences that was used by prior investigators.

Experiment 4 Summary

A self-paced reading experiment and a subsequent memory test found faster reading times and enhanced memory for words in plausible than implausible continuations of texts. In the reading experiment, this processing facilitation was evident both in the region of the verb that provided an initial cue to (im)plausibility, and in the region of the noun which clarified the nature of the (im)plausible event. The lack of effects on the 2nd verb spillover word and on the following noun suggested that this updating of the discourse model to accommodate the implausibility was neither a single punctate event nor a single process drawn out across 5 words. Rather, an adjustment of the situation model occurred in response to the verb, reflected on the verb’s 1st spillover word, and then again at the noun, reflected on the noun’s two spillover words. The memory experiment reinforced the conclusion that participants’ processing benefited from the stronger causal relationships in the *plausible* condition, leading to superior memory accuracy for those critical words. In summary, results indicated that participants activated contextual relevant causal information to connect the events described in the text, and that the activation and use of this material was temporally extended process.

Experiment 5

When readers encounter words and phrases that are to some degree incompatible with the causal structure established up to that point in the discourse, it becomes necessary to update the discourse model in order to accommodate this new information. Experiment 5 focused on two aspects of the time-course of that process. The first is the question of whether this discourse updating happens all in one fell swoop, as it were – perhaps early in the sentence at the first hint of a discourse incongruity, or perhaps later when the discourse incongruity is confirmed – or whether this updating process is temporally extended across several words. The second issue is which aspects of word processing are influenced by this causal information: early aspects of word processing, such as the semantic activation processes indexed by the N400 component (Kutas & Federmeier, 2011), or perhaps later, integrative processes indexed by post-N400 components like the P600 (Brouwer et al., 2012).

With respect to the first issue, Experiment 4 suggested that discourse updating was spread across several words: slowed RTs in the *implausible* condition were observed in the spillover regions of both the mildly implausible verb and the more implausible noun. However, self-paced reading data do not readily permit inferences about stages or aspects of word processing. For instance, Bicknell and colleagues (2010) compared ERP and self-paced reading data recorded as different groups of participants read stories that varied the typicality of a verb argument, in light of that verb's agent (e.g., "The mechanic carefully checked the *brakes*/*spelling*…"). Although N400 effects emerged on the critical noun, a relatively early effect of semantic context, RT effects emerged only on the two spillover words following the critical noun.

Van Berkum and colleagues (2005) observed an even more extreme mismatch between ERP and self-paced reading data. Participants in the ERP experiment (Experiment 1) listened to short stories in Dutch that ended with either an inappropriate noun, the presence of which was foreshadowed by gender agreement morphology on an earlier adjective, indicated with subscript text (e.g., "The burglar had no trouble locating the secret family safe. Of course it was situated behind a big$_{\text{neutral/common}}$ but rather unobtrusive painting$_{\text{neutral}}$/bookcase$_{\text{common}}$.") ERPs timelocked to the onset of the predictive adjectival suffix elicited an effect of appropriateness peaking before 250 ms, yet effects of appropriateness on RTs in the self-paced experiment did not emerge until 3 words later ("unobtrusive" in the example given). The latency and duration of spillover effects does not map particularly cleanly onto the neurocognitive processes indexed by ERPs.

Indeed, within the ERP literature itself, there is some inconsistency as to which aspects of word processing are affected by causal information. The most common finding is an N400 effect. The N400 is a negative-going ERP component that peaks at ~400 ms after stimulus presentation, and is elicited by every meaningful stimulus, including words (Kutas & Hillyard, 1980). The amplitude of the N400 response to a word in context varies inversely with the predictability of that word, given its context (Kutas & Hillyard, 1984). As such, the N400 is also sensitive to manipulations of discourse coherence including causal relatedness. For example, Kuperberg and colleagues (2011) conducted an ERP study in which discourse stimuli varied in the strength of causal relatedness between sentences, but not in the strength of lexical association between the critical word and prior context. They found that N400 amplitude on the critical word varied inversely with the degree of causal relatedness. The results suggest that the situation model built up across multiple sentences can facilitate processing of a compatible incoming word, resulting in reduced N400 amplitude (Van Berkum, 2009).

However, other ERP studies of causal inference have revealed late effects indexing processing mechanisms that follow the semantic retrieval operations indexed by the N400. For instance, Burkhardt (2007) found only P600 effects on the critical word (e.g., "pistol") in a story like, "A PhD student was found dead/killed/shot downtown. The press reported that the ***pistol*** was…" She suggested that the enhanced P600 in the less related conditions (following "dead" or "killed") indexed the cost of adding new information to the discourse representation in working memory (*cf.* Brouwer et al., 2012), an operation that was not as difficult when the new information ("pistol") was implied by a previous word ("shot").

On the other end of the spectrum, in Chapter 2 we observed very early effects of causal relatedness on visual probe words ("WIND" or "HUNTER") presented after story primes that encouraged a causal inference (e.g., "The stack of papers was sitting next to the open window. A moment later, the papers were fluttering into the yard"). Effects of causal relatedness began on the P2 component, which is thought to index perceptual properties of visual stimulus evaluation (Potts, 2004), and continued to influence the N400 and a post-N400 sustained frontal negativity. The sustained negativity was larger in the *causal unrelated* condition (“HUNTER”), in keeping with observations in other ERP inference experiments showing that critical words that disconfirm a previously made inference can elicit an enhanced sustained negativity (e.g., Pijnacker, Guerts, Van Lambalgen, Buitelaar & Hagoort, 2010). Van Berkum (2009) has suggested that sustained negativities index working memory operations supporting inference processes that attempt to resolve a discourse incongruity or ambiguity. However, much more research is needed to understand even the conditions that elicit this component, to say nothing of its functional significance.

Although these experiments provide valuable insight into the time-course of causal relatedness effects on single words, they do not address our other question about the time-course of causal relatedness effects over multiple words. The only extant ERP study of causal relationships in which ERPs were measured on multiple words in the target sentence unfortunately did not produce suitable data for addressing this question, because ERPs were averaged across all critical words (St. George et al., 1997). Experiment 5 examined both of these issues using the ERP technique and the same stimuli as in Experiment 4. Importantly for this study, the N400 component is sensitive to mismatches between the semantic information active in working memory and the semantic information that would be activated by the incoming word (Van Berkum, 2009). An N400 effect of plausibility on the critical verb would therefore indicate that the semantic activation state prevailing before the verb was encountered differed between conditions, or in other words that participants pre-activated some information relevant to the critical verb.

Methods

Participants

Participants were 32 UC San Diego undergraduates (17 female), who were compensated in course credit and/or payment. All participants gave informed consent to participate. All were right-handed, as assessed by the Edinburgh handedness inventory (Oldfield, 1971). Seven participants reported left-handed members of their immediate family. All participants were native English speakers, had normal or corrected-to-normal vision and reported no history of either neurological disorders or psychiatric medications. Their ages ranged from 18 to 33, with a mean age of 20.2 years. An additional 5 participants were excluded from analysis due to excessive movement and/or blocking artifacts.

Materials

Materials for the ERP experiment were identical to those used in the self-paced reading experiment, with the exception that participants in Experiment 5 read twice as many stories as did individual participants in Experiment 4 (40 *plausible*, 40 *implausible*, and 80 *fillers*). Therefore, only two stimulus lists were used, rotating each experimental story through the *plausible* and *implausible* conditions. The filler items in Experiment 5 were composed of the 40 fillers from Experiment 4, as well as an additional 40 filler items similar in structure to the others.

Procedure

After filling out a consent form, a demographic questionnaire, an Edinburgh handedness inventory (Oldfield, 1971), and an autism quotient questionnaire (Baron-Cohen, Wheelwright, Skinner, Martin & Clubley, 2001), participants completed a listening variant of the reading span test (Daneman & Carpenter, 1980) on a computer. After the listening span test, participants were fitted with an EEG cap and seated in a dimly lit, electrically shielded chamber, one meter away from the computer monitor on which stimuli appeared. Between trials, the monitor displayed a black field with a yellow fixation dot in the center.

In the main experiment, “the reading experiment,” participants read the two-sentence passages, the first sentence presented in its entirety on the screen for participants to read at their own pace. Participants held a 16-key numerical keypad and were told to focus on the center of the screen and press ‘0’ when they were ready to continue to the second sentence. The second sentence, containing the critical words, was presented in rapid serial visual presentation (RSVP) format, with each word appearing for 250 ms with a 200 ms inter-stimulus interval (ISI). This reading experiment was divided into 8 blocks of twenty stories each, with each block containing 5 *plausible* stories, 5 *implausible* stories, and 10 *filler* stories presented in random order. Participants were instructed to read the stories carefully and try to understand their meanings, because they would be given a memory test at the end of the reading experiment. They were also instructed to refrain from bodily movements throughout the experiment, and to refrain from eye blinks and movements during the RSVP portion of each trial. The total EEG recording time was about 50 minutes.

The memory test at the end (hereafter, “the memory experiment”) was identical to the memory experiment in Experiment 4, with a few key exceptions. First, it was conducted in the EEG chamber, with EEG being recorded. Second, responses were not speeded as they were in Experiment 4; participants read visually presented words and then, after waiting 1500 ms for a cue to respond, indicated with a key press whether the word was familiar from the reading experiment or not. This delay was introduced to avoid contaminating ERPs with movement artifacts. Responses were made with the thumbs on the ‘1’ and ‘3’ keys of a numerical keypad, rather than a full keyboard. The keys associated with each response were counterbalanced across participants. Participants were instructed to refrain from blinking in each trial until after making the familiarity judgment.

EEG Recording

EEG was recorded with 29 tin electrodes in an Electrocap mesh cap, organized in the international 10-20 configuration. Recordings were taken from 8 lateral sites: T5/6, TP 7/8, FT 7/8, F7/8; 10 medial sites: P3/4, CP3/4, C3/4, FC 3/4, and F3/4; 5 midline sites: Pz, CPz, Cz, FCz, and Fz; 3 frontal polar sites: FP1/2 and FPz; and 3 occipital sites: O1/2 and Oz. Three additional electrodes were placed at the outer canthi of the eyes and below the left eye, to record eye movements and blinks. At all sites, electrical impedance was reduced below 5kΩ through gentle abrasion and by applying Electrocap conductive gel. All EEG recorded was referenced on-line to a single electrode on the left mastoid, and re-referenced to an average of the left and right mastoid electrodes before data analysis. The EEG was amplified with half-amplitude cutoffs at 0.1 Hz and 100 Hz using a SA Instrumentation bioelectric amplifier. The data were digitized online at 250 Hz.

Analysis

In the reading experiment, I analyzed ERPs to all of the words analyzed in Experiment 4. In the memory experiment, ERPs to the visually presented words were analyzed, as well as the accuracy of the button response. ERPs were time-locked to word onset and averaged in a time window spanning 200 ms pre-onset to 900 ms post-onset. The period from 200 ms pre-onset to stimulus onset served as the baseline. Critical EEG epochs were examined and trials containing blocking, drift, movement artifacts, horizontal eye movements, and eye blinks overlapping word presentation were rejected by hand. If more than 25% of a participant's critical epochs were rejected in this way, that participant was excluded from analysis. As noted above, five participants were rejected for this reason. In the data reported below, an average of 9% of trials (SD: 2.7%) were rejected due to artifacts.

Analysis of ERP components focused on two time windows: 300-500ms (N400 component) and 500-800ms (post-N400 effects). Mean amplitudes were analyzed with a repeated measures ANOVA, using a 2 (plausibility: plausible vs. implausible) x 29 (electrode) design. In cases when the ANOVA yielded a significant or marginal ($p < 0.1$) Plausibility effect or Plausibility x Electrode interaction, follow-up analyses were conducted in order to examine the scalp topography of the ERP effect of probe type. In these analyses we analyzed the midline (FPz, Fz, FCz, Cz, CPz, Pz, Oz) and medial electrodes (FP1/2, F3/4, FC3/4, C3/4, CP3/4, P3/4, and O1/2) with additional repeated measures ANOVAs. The midline analyses used a 2 (Plausibility) x 7 (Anteriority: prefrontal through occipital) design. The medial analyses used a 2 (Plausibility) x 2 (Hemisphere: left vs. right) x 7 (Anteriority: prefrontal through occipital) design. To compensate for violation of the sphericity assumption, the Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied to all reported p-values.

The memory experiment was analyzed in a similar way, except that the time windows analyzed were chosen based on prior ERP experiments testing memory for words (specifically Paller, 1990). Thus, the 200-400ms, 400-600ms, and 600-800ms time windows were analyzed. Initial omnibus analyses included *plausible, implausible,* and *lure* words with all electrode sites. Further planned analyses were made with a 2(Condition) x 29(Electrode) design for each of the pair-wise comparisons between the 3 conditions. Each of those analyses was used as criteria for whether to analyze the medial and midline electrode sites for those two conditions. Only the stimuli that were correctly identified as old or new were analyzed. Because of the low number of stimuli per bin (20 each for the *plausible* and *implausible* words and 40 for the *lures*), and because some of these stimuli were rejected due to incorrect responses, a fairly permissive threshold was set for retaining a subject’s data. A subject’s data was included provided that at least 50% of the trials in each bin could be analyzed. Altogether, 26 subjects were included in the analysis of the memory experiment, compared to 32 in the reading experiment.

ERP Results

Reading Experiment

Grand average ERPs to critical verbs are shown in Figure 12, and ERPs to critical nouns are shown in Figure 13. Figure 12 shows that *implausible* verbs elicited larger N400 over posterior sites. This relative negativity continued in the 500-800 ms window with a similar scalp distribution. Figure 13 shows that grand average ERPs to the critical nouns did not differ by condition in any time window. ERP responses to the spillover words and the sentence-final word are plotted in Figure 14, showing that positivities occurred during the N400 window on the first spillover word in each region. To help the reader remember the positions of each word, results are described below for the underlined words in the example stimulus, "Without warning, the cowboy was <u>bitten</u> <u>by</u> <u>a</u> <u>snake,</u> <u>and</u> <u>he</u> had to call for <u>help</u>."

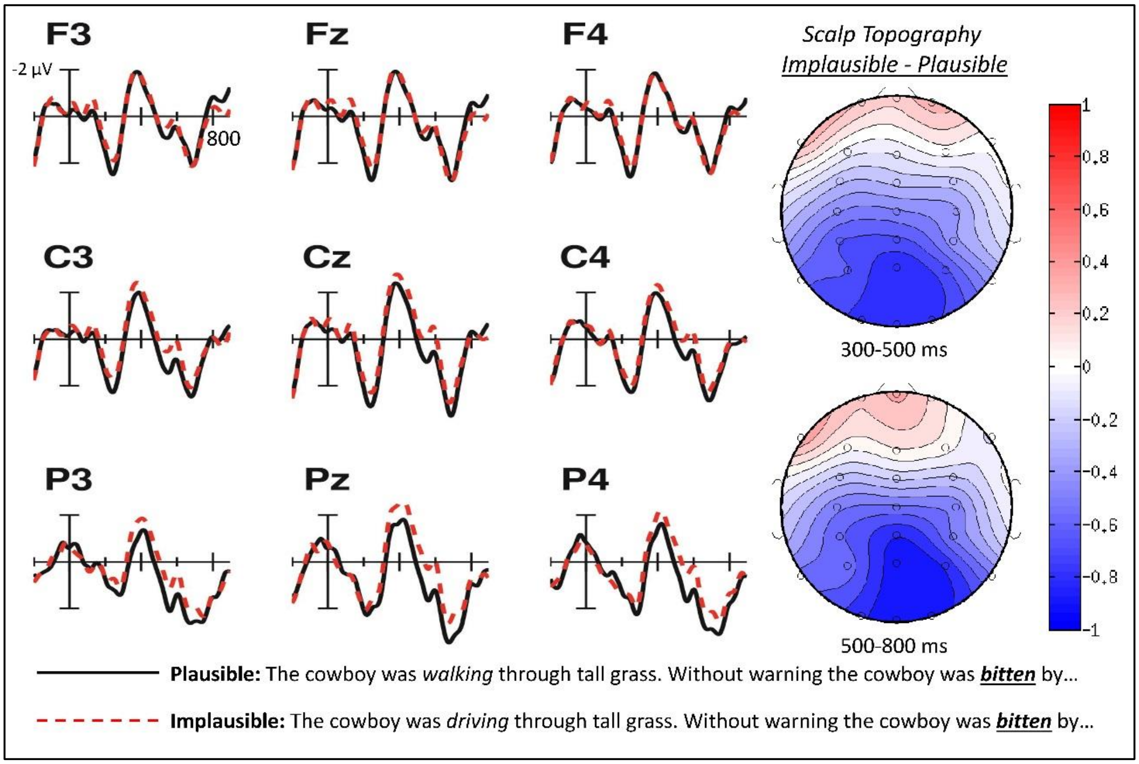

Figure 12: Waveforms and scalp voltage maps of grand average ERP responses to the critical verbs (e.g., "bitten") in the reading portion of Experiment 5. Solid black lines represent the *plausible* condition, and dashed red lines represent the *implausible* condition. Waveforms are low-pass filtered at 10 Hz for presentation purposes, and negative voltage is plotted upwards. Scalp voltage maps were calculated by subtracting the *plausible* condition from the *implausible* condition.

Effects on the Critical Verb ("bitten")

Verb N400 (300-500 ms)

The initial analysis of brain responses to the critical verbs found no main effect of Plausibility ($F[1,31] = 1.4$, $p = 0.2$). Plausibility interacted significantly with the Electrode factor, however ($F[28,868] = 2.1$, $p = 0.02$), apparently reflecting greater N400 reduction in the *plausible* condition focused on posterior sites. However, the Plausibility

X Anteriority interaction was not significant on the midline (F[6,186] = 2.0, p = 0.15) and only marginal on the medial sites (F[6,186] = 2.7, p = 0.08). Medial analyses involving the Hemisphere factor were non-significant, indicating that the scalp focus of the Plausibility effect was bilaterally symmetric (Plausibility X Hemisphere: F[1,31] < 1; Plausibility X Hemisphere X Anteriority: F[6,186] < 1).

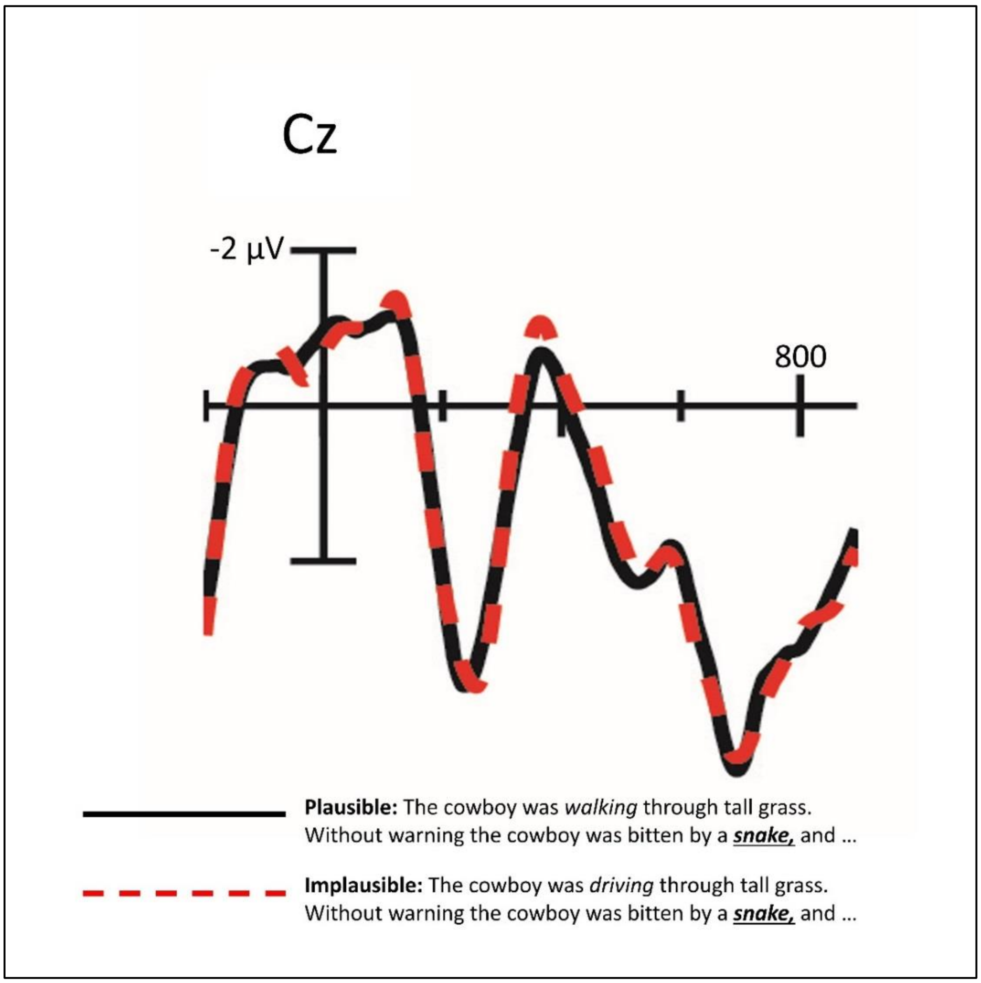

Figure 13: Grand average ERP responses to the critical nouns (e.g., “snake”) in the reading portion of Experiment 5, demonstrating the null effect of Plausibility, shown only at Cz. The solid black line represents the *plausible* condition, and the dashed red line represents the *implausible* condition. Waveforms are low-pass filtered at 10 Hz for presentation purposes, and negative voltage is plotted upwards.

Verb Post-N400 effects (500-800 ms)

The critical verbs elicited no significant main effect of Plausibility (F[1,31] = 1.1, p = 0.3). Again, however, there was a significant Plausibility X Electrode interaction (F[1,31] = 3, p = 0.01). Follow-up analyses on the medial and midline sites revealed that the *implausible* condition elicited more negative ERPs than the *plausible* condition, primarily over posterior sites (see **Figure 12**), reflected by significant Plausibility X Anteriority interactions in both the midline analysis (F[6,186] = 4.1, p= 0.02) and in the medial analysis (F[6,186] = 4.0, 0.02).

First Verb Spillover Word Effects (“by”)

First Verb Spillover Word N400 (300-500 ms)

In the analysis of the first verb spillover word, words in the *plausible* condition elicited larger N400 (F[1,31] = 5.7, p = 0.02). The Plausibility X Electrode interaction was not significant (F[28,868] = 1.4, p = 0.2). Follow-up analyses on the medial and midline sites confirmed the main effect of Plausibility (Midline: F[1,31] = 6.5, p = 0.02; Medial: F[1,31] = 5.4, p = 0.03). There were no significant interactions between Plausibility and Anteriority/Hemisphere (all p-values > 2).

First Verb Spillover Word Post-N400 effects (500-800 ms)

On the first verb spillover word, there was no main effect of Plausibility ($F[1,31] < 1$) after the N400, nor was there an interaction effect of Plausibility X Electrode ($F[28,868] < 1$).

Second Verb Spillover Word Effects (“a”)

Second Verb Spillover Word N400 (300-500 ms)

In the analysis of the second verb spillover word, no main effects or interaction effects of plausibility on the N400 were observed (all $Fs < 2$, all p-values > 0.1).

Second Verb Spillover Word Post-N400 effects (500-800 ms)

In the analysis of the second verb spillover word, no main effects or interaction effects of plausibility after the N400 were observed (all $Fs < 1$).

Noun Effects (“snake,”)

Noun N400 (300-500 ms)

In the analysis of the critical nouns, neither the main effect of Plausibility ($F[1,31] < 1$) nor the Plausibility X Electrode interaction were significant ($F[28,868] < 1$).

Noun Post-N400 effects (500-800 ms)

On the critical nouns, there was no main effect of Plausibility ($F[1,31] < 1$) after the N400 nor was there an interaction effect of Plausibility X Electrode ($F[28,868] < 1$).

First Noun Spillover Word Effects (“and”)

First Noun Spillover Word N400 (300-500 ms)

In the analysis of the first noun spillover word, neither the main effect of Plausibility ($F[1,31] = 1.9$, $p = 0.2$) nor the Plausibility X Electrode interaction were significant ($F[28,868] = 1.3$, $p = 0.3$).

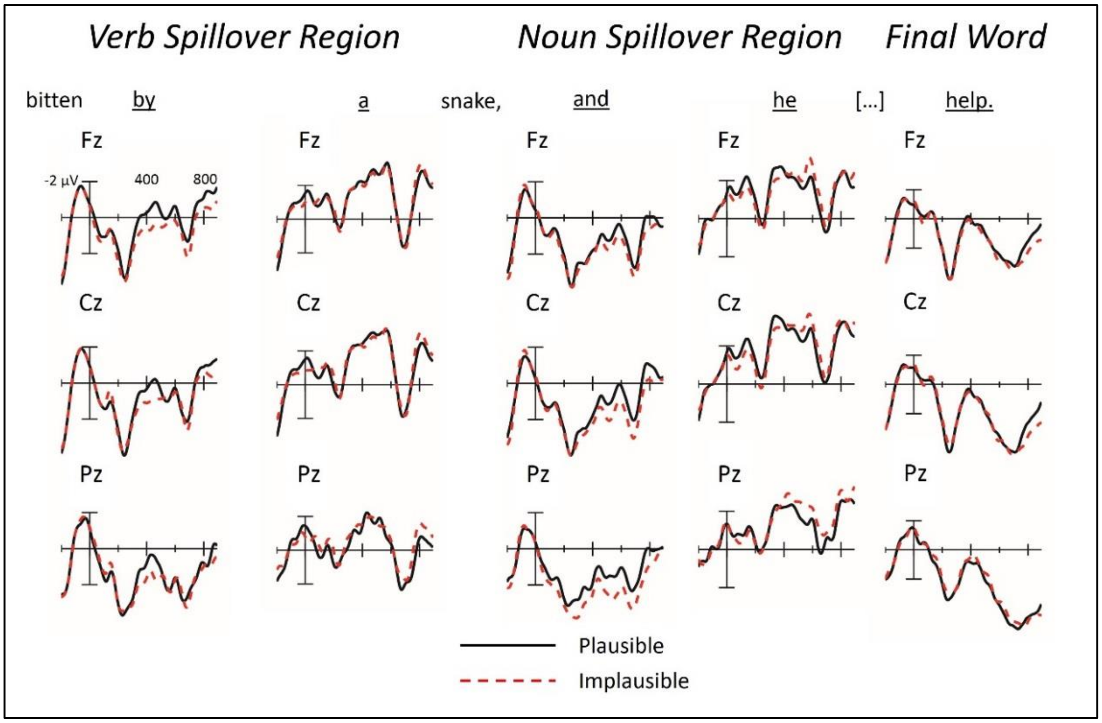

Figure 14: Grand average ERP responses to the spillover words in Experiment 5. Solid black lines represent the *plausible* condition, and dashed red lines represent the *implausible* condition. Waveforms are low-pass filtered at 10 Hz for for presentation purposes, and negative voltage is plotted upwards.

First Noun Spillover Word Post-N400 effects (500-800 ms)

In the analysis of the first noun spillover word, the *implausible* words were significantly more positive than the *plausible* words. This late positivity caused a main effect of Plausibility (F[1,31] = 7.7, p = 0.01) but no interaction with the Electrode factor (F[28,868] = 1.6, p = 0.2). Follow-up testing found the same main effect of Plausibility on the midline electrodes (F[1,31] = 5.5, p = 0.03) and on the medial electrodes (F[1,31] = 7.7, p = 0.01). Although there were no significant interactions with scalp location factors, a marginal Plausibility X Anteriority interaction on the midline electrodes reflected a tendency for a larger Plausibility effect on centro-parietal sites ($F[6,186] = 2.8$, $p = 0.07$).

Second Noun Spillover Word Effects (“he”)

Second Noun Spillover Word N400 (300-500 ms)

In the analysis of the second noun spillover word, no main effects or interaction effects of plausibility on the N400 were observed (all $Fs < 1$).

Second Noun Spillover Word Post-N400 effects (500-800 ms)

In the analysis of the second noun spillover word, there was no main effect of plausibility ($F[1,31] = 2.5$, $p = 0.12$) nor Plausibility x Electrode interaction after the N400 ($F < 1$).

Sentence-Final Words (“help.”)

Sentence-Final Words N400 effects (300-500 ms)

In the analysis of the sentence-final words, no effects were observed on the N400 (both $Fs < 1$).

Sentence-Final Words Post-N400 effects (500-800 ms)

No effects were observed after the N400 (both Fs < 1), either.

Memory Experiment Results

Grand average ERPs to *plausible*, *implausible* and *lure* words in the memory experiment are shown in Figure 15. For three time windows spanning 200-400ms, 400-600ms, and 600-800ms, repeated measures ANOVAs were conducted on all 3 word type conditions and 29 electrode sites. Planned follow-up comparisons were also conducted to examine pair-wise differences within all three pairs of word type conditions.

Response accuracy

Accuracy rates for the *plausible*, *implausible*, and *lure* words were 79%, 80%, and 76%, respectively. Accuracy did not differ by word type ($F[2,50] = 1.2$, $p = 0.3$), as confirmed by pair-wise follow-up comparisons ($F[1,25]$, all p-values > 0.2). Because responses were not speeded, RTs were not analyzed.

ERP Effects (200-400 ms)

The initial ANOVA encompassing all three conditions revealed that word type did not affect ERP amplitudes ($F[2,50] = 1.1$, p 0.4; Condition x Electrode: $F[56, 1400]$ <

1). Pairwise analyses showed no difference between any of the three conditions (all p-values > 0.19).

ERP Effects (400-600 ms)

The initial ANOVA encompassing all three conditions revealed a marginal effect of word type on ERP amplitude in the 400-600 ms window ($F[2,50] = 2.3$, $p = 0.1$), with the largest ERP response elicited by the *lure* words. This effect of word type did not interact with the Electrode factor ($F[56,1400] = 1.1$, $p = 0.4$). Follow-up pair-wise analyses revealed significant differences between the *lure* conditions and the familiar word types: (Plausible vs. Lure: $F[1,25] = 14.6$, $p = 0.0008$; Implausible vs. Lure: $F[1,25] = 4.3$, $p = 0.049$). There was no difference between the *plausible* and *implausible* conditions ($F[1,25] = 1.2$, $p = 0.3$), nor were any of the Condition x Electrode interactions significant or marginal (all p-values > 0.1).

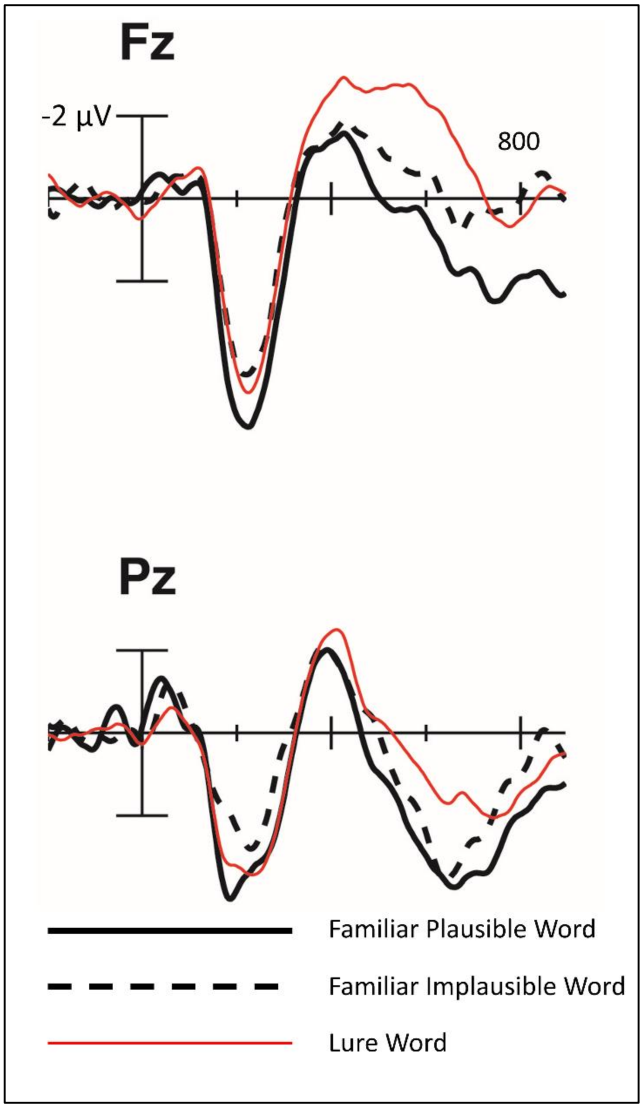

Figure 15: ERP responses to test words in the memory experiment on one anterior midline channel and one parietal midline channel. Waveforms are low-pass filtered at 10 Hz for for presentation purposes, and negative voltage is plotted upward. The thick black line indicates *plausible* familiar words, the dashed black line *implausible* familiar words, and the thin red line *lure* words.

ERP Effects (600-800 ms)

The initial ANOVA encompassing all three conditions revealed that word type significantly influenced ERP responses after the N400 ($F[2,50] = 5.2$, $p = 0.009$). Again, the *lure* words elicited less positive responses than either of the familiar conditions, and responses to the *implausible* words were slightly less positive than those to *plausible* words. This effect of word type interacted marginally with the Electrode factor ($F[56,1400] = 1.7$, $p = 0.1$), reflecting larger voltage differences over anterior electrodes.

The *implausible* condition elicited an ERP response that was intermediate between the more positive *plausible* condition and the more negative *lure* condition. *Implausible* ERPs also did not differ from either the *plausible* ERPs (Condition: $F[1,25] = 2$, $p = 0.2$, Condition x Electrode: $F[28,700] = 1.7$, $p = 0.2$) or the *lure* ERPs (Condition: $F[1,25] = 2.6$, $p = 0.1$, Condition x Electrode: $F[28,700] = 1.3$, $p = 0.3$). *Lures* did elicit a response about 1.6 µV less positive than *plausible* words (Condition: $F[1,25] = 14.3$, $p = 0.0009$, Condition x Electrode: $F[28,700] = 2.1$, $p = 0.09$). Follow-up comparisons on medial and midline electrodes revealed no significant or marginal interactions with the Anteriority and/or Hemisphere factors (all p-values > 0.1).

Post-hoc Analyses

Although grand average effects in the reading experiment were small on the verb and spillover words and nonexistent on the nouns, visual inspection of waveforms elicited from individual participants suggested a great deal of individual variability in effect size and in the direction of the voltage difference. Unplanned correlations were conducted on individual effect sizes at the verb vs. individual effect sizes at the noun, for both the N400 window and post-N400 window. These analyses revealed significant negative correlations ($p< 0.05$, $-0.35 <$ Pearsons's $r < -0.5$) between the post-N400 effect size on the verb and the post-N400 effect size on the noun on 6 centro-parietal electrodes.

To investigate further, we conducted a median split based on the mean size of the post-N400 ERP effect on the verb, thereby dividing the 32 participants in the grand average into two subgroups of 16. These groups did not differ on any of the individual difference measures that we collected (summarized in **Table 7**), except on the Attention to Detail subscore of the autism quotient test. As shown by the scalp plots and grand average ERPs in **Figure 16**, one subgroup showed an N400 effect of plausibility as well as a net post-N400 negativity on verb and a late positivity on the noun (hence, the "verb negativity group"); the other subgroup ("the verb positivity group") showed a nearly opposite pattern: no N400 effect on either word, a right frontal-central late positivity on the verb, followed by a sustained negativity of similar distribution on the noun.

Table 7: Individual difference scores in the two post-hoc participant groups. T-tests were two-tailed and unpaired, with 30 degrees of freedom. The tests that were significant at the 0.05 alpha level are marked with a *. No tests were significant at the Bonferroni-corrected alpha level of 0.00625.

	Verb Negativity Group	Verb Positivity Group	t-score (p-value)
# females, males	7F, 9M	10F, 6M	-
Age	19.5	20.8	1.2 (*n.s.*)
Verbal Span	3.6	3.5	0.3 (*n.s.*)
Overall Autism Quotient	15.6	15.9	0.1 (*n.s.*)
AQ: Social Skills	1.9	1.4	0.7 (*n.s.*)
AQ: Attention Switching	4.8	4.8	0.1 (*n.s.*)
AQ: Attention to Detail	4.9	6.5	2.1 (0.046) *
AQ: Communication	2	1.4	1.1 (*n.s.*)
AQ: Imagination	2	1.8	0.5 (*n.s.*)

Repeated measures ANOVAs examining these effects by group used only the midline and medial electrode columns, in order to examine interactions with the hemisphere and anteriority factors while minimizing the number of separate tests. All tests reported used a 2 (Plausibility) x 7 (Anteriority) x 3 (Laterality) design.

Analyses of ERP responses to the Verb (“bitten”) revealed an N400 effect only in the Predictive (Verb Negativity) group ($F[1,15] = 10.7$, $p = 0.005$), which was focused over more posterior sites, causing a Plausibility x Anteriority interaction ($F[6,90] = 7.2$, $p = 0.003$). The plausibility manipulation elicited a significant voltage difference on ERPs to verbs measured 500-800ms in both groups. In the Predictive (Verb Negativity) group, *implausible* verbs were reliably more negative than *plausible* ones ($F[1,15] = 15.2$, $p = 0.001$), especially over more posterior sites ($F[6,90] = 6.2$, $p = 0.006$). In the Bridging (Verb Positivity) group, *implausible* verbs were marginally more positive than *plausible* ones ($F[1,15] = 4.3$, $p = 0.056$), and this effect was significant in the Plausibility x Anteriority interaction, reflecting its frontal-central focus ($F[6,90] = 3.3$, $p < 0.05$).

On the First Verb Spillover word (“by”), recall that in the grand average the *implausible* condition elicited more positive ERPs than the *plausible* condition in the N400 time window. Subgroup analyses revealed that this effect was driven by the Bridging (Verb Positivity) group, in which a main effect of Plausibility reflected significantly more positive ERPs for the *implausible* condition ($F[1,15] = 12.2$, $p = 0.003$). Neither this main effect nor any interaction effects were significant in the Predictive (Verb Negativity) group (all $Fs < 1$). No significant effects were observed in the post-N400 response to the first verb spillover word (all $Fs < 2$, all p-values > 0.2).

At the Second Verb Spillover word (“a”), the Plausibility did not affect the N400 in either group (all Fs < 3.0, all p-values > 0.1), nor did the post-N400 components (all Fs < 2, all p-values > 0.1).

No effects at the Noun (“snake”) approached significance in the Predictive group, but in the Bridging group the *implausible* nouns elicited marginally more negative ERPs 500-800ms than the *plausible* nouns over frontal-central sites (F[6,90] = 2.8, p = 0.07).

Among the noun spillover words (“and” & “he”), significant effects emerged only on the Second Noun Spillover word (“he”), and only in the Predictive group. In the N400 window, the *implausible* condition elicited marginally more negative N400 than the *plausible* condition (F[1,15] = 3.6, p = 0.08), and interactions between Plausibility and scalp location factors were not significant (all Fs < 3, all p-values > 0.1). In the post-N400 window, *implausible* words elicited significantly more negative ERPs among the Predictive group (F[1,15] = 6.2, p = 0.03). Interaction with scalp location factors in the Predictive group were non-significant, as were all statistical tests in the Bridging group (All Fs < 3, all p-values > 0.1).

No significant effects were observed on the final words (“help.”) in either group, although in the Bridging group a marginal Plausibility x Laterality interaction reflected a trend towards a late positivity that was largest over left medial sites (F[2,30] = 3.1, p = 0.07).

In sum, large plausibility effects were observed in the Predictive (Verb Negativity) group’s ERPs to the verb (“bitten”) 300-800ms, and none at the noun (“snake”), although a negativity emerged on the second noun spillover word (“he”). In the Bridging (Verb Positivity) group, verbs elicited a larger fronto-central positivity 500-

800ms in implausible contexts than plausible ones, which continued in the first verb spillover word ("by"), whereas the nouns elicited a marginally larger late negativity in *implausible* contexts than in *plausible* contexts.

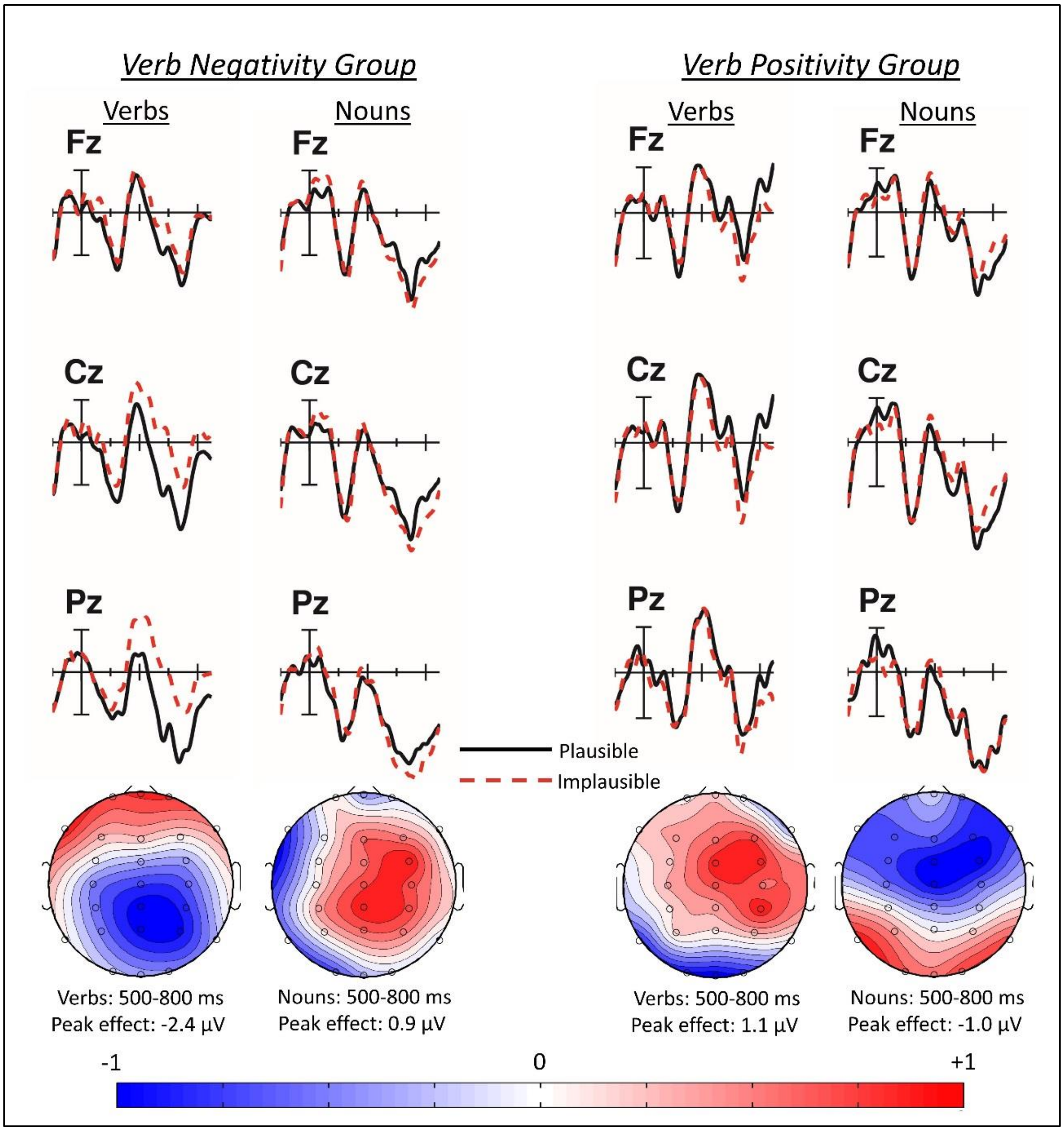

Figure 16: ERP waveforms and scalp voltage plots of ERP responses to the nouns and verbs in the two subject groups. ERPs to *plausible* words are represented with a solid black line and ERPs to *implausible* words are represented with a dashed red line. Negative voltage is plotted upwards. Scalp voltage maps were calculated by subtracting the *plausible* condition from the *implausible* condition.

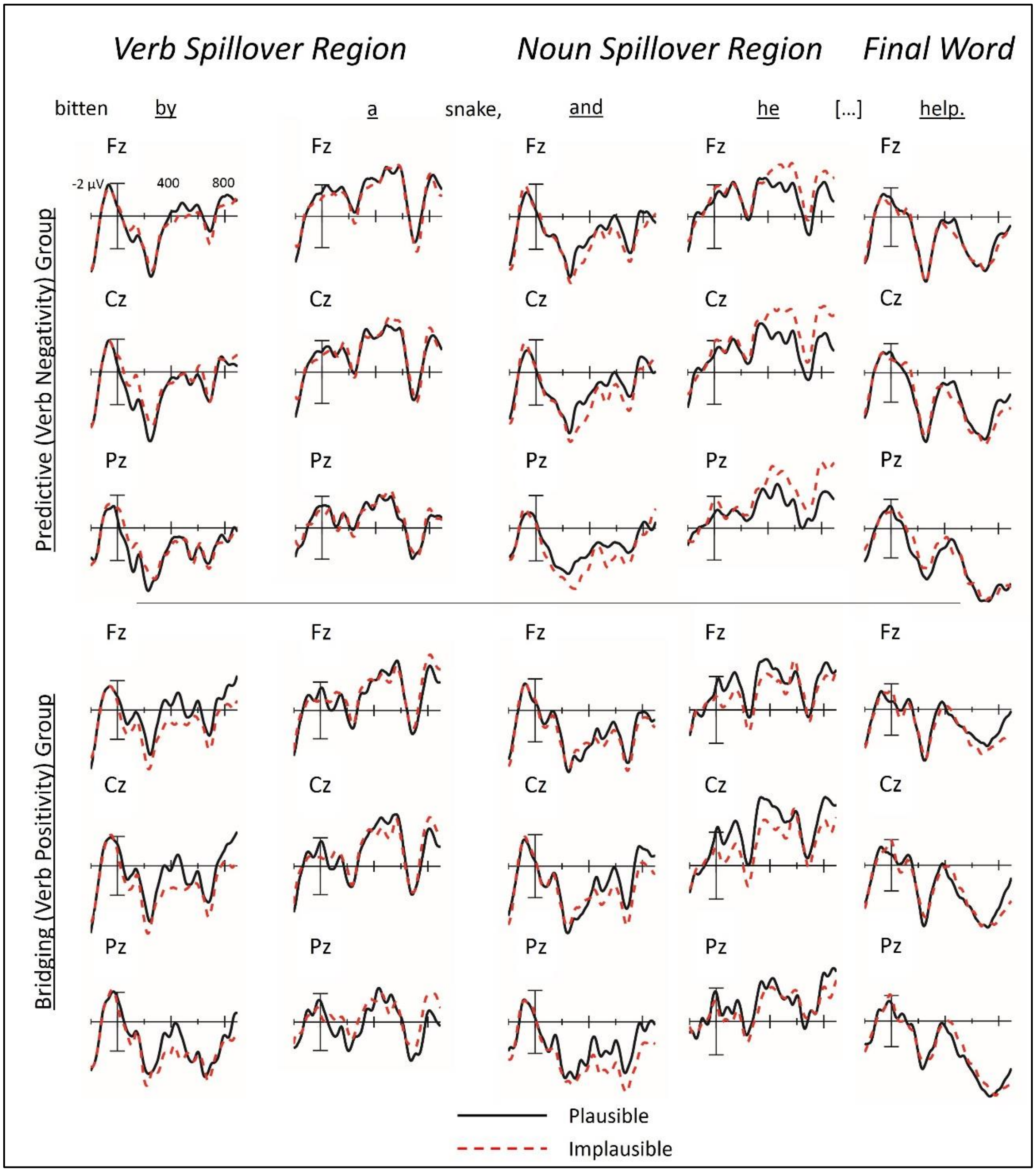

Figure 17: Grand average ERP responses to the spillover words in Experiment 5. Solid black lines represent the *plausible* condition, and dashed red lines represent the *implausible* condition. Waveforms are low-pass filtered at 10 Hz for for presentation purposes, and negative voltage is plotted upwards.

Discussion

ERPs were recorded as participants read two-sentence stories beginning with a discourse context (a cowboy walking through tall grass, or a cowboy driving through tall grass) that rendered the second sentence a *plausible* or *implausible* continuation ("Without warning, the cowboy was ***bitten*** by a ***snake***, and he had to call for help."). Analysis of the ERP reading experiment indicated that processing difficulty in the *implausible* condition emerged on the earliest cue to the nature of the event, the verb "bitten" and on the following function word, but did not affect the second cue, the noun "snake." This finding of processing effects only on the verb is intriguing in light of the fact that only the noun differed in cloze probability between conditions. Because differences in cloze probability are a reliable correlate with N400 amplitude (Kutas & Hillyard, 1984; Wlotko & Federmeier, 2012), the lack of an N400 effect on the noun is an unusual finding, suggesting that the additional semantic activation that would be normally be triggered by the less predictable, *implausible* noun that had already occurred earlier in the sentence. The grand average ERP pattern in Experiment 5 appeared to contradict the findings of Experiment 4, in which slowed reading times were observed on the spillover regions of both the verb and the noun.

Post-hoc analyses revealed that the post-N400 ERP responses to the verb correlated negatively with those to the noun, suggesting that individual participants differed in how they handled these two redundant discourse cues. ERP results contrast with those reported for the reading time data from Experiment 4 in which positive correlations were found between reading times for the verb and reading times for the noun. So in Experiment 4, processing costs of all kinds manifested as slowed RTs. However, in Experiment 5, some processing costs manifested as an enhanced late positivity and others as an enhanced late negativity. This comparison highlights an important tradeoff between the self-paced reading and ERP measurements. On one hand, ERP is more sensitive because it can differentiate between various types of neurocognitive effects: early vs. late in the time-course of word-processing, positive vs. negative voltage fluctuations, and so on. On the other hand, overlapping ERP effects of opposite voltage polarity can sum to 0μV and cancel each other out, leading to an apparent null effect when context actually is affecting processing. Taken together, the individual differences results in the two experiments suggest that processing costs for implausibility occurred on both the verb and the noun, but that the nature of them was different, both between the two words and between different participants.

Further analyses of the Experiment 5 ERP reading data suggested participants could be divided into two subgroups with distinct processing patterns. In one group, dubbed the Predictive (Verb Negativity) group, the verb elicited a relatively large N400 effect that continued into the post-N400 epoch. This pattern was qualitatively similar to the effects observed in the grand average, although the voltage differences were larger. Participants in the Predictive group did not show any plausibility effects at the noun, although a N400 effect did emerge on the second spillover word following the noun. Participants in the other subgroup, the Bridging (Verb Positivity) group, showed a pattern of effects that was in some ways the opposite: no N400 effect in either word position, a frontal-central positivity on the implausible verb and on the subsequent spillover word, and a frontal-central late negativity on the implausible noun (see Figure 16).

A comparison of the N400 effects on the verb in these two groups – that is, a large effect in one group, and no effect in the other – suggests that in the Predictive group, the state of memory activation differed between the *plausible* and *implausible* conditions at the time the verb was encountered. The result was facilitated semantic retrieval in the *plausible* condition, reflected by reduced N400 amplitude (Kutas & Federmeier, 2011). This result is interesting because cloze probability, the best predictor of N400 amplitude (DeLong, Urbach & Kutas, 2005), did not differ by condition on the verb. This suggests that the N400 reduction on the *plausible* verb was driven by predictive inference based on the scene that was set in the first sentence. In the Bridging group, however the activation state of semantic memory activation was not affected by the plausibility of the current sentence in relation to the context. It is possible that the primary difference between the two groups, causing the difference in N400 effects as well as the other post-N400 differences, was that one group was more able and/or inclined to pre-activate semantic material related to the upcoming scenario.

Another possibility is that the difference between the groups was a depth of processing effect. That is, participants in the Verb Negativity group (who also showed an N400 effect) were simply paying more attention to the story stimuli than those in the Verb Positivity group, who showed rather small effects in general. If that is the case, then the participants in the Verb Negativity group should also have shown superior recognition performance during the memory experiment. However, this possibility was not supported by the memory experiment results. Because not all of the participants in the reading experiment met the accuracy criteria for ERP analysis in the memory experiment, it was not possible to compare ERP results on the memory experiments between exactly the same two subgroups of participants. However, comparison of those participants in both the reading and the memory experiment revealed no differences in accuracy rates of the two subgroups ($t[15] = -1.4$, $p = 0.2$).

General Discussion

Reading times and ERP responses were measured on the several target words of experimental stories that varied in plausibility due to a manipulation of causal coherence. On the verbs, grand average results in the two experiments were consistent. Slowed reading times were observed in the spillover region in Experiment 4, and in Experiment 5 ERP effects of plausibility were observed on the verb and on the subsequent function word. These findings included an N400 effect and a continuing negativity on the verb, which may have been a sustained negativity effect of the kind associated with pragmatic processing triggered by a discourse incongruity (Van Berkum, 2009; Chapter 2) or an enhanced P600 in the *plausible* condition. Enhanced P600 in the more semantically congruent condition is hardly a common finding. However, it is not outside the realm of possibility that a plausible but unexpected piece of information could provide more information for updating a discourse representation than an unexpected but implausible piece of information. In such a case, the verb in the *plausible* condition could elicit a larger P600 than the verb in the *implausible* condition.

On the noun, effects of plausibility were found on the spillover region only in both Experiments 4 and 5. In Experiment 5, there was no effect of plausibility on the N400 to any word in the noun region, but the first noun spillover word elicited an enhanced late positivity in the *implausible* condition. It seems unlikely that this late positivity was elicited by a semantic conflict between the discourse context and the first noun spillover word (which is “and” in all critical items), therefore it is more likely that this late positivity represents a delayed response to the noun itself.

Furthermore, inspection of individual subject waveforms indicated that while the net effect size at the noun did not differ reliably from zero, there was considerable variation among subjects, with some showing a late positivity and some a late negativity. Because both of those types of effects can reflect processing difficulty of some kind, one or both of those effects may have manifested in Experiment 4 as slowed reading times on the noun spillover region. A between-groups comparison of memory accuracy for the stimulus materials found no difference (79% vs. 80%, $t(30) < 1$), suggesting that this difference in processing styles did not arise from one group being more attentive than the other.

Individual Differences in Discourse Comprehension

Individual differences in discourse processing reported previously in the literature have generally correlated with an independently testable measure of comprehension skill, in particular verbal span (e.g., St. George et al., 1997; Nieuwland & Van Berkum, 2006a) or performance on a comprehension task (e.g., Coulson & Kutas, 2001). In the current study, however, processing styles did not differ by verbal span, autism quotient, or task performance; nevertheless, there was considerable individual variance in the size and even the direction of ERP effects observed in Experiment 5. Post-hoc comparisons discovered an overall negative correlation between the plausibility effects on the verb at 500-800 ms and on the noun at 500-800 ms. Subjects were therefore divided into two groups on the basis of the plausibility effect on the verb at 500-800 ms. In one group, the verb elicited a large N400 effect followed by a continuing negativity, and the noun elicited a late positivity. In the other group, no N400 effect was evident on either critical word, and a late positivity on the verb was followed by a late negativity on the noun. This experiment does not provide the information necessary to distinguish the functional significance of all of these effects, but the qualitative difference in ERP patterns likely reflects a difference in processing styles. Perhaps the most significant difference is in the N400 effect on the verb. N400 reduction unrelated to lexical factors like word frequency (which were identical between plausibility conditions) indicates a degree of semantic pre-activation (Kutas & Federmeier, 2011; Van Berkum, 2009). This is not necessarily lexical prediction, but simply a case in which some of the information that would normally be activated by that word in that context has already been activated by earlier material, leading to less effortful semantic activation on the word itself, in this case the critical verb.

This between-groups difference in semantic pre-activation recalls the long-standing debate over whether comprehenders make predictive inferences during reading. Some studies have suggested that participants do not make predictive inferences (e.g., Potts et al., 1988; Duffy, 1986), other studies have suggested that in appropriately constraining contexts predictive inferences are drawn (Klin, Guzman & Levine, 1999; Calvo et al., 2001). Theories of inference generation such as the constructionist framework (Graesser et al., 1994) and the minimalist hypothesis (McKoon & Ratcliff,

1992) state that predictive inferences will not be drawn in most cases, but are possible in a sufficiently constraining context with the necessary information easily available. The stark difference between the two groups of participants in the present study – in a weakly constraining context, no less – suggests that not only the discourse situation determines whether predictive semantic activation occurs, but also factors internal to the comprehenders.

The important roles of both discourse context and individual differences in these results shed light on the conflicting results obtained in prior studies of causal inference. Some studies have reported N400 effects of causal relatedness (Kuperberg et al., 2011; St. George et al., 1997, Yang et al., 2007), in line with a memory- or expectation-based mechanism in which causal relationships are activated during lexico-semantic activation (see e.g., Coulson, 2006). However, others have reported later effects on the P600 component (Burkhardt, 2007), pointing to a slower, integrative causal inference process. Based on our individual differences data, it appears that both camps in this debate were correct: some comprehenders make predictive inferences even in weakly constraining contexts, and others do not. Of course, due to the post-hoc nature of these comparisons, further controlled testing is warranted.

Redundant Semantic Cues in Discourse Processing

A key motivation for this study was to examine how semantic processing changed over successive critical words that provided complementary discourse cues. For example, the critical verb and the critical noun in the second sentence of the following story accomplish complementary discourse functions. "The cowboy was walking/driving through tall grass. Without warning, the cowboy was ***bitten*** by a ***snake***, and he had to call for help." The verb "bitten" establishes the nature of the event, which might be more likely to occur in the *plausible* context than in the *implausible* context. The noun "snake" makes it clear what kind of event is being described and whether it could plausibly occur given the preceding context.

The processing of successive, complementary discourse cues is a little-studied topic, as most studies of discourse processing have focused either on coarse measures of online processing, such as reading times for entire sentences, or on fine-grained analyses of context effects on a single critical word. One of the few discourse processing studies to address the impact of multiple, complementary cues in building a discourse representation was conducted by Nieuwland and Van Berkum (2006b, Experiment 1). In an ERP experiment, participants listened to cartoon-like stories that described the interactions between a human woman and an inanimate object (the anomalous *inanimate* condition) or between a human woman and a human man (the semantically well-formed *animate* condition). For example:

> *Once upon a time a psychotherapist was consulted in her home office by a* **yacht/sailor** *with emotional problems. The yacht/sailor confided in her that everything in life had gone wrong and started crying. The psychotherapist* <u>*consoled*</u> *the* **yacht/sailor** *by stating that everybody experiences these kinds of troubles every now and then. But the yacht/sailor doubted whether to continue outlining his problems to her. The psychotherapist* <u>*advised*</u> *the* **yacht/sailor** *to be honest not only with her, but especially with himself. At that moment the yacht/sailor cried out that he was absolutely terrified of water.*

ERP data were reported from the bolded critical nouns in the first, third, and fifth sentences. In the first sentence of the story, the *inanimate* critical noun elicited a large

N400 compared to the more conventional *animate* noun. However, this effect was not present on the critical nouns in either the third or fifth sentences, indicating that participants quickly adjusted their understanding of the discourse situation to accommodate the impossible scenario of a boat consulting a therapist. The authors suggested that the influence of the global discourse context overwhelmed the countervailing influence of local animacy constraints on word processing. ERP responses to the underlined verbs in the third and fifth sentences (both of which conventionally require animate patients) were also reported. Interestingly, the critical verb in the third sentence (*consoled*) also elicited a larger N400 in the *inanimate* condition, an effect that was not present either on the critical noun in that sentence or on the critical verb in the fifth sentence. The effect on the third-sentence verb was attributed to predictive activation of the type of arguments encoded in that verb's lexical entry (*viz.* animate nouns) (Kamide, Altmann & Haywood, 2003) in conflict with the discourse-based expectation for an *inanimate* patient noun. However, this local constraint was also overwhelmed by the global discourse meaning by the time the fifth-sentence verb was encountered.

Our results cannot be explained in terms of global discourse information overwhelming local constraints arising from lexical entries, because no such constraints were violated in our materials. In Experiment 2 of the present study, at least in the verb negativity group, the difficulty of processing the *implausible* condition was reflected in an N400 effect on the critical verb ("bitten"), indicating that the semantics of the verb overlapped with the semantic information already more active in the *plausible* condition than in the *implausible* condition. However, this discrepancy in semantic readiness was resolved within a span of three words, as no N400 effect was observed on the noun ("snake") in either subject group. Our results suggest instead that in a relatively short time, participants were able to recover from a surprising turn of events in the narrative by updating their discourse model sufficiently to accommodate further implausible elaborations of that event about equally well in the two conditions.

Conclusion

The results underscore the immense flexibility of language comprehension, as well as the speed at which unexpected information can be incorporated into the evolving discourse model. Our initial hypothesis, that comprehenders activate more detailed semantic information than needed, which can aid in the processing of later unexpected words, was partially correct. This result is compatible with the notion of frame- or schema-based semantic activation, meaning that lexical access is not limited solely to the context-free meaning of a word (Elman, 2009), but also includes information about the kinds of situations that occur around that sort of event or entity, including their likely participants (Metusalem et al., 2012; Bicknell et al., 2010). Semantic pre-activation, rooted in both the specific context and in general world knowledge, occurs automatically as part of the semantic activation process triggered by each content word (Coulson, 2001; Coulson 2006).

Further research is needed to understand the constraints on how quickly different types of information can be added to the discourse. Unnatural information such as unexpected animacy properties appears to be completely incorporated into the discourse model somewhat gradually, as suggested by Nieuwland and Van Berkum (2006b) finding of N400 effects on verbs as late as their third sentence, but information about different degrees of plausibility in realistic events appears to be incorporated into the discourse model much more quickly. This is not to say that all subsequent processing of the description of that event occurs “for free,” however. Adjustment of the situation model continued in the noun region, indexed by slowed processing in a self-paced reading task and by post-N400 ERP effects. We can conclude that even in implausible discourse situations, an initial cue to the nature of the event triggers an updating of the discourse model strong enough that further cues to implausibility can be incorporated without additional demands on semantic activation.

References

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. Journal of autism and developmental disorders, 31(1), 5-17.

Bicknell, K., Elman, J., Hare, M., McRae, K. & Kutas M., (2010). Effects of event knowledge in processing verbal arguments. Journal of Memory and Language 63 (4), 489-505.

Brouwer, H.., Fitz, H. & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. Brain Research 1446, 127-143.

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

Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. Memory, 9(4-6), 365-381.

Calvo, M. G., Meseguer, E., & Carreiras, M. (2001). Inferences about predictable events: Eye movements during reading. Psychological research, 65(3), 158-169.

Camblin, C., Gordon, P. & Swaab, T. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. Journal of Memory and Language 56, 103–128.

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

Coulson, S. (2001). Semantic leaps: Frame-shifting and conceptual blending in meaning construction. Cambridge: Cambridge University Press.

Coulson, S. (2006). Constructing meaning. Metaphor Symb. 21 (4), 245-266.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of verbal learning and verbal behavior, 19(4), 450-466.

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

Duffy, S. A. (1986). Role of expectations in sentence integration. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12(2), 208.

Elman, J. (2009). On the Meaning of Words and Dinosaur Bones: Lexical Knowledge Without a Lexicon. Cogn., Sci., 33, 547–582

Graesser, A. C., & Bertus, E. L. (1998). The construction of causal inferences while reading expository texts on science and technology. Scientific Studies of Reading, 2(3), 247-269.

Graesser, A., Singer, M. & Trabasso, T. (1994). Constructing Inferences During Narrative Text Comprehension. Psychological Review 101 (3), 371-395.

Greenhouse, S. & Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, 2. 95-112.

Gygax, P., Tapiero, I., & Carruzzo, E. (2007). Emotion inferences during reading comprehension: What evidence can the self-pace reading paradigm provide?. Discourse processes, 44(1), 33-50.

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

Keenan, J., Baillet, S. & Brown, P. (1984). The Effects of Causal Cohesion on Comprehension and Memory. Journal of Verbal Learning and Verbal Behavior 23, 115-126.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review, 95 (2), 163-182.

Klin, C. M. (1995). Causal inferences in reading: From immediate activation to long-term memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(6), 1483.

Klin, C. M., Guzmán, A. E., & Levine, W. H. (1999). Prevalence and persistence of predictive inferences. Journal of Memory and Language, 40(4), 593-604.

Koornneef, A. W., & Van Berkum, J. J. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. Journal of Memory and Language, 54(4), 445-465.

Kuperberg, G., Paczynski, M. & Ditman, T. (2011). Establishing Causal Coherence across Sentences: An ERP Study. J. Cogn. Neuro. 23 (5), 1230-1246.

Kutas, M. & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205.

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

Kutas, M. & Federmeier, K. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). Annual Review of Psychology 62, 621–47

Landauer, T., Foltz, P. & Laham, D. (1998). Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.

Mason, R. A., & Just, M. A. (2004). How the Brain Processes Causal Inferences in Text A Theoretical Account of Generation and Integration Component Processes Utilizing Both Cerebral Hemispheres. Psychological Science, 15(1), 1-7.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. Journal of Experimental Psychology: Learning, memory, and cognition, 12(1), 82.

McKoon, G. & Ratcliff, R. (1992). Inference during reading. Psychological Review 99 (3), 440-466.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. Journal of memory and language, 66(4), 545-567.

Murray, J. D., Klin, C. M., & Myers, J. L. (1993). Forward inferences in narrative text. Journal of Memory and Language, 32(4), 464-473.

Myers, J., Shinjo, M. & Duffy, S. (1987). Degree of causal relatedness and memory. Journal of Memory and Language 26, 453-465.

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

Nieuwland, M. S., & Van Berkum, J. J. (2006a). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. Brain Research, 1118(1), 155-167.

Nieuwland, M. S., & Van Berkum, J. J. (2006b). When peanuts fall in love: N400 evidence for the power of discourse. Journal of cognitive neuroscience, 18(7), 1098-1111.

Noordman, L. G., Vonk, W., & Kempff, H. J. (1992). Causal inferences during the reading of expository texts. Journal of Memory and Language, 31(5), 573-590.

Oldfield, R. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

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

Pijnacker, J., Guerts, B., Van Lambalgen, M., Buitelaar, J. & Hagoort, P. (2010). Reasoning with Exceptions: An Event-related Brain Potentials Study. J. Cogn. Neuro. 23 (2), 471-480.

Potts, G. F. (2004). An ERP index of task relevance evaluation of visual stimuli.Brain and cognition, 56(1), 5-13.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. Journal of Memory and Language, 27(4), 399-415.

Prat, C. S., Mason, R. A., & Just, M. A. (2011). Individual differences in the neural basis of causal inferencing. Brain and language, 116(1), 1-13.

Rohde, D. (2005). Linger experiment presentation software.

St. George, M., Mannes, S. & Hoffman, J. (1997). Individual Differences in Inference Generation: An ERP Analysis. J. Cogn. Neuro. 9 (6), 776-787.

Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time-course of the influence of implicit causality information: Focusing versus integration accounts. Journal of Memory and Language, 42(3), 423-443.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(3), 443.

Van Berkum, J. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In U. Sauerland, & K. Yatsushiro (Eds.), Semantics and pragmatics: From experiment to theory (pp. 276-316). Basingstoke: Palgrave Macmillan.

van den Broek, P. (1990). The causal inference maker: Towards a process model of inference generation in text comprehension. Comprehension processes in reading, 423-445.

Van den Broek, P., & Lorch Jr, R. F. (1993). Network representations of causal relations in memory for narrative texts: Evidence from primed recognition. Discourse Processes, 16(1-2), 75-98.

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. Neuroimage, 62(1), 356-366.

Yang, C., Perfetti, C. & Schmalhofer, F. (2007). Event-Related Potential Indicators of Text Integration Across Sentence Boundaries. Journal of Experimental Psychology: Learning, Memory, and Cognition 33 (1), 55–89.

Chapter 4:

Hemispheric Asymmetry in the Influence of Causal and Lexical Associative Information on Word Processing

Introduction

The most striking fact about the organization of language processing in the brain is its apparent lateralization, with "core" language functions such as speech production and syntactic knowledge concentrated in the left hemisphere (LH) (Broca, 1865, *trans.* Berker, Berker & Smith, 1986), relegating the right hemisphere (RH) to a supporting role. This classical view of lateralization has held up reasonably well over time: at least in right-handed patients, severe linguistic syndromes including Broca's aphasia and Wernicke's aphasia overwhelmingly result from LH lesions (for a review see Ross, 2010). Although some basic linguistic functions such as phonological analysis and lexical activation have been shown to activate bilateral areas of cortex, these functions generally remain intact in the event of RH brain damage (RHD) (Hickok & Poeppel, 2007).

Symptoms of RHD are frequently described as affecting the pragmatic aspects of language use that involve discourse inference and/or social cognition. A rich literature on RHD patients describes impairments in understanding sarcasm (Giora, Zaidel, Soroker, Batori & Kasher, 2000), emotional prosody (Bowers, Coslett, Bauer, Speedie & Heilman, 1987), jokes (Brownell, Michel, Powelson & Gardner, 1983), indirect requests (Kaplan, Brownell, Jacobs & Gardner, 1990), theory of mind inferences (Siegal, Carrington & Radel, 1996; Happé, Brownell & Winner, 1999), as well as other discourse inferences

(Brownell, Potter, Bihrle & Gardner, 1986). Accordingly, some researchers have proposed that language-affecting RH lesions compromise some sort of generalized inferential or semantic processing system, leading to a broad range of pragmatic and semantic deficits (e.g., Happé et al., 1999).

For example, based on findings that RHD patients are poorer at answering questions about stories and cartoons involving mental state inferences compared to "non-mental" stimuli, Happé and colleagues have proposed that RHD pragmatic difficulties are related to a specific impairment in theory of mind processing (Happé et al., 1999). However, a replication attempt based on Happé et al.'s story stimuli, but also including a matching set of control stories that involved non-mental causal inferences, found that RH lesion patients were actually less impaired on mental state inferences than they were on non-mental causal inferences (Tompkins, Scharp, Fassbinder, Meigh & Armstrong, 2008). Based on this result, Tompkins and colleagues suggested that a theory of mind deficit is unlikely to be at the root of the many pragmatic and social deficits associated with RH brain damage. Tompkins and colleagues' data are also compatible with the view that the pragmatic/social deficits in RH patients arise from a general impairment having to do with the maintenance and selection of contextually appropriate inferences (Blake & Lesniewicz, 2005; Lehman-Blake & Tompkins, 2001).

A crucial role for RH in some aspect of discourse inference is also consistent with findings from neurologically intact comprehenders. Hemispheric contributions to online language processing can be assessed using hemifield presentation, a technique in which a visual stimulus (in this case a word) is presented outside of the fovea to the left or right of where the participant is currently fixating their gaze. Information about a stimulus presented in right visual field is transmitted to LH first (rvf/LH presentation), and vice versa (lvf/RH). Although inter-hemispheric transfer of information begins very rapidly (Banich, 2002), the hemisphere contralateral to stimulus presentation nevertheless represents the stimulus with greater fidelity for a second or more after presentation, as shown with response time and event-related brain potential (ERP) measures. Experiments using the hemifield technique have suggested that RH's contribution to discourse processing arises from an advantage in activating distant semantic associates (Beeman et al., 1994; Beeman, Bowden & Gernsbacher, 2000), activating novel or unexpected meanings (Giora, Zaidel, Soroker, Batori & Kasher, 2000) or activating complex structures of event knowledge (Coulson & Williams, 2005; Coulson & Wu, 2005). Differences in semantic activation have been claimed to underlie the RH processing advantage observed for joke comprehension (Coulson & Williams, 2005; Coulson & Wu, 2005), insight problem solving (Beeman & Bowden, 2000), exploiting indirect lexical associations (Beeman et al., 1994), and predictive inference construction (Beeman et al., 2000).

Neuroimaging work, however, has tended to support the notion that the activation of inference information is a bilateral process. Mason and Just (2004) reported fMRI measurements on participants who read two-sentence stories varying in the degree of causal relatedness between the two sentences. The stimuli fell into three categories of causal relatedness: a highly related condition in which the causal relationship between the two sentences was obvious, a moderately related condition in which the causal relationship was available but non-obvious, and a distantly related condition in which there was no apparent causal relationship. The moderately related condition was considered the most likely to provoke the activation of a complex causal bridging inference because in prior behavioral studies, this condition was associated with the strongest recall of the story stimuli (Myers, Shinjo & Duffy, 1987). Indeed, although activation in bilateral dorsolateral prefrontal cortex correlated inversely with the degree of relatedness, an effect interpreted as reflecting the generation of an inference, the moderately related condition elicited the largest response in RH homologues of language areas, interpreted as reflecting the integration of the inference into the active representation of recent context.

A replication with different stimuli showed that intermediately related sentences activated a bilateral network comprising classical language areas (Wernicke's and Broca's) in LH as well as their RH homologues, with RH activation largest in the least skilled readers (Prat, Mason & Just, 2011). An fMRI study using similar stimuli also found bilateral activation for intermediately related sentences in classical language areas and their RH homologues, as well as in bilateral superior medial prefrontal gyrus and LH middle frontal gyrus (Kuperberg, Lakshaman, Caplan & Holcomb, 2006). These results implicating both hemispheres in causal inference raise an obvious conflict with the lesion study findings that ascribe a unilateral RH basis to these processes. On balance, the data suggest that both hemispheres are involved in the inferential aspects of text comprehension, but that the RH plays a crucial role in this function that is impaired with RHD. It is possible, as Mason and Just's (2004) data suggest, that activating inferential information is a bilateral function but that RH is responsible for integrating that inference into the representation of context, or perhaps for selecting the appropriate inferential information and suppressing the irrelevant aspects of the inference (Blake & Lesniewicz, 2005; Lehman-Blake & Tompkins, 2001).

The experiments described in this chapter tested the hypothesis that the RH involvement in making discourse inferences depends on a RH advantage for activating and/or maintaining inference-related information. We combined the temporal sensitivity of event-related potentials (ERPs) with the hemifield presentation method of assessing the lateralization of stimulus processing. This allowed us to observe the time-course of processing a word that either harmonizes or conflicts with a recently made causal inference, and to examine how these effects differ as a function of hemifield presentation. To our knowledge, this study represents the first use of ERPs to study the time-course and lateralization of causal inference processing. As in previous studies (Chapter 2) our stimuli consisted of short (2-sentence) stories intended to elicit a causal bridging inference by virtue of a coherence gap (e.g., *The surfer took his board out into the waves. He emerged screaming with only one arm.*). These stimuli are described comprehensively in Chapter 2. In three experiments, we recorded brain responses to visually presented probe words that occur after the entire stimulus story had been presented. Experiment 6 compared the causally related *SHARK* vs. a lexical associate of the story-final word, *LEG*. Prior research with these materials suggests that *lexical related* words are treated as unrelated items in relation to *causal related* words (Chapter 2), and prior comparisons of lexical association and discourse relatedness effects show that the influence of lexical association on processing is minimized in discourse contexts (e.g., Camblin, Gordon & Swaab, 2007; Van Petten, 1993). Experiment 7 was designed to test whether those assumptions about lexically related control words were valid, by comparing causally related probe words (SHARK) with unrelated probe words that were chosen from the causal related condition of other stories (e.g., HUNTER). Finally, Experiment 8 used all 4 conditions (causal related, causal unrelated, lexical related, and lexical unrelated), but with half as many items per bin (20), in order to compare effects of causal relatedness and lexical relatedness between visual fields.

The predictions for these experiments are based on the results of Coulson and Wu's (2005) study of joke comprehension. They used a similar paradigm involving a joke or non-joke prime (*Everyone had so much fun jumping into the swimming pool, we decided to put in a little water/platform.*) followed by a probe word related only to the joke meaning (*CRAZY*) to which ERPs were recorded. They observed larger facilitation of the N400 – an ERP component linked to semantic activation processes – in lvf/RH for the probe word in a joke context, as well as a large anterior positivity rvf/LH. Although the N400 asymmetry indicated that the joke's meaning was more active in RH, the LH played a role in selecting the contextually relevant elements of the joke's meaning. In the present study, a RH advantage for activating causal inference-related information should also be associated with larger N400 effects of causal relatedness in lvf/RH. Likewise, if one hemisphere contributes more than another to selecting the appropriate aspects of the inference, then we should observe an enhanced frontal positivity in the control conditions (lexical related in Experiment 6; causal unrelated in Experiment 7; all conditions except causal related in Experiment 8) in rvf/LH.

Methods

Participants

Participants were 16 UC San Diego undergraduates (10 female), who were compensated in course credit and/or payment. An additional 3 participants were excluded from analysis due to excessive movement and/or blocking artifacts. All participants gave informed consent to participate. All were right-handed, as assessed by the Edinburgh handedness inventory (Oldfield, 1971). No participants reported left-handed members of their immediate family. All were native speakers of English, had normal or corrected-to-normal vision and were free of neurological disorders and psychiatric medications. Their ages ranged from 18 to 32, with a mean age of 21.1 years.

Materials

Experimental stimuli consisted of 160 short spoken narratives consisting of two sentences each, as well as visually presented probe words. In each audio narrative, the first sentence established the scene of the described event. The second sentence described a subsequent event, the direct cause of which was left unstated, thus creating a causal coherence gap between the two sentences. The audio narratives were recorded with Adobe Audition software by a female speaker, reading the sentences at a natural rate of speech and with neutral intonation. The sound files were amplitude-normalized to an average of 70 dB.

Each audio narrative was associated with two visual probe words: the *causal related* and *lexical related* probe words used in Chapter 2. For example stimuli, see **Table 8**. For a summary of word-level statistical factors, see Chapter 2, **Table 2.**

Table 8: Sample stimuli, with the probe words used in Experiments 6, 7, and 8.

	Experiment 6 Probe Words	Experiment 7 Probe Words	Experiment 8 Probe Words
The pirate ship pulled up alongside the Spanish galleon. Soon, the galleon sank in a cloud of gunpowder and caustic smoke.	Causal Related: CANNON Lexical Related: VAPOR	Causal Related: CANNON Causal Unrelated: TREE	Causal Related: CANNON Causal Unrelated: TREE Lexical Related: VAPOR Lexical Unrelated: ABSENT
A stack of papers was sitting next to the open window. A moment later, the pages were fluttering into the yard.	Causal Related: WIND Lexical Related: GRASS	Causal Related: WIND Causal Unrelated: DRAIN	Causal Related: WIND Causal Unrelated: DRAIN Lexical Related: GRASS Lexical Unrelated: FLOOR
The fly was buzzing along over the swamp. Suddenly, it was struck by something sticky, and pulled to its death.	Causal Related: FROG Lexical Related: LIFE	Causal Related: FROG Causal Unrelated: SMOKE	Causal Related: FROG Causal Unrelated: SMOKE Lexical Related: LIFE Lexical Unrelated: FAR

Procedure

After filling out a consent form, a demographic questionnaire and an Edinburgh handedness inventory (Oldfield, 1971), participants were fitted with an EEG cap and seated in a dimly lit, electrically shielded chamber, one meter away from the computer monitor on which stimuli appeared. Audio narratives were played through a pair of external speakers flanking the monitor. Throughout the experiment, the monitor displayed a black field with an orange fixation dot in the center.

Participants were told that they would hear two-sentence-long short stories, and that at the end of each story a word would be flashed on the monitor slightly to the left or right of their fixation point (marked by a dot in the center of the monitor). The participant's task was to listen to the stories and try to understand them, and to pay attention to the words that appeared in the center of the screen. Participants were instructed to fixate on the central dot, to avoid blinking whenever the audio was not playing, and to refrain from body movements. Before beginning the experiment, participants were given several practice trials to get used to inhibiting blinks and eye movements at the proper times. Throughout the experiment, participants were given feedback on their blinking behavior. The total EEG recording time was about 45 minutes.

The experimental paradigm is represented schematically in Figure 18. In each trial, participants fixated on the center of the monitor while listening to the audio narrative stimulus. Each sentence of the narrative lasted between 3 and 8 seconds, and there was a pause of 300 ms between sentences. Following the second sentence of each trial, there was an additional pause of 400 ms before a laterally presented probe word was displayed. This latency was chosen because the duration of the final spoken word ranges from 300-450 ms, producing a total stimulus onset asynchrony (SOA) of 700-850ms. This approximates the SOA of 800ms at which Holcomb & Anderson (1993) found N400 effects of auditory-to-visual cross-modal lexical priming. The probe word was displayed for 250 ms, with its inner edge 2° to the left or right of the center of the monitor, the same degree of lateralization used in previous divided visual field studies in our lab (cf. Coulson & Van Petten, 2007; Davenport & Coulson, 2013).

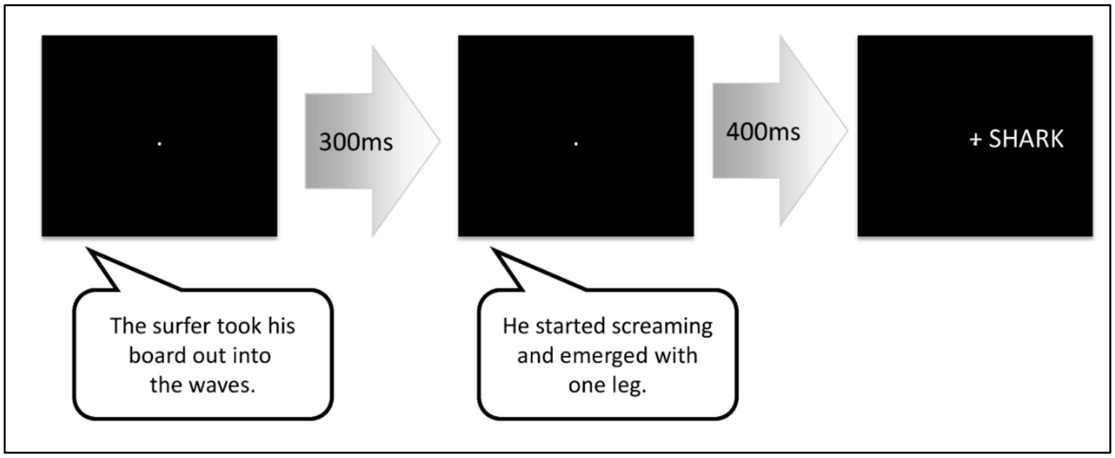

Figure 18: Schematic depiction of a Causal rvf/LH trial in Experiment 6.

EEG Recording

EEG was recorded with 29 tin electrodes in an Electrocap mesh cap, organized in the international 10-20 configuration. Recordings were taken from 8 lateral sites: T5/6, TP 7/8, FT 7/8, F7/8; 10 medial sites: P3/4, CP3/4, C3/4, FC 3/4, and F3/4; 5 midline sites: Pz, CPz, Cz, FCz, and Fz; 3 frontal polar sites: FP1/2 and FPz; and 3 occipital sites: O1/2 and Oz. Three additional electrodes were placed at the outer canthi of the eyes and below the left eye, to record eye movements and blinks. At all sites, electrical impedance was reduced below 5kΩ through gentle abrasion and by applying Electrocap conductive gel. All EEG recorded was referenced on-line to a single electrode on the left mastoid, and re-referenced to an average of the left and right mastoid electrodes before data analysis. The EEG was amplified with half-amplitude cutoffs at 0.1 Hz and 100 Hz using a SA Instrumentation bioelectric amplifier. The data were digitized online at 250 Hz.

Analysis

Only ERPs to the visual probe words were analyzed. ERPs to visual probe words were time-locked to word onset and averaged in a time window spanning 100ms pre-onset to 920 ms post-onset. The period from 100ms pre-onset to stimulus onset served as the baseline. During data analysis, critical EEG epochs were examined and trials containing blocking, drift, and movement artifacts were rejected. If more than 25% of a participant's critical epochs were rejected in this way, that participant was excluded from analysis. As noted in section 2.1, three participants were rejected for this reason. In the data reported below, 16% of trials were rejected due to artifacts. Artifact rejection rates were identical across presentation sides (rvh/LH: 16%, lvf/RH: 16%); *lexical* and *causal* probe trials also did not differ significantly in rejection rates (lexical: 15%, causal: 17%; two-tailed t_{15}=-0.85, p=0.4).

Effects of Presentation Side

Presentation side analyses were intended to assess whether our divided visual field paradigm adequately shifted the balance of processing of lateralized probe words to the targeted hemisphere. Accordingly, brain responses to all probe words presented to rvf/LH were averaged together, and compared to those presented to lvf/RH with repeated measures ANOVAs. In keeping with prior work combining divided visual field presentation with ERP recordings (e.g., Coulson, Federmeier, Van Petten & Kutas, 2005), presentation side analyses targeted the early visual N1 component, as well as the selection negativity (SN) that indexes attentional selection of task-relevant stimuli for further processing (see Hillyard & Anllo-Vento, 1998 for a review). Measurements were made at lateral posterior electrode sites T5/T6 and TP7/TP8, chosen first, because early visual potentials are often the largest there, and, second, because their lateral location affords easy measurement of the SN. The early visual N1 potential was analyzed for mean amplitude in the 100-200ms time window. Analysis of the SN involved only mean amplitude measurements of ERPs 300-800ms because, unlike the visual potentials, this component has no clear peak. Repeated measures ANOVAs involved factors Presentation Side (lvf/rvf), Hemisphere (left, right), and Electrode Site (Temporal, Temporal-Parietal).

Effects of Probe Type

Analysis of ERPs to experimental manipulations focused on three time windows: 150-300ms post-onset (P2 component), 300-500ms (N400 component), and 500-800ms (post-N400 effects). In each time window, mean voltage amplitudes were analyzed with an initial repeated measures ANOVA using a 2 (Presentation Side: rvf/LH vs. lvf/RH) x 2 (Probe Type: causal vs. lexical) x 29 (Electrode). Within each level of the Presentation Side factor, planned follow-up analyses were conducted on the midline (FPz, Fz, FCz, Cz, CPz, Pz, Oz) and medial electrodes (FP1/2, F3/4, FC3/4, C3/4, CP3/4, P3/4, and O1/2) with additional repeated measures ANOVAs. The midline analyses used a 2 (Probe Type: causal vs. lexical) x 7 (Anteriority: prefrontal through occipital) design. The medial analyses used a 2 (Probe Type: causal vs. lexical) x 2 (Hemisphere: left vs. right) x 7 (Anteriority: prefrontal through occipital) design. To compensate for violation of the sphericity assumption, the Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied to all reported p-values. For clarity, we report the uncorrected degrees of freedom.

Results

Grand average ERPs for Experiment 6 are shown in Figures 19, 20 and 21. Figure 19 highlights the impact of the presentation side manipulation on the early visual components, particularly the N1 (100-200ms) at lateral posterior sites, as well as the SN (300-800ms) that follows. Figure 20 shows ERPs to both probe type conditions in each presentation side. In Figure 21, ERPs from both presentation sides and both conditions are overlaid on 4 channels, allowing comparison between presentation sides.

Effects of Presentation Side (rvf/LH vs. lvf/RH)

As shown in Figure 19, the N1 and Selection Negativity were both larger over sites contralateral to word presentation.

N1 (100-200 ms)

Analysis of the N1 component, as described in section 2.5.1, revealed neither a significant main effect of Presentation Side ($F[1,15] = 1.9$, $p = 0.2$), nor of Hemisphere ($F[1,23] = 2.5$, $p = 0.1$). As hypothesized, there was a robust Hemisphere X Presentation

Side interaction, reflecting N1 responses an average of 1.3 μV larger after contra-lateral word presentation, compared to ipsi-lateral presentation ($F[1,15] = 23.8$, $p = 0.0002$). Presentation Side effects are shown in Figure 19.

Selection Negativity (300-800 ms)

The Selection Negativity (SN) is an extended negativity observed in divided visual field presentation, which is generally more negative over lateral temporal sites in the targeted hemisphere (Coulson et al., 2005). As with the N1, there were no main effects of Presentation Side ($F[1,15] = 1.3$, $p = 0.3$) or Hemisphere ($F[1,15] < 1$). The SN was more negative over contra-lateral than ipsi-lateral sites, as reflected by the Presentation side x Hemisphere interaction ($F[1,15] = 26.0$, $p < 0.0001$).

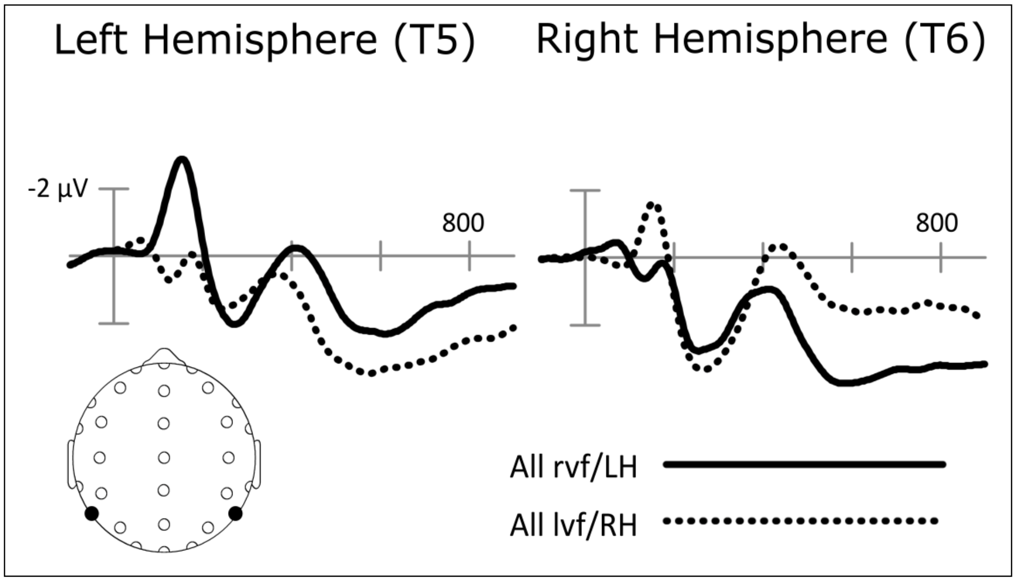

Figure 19: Grand average ERP responses to lateralized words categorized by presentation side. Two posterior temporal sites are displayed, showing that contra-lateral presentation leads to enhanced N1 potentials as well as enhanced selection negativity. ERPs are presented from a 100 ms pre-stimulus baseline to 900 ms post-stimulus. Negative voltage is plotted upwards. ERPs are low-pass filtered at 7Hz for for presentation purposes.

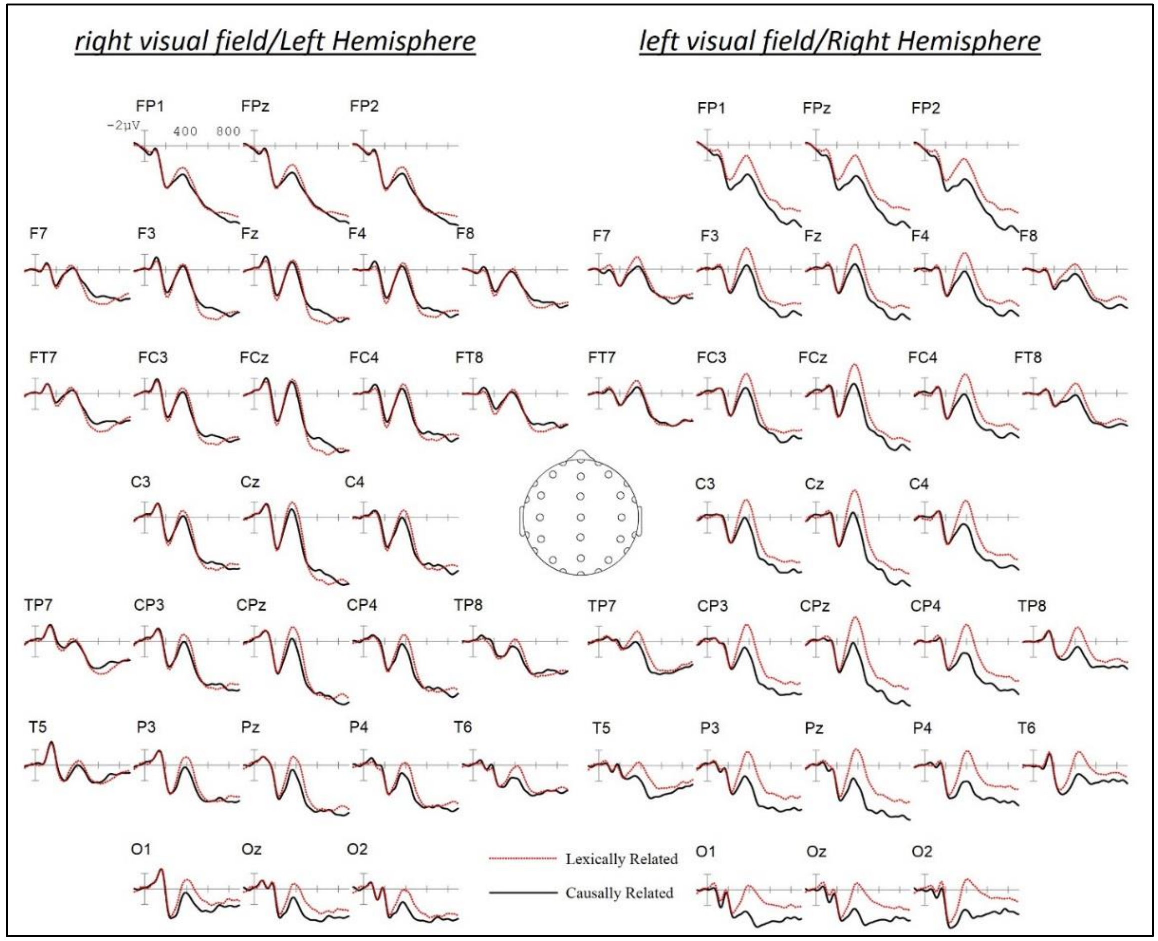

Figure 20: Grand average ERP results of Experiment 6 by probe type and presentation side. The solid red lines represent causal probes, and the dotted black lines represent lexical probes. Negative voltage is plotted upward. Waveforms are low-pass filtered at 7 Hz for for presentation purposes.

Effects of Probe Type

Figure 20 shows that in both presentation sides, the *causal* condition elicited a smaller N400 (300-500ms) response than the *lexical* condition. For both presentation sides, N400 effects were broadly distributed, though largest at central and posterior sites. N400 effect sizes appear larger and more broadly distributed with lvf/RH presentation. Figure 20 also shows that post-N400 (500-800ms) effects differed as a function of presentation side: in rvf/LH the *lexical* condition elicited a late frontal positivity, but in lvf/RH the *lexical* condition elicited a late negativity with a distribution similar to that of the N400 effect.

Effects of Probe Type on the P2 (150-300 ms)

The initial ANOVA yielded no main effects on the P2 of Presentation Side ($F[1,15] < 1$) nor Probe Type ($F[1,15] < 1$). Planned follow-up analyses in both presentation sides revealed no significant effects or interactions involving the Probe Type factor (all p's > 0.1).

Effects of Probe Type on the N400 (300-500 ms)

In contrast to the P2, effects of probe type on the N400 were large and significant. The initial ANOVA revealed that the *causal* probes elicited smaller N400 amplitude overall than the *lexical* probes ($F[1,15] = 18.4$, $p = 0.0006$). The significant Probe Type x

Presentation Side interaction reflected the larger and more robust mean voltage difference in lvf/RH (2.4 µV) compared to rvf/LH (0.9 µV).

Follow-up comparisons were conducted on the medial and midline electrodes. In rvf/LH, both of these analyses suggested the *lexical* probes elicited larger N400 overall than the *causal* probes (midline Probe Type: $F[1,15] = 5.3$, $p = 0.04$; medial Probe Type: $F[1,15] = 4.6$, $p < 0.05$), and the effect was evenly distributed over the scalp, as Probe Type did not interact with any of the factors indexing electrode site.

In lvf/RH, follow-up analyses also revealed that the *causal* probes elicited significantly reduced N400 amplitude compared to the *lexical* probes (midline Probe Type: $F[1,15] = 31.0$, $p = 0.0001$; medial Probe Type $F[1,15] = 27.7$, $p = 0.0001$). Probe Type effects were evenly distributed over the scalp, as the Probe Type factor did not interact with factors indexing scalp distribution (midline Probe Type x Anteriority: $F[6,90] = 2.5$, $p = 0.1$; medial Probe Type x Anteriority: $F[6,90] = 2.4$, $p = 0.1$), medial Probe Type x Hemisphere ($F[1,15] = 1.8$, $p = 0.2$), medial Probe Type x Anteriority x Hemisphere ($F[6,90] < 1$).

Effects of Probe Type on post-N400 components (500-800ms)

The initial repeated measures ANOVA revealed that *causal* probes were marginally more positive than *lexical* probes across both presentation sides ($F[1,15] = 4.2$, $p = 0.06$). Probe type did interact with the Electrode factor, suggesting localized voltage differences in one or both presentation sides ($F[28,420] = 4.7$, $p = 0.003$). Additionally, a significant Probe Type x Presentation Side interaction indicated different probe type effects in each presentation side ($F[1,15] = 6.2$, $p = 0.03$), though these factors did not interact with Electrode ($F[28,420] < 1$).

In rvf/LH, follow-up comparisons revealed a focal effect of Probe Type over fronto-central electrodes, suggested in the analysis by a significant interaction between Probe Type and Anteriority in the midline analysis ($F[6,90] = 4.7$, $p = 0.02$), and a marginal one in the medial analysis ($F[6,90] = 2.7$, $p = 0.08$). Relative to *causal* probes, the *lexical* probes elicited a larger positivity over frontal-central electrodes.

Probe type effects were also observed with lvf/RH presentation, as the *lexical* condition remained significantly more negative than the *causal* condition after the N400 window (midline: $F[1,15] = 13.3$, $p = 0.002$; medial: $F[1,15] = 13.2$, $p = 0.003$). This negativity was broadly distributed on the scalp, reflected statistically by the absence of interactions with factors indexing electrode site.

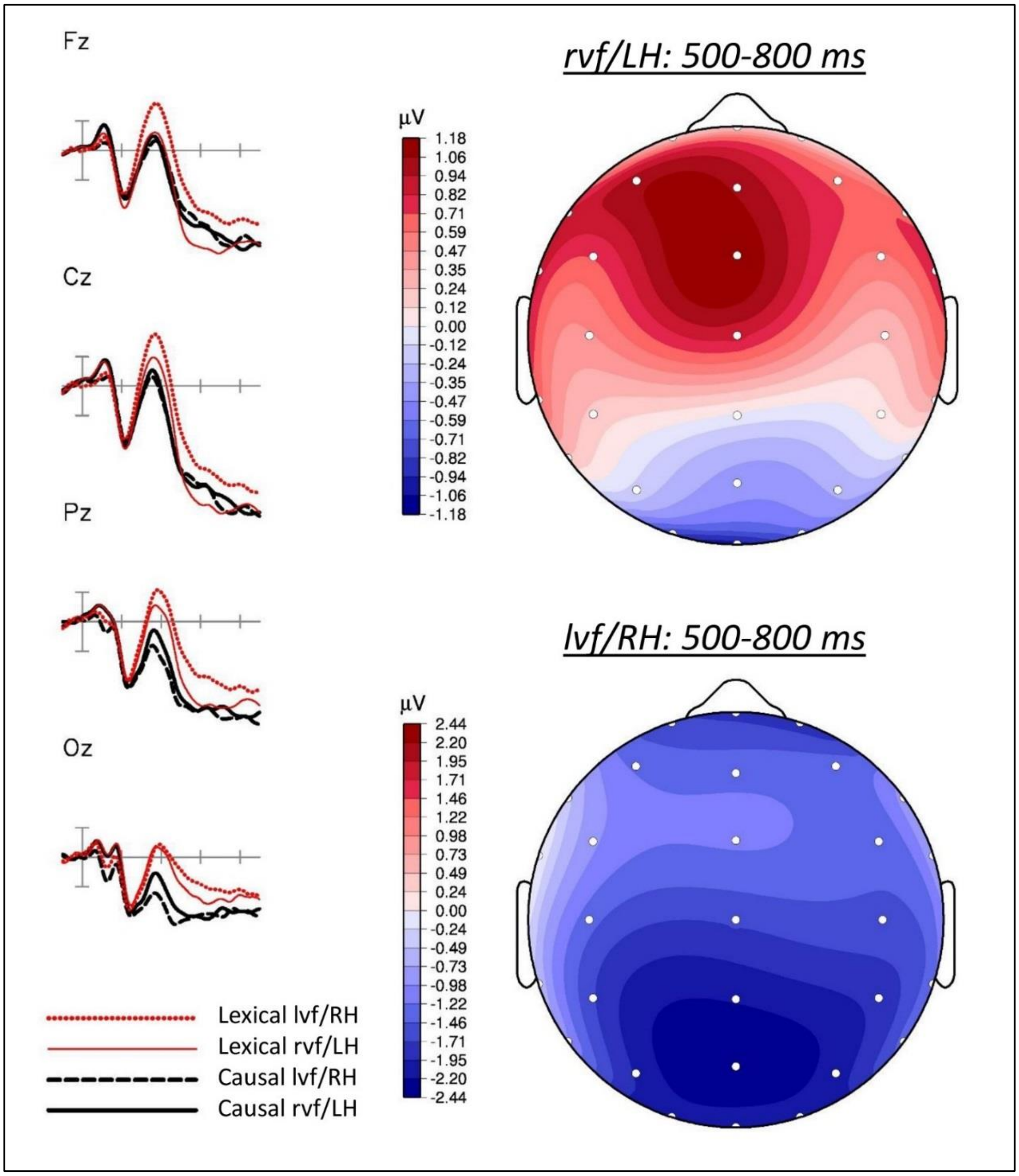

Figure 21: Waveform plots of all conditions at 4 electrode sites and scalp maps of voltage differences in the 500-800 ms window. In the waveform plots, black lines represent causal probes and red lines represent lexical probes; solid lines indicate presentation to rvf/LH and dotted or dashed lines indicate presentation to lvf/RH. Negative voltage is plotted upwards. Scalp voltage maps both indicate the *lexical* - *causal* subtraction, with red shading denoting a net positivity and blue shading a net negativity.

Summary of Probe Type Effects

Effects of probe type are shown in Figure 20. In both visual fields, the *causal* probes elicited a smaller N400 than the *lexical* probes, although this effect was larger and more broadly distributed in lvf/RH. Post-N400 effects differed more dramatically as a function of presentation side. The *lexical* probe words elicited a transient frontal positivity in relation to the *causal* probe words in rvf/LH, and a prolonged negativity in lvf/RH.

Discussion

The aim of this experiment was to compare hemispheric processing abilities for two types of semantic relationship: lexical association and causal relatedness. Participants listened to short stories containing causal coherence gaps, designed to provoke causal bridging inferences. Shortly after the story ended, a word that was either related to the missing causal element *or* was lexically associated with the story's final word appeared in left or right visual field. Selective enhancement of contra-lateral N1 and selection negativity amplitudes indicated that parafoveal presentation succeeded in targeting the contralateral hemisphere.

Results suggest a right hemisphere advantage for the activation of causal information. Although *causal* probes elicited reduced N400 relative to the *lexical* probes in both visual fields, suggesting the facilitative impact of context on semantic retrieval of causal information, the N400 effect was larger with presentation to the lvf/RH. After the

N400, *lexical* probes elicited a sustained, broadly distributed negativity in lvf/RH and a late frontal positivity with rvf/LH presentation. This indicates that RH was more successful in processing the inference-related *causal* probes, but that the two hemispheres pursued qualitatively different strategies in processing the inference-irrelevant *lexical* probes.

Experiment 7

In Experiment 6, the interpretation of the N400 results may have been clouded by possible overlap between the N400 component and the post-N400 components, which would have made the N400 effect appear smaller in rvf/LH and larger in lvf/RH. In addition, it is important to understand whether the post-N400 effects observed in Experiment 6 index different strategies for processing inference-irrelevant words in general, or if it was specifically the high degree of lexical association that was the determining factor in this hemispheric asymmetry. Of particular interest is the rvf/LH frontal positivity. Frontal positivities are usually elicited by highly unexpected items when there is some degree of local coherence (see Van Petten & Luka, 2012 for a review), leading to the prediction that the rvf/LH frontal positivity will not occur if the *causal* probe word is compared to an unrelated probe word that is equally lexically associated with the context.

The intent of Experiment 7 was to resolve this ambiguity and determine whether the apparent RH advantage for causal inference processing persists when all of the stimuli are matched for lexical association. This was tested by replacing the lexical condition with an *unrelated* condition that is matched with the *causal* condition on lexical association to the context story (Landauer, Foltz & Laham, 1998). On one hand, it is possible that using stimuli matched for lexical association will eradicate the conditions that elicit the rvf/LH frontal positivity, leading to more similar effects in each presentation side. A finding of similar causal relatedness effects in both presentation sides would indicate that the results of Experiment 6 were driven primarily by the highly lexically related control condition, undermining the hypothesis of a RH advantage for activating causal information. On the other hand, a rvf/LH frontal positivity effect has been observed in a hemifield experiment comparing discourse-related and unrelated probe words (Coulson & Wu, 2005), suggesting that local relatedness may in fact not be necessary to elicit this effect, and that the frontal positivity is a standard LH response to unexpected events. Such an outcome in the present experiment would indicate that the *lexical* probes in Experiment 6 were treated simply as unrelated by comprehenders, supporting the hypotheses of a RH advantage for activating causal information with a hemispheric asymmetry in how inference-irrelevant items are processed.

Methods

Participants

Participants were 16 UC San Diego undergraduates (9 female), who were compensated in course credit and/or payment. An additional 4 participants were excluded from analysis due to excessive movement and/or blocking artifacts. All participants gave informed consent to participate. All were right-handed, as assessed by the Edinburgh handedness inventory (Oldfield, 1971). No participants reported left-handed members of their immediate family. All were native speakers of English, had normal hearing, had normal or corrected-to-normal vision, and were free of neurological disorders and psychiatric medications. Their ages ranged from 18 to 27, with a mean age of 20.8 years.

Materials

Experimental stimuli consisted of 160 short spoken narratives consisting of two sentences each, as well as visually presented probe words. These were the same narrative stimuli used in Experiment 6 and Experiment 1, with the exceptions that the *lexical related* and *lexical unrelated* probe words were not used, and that the *causal unrelated* probes were matched with the *causal related* probes for lexical association with their narrative contexts. For example stimuli, see **Table 8**. Because the probe word could be presented either in lvf/RH or rvf/LH, four stimulus lists were constructed such that in each list, each audio narrative was presented with only one probe word appearing in one visual field. A four-cell Latin square design ensured that across lists, all narrative/probe word/presentation side combinations occurred.

Because the same list of words was used in the *causal* and *unrelated* conditions, the probe words were identical on all word-level statistical factors, summarized in **Table 2,** Chapter 2**.** Using a modified version of the Hungarian algorithm (Kuhn, 1955), the narrative-probe in the *causal unrelated* condition were matched as closely as possible with the *causal related* condition for associative strength as measured with LSA. LSA ratings between conditions did not differ in relatedness to the final word of the narrative (t = 1.3, p = 0.2), but the two conditions differed marginally in their relatedness to the entire narrative (0.27 vs. 0.24; t = 1.9, p = 0.053).

Procedure

The procedure for Experiment 7 was identical to that of Experiment 6.

EEG Recording

EEG was recorded in a manner identical to that of Experiment 6.

Analysis

ERPs to visually presented probe words were recorded as in Experiment 6. Artifacts were rejected using the same procedures and criteria as in Experiment 6. As noted in section 5.1, four participants were rejected for unacceptable numbers of artifacts. In the data reported below, 15% of trials were rejected due to artifacts. Artifact rejection rates did not significantly differ by presentation side (rvh/LH: 14%, lvf/RH: 16%; two-tailed paired-sample t_{15}=1.2, p = 0.2); *causal* and *unrelated* trials also did not differ significantly in rejection rates (causal: 14%, unrelated: 16%; two-tailed paired-sample t_{15}=1.5, p = 0.14).

Analyses of Presentation Side Effects

Analyses of Presentation Side effects on the N1 and selection negativity were conducted in a manner identical to those in Experiment 6.

Analyses of Relatedness Effects

As in Experiment 6, analysis of ERP components focused on three time windows: 150-300ms post-onset (P2 component), 300-500ms (N400 component), and 500-800ms (post-N400 effects). In each time window, mean voltage amplitudes elicited by the causal and lexical conditions were analyzed on the midline (FPz, Fz, FCz, Cz, CPz, Pz, Oz) and medial electrodes (FP1/2, F3/4, FC3/4, C3/4, CP3/4, P3/4, and O1/2) with a repeated measures ANOVA using a 2 (Presentation Side: rvf/LH vs. lvf/RH) x 2 (Relatedness: related vs. unrelated) x 7 (Anteriority: Prefrontal to Occipital) x 3 (Laterality: Left Medial, Midline, Right Medial). Planned follow-up analyses were conducted on medial electrodes only. Within each presentation side, interactions between relatedness and scalp location factors were analyzed using a 2 (Relatedness) x 7 (Anteriority) x 2 (Hemisphere) design. To compensate for violation of the sphericity assumption, the Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied to all reported p-values. For clarity, we report the uncorrected degrees of freedom.

Results

Effects of Presentation Side

Posterior temporal N1 responses (100-200 ms) were enhanced by contralateral word presentation (see figure 22). This was confirmed by a significant Presentation Side X Hemisphere interaction effect ($F[1,15] = 26.8$, $p < 0.0001$). The selection negativity (300-800 ms) was also larger with contralateral word presentation on those sites (Presentation Side X Hemisphere: $F[1,15] = 31.5$, $p < 0.0001$). Unlike the N1 effect, however, this interaction appeared to be driven by the much larger voltage difference over RH sites (see figure 22).

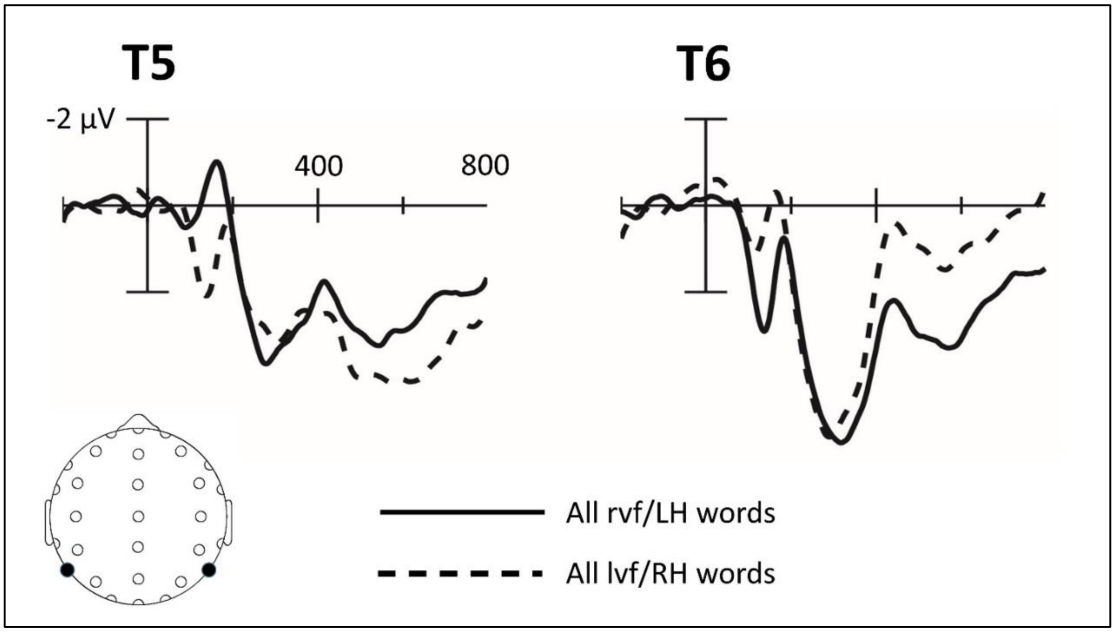

Figure 22: Grand average ERP responses in Experiment 7 to lateralized words categorized by presentation side. Two posterior temporal sites are displayed, showing that contra-lateral presentation leads to enhanced N1 potentials as well as enhanced selection negativity. ERPs are presented from a 100 ms pre-stimulus baseline to 900 ms post-stimulus. Negative voltage is plotted upwards. ERPs are low-pass filtered at 10 Hz for for presentation purposes.

Effects of Relatedness

P2 Effects (150-300)

The initial repeated measures ANOVA did not reveal any main effect of Relatedness in the P2 window ($F[1,15] = 2.9$, $p = 0.1$), nor was there any evidence of a modulation of effect size by presentation side ($F[1,15] < 1$), nor of interactions between Relatedness, Presentation, and either scalp location factor (all $Fs < 1$). Planned comparisons within each presentation side also failed to reveal any main or interaction effects involving the Relatedness factor (all p-values > 0.18).

N400 Effects (300-500 ms)

In both presentation sides, the *causal* probes elicited sharply reduced N400 compared to the *unrelated* probes (F[1,15] = 47.9, p < 0.0001). However, relatedness did not interact with Presentation Side [F1,15] = 1.1, p 0.3). The N400 effects were, however, distributed differently between presentation sides (Relatedness X Presentation Side X Anteriority X Laterality: F[12,180] = 2.4, p = 0.03; all other interactions non-significant).

Planned comparisons were also conducted within each presentation side. *Related* probes elicited significantly smaller N400 amplitude in rvf/LH (F[1,15] = 26.4, p = 0.0001). This effect was focused on parietal sites, causing a Relatedness x Anteriority interaction (F[6,90] = 4.4, p = 0.049) but no interactions with the Hemisphere factor (both Fs < 1). The lvf/RH N400 effect was also significant (F[1,15] = 35.5, p < 0.0001). However, it was more broadly distributed with respect to anteriority (F[6,90] = 1.6, p = 0.2). A marginal Relatedness X Anteriority X Hemisphere interaction reflected larger effects over right central scalp (F[6,90] = 2.2, p = 0.08). Overall, the rvf/LH N400 effect (2.6 µV on average) was numerically larger than the lvf/RH N400 effect (2.1 µV), but the sizes of the N400 effects were statistically similar between presentation sides.

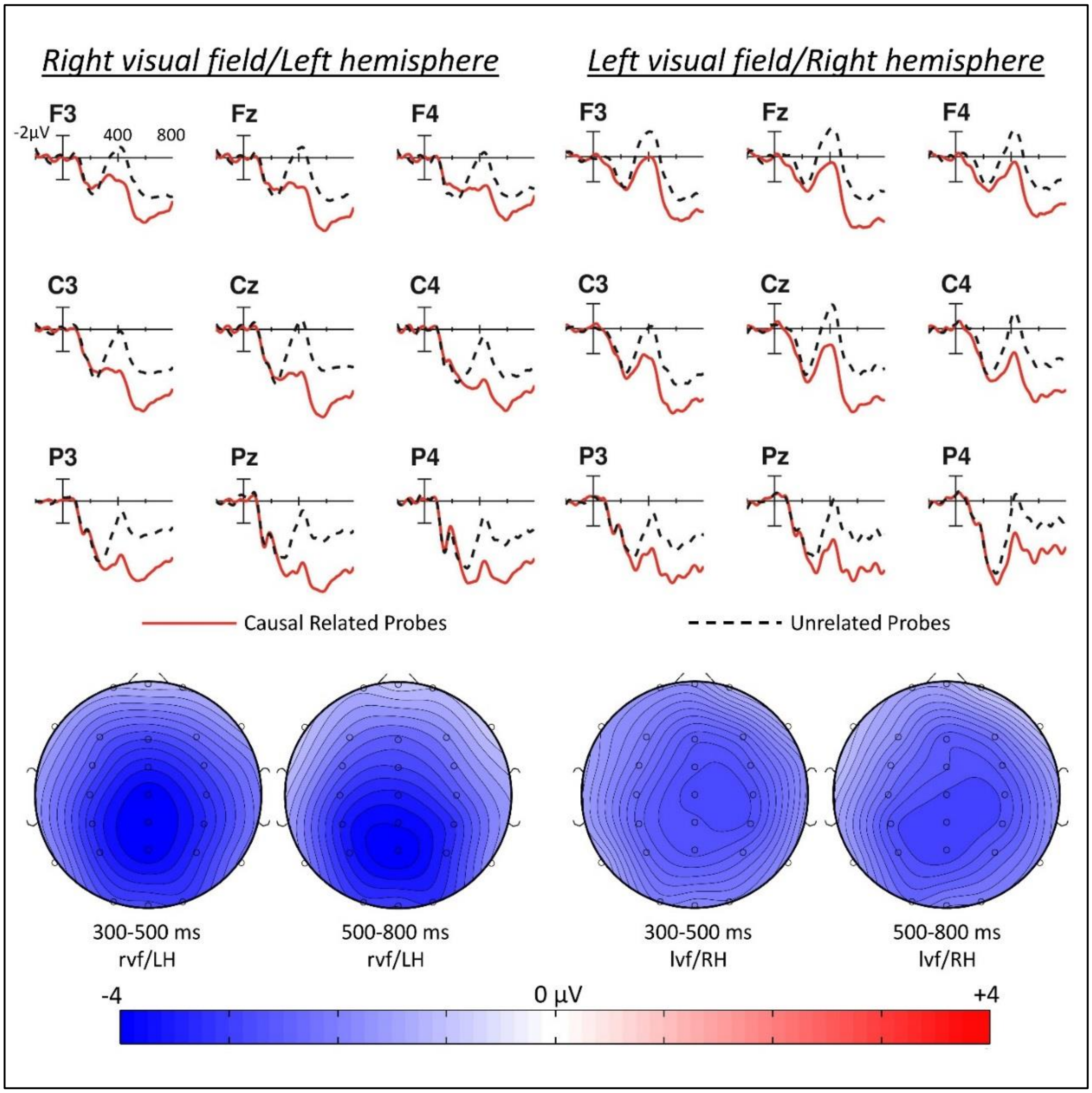

Figure 23: Experiment 7 results. Nine electrode channels in each presentation side are shown, with solid red lines denoting the *causal related* condition and dashed black lines denoting the *unrelated* condition. Below, scalp voltage maps of the N400 and sustained negativity (500-800 ms) are shown for each presentation side.

Post-N400 Negativity Effects (500-800 ms)

In both presentation sides, the *unrelated* probes elicited an enhanced negativity after the end of the traditional N400 window (see Figure 23). The *unrelated* condition was more negative overall than the *causal* condition ($F[1,15] = 39.1$, $p < 0.0001$). The mean voltage difference was virtually identical in two presentation sides (2.1 μV in both visual fields), however an interaction effect with both scalp factors indicated differences in scalp distribution dependent on presentation side (Relatedness X Presentation Side X Anteriority X Laterality: $F[6,90] = 2.8$, p 0.01; all other interactions: $p > 0.3$).

Once again, planned follow-up comparisons confirmed that the *unrelated* words elicited more negative ERPs in both rvf/LH ($F[1,15] = 31.1$, $p < 0.0001$) and in lvf/RH ($F[1,15] = 17.6$, $p = 0.0008$). Topographical analyses showed a left parietal concentration in rvf/LH (Relatedness X Anteriority X Hemisphere: $F[6,90] = 2.9$, $p = 0.047$), but no such interaction was present in lvf/RH. In both presentation sides, a marginal Relatedness X Anteriority interaction attested to the parietal or centro-parietal focus of the sustained negativity effect (rvf/LH: $F[6,90] = 3.1$, $p = 0.06$; lvf/RH: $F[6,90] = 4.2$, $p = 0.054$).

Discussion

EEG was recorded as participants listened to short narratives that encouraged a causal bridging inference, then read lateralized probe words that were either related or unrelated to the inference that the stories were intended to trigger. Probe words in the *causal related* and *unrelated* conditions consisted of the same set of words paired with different narrative stimuli. In addition, *causal related* and *unrelated* probe words were matched for the degree of lexical association with their narrative contexts. The effects of causal relatedness were remarkably similar in each visual field. The *unrelated* probes elicited larger N400 than the *causal* probes did, and this negative going effect continued 500-800ms post-onset. Presentation side modulated the scalp distribution of relatedness effects, consistent with the claim that slightly different neural generators were active with presentation to the left versus the right visual field. However, the overall size and pattern of relatedness effects did not reliably differ as a function of presentation side.

These results argue against activation-based accounts of hemispheric asymmetry in causal inference (Beeman, 1993; Beeman et al., 1994; Coulson & Wu, 2005). The results are more consistent with bilateral inference activation, with RH possibly dominating in later aspects of inference processings such as inference selection and maintenance (Lehman-Blake & Tompkins, 2001; Tompkins et al., 2004). These results also suggest that the results of Experiment 6, which argued for a RH advantage in causal inference processing, were an artifact of hemispheric asymmetry in attentiveness to lexical association in discourse contexts that render those associations irrelevant. Experiment 8 will test this interpretation by comparing both *causal* and *lexical* probe words to their matched control words in a similar paradigm.

Experiment 8

Experiment 8 further investigates the role of RH in processing causally related words by comparing all four conditions developed in Chapter 2, using the hemifield paradigm employed in Experiments 6 and 7. A straightforward extrapolation from our previous two hemifield experiments suggests that no evidence for a RH causal inference advantage will be found: effects of causal relatedness should be similar in the two hemifields, as was observed in Experiment 7. In keeping with Experiment 6, we also predict that the lexical relatedness manipulation will elicit a frontal positivity in rvf/LH, suggesting a hemispheric asymmetry in lexical associative processing but not in causal inference activation.

Methods

Participants

Participants were 21 UC San Diego undergraduates (11 female), who were compensated in course credit and/or payment. An additional 4 participants were excluded from analysis due to excessive movement and/or blocking artifacts. All participants gave informed consent to participate. All were right-handed, as assessed by the Edinburgh handedness inventory (Oldfield, 1971). Five participants reported left-handed members of their immediate family. All participants were native speakers of English, had normal hearing, had normal or corrected-to-normal vision and were free of neurological disorders and psychiatric medications. Their ages ranged from 18 to 32, with a mean age of 21.1 years.

Materials

Experimental stimuli consisted of 160 short spoken narratives consisting of two sentences each, as well as visually presented probe words. These were the same narrative stimuli used in Experiment 1, with with all four probe conditions used. For example stimuli, see **Table 8**. Because the probe word could be presented either in lvf/RH or rvf/LH, eight stimulus lists were constructed such that in each list, each audio narrative was presented with only one probe word appearing in one visual field. A four-cell Latin square design ensured that across lists, all narrative/probe word/presentation side combinations occurred. Each participant was exposed to 20 items per condition. For a summary of the statistics of the stimuli, see **Table 2**, Chapter 2, and **Table 9**.

Procedure

The procedure for Experiment 8 was identical to that of Experiments 6 and 7, except that 1000 ms after the offset of the visual probe word, participants indicated with a button press whether they were able to identify the visual probe word. A schematic depiction of a *causal related* rvf/LH trial is shown in Figure 24.

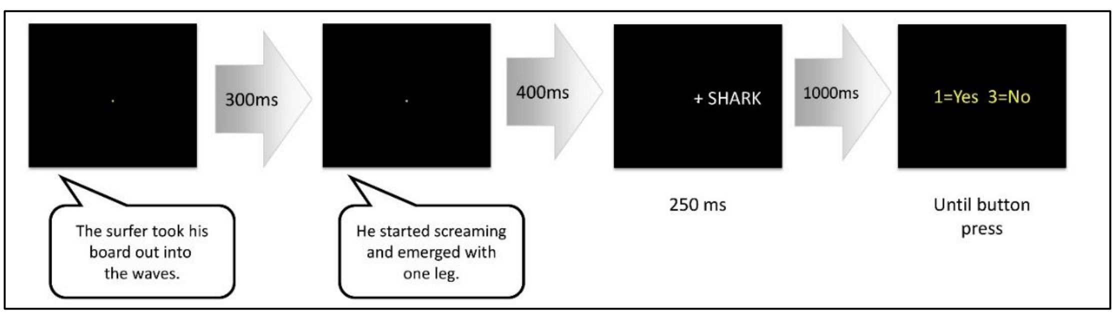

Figure 24: Schematic depiction of a *causal related* rvf/LH trial in Experiment 8.

EEG Recording

EEG was recorded in a manner identical to that of Experiments 6 and 7.

Analysis

As in Experiments 6 and 7, only ERPs to the visual probe words were analyzed. ERPs to visual probe words were time-locked to word onset and averaged in a time window spanning 200ms pre-onset to 900ms post-onset. The period from 200ms pre-onset to stimulus onset served as the baseline. The rejection procedures and criteria were the same as those used in Experiments 6 and 7. As noted in section 8.1, four participants were rejected for having an unacceptable number of artifacts. In the data reported below, 11% of trials were rejected due to artifacts. Artifact rejection rates did not significantly differ by presentation side (rvh/LH: 11%, lvf/RH: 10%; two-tailed paired-sample $t_{20} < 1$). However, the percentage of probe words that were rejected because participants reported being unable to read them was significantly lower in rvf/LH (8%) than in lvf/RH (18%) ($t_{20} = 4.3$, $p = 0.0002$). This difference reflects the well-known rvf advantage for reading in right-handed individuals (Bradshaw & Nettleton, 1983).

Analyses of Presentation Side Effects

Presentation side effects on the N1 and selection negativity were analyzed just as in Experiments 6 and 7.

Analyses of Relatedness Effects

Analysis of ERP components focused on three time windows: 150-300ms post-onset (P2 component), 300-500ms (N400 component), and 500-800ms (post-N400 effects). In each time window, three planned analyses of mean voltage amplitudes were conducted. Mean voltage amplitudes elicited by all four probe conditions were analyzed on the midline (FPz, Fz, FCz, Cz, CPz, Pz, Oz) and medial electrodes (FP1/2, F3/4, FC3/4, C3/4, CP3/4, P3/4, and O1/2) with a repeated measures ANOVA using a 2 (Presentation Side: rvf/LH vs. lvf/RH) x 2 (Probe Type: Causal vs. Lexical) x 2 (Relatedness: related vs. unrelated) x 7 (Anteriority: Prefrontal to Occipital) x 3 (Laterality: Left Medial, Midline, Right Medial). Planned follow-up analyses were conducted on the condition comparisons from Experiments 6 and 7: *causal related* vs. *lexical related* and *causal related* vs. *causal unrelated*, respectively. Both of those condition comparisons were analyzed using a 2 (Condition) x 2 (Presentation Side) x 7 (Anteriority) x 3 (Laterality) design. When a significant interaction effect between the Probe Type and/or Relatedness factor and the Presentation Side factor was observed, a follow-up analysis was conducted within each presentation side to assess hemispheric asymmetries in the brain responses to those conditions. Follow-up analyses were repeated measures ANOVAs involving the relevant probe factors, 7 levels of Anteriority, and 3 levels of Laterality. To compensate for violation of the sphericity assumption, the Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied to all reported p-values. For clarity, we report the uncorrected degrees of freedom.

Results

Effects of Presentation Side

Hemifield presentation was associated with larger posterior temporal N1 responses (100-200ms) over the contrateral hemisphere (see figure 25). This was confirmed by a significant Presentation Side X Hemisphere interaction effect (F[1,20] = 17.5, p = 0.0005). The selection negativity (300-800ms) was also larger with contralateral word presentation on those sites (Presentation Side X Hemisphere: F[1,20] = 23.5, p = 0.0001).

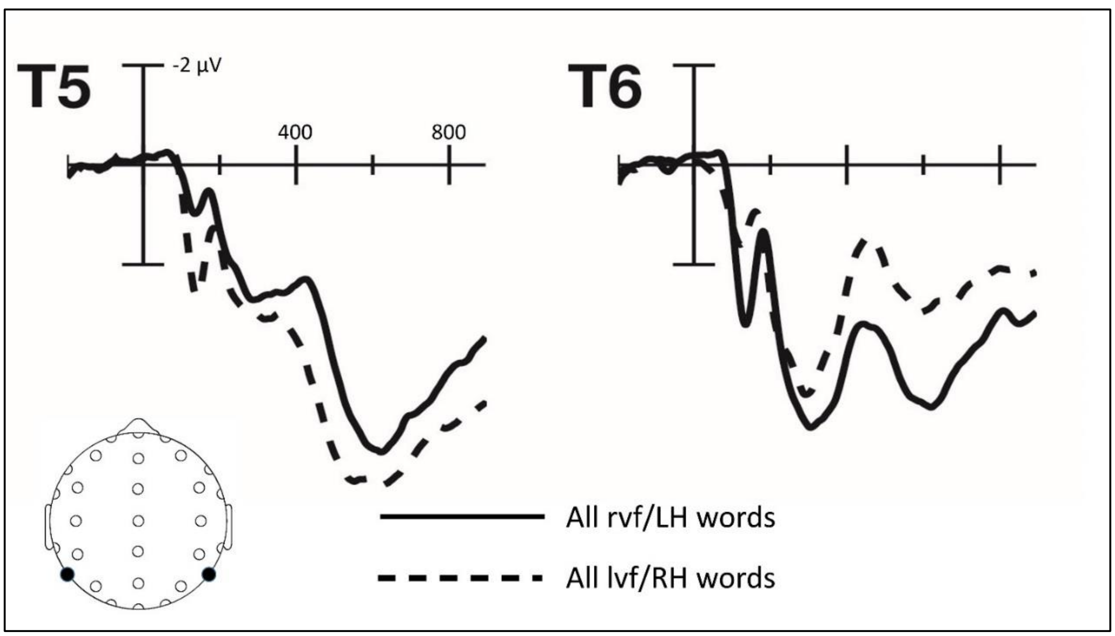

Figure 25: Grand average ERP responses to lateralized words categorized by presentation side. Two posterior temporal sites are displayed, showing that contra-lateral presentation leads to enhanced N1 potentials as well as enhanced selection negativity. ERPs are presented from a 200ms pre-stimulus baseline to 900ms post-stimulus. Negative voltage is plotted upwards. ERPs are low-pass filtered at 10 Hz for for presentation purposes.

Effects of Relatedness

P2 Effects (150-300)

Relatedness effects are depicted in Figures 26, 28 and 29. The initial omnibus analysis in the P2 window revealed a Presentation Side x Probe Type x Anteriority interaction ($F[6,120] = 6.1$, $p = 0.005$). This interaction was followed up with a Probe Type x Anteriority x Laterality test within each presentation side. In rvf/LH, a significant Probe Type x Anteriority interaction pointed to more positive early ERPs in the Causal conditions over anterior sites ($F[6,120] = 3.8$, $p = 0.03$). In lvf/RH, only a Probe Type x Laterality interaction was found, reflecting more positive early ERPs to the Lexical conditions over left medial sites ($F[2,40] = 5.4$, $p = 0.008$). The other planned comparisons in the P2 window revealed no significant effects or interactions involving either the Relatedness or Probe Type factors (all p-values > 0.1).

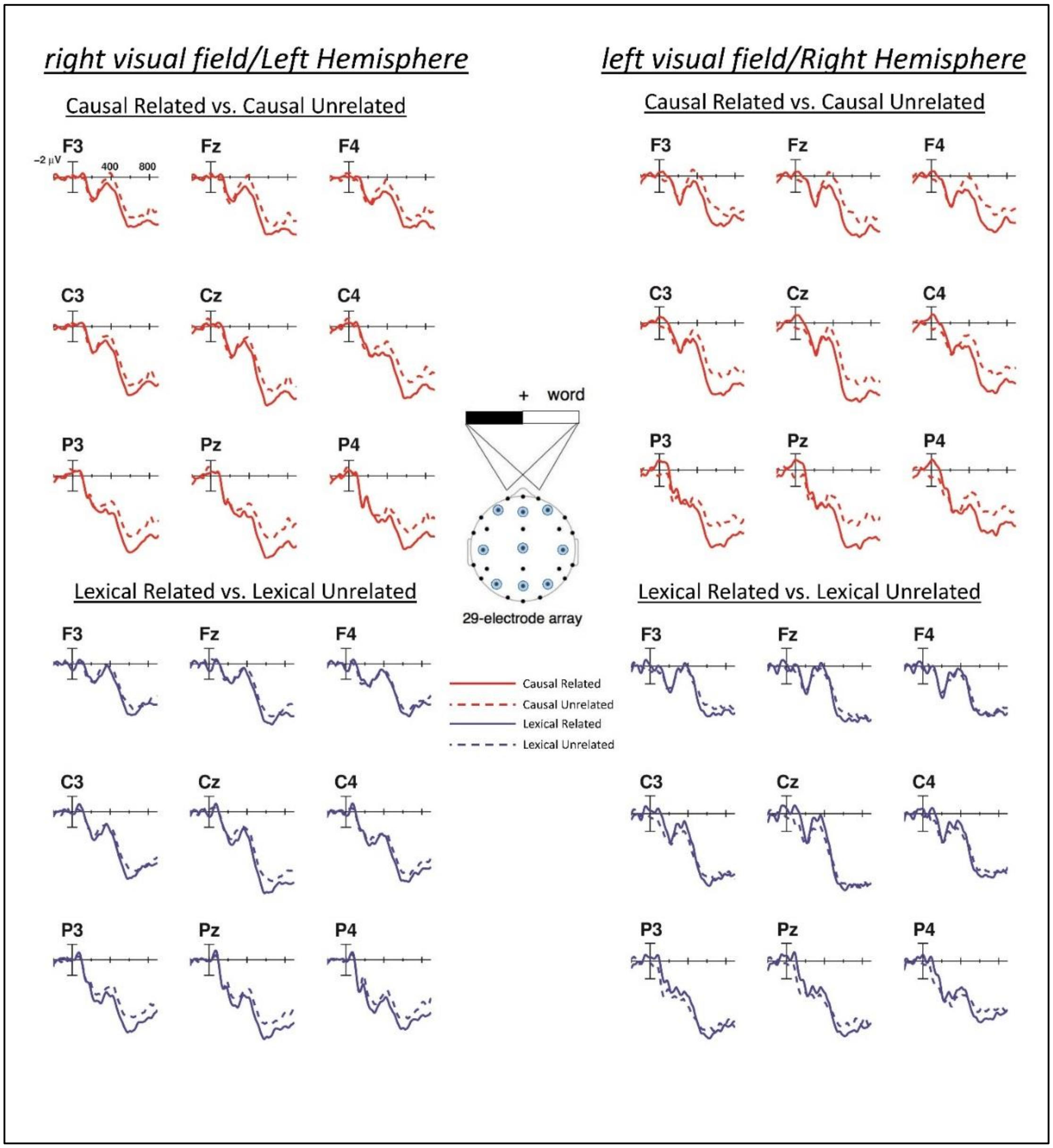

Figure 26: Grand average ERPs for Experiment 8. Negative voltage is plotted upwards, and all waveforms are low-pass filteres at 10Hz for for presentation purposes.

N400 Effects (300-500ms)

The initial N400 analysis revealed that *related* probes elicited reduced N400 compared to *unrelated* probes, reflected by a main effect of Relatedness (F[1,20] = 6, p =

0.02). Planned analyses involving the Probe Type factor manipulated in Experiment 6 and the Causal Relatedness factor manipulated in Experiment 7 confirmed that this Relatedness effect was driven by the contrast between the *causal related* and *causal unrelated* conditions ($F[1,20] = 7.2$, $p = 0.01$), and that the *causal related* and *lexical related* conditions did not differ in the N400 window ($F < 1$). There were no significant interaction effects involving the Presentation Side factor, indicating that this causal relatedness effect was of similar size in both hemifields.

Post-N400 Negativity Effects (500-800ms)

The initial analysis of the post-N400 window indicated that across the board, *related* probes elicited more positive voltage responses than *unrelated* probes ($F[1,20] = 12.7$, $p = 0.002$). Relatedness did not interact with Presentation Side (all p-values > 0.3), however there was a significant Presentation Side x Probe Type x Laterality interaction ($F[2,40] = 3.6$, $p = 0.04$), which triggered follow-up analyses of the Probe Type factor within each level of the Presentation Side factor. However, neither of these follow-up tests yielded significant interaction or effects of Probe Type.

Two additional planned analyses, however, compared the two *related* conditions and the two *causal* conditions, replicating the comparisons of Experiments 6 and 7, respectively. Again, the Relatedness effect observed in the omnibus analysis reflected more positive responses to the *causal related* condition than to the *causal unrelated* condition ($F[1,20] = 10.3$, $p = 0.004$), however, this causal relatedness effect did not interact with the Presentation Side factor (all p-values > 0.3).

Post-hoc analysis of the 200-400ms window

Visual inspection of the Experiment 8 waveforms suggested an early hemispheric asymmetry between the *lexical related* and *causal related* probes that was not captured by planned analyses. Post-hoc analyses were conducted to examine these voltage differences. Initially, a 2 (Probe Type) x 2 (Presentation Side) x 3 (Laterality) x 7 (Anteriority) repeated measures ANOVA was conducted on mean voltage amplitudes elicited by *related* probe words. No significant effects involving the Probe Type factor were found in this omnibus analysis (all p-values > 0.1). However, in the follow-up analyses conducted within each presentation side (and therefore excluding Presentation Side as an analysis factor), *lexical related* probes were found to elicit more negative voltages than *causal related* probes over anterior sites, and more positive voltages over posterior sites with rvf/LH presentation. This pattern caused a significant Probe Type x Anteriority interaction ($F[6,120] = 4.6$, $p = 0.02$). No other significant results were found in rvf/LH, nor in lvf/RH. However, there was a marginal Probe Type x Laterality interaction in lvf/RH, reflecting more negative responses to the *lexical related* probes of right hemisphere sites ($F[2,40] = 3$, $p = 0.09$).

Summary of Relatedness Effects

The effects of causal relatedness were quite similar in each visual field: the *causal unrelated* probes elicited larger N400 and sustained negativity responses than the *causal* *related* probes did. Effects of probe type (viz. *causal* vs. *lexical* conditions), however, tended to be small and localized.

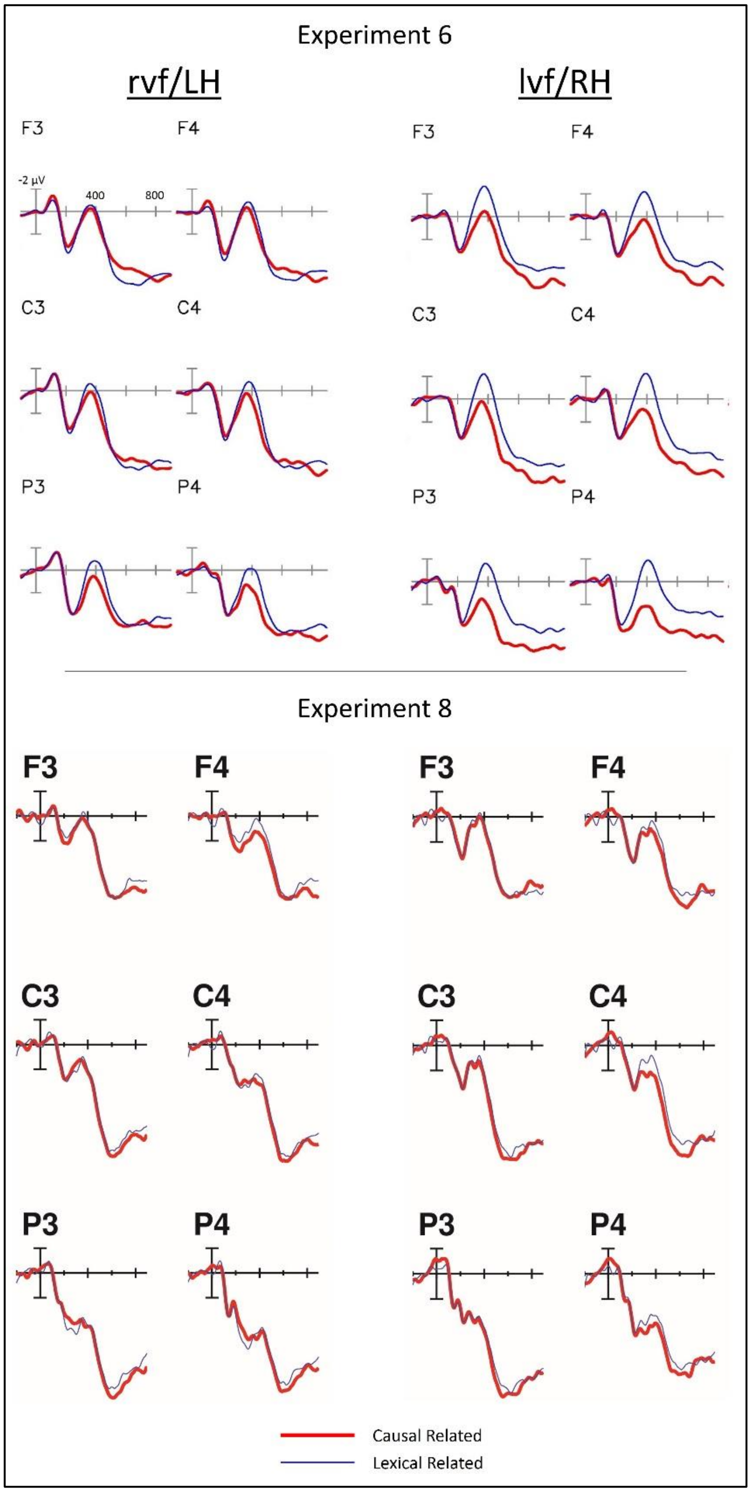

Figure 27: Comparison of probe type effects on *causal related* and *lexical related* probes in Experiments 6 and 8 (top and bottom panes, respectively). *Causal related* probes are indicated by the thick red line and *lexical related* by the thin blue line.

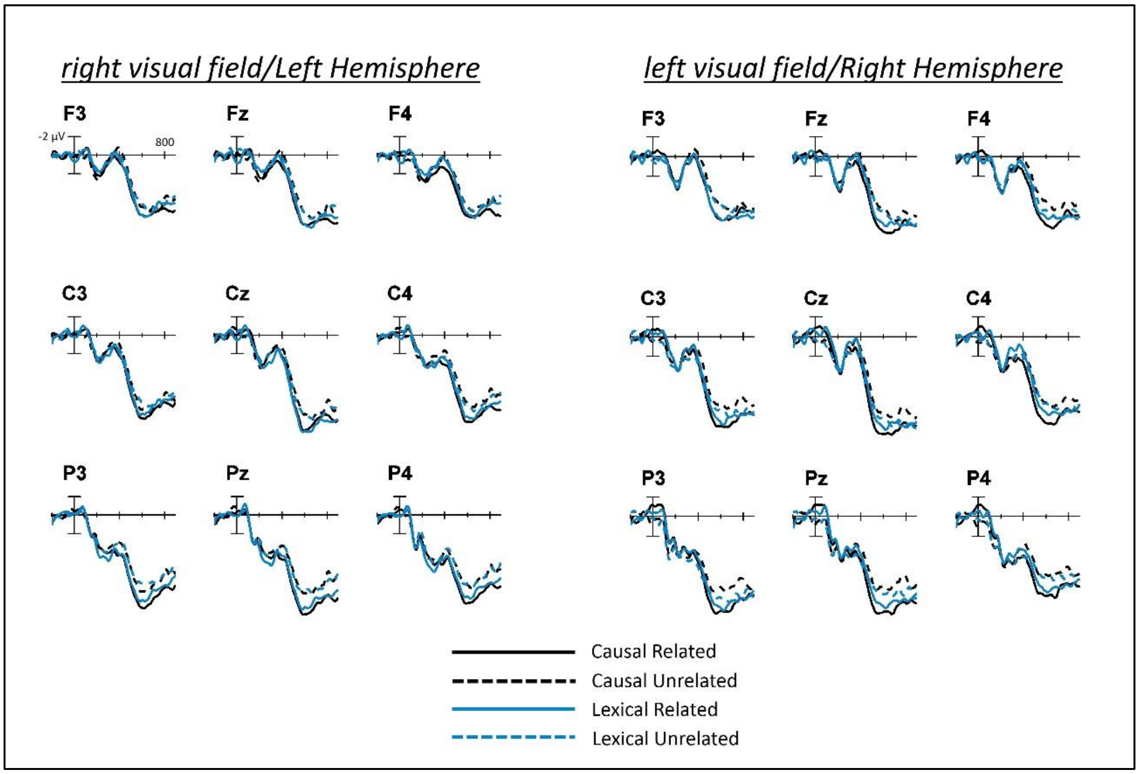

Figure 28: ERPs to all four conditions in Experiment 8, in both presentation sides. Related conditions are denoted by solid lines and Unrelated conditions by dashed lines. Causal conditions are denoted by black lines and Lexical conditions by blue lines.

Discussion

The intent of Experiment 8 was to compare the effects of causal and lexical relatedness using a hemifield paradigm, in order to test hypotheses about the relative contributions of the two hemispheres to the activation of information related to a causal bridging inference. A secondary purpose of the experiment was to test whether effects of probe type and causal relatedness observed in Experiments 6 and 7 were replicated in a hemifield experiment that included more conditions (Experiments 6 and 7).

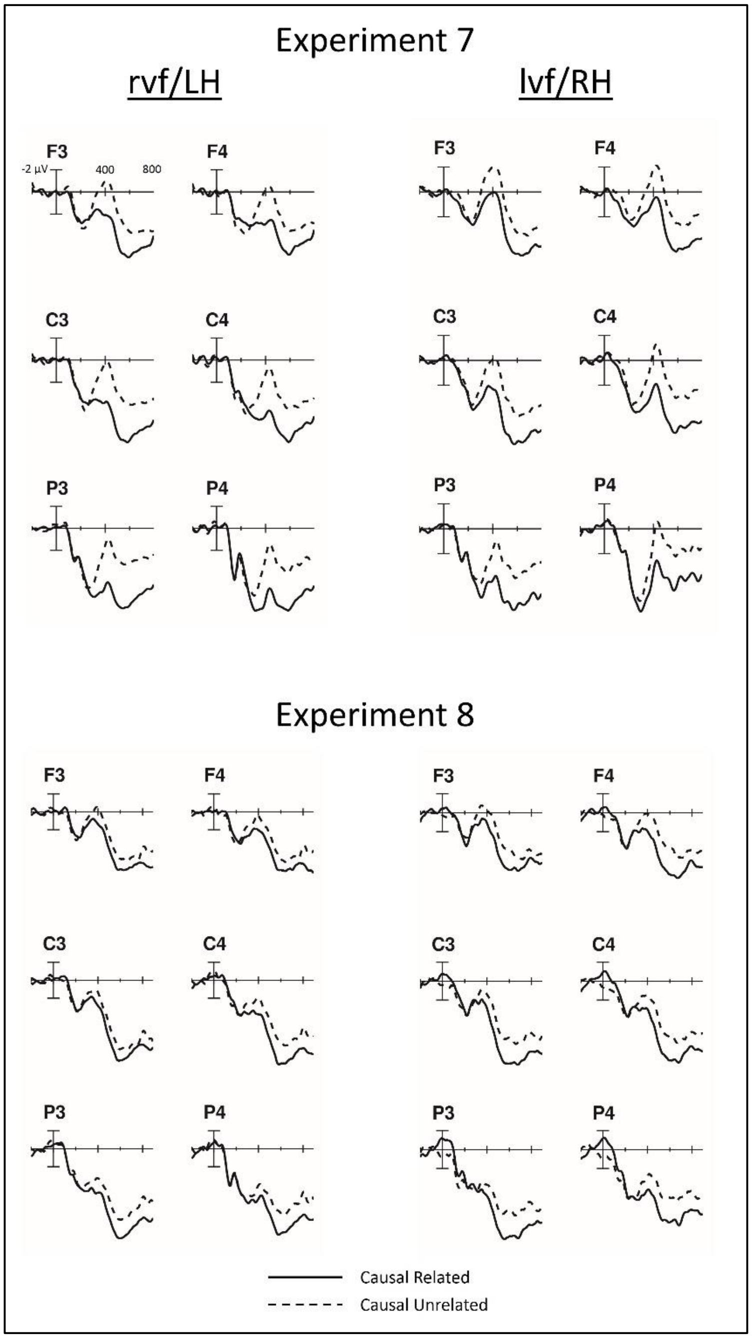

Figure 29: Comparison of relatedness effects on *causal* probes in Experiments 7 and 8 (top and bottom panes, respectively). *Causal related* probes are indicated by solid black lines and *causal unrelated* probes by the dashed black line.

In Experiment 8, causal relatedness ERP effects of similar size were observed in both hemifields, a result analogous to that of Experiment 7. *Causal related* probes elicited less negative ERPs than *causal unrelated* probes from 300ms post-onset until 800ms, indicating that causal information after an inference was active in both hemispheres. In marked contrast to Experiment 6, probe type effects were largely absent in both hemifields, although an early (200-400 ms) posterior positivity in the *lexical related* condition compared to the *causal related* condition was observed in rvf/LH, as well as a marginal negativity in the same time window in lvf/RH.

Comparison to Prior Results

This pattern of effects differed in some interesting ways from ERP results obtained with the same stimuli in Experiment 1, in which probe words were presented centrally in order to promote more normal patterns of hemispheric participation in language processing. With central presentation, causal relatedness effects were characterized by a frontal P2 effect, followed by a broadly distributed N400 effect and then a frontal sustained negativity. Throughout the epoch, *causal related* probes elicited less negative ERPs than *causal unrelated* probes. In the present study, however, the P2 effect of causal relatedness was absent in both hemifields, and the effects that followed were smaller than they were with central presentation: a 1.1 µV N400 effect in the present experiment vs. 2.6 µV with central presentation; the sustained negativity effect averaged 1.5 µV in the present study vs. 2.0 µV in Experiment 1. We attributed the P2 effect in Experiment 1 to successful word form prediction, so its absence in the present study could mean that the attentional demands of the hemifield paradigm prevented participants from forming predictions of the *causal related* probe words. Apart from the elimination of the P2 effect, hemifield presentation did not qualitatively change ERP responses to *related* and *unrelated* probes, but rather reduced effect sizes across the board.

The paradigm of Experiment 8 is quite similar to Experiment 7, which compared ERPs to causal related and causal unrelated probes presented in the visual hemifields. In fact, the only difference between Experiments 8 and 7 was that Experiment 8 also included materials presented with *lexical related* and *lexical unrelated* probes. Despite the inclusion of these additional conditions, causal relatedness effects were quite similar in the two experiments. In both studies, relatedness effects were observed 300-800ms with *causal related* probes eliciting less negative ERPs than *causal unrelated* probes. Moreover, in both studies causal relatedness effects were similar in both presentation sides, suggesting that causal relatedness information was equally active in both hemispheres.

Experiment 8, however, conflicted somewhat with results of Experiment 6, which ERPs to the *causal related* and *lexical related* probes presented in the visual hemifields. In Experiment 6, results indicated a RH advantage for processing the *causal related* probes relative to *lexical related* probes, particularly evident in the N400 component. Additionally, we observed in Experiment 6 a large rvf/LH frontal positivity elicited by *lexical related* probes. In Experiment 8 we observed much smaller, more localized differences between the *causal related* and *lexical related* probes. The pattern of these probe type effects – a posterior positivity in rvf/LH and a marginal negativity in lvf/RH – offers modest support for the idea of qualitatively different LH processing strategies for lexically related items, which was motivated by the unreplicated finding of an enhanced rvf/LH late frontal positivity in Experiment 6. Instead of a hemispheric asymmetry for causal processing in particular, the asymmetries in these data to be a driven by differing hemispheric strategies for processing the *lexical related* probes. That is, LH classified the *lexical related* probes as unexpected but locally congruent, a condition that has been shown to elicit enhanced rvf/LH frontal positivity (Coulson et al., 2005). However, RH classified them simply as unrelated.

Taken together, all of these results argue against activation-based accounts of hemispheric asymmetry in causal inference (Beeman, 1993; Beeman et al., 1994; Coulson & Wu, 2005). Our data are most consistent, however, with a class of hypotheses suggesting that discourse inferences are activated and maintained in a collaborative process involving both hemispheres. Several fMRI studies have reported bilateral activation in classical language areas and their RH homologues for short stories requiring causal bridging inferences (Kuperberg, Lakshmanan, Caplan & Holcomb, 2006; Mason & Just, 2004; Prat, Mason & Just, 2011). It is still possible that RH cortical areas play a specialized role in the inference process that does not strongly affect the amount of inference-related information kept active in each hemisphere (Kuperberg et al., 2006; Mason & Just, 2004; Prat et al., 2011) but nevertheless leads to an impairment in causal inference when that function is knocked out by RHD (Beeman, 1993; Tompkins et al., 2004; Lehman-Blake & Tompkins, 2001).

Reconciliation with Lesion Findings

Our findings of similar effects of causal relatedness in both presentation sides are difficult to reconcile with the lesion and functional neuroimaging findings that point to a RH advantage for causal inference. One possible resolution to this conflict is the evidence from a variety of neuropsychological methods suggesting that RH is most dominant in the early stages of the causal inference process (Beeman et al., 2000; Virtue Haberman, Clancy, Parrish & Jung-Beeman, 2006). These data suggest that RH has an advantage in predictive inference construction, activating information that is likely to be useful for understanding upcoming text. Post-hoc inferences are processed bilaterally or even with a LH advantage (Beeman et al., 2000; Virtue et al., 2006; Virtue, Parrish & Jung-Beeman, 2008). This hypothesis of the RH role in causal inference shall be termed "the early RH activation account." Under an early RH activation analysis of our own data, RH would have dominated processing during the first sentence or so of our narrative stimuli, but after the coherence gap became apparent, the role of LH in processing would have increased, resulting in balanced bilateral representation of causal information by the time the probe word appeared, despite the actual existence of a RH advantage for causal inference processing. Under this view, the RH advantage for joke comprehension (Coulson & Williams, 2005; Coulson & Wu, 2005) persisted after the punchline had already been delivered (e.g., Coulson & Wu, 2005) only because the crucial manipulation that made the joke meaning clear occurred right at the end of the stimulus sentence (e.g., "A replacement player with a home run with my girl/ball.").

The early RH activation account has been criticized based on lesion study evidence that RHD patients actually do activate causal inferences as well as neurologically normal participants, and their deficit lies primarily in how they apply that information to comprehension of later text. For example, the “maintenance account” posits that discourse inferences are initially activated in both hemispheres but are kept active longer in RH than in LH. This “maintenance account” is supported by a lesion study showing that when RHD patients and control participants read stories that required a bridging inference, with the biasing information critical to the bridging inference occurring either 1 sentence or 3 sentences before the coherence gap, the RHD patients showed larger recency effects and smaller effects of inference vs. control condition in a lexical decision task on probe words (Lehman-Blake & Tompkins, 2001).

It is also possible that inferential information is initially activated and maintained about equally well in both hemispheres, but that RH has an advantage in selecting the appropriate inference from among competing alternative interpretations. This “RH selection account” is supported by a follow-up study to that of Lehman-Blake and Tompkins (2001). When the final sentence of the story stimulus was, in context, consistent with the appropriate inference, but in isolation more consistent with an alternative interpretation, RHD participants’ performance on lexical decision to probe words suggested that they had activated and failed to suppress the inappropriate interpretation (Tompkins, Fassbinder, Lehman-Blake, Baumgaertner & Jayaram, 2004).

The results of Experiment 8 offer some support for this hypothesis. Even though both presentation sides had similar causal relatedness effects, suggesting balanced bilateral processing of causal information, the ERP responses to the *causal related* and *lexical related* conditions differed as a function of presentation side in the time window just prior to the N400. In particular, the enhanced negativity elicited by the *lexical related* condition in lvf/RH suggests that RH was somewhat more able to prioritize global context over local context. When this capacity is lost due to RHD, patients may as a result be more distracted from the global meaning by local relationships.

Conclusion

When strength of lexical association was controlled, causal relatedness ERP effects of similar size were observed in both hemifields, indicating that causal information after an inference was active in both hemispheres. The results of Experiments 6, 7, and 8 argue against activation-based accounts of hemispheric asymmetry in causal inference (Beeman, 1993; Beeman et al., 1994; Coulson & Wu, 2005). The results are more consistent with bilateral inference activation, with RH cortical areas perhaps playing a specialized role in the inference process that does not strongly affect the amount of inference-related information kept active in each hemisphere (Kuperberg et al., 2006; Mason & Just, 2004; Prat et al., 2011) but nevertheless leads to an impairment in causal inference when that function is knocked out by RHD (Beeman, 1993; Tompkins et al., 2004; Lehman-Blake & Tompkins, 2001).

References

Baggio, G., Van Lambalgen, M. & Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. J. Mem. Lang. 59, 36-53.

Banich, M. (2002). The divided visual field technique in laterality and interhemispheric integration. Experimental Methods in Neuropsychology 21, 47–63.

Beeman, M. (1993). Semantic processing in the right hemisphere may contribute to drawing inferences from discourse. Brain and language, 44(1), 80-120.

Beeman, M. J., & Bowden, E. M. (2000). The right hemisphere maintains solution-related activation for yet-to-be-solved problems. Memory & cognition, 28(7), 1231-1241.

Beeman, M., Bowden, E. & Gernsbacher, M. (2000). Right and Left Hemisphere Cooperation for Drawing Predictive and Coherence Inferences during Normal Story Comprehension. Brain Lang. 71, 310-336.

Beeman, M., Friedman, R., Grafman, J., Perez, E., Diamond, S. & Lindsay, M. (1994). Summation priming and coarse semantic coding in the right hemisphere. J. Cogn. Neurosci. 6 (1), 26-45.

Berker, E., Berker, A. & Smith, A. (1986). Translation of Broca's 1865 report: Localization of speech in the third left frontal convolution. Arch. Neurol., 43 (10), 1065-1072.

Blake, M. L., & Lesniewicz, K. (2005). Contextual bias and predictive inferencing in adults with and without right hemisphere brain damage. Aphasiology, 19(3-5), 423-434.

Borovsky, A., Kutas, M., & Elman, J. (2013). Getting it right: Word learning across the hemispheres. Neuropsychologia, 51 (5), 825-837.

Bowers, D., Coslett, H., Bauer, R. M., Speedie, L. J., & Heilman, K. M. (1987). Comprehension of emotional prosody following unilateral hemispheric lesions: processing defect versus distraction defect. Neuropsychologia, 25(2), 317-328.

Breasted, J. H. (Ed.). (1930). The Edwin Smith Surgical Papyrus: Published in Facsimile and Hieroglyphic Transliteration with Translation and Commentary in 2 Vol. University of Chicago Press.

Brookshire, R. E., & Nicholas, L. H. (1993). Discourse Comprehension Test. Tucson, AZ: Communication Skill Builders.

Brownell, H., Michel, D., Powelson, J. & Gardner, H. (1983). Surprise But Not Coherence: Sensitivity to Verbal Humor in Right-Hemisphere Patients. Brain Lang 18, 20-27.

Brownell, H., Potter, H., Bihrle, A. & Gardner, H. (1986). Inference deficits in right brain-damaged patients. Brian Lang. 27, 310-321.

Coltheart, M. (1981). The MRC psycholinguistic database. The Quarterly Journal of Experimental Psychology, 33(4), 497-505.

Coney, J. (2002). The effect of associative strength on priming in the cerebral hemispheres. Brain Cogn., 234-241.

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

Coulson, S. & Williams, R. (2005). Hemispheric asymmetries and joke comprehension. Neuropsychologia 43, 128-141.

Coulson, S. & Wu, Y. (2005). Right hemisphere activation of joke-related information: An event-related potential study. J. Cogn. Neurosci. 17(3), 494–506.

Coulson, S., Federmeier, K., Van Petten, C. & Kutas, M. (2005). Right Hemisphere Sensitivity to Word- and Sentence-Level Context: Evidence From Event-Related Brain Potentials. J. Exp. Psych.: Learn. Mem. Cogn. 31 (1), 129-147.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of verbal learning and verbal behavior, 19(4), 450-466.

Davenport, T. & Coulson, S. (2013, in press). Hemispheric asymmetry in interpreting novel literal language an event-related potential study. Neuropsychologia.

Faust, M., & Chiarello, C. (1998). Sentence context and lexical ambiguity resolution by the two hemispheres. Neuropsychologia, 36(9), 827-835.

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

Federmeier, K., Wlotko, E., De Ochoa-Dewald, E. & Kutas M. (2007). Multiple effects of sentential constraint on word processing. Brain Research 1146, 75–84

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. Cogn. Ling. 7, 183–206.

Giora, R. (2003). On Our Mind: Salience, Context, and Figurative Language. Oxford Univ. Press, New York.

Giora, R., Zaidel, E., Soroker, N., Batori, G., Kasher, A., (2000). Differential effects of right- and left hemisphere damage on understanding sarcasm and metaphor. Metaphor Symb. 15, 63–83.

Goel, V. & Dolan, R. (2001). The functional anatomy of humor: segregating cognitive and affective components. Nature Neurosci. 4 (3), 237-238.

Greenhouse, S. & Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, 2. 95-112.

Happé, F., Brownell, H., & Winner, E. (1999). Acquired 'theory of mind' impairments following stroke. Cognition, 70(3), 211-240.

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

Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. Language and Cognitive Processes, 8(4), 379-411.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. Trends Cogn. Sci. 9 (11), 512-518.

Kaplan, J., Brownell, H., Jacobs, J. & Gardner, H. (1990). The effects of right hemisphere damage on the pragmatic interpretation of conversational remarks. Brian Lang. 315-333.

Koivisto, M. (1997). Time-course of semantic activation in the cerebral hemispheres. Neuropsychologia, 35(4), 497-504.

Kučera, H., & Francis, W. N. (1967).Computational analysis of present day American English. Providence, RI: Brown University Press.

Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of discourse: An fMRI study of causal inferencing across sentences. Neuroimage, 33(1), 343-361.

Kutas, M. & Federmeier, K. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). Annual Review of Psychology 62, 621–47

Landauer, T., Foltz, P. & Laham, D. (1998). Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.

Lehman-Blake, M. T., & Tompkins, C. A. (2001). Predictive inferencing in adults with right hemisphere brain damage. Journal of Speech, Language and Hearing Research, 44(3), 639.

Mashal, N. & Faust M. (2008). Right hemisphere sensitivity to novel metaphoric relations: Application of the signal detection theory. Brain Lang. 104, 103-112.

Mason, R. A., & Just, M. A. (2004). How the Brain Processes Causal Inferences in Text A Theoretical Account of Generation and Integration Component Processes Utilizing Both Cerebral Hemispheres. Psychological Science, 15(1), 1-7.

McKoon, G. & Ratcliff, R. (1992). Inference during reading. Psychological Review 99 (3), 440-466.

Myers, J., Shinjo, M. & Duffy, S. (1987). Degree of causal relatedness and memory. Journal of Memory and Language 26, 453-465.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). The University of South Florida word association norms. Tampa, FL, University of South Florida.

Nieuwland, M. S., & Van Berkum, J. J. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. Brain Research, 1118(1), 155-167.

Oldfield, R. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Pijnacker, J., Guerts, B., Van Lambalgen, M., Buitelaar, J. & Hagoort, P. (2010). Reasoning with Exceptions: An Event-related Brain Potentials Study. J. Cogn. Neuro. 23 (2), 471-480.

Prat, C. S., Mason, R. A., & Just, M. A. (2011). Individual differences in the neural basis of causal inferencing. Brain and language, 116(1), 1-13.

Ross, E. D. (2010). Cerebral localization of functions and the neurology of language: Fact versus fiction or is it something else?. The Neuroscientist, 16(3), 222-243.

Saron, C. D., & Davidson, R. J. (1989). Visual evoked potential measures of interhemispheric transfer time in humans. Behavioral Neuroscience, 103(5), 1115.

Siegal, M., Carrington, J., & Radel, M. (1996). Theory of mind and pragmatic understanding following right hemisphere damage. Brain and Language, 53(1), 40-50.

Tompkins, C. A., Fassbinder, W., Blake, M. L., Baumgaertner, A., & Jayaram, N. (2004). Inference generation during text comprehension by adults with right hemisphere brain damage: Activation failure versus multiple activation. Journal of Speech, Language and Hearing Research, 47(6), 1380.

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.

Tompkins, C., Scharp, V., Fassbinder, W., Meigh, K. & Armstrong, E. (2008). A different story on "theory of mind" deficit in adults with right hemisphere brain damage. Aphasiology 22 (1), 42-61.

Van Berkum, J. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In U. Sauerland, & K. Yatsushiro (Eds.), Semantics and pragmatics: From experiment to theory (pp. 276-316). Basingstoke: Palgrave Macmillan.

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

Virtue, S., Haberman, J., Clancy, Z., Parrish, T. & Jung-Beeman, M. (2006). Neural activity of inferences during story comprehension. Brain Res. 1084, 104-114.

Virtue, S., Parrish, T. & Jung-Beeman, M. (2008). Inferences during Story Comprehension: Cortical Recruitment Affected by Predictability of Events and Working Memory Capacity. J. Cogn. Neuro. 20 (12) 2274-2284.

Wlotko, E. & Federmeier, K. (2007). Finding the right word: Hemispheric asymmetries in the use of sentence context information. Neuropsychologia 45, 3001-3014.

Acknowledgements

Chapter 4, in part, is currently being prepared for submission for publication of the material. Davenport, Tristan S.; Coulson, Seana. The dissertation author was the primary investigator and author of this material.

Chapter 5

Conclusion

This dissertation has explored three theoretical issues surrounding the role of causal inference in language comprehension. In Chapter 2, I investigated the ordering of context effects on individual words that varied in the strength and nature of their relatedness to their discourse context. The latency at which different forms of contextual information begin to affect word processing is a key source of evidence for distinguishing between different theories of language comprehension in general. The two experiments described in Chapter 3 tested hypotheses about the time-course of causal inference activation across several partially redundant semantic cues. These experiments shed light not only on how long it takes for an inference to become available to aid in language comprehension, but also on whether comprehenders are capable of making predictive inferences in a weakly constraining discourse context. Finally, the three ERP experiments described in chapter 4 investigated the neural underpinnings of causal inference more directly, by testing the long-standing hypothesis of a unique RH advantage for constructing causal inferences and applying them to language comprehension. These three theoretical areas, and what I have shown about each of them in this dissertation, are discussed in order below.

Ordering of Context Effects on Word Processing

The first of these issues concerns the difference between “high” and “low” levels of information available in language processing, and what can be learned about them from the order in which “high-level” and “low-level” relationships between a word and its context begin to affect word processing. This is an important issue because theories of language comprehension make predictions about the possible orderings of different types of context effects on word processing, as part and parcel of their explanations of how different levels of information interact during incremental language comprehension. Therefore, comparing the latencies at which a “high-level” contextual manipulation (viz., relatedness to an available causal inference) begins to affect processing, compared to a “low-level” contextual manipulation (lexical association to the most recent word), allowed me to compare the predictions of three prominent theories of language comprehension.

Theories of hierarchical or serial processing (e.g., Kintsch, 1988) predict that the manipulation of lexical association, being a “lower”-level aspect of word processing, should begin to affect the ERP response to a word earlier than a “higher”-level variable such as the manipulation of causal relatedness. Memory-based processing theories, on the other hand, explain semantic processing in terms of overlap between the input and items in memory that are active to different degrees (McKoon & Ratcliff, 1992; Myers & O’Brien, 1998). In this view, manipulations of causal relatedness and lexical association both affect the semantic activation processes that are indexed by the N400 (*cf.* Kuperberg et al., 2011; Van Berkum, 2009). Finally, dynamic processing accounts based on the connectionist framework (e.g., Elman, 1990; Altmann & Mirkovic, 2009; McRae, Spivey-Knowlton & Tanenhaus, 1998) are more sensitive to the relative strengths of different types of information, and therefore it is possible that if the causal bridging inference made during the story is a stronger cue to the meaning of the incoming word than lexical association is, effects of causal relatedness will actually begin earlier than those of lexical relatedness.

To distinguish among these theories, in Chapter 2 I developed and normed a set of stimuli intended to reliably evoke causal antecedent inferences. These materials consisted of two-sentence stories containing a causal coherence gap, such as, "The surfer took his board out into the waves. He emerged screaming and missing an arm." Paired with each story were four probe words, of which the *causal related* condition (SHARK) was related to a likely causal inference and the *lexical related* condition (LEG) was unrelated to that inference but strongly associated with the story's final word. The *causal unrelated* (PUNCH) and *lexical unrelated* (SLEEP) probe words were control items that were drawn from the *causal related* and *lexical related* probe words associated with other stories in the stimulus set. In Experiment 1 of Chapter 2, EEG was recorded as participants listened to 160 such stories and then read one of the 4 possible probe words associated with each.

In Experiment 1, lexical relatedness affected only the N400. Causal relatedness effects established a new lower bound in the literature for how early causal inference can begin to impact word processing by affecting the P2, a perceptual processing component associated with stimulus evaluation and linked in the language literature to successful word form prediction (Wlotko & Federmeier, 2007). The earlier onset of causal relatedness effects supported the dynamic theory. Additionally, a post-N400 sustained negativity that was also affected by the causal relatedness manipulation was observed. This late component was interpreted in light of Van Berkum's (2009) hypothesis that this component reflects inferential operations on the contents of working memory. Noting that sustained negativity effects are observed in contexts where an inference must be revised or turns out to be wrong (e.g., Pijnacker et al., 2011; Baggio et al., 2008), we suggested that the sustained negativity enhanced by *causal related* probes reflected the reinterpretation of context in an attempt to reconcile the unexpected probe word with the discourse meaning established so far.

Experiment 2 of Chapter 2 was conducted to answer the possible objection that despite our lexical association norming with latent semantic analysis (Landauer et al., 1998), the *causal related* probes were in fact more strongly primed at a purely lexical level than the *lexical related* probes, accounting for the much greater N400 reduction elicited by the *causal related* probes. We tested this possibility by presenting *causal related* and *lexical related* probe words in the context of only the final word of their associated story. Not only did the *lexical related* probes show much larger N400 reduction than the *causal related* probe words (the reverse of what was observed in Experiment 1), this difference was actually larger than the lexical relatedness effect observed in Experiment 1. The latter observation vindicated another prediction of the dynamic framework, namely that the prioritization of one aspect of the context (the causal inference) led to suppressed processing of a weaker aspect of the context. Therefore, lexical relatedness N400 effects in Experiment 1 were smaller than would otherwise be the case.

Experiment 3 of Chapter 2 presented evidence that lexical prediction based on causal inference works somewhat differently than lexical prediction that occurs within a sentence. Specifically, ERP effects that could be associated with lexical prediction, the P2 (Wlotko & Federmeier, 2007) and the P600 (Van Petten & Luka, 2012), were attenuated when the presentation latency of the inference-related probe word was reduced from 400ms to 0ms, in order to interrupt whatever processes occurred at the end of the sentence. Our original hypothesis was that interrupting sentence wrap up with probe word presentation would affect processes of updating the situation model with inferred semantic information, as suggested by Just and Carpenter (1980) in their original discussion of sentence wrap-up effects. Therefore, we expected a smaller N400 effect of causal relatedness in the immediate presentation condition than in the delayed presentation condition. However, the N400 effect of causal relatedness was completely unaffected by this latency manipulation: only the P2 effect and a P600 effect were knocked out in the immediate presentation condition. I took this to indicate that lexical prediction had not occurred in the immediate condition, although the strength of semantic information available to aid in processing the probe word was the same at both latencies. Since the only difference between the two conditions was how early the probe word was presented, I suggest that the period of time just after the sentence boundary was used to convert inferred semantic information into a word-form prediction. This hypothesis would explain the dissociation observed between the N400 effects (which reflect semantic preactivation and were unaffected by Experiment 3’s latency manipulation) on one hand, and the P2 effects (which reflect word-form preactivation and *were* affected by the latency manipulation. This dissociation indicates that different neural systems and/or processes are responsible for word-form prediction and semantic preactivation.

Together, these results strongly refuted the serial processing theory: not only did effects of the purportedly “higher-level” variable occur earlier in word process than those of the “lower-level” variable, but comprehenders displayed a sensitivity to the overall composition of the stimulus set, which runs directly counter to the primary assumptions of serial processing theories. In fact, these results highlighted the shortcomings of the serial processing framework in the general question of high-level meaning construction: comprehenders engage in preactivation of linguistic representations at many different levels (DeLong et al., 2005; Levy, 2008; Van Berkum et al., 2005), and the time course of processing is not divided into stages conforming to theoretical constructs such as syntax and the lexicon (see e.g., Coulson, 2006).

Instead, these data supported the dynamic processing theory, because they showed in Experiment 1 that the causal relatedness manipulation affected word processing earlier than the lexical relatedness manipulation, a pattern predicted only by the dynamic theory. The P2 effect of causal inference, furthermore, supports the dynamic picture of a comprehender who is constantly attempting to predict upcoming input. Experiment 2 showed further that the lexical relatedness effect was larger and began earlier in the absence of a rich discourse context. This suggests that the processing resources devoted to different types of contextual information is determined dynamically according to what other forms of information are available, and that a form of context that receives low priority when competing with a more informative signal can receive higher priority when it occurs in isolation, as lexical relatedness did in the single-word contexts of Experiment

2. Experiment 3 shed more light on word-form prediction, indicating a dissociation between semantic pre-activation (which reduces N400 amplitude) and word-form pre-activation (which enhances P2 amplitude), with the latter being activated during sentence wrap-up processing with these materials.

The evidence that I use to distinguish the memory-based theory from the dynamic theory hinges on word-form prediction. Word-form prediction is fundamental to the dynamic framework, which is essentially a connectionist network that is constantly attempting to predict upcoming input, adjusting its connection weights according to the disparity between the predicted and actual inputs. The dynamic framework is therefore not a parsing model so much as a connectionist model of language learning that exhibits some comprehension-like behavior after extensive training on corpora (see e.g., Elman, 2009; Altmann & Mirkovic, 2009). The memory-based theory, on the other hand, models the dynamics of working memory during comprehension. Crucially, the network of words and propositions comprising the model is a static construction: activation levels of particular words and phrases change over time, but the connection weights between those elements are not trained with prediction error and do not change with experience. Thus, the two models suffer from complementary flaws: the dynamic model does not represent linguistic structures in the way that actual comprehenders appear to do (e.g., in terms of hierarchical structures of words and phrases), nor in any way that is obvious at all from examining the network state. On the other hand, the memory-based theory does not include an explanation of how the network attained its “mature” state evidenced by the processing behavior of undergraduate research subjects. Perhaps as a result, the memory-based theory is unable to account for word-prediction data such as those presented in Chapter 2.

Indeed, word form prediction so far has not figured prominently in any version of memory-based processing. Van Berkum's (2009) Memory in Readiness (MIR) theory is a version of the memory-based processing framework that is applied specifically to ERP data. Although the N400 plays a prominent role in the MIR hypothesis as a marker of semantic pre-activation, prediction of specific words is not discussed in this theory, nor in other theoretical descriptions of the memory-based processing framework. I therefore interpreted the P2 effects observed in Experiments 1 and 3 as evidence in favor of the dynamic processing model over the memory-based theory, which in its silence on the matter of word-form prediction is either wrong or incomplete. However, I also noted that if the memory-based prediction model were extended to include word forms among the memory representations that could resonate with incoming words, it would account for the P2 effects observed in Experiments 1 and 3.

Another possibility for reconciling the two theories is that they describe different aspects of the language comprehension system. The dynamic model is primarily a theory of learning, whereas the memory-based model is entirely a theory of real-time processing. Both models are network-based – a connectionist network of distributed representations in the case of the dynamic model and a network of discrete symbol nodes in the case of the memory-based model. I suggest that they could each model different aspects of the language faculty: the dynamic model describes how brain networks come to embody the structure of linguistic input, including the development of the situation models that lead inferences to arise automatically out of coherent discourses; the memory-based model, on the other hand, describes how those networks are used in language comprehension. Of course, learning continues throughout the lifespan, and as prediction is the primary engine of learning in the dynamic view, comprehenders continue to attempt to predict upcoming input.

Time-course of Causal Inference Activation

In Chapter 3, I turned to the question of the time-course of inference activation. I conceptualized this question in terms of how much relevant information needs to be available from context in order for an inference to affect processing. That is, if a reader of a story encountered a word that was weakly related to a possible inference, followed shortly thereafter by a word that was strongly related enough to confirm the inference, would the inference be available at the first, weaker cue, only at the later cue, or at some point in between?

Hypotheses related to the time-course of inference activation have been tested in the past using the narrative prime/probe word paradigm, like that used in Chapters 2 and 4 of this dissertation. However, these studies have typically been designed and analyzed with an eye towards finding an effect on a single critical word. Therefore, while a number of studies have explored the timing at which an inference becomes available to affect the processing of a single, strongly related probe word (e.g., Till et al., 1988; Millis & Graesser, 1994), little is known about the time course of an inference's development across multiple words in a sentence.

The experiments in Chapter 3, then, tested the effect of a causal relatedness manipulation on multiple words embedded in a discourse context instead of on isolated probe words. In Chapter 3, I conducted a self-paced reading experiment (Experiment 4) and an ERP experiment (Experiment 5) on the use of multiple, semi-redundant discourse cues in the context of a causal relatedness manipulation. In both experiments, participants read short stories that varied in whether the initial sentence created a plausible or implausible setting for the event described in the second sentence:

1a. Plausible: The cowboy was walking through tall grass. Without warning, the cowboy was *bitten* by a *snake*, and he had to call for help.

1b. Implausible: The cowboy was driving through tall grass. Without warning, the cowboy was *bitten* by a *snake*, and he had to call for help

In each of these stimuli, the main verb ("bitten") in the second sentence defined the nature of the event but was only a weak cue as to whether it was plausible in light of the first sentence or not. The agent noun ("snake") provided a stronger cue to implausibility by naming an entity that was likely or unlikely to be present in the scene described earlier. A cloze norming task determined that there was no difference in predictability between the plausible and implausible verbs (both conditions < 0.1 mean cloze probability), but the plausible noun was significantly more predictable (mean cloze = 0.5) than the implausible noun (0.3).

I reasoned that if sequential, partially redundant discourse cues have independent effects on comprehension, then relative processing difficulty in the implausible condition would manifest on both the critical verb ("bitten") and on the critical noun ("snake"). However, if participants fill in a more detailed discourse representation on the basis of incomplete information, perhaps by activating frames or schemas about biting events (Coulson, 2006; Coulson, 2001) or event knowledge that includes typical participants in biting events (e.g., Metusalem et al., 2012; McRae, Hare, Elman & Feretti, 2005; Bicknell et al., 2010), then this activation might facilitate later processing of a word describing the agent of a biting event (e.g., "snake") even in the implausible condition. In this case, plausibility would affect processing on the verb but not on the noun. Finally, it was possible that the verb, being a weak cue to discourse coherence, would not affect processing differently in the two conditions, and effects of plausibility would only appear on the noun.

Overall, both experiments supported the first hypothesis. The self-paced reading data in Experiment 4 supported the first hypothesis, with implausibility causing slowed reading times (RTs) on the spillover regions of both the verb and the noun. The grand average ERP results in Experiment 5 initially appeared to support the second hypothesis: plausible verbs elicited smaller N400 and smaller sustained negativity than implausible verbs, but ERP responses to the nouns were identical in both plausibility conditions.

The apparent contradiction between the two experiments was resolved by an investigation of individual differences. The avenue I chose was to examine within-subject differences between the plausibility effects on the verbs and those on the nouns. These analyses showed a second striking difference between the two experiments. Self-paced reading effects in the verb region ("bitten by a") correlated positively with those in the noun region ("snake, and he"). However, a negative correlation was observed in the ERP experiment: participants showed a late negativity on the verb ("bitten") also tended to show a late positivity on the noun ("snake") and vice-versa. A median split on the size of the ERP verb effect (500-1000ms) showed that two qualitatively different processing styles were represented in different groups of participants. In some participants (dubbed the Predictive group), a large N400 effect at the verb was followed by a sustained negativity effect; a small late positivity on the noun did not approach significance. In the other participants (the Bridging group), there was no N400 effect at all on either word, and the pattern of late effects was precisely the opposite of that seen in the Predictive group.

This pattern of results suggested two distinct processing strategies at work in my participant set. First, some participants engaged in predictive semantic activation or discourse model updating at the implausible verb, leading to similar responses to the noun in both conditions. Other participants adopted a wait-and-see strategy, adjusting their situation model (reflected by post-N400 ERP effects) on each word that was relevant to the plausibility manipulation. Second, we assume that the amplitude of the N400 response indexes the degree to which a word triggers the activation of semantic material that was not already active (Van Berkum, 2009; Kutas & Federmeier, 2011). In light of that assumption, the N400 effect on the verb (among those participants who had an N400 effect) indicates that in the plausible condition, more information related to the verb was active than in the implausible condition, despite no difference in the verb's cloze probability or lexical association to prior context. This suggests that those participants engaged in some degree of causal consequence inference, despite the weakly constrained nature of the contexts. This conclusion, which would not necessarily be licensed by spillover effects in a self-paced reading study, highlights the future usefulness of the ERP technique for studying predictive inference activation.

Finally, it is worth considering these two processing styles in light of the theoretical models that have been discussed so far in this dissertation. I've previously argued that the N400 effect observed on the critical verb indicated that semantic information about upcoming discourse elements was already active at the time the verb was encountered (e.g., information about possible biting agents like snakes in advance of encountering "bitten"). This possibility is precluded by serial processors, which do not make predictive inferences at all, and by memory-based models, in which the resonance processes responsible for semantic preactivation do not lead to predictive inference except under the most constraining contexts (McKoon & Ratcliff, 1986; 1992). However, Altmann & Mirkovic's (2009) dynamic model demonstrates the possibility of making predictive inferences with a prediction-driven network that is trained on both linguistic input and on descriptions of the events described by those sentences. Initially, the results of Experiments 4 and 5 support the dynamic model overall. However, the subgroup data in Experiment 5 tell a more nuanced story. Recall that the Predictive subgroup showed a pattern of effects similar to that found in the grand average, albeit with larger effect sizes at the verb. However, the Bridging subgroup showed only small, late effects at the verb, and larger effects at the later noun, where the nature of the situation was clarified. This pattern of effects, which indicates a post-hoc bridging inference made at the noun (possibly beginning at the semantically unexpected verb), is more compatible with a memory-based or even serial view of discourse processing. Therefore, the problem of how best to model a comprehender's performance may not depend just on the nature of the linguistic input, but also on the comprehender's goals and on individual traits.

Still left unresolved, however, is *which* individual traits determine these processing differences. The pattern of ERP effects observed in a given participant was not explained by either of the individual difference measures I collected: verbal working memory (Daneman & Carpenter, 1980) or the autism quotient (Baron-Cohen et al., 2001). Further research into the role of individual differences in real-time discourse processing is strongly warranted.

Hemispheric Basis of Causal Inference

The third major theoretical question considered by this dissertation was an aspect of the neural underpinnings of causal inference – specifically the hypothesis that in right-handed adults, the right hemisphere (RH) has an advantage for activating causal inferences, in addition to other pragmatic functions. This hypothesis emerged from neuropsychological tests of lesion patients (see e.g., Beeman, 1993; Tompkins et al, 2008; Lehman-Blake & Tompkins, 2001; Tompkins, Fassbinder, Lehman-Blake, Baumgaertner & Jayaram, 2004). Tests on neurologically normal adults using brain imaging have returned conflicting results, with some experiments indicating a RH specialization (Mason & Just, 2004), some indicating basically balanced bilateral processing of causal inferences (Kuperberg et al., 2006), and some indicating that the two hemispheres specialize in different aspects of causal inference processing, with RH specializing in the earliest stages of inference activation (Virtue Parrish & Jung-Beeman, 2008; Virtue et al., 2006).

The experiments described in Chapter 4 of this dissertation (Experiments 6, 7 and 8) used the stimuli introduced in Chapter 2 combined with hemifield presentation of the probe words in order to investigate the long-standing hypothesis of a right hemisphere (RH) advantage for activating causal inferential information. In these three experiments, participants listened to stories containing causal coherence gaps and then read a probe word that was presented either in right visual field targeting left hemisphere (rvf/LH) or in left visual field targeting right hemisphere (lvf/RH). The experiments differed in which probe words were used: in Experiment 6, the *causal related* and *lexical related* probe words were compared; in Experiment 7, the *causal related* and *causal unrelated*; and in Experiment 8 all four conditions were used, at the cost of fewer trials per condition. The results of Experiment 6 supported the RH activation theory: N400 to the *causal related* condition was reduced in both hemifields, but to a greater degree in lvf/RH. The interpretation of this effect was clouded, however, by a late positivity elicited by the *lexical related* items in rvf/LH.

The manipulation of probe type made in Experiment 6 was licensed by the observation in Chapter 2 that participants appeared to treat the *lexical related* probes as unrelated in comparison to the *causal related* probes. The follow-up experiment 7 was conducted in order to test whether the stronger lexical association in the *lexical related* condition had triggered the rvf/LH frontal positivity, which may have artificially reduced the size of the rvf/LH N400 effect. In Experiment 7, the *lexical related* condition was replaced with the *causal unrelated* condition, with probe words rotated between conditions so that the *causal related* and *causal unrelated* conditions were matched with each other for strength of lexical association to its context. This experiment showed very similar effects of causal relatedness in both hemifields, yielding no evidence of a RH advantage for causal inference processing.

It was possible, however, that the rather obvious relatedness manipulation used in Experiment 7 (50% of probe words were causally related, and 50% were completely unrelated) encouraged strategic processing on the part of participants who had figured out the structure of the experiment, failing to reveal more naturalistic inference processes engaged when reading texts. In Experiment 8, therefore, all four conditions from Chapter 2, Experiment 1 were used, to attempt to replicate the results of Experiment 7 and to test whether a hemispheric asymmetry in lexical associative processing might have confounded the results of Experiment 6. Again, similar causal relatedness effects were observed in both presentation sides. Secondly, the attenuation of lexical relatedness effects in a discourse context (*cf.* Chapter 2) was even more extreme, with no significant N400 effects of lexical relatedness in either hemifield. It is worth noting, however, that the N400 difference between the *causal related* and *lexical related* conditions was significant only in lvf/RH, where the *causal related* condition elicited smaller N400, indicating facilitated semantic activation. The rvf/LH frontal positivity, enhanced in the *lexical related* condition in Experiment 6, was not replicated. Thus, Experiments 7 and 8 offered little support for any hypothesis of a causal advantage in RH, and suggested that the asymmetric result found in Experiment 6 was an artifact of strategic processes carried out in LH in response to strong lexical association in the inference-unrelated condition.

Comparing these three experiments, it is interesting to note that ERP patterns across experiments 6-8 were similar in lvf/RH, but differed in rvf/LH depending on which control conditions were included. Regardless of the proportion of stimuli belonging to each condition, or even the nature of the control condition, causal relatedness effects in lvf/RH were the same in each experiment: an N400 effect followed by a parietally focused sustained negativity. However, the brain responses in rvf/LH seemed to depend much more heavily on the composition of the stimulus set in each experiment. It appears that RH responded to the various conditions based simply on the strength of their relationship to the global linguistic context. LH, on the other hand seemed to be more sensitive to strategic processing factors influenced by the design of the experiment, as well as to local relationships such as lexical association.

This is a noteworthy finding, because one hallmark of RH lesion patients is that even when they are able to activate a causal inference, they are often distracted from it by locally coherent meanings (Tompkins et al., 2004). I offer the speculative possibility that the poor performance of RH lesion patients on causal inference tasks may not only be a reflection of a missing RH competence, but also of a now-unrestrained LH tendency to focus on statistics-driven, local semantic relationships. Further research, both with lesion patients and with healthy adults, is necessary to understand the contribution of LH to inference-making as well as to interference from strong local semantic relationships.

These results point to fundamentally different processing styles in the two hemispheres. As Federmeier (2007) characterizes the hemispheres in her PARLO model (Production Affects Reception in Left hemisphere Only), LH tends to be more sensitive to top-down influences on comprehension, and appears to be the only hemisphere that engages in word form prediction (Wlotko & Federmeier, 2007). RH, on the other hand, appears to prioritize more bottom-up processing, in the manner of a traditional serial parser (Federmeier, 2007; Federmeier & Kutas, 1999). Extending my investigation of causal inference processing into the realm of hemispheric asymmetry therefore complicates the question of which processing model is the best suited to explain the language processing architecture. LH's sensitivity to the makeup of the stimulus set in Experiments 6-8 recalls Experiments 1 and 2, in which it was found that lexical relatedness effects were attenuated when the stimulus set included a discourse-level manipulation. However, RH displayed no such sensitivity. Thus, LH appears to function more in line with the dynamic theory. RH, however, more closely matches the predictions of the memory-based theory: as Experiment 8 shows, causal and lexical relatedness effects occur at similar latencies, and RH is also not sensitive to the overall makeup of the stimulus set.

Conclusion

Taken together, the experiments described in this dissertation advance the study of causal inference in language comprehension in several ways. In Chapter 2, I offered evidence of lexical prediction driven by causal inference, which, unlike lexical prediction in other circumstances (see e.g., DeLong et al., 2005), appears to depend on sentence wrap-up in order to occur. This is, to my knowledge, a completely novel finding and therefore a ripe topic for future research. The results of Chapter 2 also supported dynamic theories of language comprehension, in which "top-down" information, such as a causal inference established in the situation model can begin to affect processing before local, "lower-level" semantic relationships such as lexical association. Indeed, the size and duration of lexical association effects is actually reduced in discourse contexts relative to single word contexts, further supporting the dynamic theory.

Two issues remain unresolved after Chapter 2. The first is what role, if any, sentence wrap-up processing plays in the dissociation between semantic activation and lexical prediction observed in Experiment 3. The finding needs to be replicated and, if successfully replicated explored further to understand how it differs from prediction based on cloze probability. The second issue is the way forward for the memory-based processing theory. In order to serve as a satisfactory account of language comprehension, it must incorporate and account for the mounting evidence of word-form prediction. In terms of Van Berkum's (2009) MIR theory, which applies memory-based processing principles to ERP data, such an extension would account for prediction-related ERP components such as the frontal positivity (Federmeier et al., 2007), and early components affected by word form predictability like the N1 and P2, and explain them in terms of the dynamics of memory activation and inhibition.

Chapter 3 also provides evidence for the use of causal inference to generate predictions, though in this case of upcoming events rather than specific words. In extremely low-constraint contexts (cloze < 0.1) some ERP participants preactivated more semantic information related to a plausible causal consequent than to an implausible causal consequent. Overall, Chapter 3's results support a dynamic processing view in which some information about an inference is available at the earliest related word. An important finding, however, is that only some of our participants appeared to engage in this preactivating strategy. This suggests that in the old debate about whether comprehenders engage in predictive inference-making or not (e.g., McKoon & Ratcliff, 1986; 1992), both sides are correct: some comprehenders make predictive inferences, and others do not, even in weakly constraining contexts that do not encourage prediction. What factor separates these two types of comprehenders remains a topic for future research, as they did not differ in this study in verbal working memory, the Communication or Imagination subscores of the Autism Quotient Test, nor on a post-hoc memory test.

Finally, my investigations of the hemispheric basis of causal inference processing in Chapter 4 found no support for the hypothesis of a RH basis for causal inference. A processing benefit for causal relatedness was observed in both hemispheres in all three experiments. However, that benefit was partially obscured in rvf/LH in Experiment 6 by an overlapping frontal positive response to the *lexical related* probes being used as a control condition, leading to an initial conclusion favoring a RH advantage for causal processing. Experiments 7 and 8, both of which included a *causal unrelated* control condition, falsified this conclusion. The collective results of these three experiments suggest instead that both hemispheres participate roughly equally in processing a causal relationship; in fact, where the hemispheres differ in these tasks is on the lexical relatedness manipulation.

In final summary, the findings described in this dissertation indicate that the effects of a causal inference on language processing can most successfully be modeled by a dynamic, prediction-based processor. Use of causal inferential information is fast and incremental, and affects comprehension at the earliest possible cue, although individual differences in processing have a considerable effect on how strongly related a word must be in order to be facilitated by an available inference. Furthermore, the impact of a causal inference on word processing is not a primarily right hemisphere function, *contra* previous research to that effect. Rather, both hemispheres participated to a roughly equal degree in the facilitation of a causally related word.

References

Altmann, G., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. Cognitive Science, 33(4), 583-609.

Baggio, G., Van Lambalgen, M. & Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. J. Mem. Lang. 59, 36-53.

Banich, M. (2002). The divided visual field technique in laterality and interhemispheric integration. Experimental Methods in Neuropsychology 21, 47–63.

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

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. Journal of autism and developmental disorders, 31(1), 5-17.

Beeman, M. (1993). Semantic processing in the right hemisphere may contribute to drawing inferences from discourse. Brain and language, 44(1), 80-120.

Beeman, M. Bowden, E. & Gernsbacher, M. (2000). Right and Left Hemisphere Cooperation for Drawing Predictive and Coherence Inferences during Normal Story Comprehension. Brain Lang. 71, 310-336.

Beeman, M. J., & Bowden, E. M. (2000). The right hemisphere maintains solution-related activation for yet-to-be-solved problems. Memory & cognition, 28(7), 1231-1241.

Beeman, M., Bowden, E. & Gernsbacher, M. (2000). Right and Left Hemisphere Cooperation for Drawing Predictive and Coherence Inferences during Normal Story Comprehension. Brain Lang. 71, 310-336.

Beeman, M., Friedman, R., Grafman, J., Perez, E., Diamond, S. & Lindsay, M. (1994). Summation priming and coarse semantic coding in the right hemisphere. J. Cogn. Neurosci. 6 (1), 26-45.

Berker, E., Berker, A. & Smith, A. (1986). Translation of Broca's 1865 report: Localization of speech in the third left frontal convolution. Arch. Neurol., 43 (10), 1065-

1072.

Bicknell, K., Elman, J., Hare, M., McRae, K. & Kutas M., (2010). Effects of event knowledge in processing verbal arguments. Journal of Memory and Language 63 (4), 489-505.

Blake, M. L., & Lesniewicz, K. (2005). Contextual bias and predictive inferencing in adults with and without right hemisphere brain damage. Aphasiology, 19(3-5), 423-434.

Borovsky, A., Kutas, M., & Elman, J. (2013). Getting it right: Word learning across the hemispheres. Neuropsychologia, 51 (5), 825-837.

Boudewyn, M. A., Gordon, P. C., Long, D., Polse, L., and Swaab,T.Y. (2012). Does discourse congruence influence spoken language comprehension before lexical association? Evidence from event-related potentials. Lang. Cogn. Process. 27, 698–733.

Boudewyn, M., Long, D. & Swaab, T. (2013). Effects of working memory span on processing of lexical associations and congruence in spoken discourse. Frontiers in Psychology 4 (60), 1-16.

Bowers, D., Coslett, H., Bauer, R. M., Speedie, L. J., & Heilman, K. M. (1987). Comprehension of emotional prosody following unilateral hemispheric lesions: processing defect versus distraction defect. Neuropsychologia, 25(2), 317-328.

Breasted, J. H. (Ed.). (1930). The Edwin Smith Surgical Papyrus: Published in Facsimile and Hieroglyphic Transliteration with Translation and Commentary in 2 Vol. University of Chicago Press.

Brookshire, R. E., & Nicholas, L. H. (1993). Discourse Comprehension Test. Tucson, AZ: Communication Skill Builders.

Brouwer, H.., Fitz, H. & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. Brain Research 1446, 127-143.

Brown, C. M., Hagoort, P., & Chwilla, D. J. (2000). An event-related brain potential analysis of visual word priming effects. Brain and Language, 72(2), 158-190.

Brownell, H., Michel, D., Powelson, J. & Gardner, H. (1983). Surprise But Not Coherence: Sensitivity to Verbal Humor in Right-Hemisphere Patients. Brain Lang 18, 20-27.

Brownell, H., Potter, H., Bihrle, A. & Gardner, H. (1986). Inference deficits in right brain-damaged patients. Brian Lang. 27, 310-321.

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

Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. Memory, 9(4-6), 365-381.

Calvo, M. G., Meseguer, E., & Carreiras, M. (2001). Inferences about predictable events: Eye movements during reading. Psychological research, 65(3), 158-169.

Camblin, C., Gordon, P. & Swaab, T. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. Journal of Memory and Language 56, 103–128.

Clark, A. (2012). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci, 1-86.

Coltheart, M. (1981). The MRC psycholinguistic database. The Quarterly Journal of Experimental Psychology, 33(4), 497-505.

Coney, J. (2002). The effect of associative strength on priming in the cerebral hemispheres. Brain Cogn., 234-241.

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

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

Coulson, S. & Williams, R. (2005). Hemispheric asymmetries and joke comprehension. Neuropsychologia 43, 128-141.

Coulson, S. & Wu, Y. (2005). Right hemisphere activation of joke-related information: An event-related potential study. J. Cogn. Neurosci. 17(3), 494–506.

Coulson, S. (2001). Semantic leaps: Frame-shifting and conceptual blending in meaning construction. Cambridge: Cambridge University Press.

Coulson, S. (2006). Constructing meaning. Metaphor Symb. 21 (4), 245-266.

Coulson, S., Federmeier, K., Van Petten, C. & Kutas, M. (2005). Right Hemisphere Sensitivity to Word- and Sentence-Level Context: Evidence From Event-Related Brain Potentials. J. Exp. Psych.: Learn. Mem. Cogn. 31 (1), 129-147.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of verbal learning and verbal behavior, 19(4), 450-466.

Davenport, T. & Coulson, S. (2013, in press). Hemispheric asymmetry in interpreting novel literal language an event-related potential study. Neuropsychologia.

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

Dikker, S., & Pylkkanen, L. (2011). Before the N400: Effects of lexical–semantic violations in visual cortex. Brain and language, 118(1), 23-28.

Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. Psychological Science, 21(5), 629-634.

Duffy, S. A. (1986). Role of expectations in sentence integration. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12(2), 208.

Elman, J. (2009). On the Meaning of Words and Dinosaur Bones: Lexical Knowledge Without a Lexicon. Cogn., Sci., 33, 547–582

Elman, J. L. (1990). Finding structure in time. Cognitive science, 14(2), 179-211.

Faust, M., & Chiarello, C. (1998). Sentence context and lexical ambiguity resolution by the two hemispheres. Neuropsychologia, 36(9), 827-835.

Federmeier, K. & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. Journal of Memory and Language 41, 469–495

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

Federmeier, K. D., & Kutas, M. (2002). Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres. Neuropsychologia, 40(7), 730-747.

Federmeier, K. D., Mai, H., & Kutas, M. (2005). Both sides get the point: Hemispheric sensitivities to sentential constraint. Memory & Cognition, 33(5), 871-886.

Federmeier, K., Wlotko, E., De Ochoa-Dewald, E. & Kutas M. (2007). Multiple effects of sentential constraint on word processing. Brain Research 1146, 75–84

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. The Quarterly Journal of Experimental Psychology, 33(4), 465-495.

Forster, K. I. Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), New Approaches to language mechanisms. Amsterdam: North-Holland, 1976.

Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. Journal of Memory and Language, 42(4), 526-544.

Gerrig, R.J., & O'Brien, E.J. (2005). The scope of memory-based processing. Discourse Processes, 39 (2 & 3), 225-242.

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. Cogn. Ling. 7, 183–206.

Giora, R. (2003). On Our Mind: Salience, Context, and Figurative Language. Oxford Univ. Press, New York.

Giora, R., Zaidel, E., Soroker, N., Batori, G., Kasher, A., (2000). Differential effects of right- and left hemisphere damage on understanding sarcasm and metaphor. Metaphor Symb. 15, 63–83.

Goel, V. & Dolan, R. (2001). The functional anatomy of humor: segregating cognitive and affective components. Nature Neurosci. 4 (3), 237-238.

Graesser, A. C., & Bertus, E. L. (1998). The construction of causal inferences while reading expository texts on science and technology. Scientific Studies of Reading, 2(3), 247-269.

Graesser, A., Singer, M. & Trabasso, T. (1994). Constructing Inferences During Narrative Text Comprehension. Psychological Review 101 (3), 371-395.

Greenhouse, S. & Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, 2. 95-112.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. Cognition, 116(1), 42-55.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. Psychophysiology, 48(12), 1711-1725.

Gygax, P., Tapiero, I., & Carruzzo, E. (2007). Emotion inferences during reading comprehension: What evidence can the self-pace reading paradigm provide?. Discourse processes, 44(1), 33-50.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. (2004). Integration of Word Meaning and World Knowledge in Language Comprehension. Science, 304, 438-441.

Happé, F., Brownell, H., & Winner, E. (1999). Acquired 'theory of mind' impairments following stroke. Cognition, 70(3), 211-240.

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

Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. Language and Cognitive Processes, 8(4), 379-411.

Huang, Y. T., & Gordon, P. C. (2011). Distinguishing the time-course of lexical and discourse processes through context, coreference, and quantified expressions. Journal of Experimental Psychology: Learning, Memory, and Cognition, 37(4), 966.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. Trends Cogn. Sci. 9 (11), 512-518.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. Psychological review, 87, 329-354.

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

Kaplan, J., Brownell, H., Jacobs, J. & Gardner, H. (1990). The effects of right hemisphere damage on the pragmatic interpretation of conversational remarks. Brian Lang. 315-333.

Keenan, J., Baillet, S. & Brown, P. (1984). The Effects of Causal Cohesion on Comprehension and Memory. Journal of Verbal Learning and Verbal Behavior 23, 115-126.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review, 95 (2), 163-182.

Kintsch, W., & Mross, E. F. (1985). Context effects in word identification. Journal of Memory and Language, 24(3), 336-349.

Klin, C. M. (1995). Causal inferences in reading: From immediate activation to long-term memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(6), 1483.

Klin, C. M., Guzmán, A. E., & Levine, W. H. (1999). Prevalence and persistence of predictive inferences. Journal of Memory and Language, 40(4), 593-604.

Koivisto, M. (1997). Time-course of semantic activation in the cerebral hemispheres. Neuropsychologia, 35(4), 497-504.

Koornneef, A. W., & Van Berkum, J. J. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. Journal of Memory and Language, 54(4), 445-465.

Kučera, H., & Francis, W. N. (1967).Computational analysis of present day American English. Providence, RI: Brown University Press.

Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of discourse: An fMRI study of causal inferencing across sentences. Neuroimage, 33(1), 343-361.

Kuperberg, G., Paczynski, M. & Ditman, T. (2011). Establishing Causal Coherence across Sentences: An ERP Study. J. Cogn. Neuro. 23 (5), 1230-1246.

Kutas, M. & Federmeier, K. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). Annual Review of Psychology 62, 621–47

Kutas, M. & Federmeier, K. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). Annual Review of Psychology 62, 621–47.

Kutas, M. & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205.

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

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

Landauer, T., Foltz, P. & Laham, D. (1998). Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.

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.

Lehman-Blake, M. T., & Tompkins, C. A. (2001). Predictive inferencing in adults with right hemisphere brain damage. Journal of Speech, Language and Hearing Research, 44(3), 639.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. Psychological review, 101(4), 676.

Mashal, N. & Faust M. (2008). Right hemisphere sensitivity to novel metaphoric relations: Application of the signal detection theory. Brain Lang. 104, 103-112.

Mason, R. A., & Just, M. A. (2004). How the Brain Processes Causal Inferences in Text A Theoretical Account of Generation and Integration Component Processes Utilizing Both Cerebral Hemispheres. Psychological Science, 15(1), 1-7.

McKoon, G. & Ratcliff, R. (1992). Inference during reading. Psychological Review 99 (3), 440-466.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. Journal of Experimental Psychology: Learning, memory, and cognition, 12(1), 82.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. Journal of Memory and Language, 38(3), 283-312.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. Journal of memory and language, 66(4), 545-567.

Murray, J. D., Klin, C. M., & Myers, J. L. (1993). Forward inferences in narrative text. Journal of Memory and Language, 32(4), 464-473.

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

Myers, J., Shinjo, M. & Duffy, S. (1987). Degree of causal relatedness and memory. Journal of Memory and Language 26, 453-465.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). The University of South Florida word association norms. Tampa, FL, University of South Florida.

Nieuwland, M. S., & Van Berkum, J. J. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. Brain Research, 1118(1), 155-167.

Nieuwland, M. S., & Van Berkum, J. J. (2006a). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. Brain Research, 1118(1), 155-167.

Nieuwland, M. S., & Van Berkum, J. J. (2006b). When peanuts fall in love: N400 evidence for the power of discourse. Journal of cognitive neuroscience, 18(7), 1098-1111.

Noordman, L. G., Vonk, W., & Kempff, H. J. (1992). Causal inferences during the reading of expository texts. Journal of Memory and Language, 31(5), 573-590.

Oldfield, R. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

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

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

Pickering, M. J., & Garrod, S. (2013). Forward models and their implications for production, comprehension, and dialogue. Behavioral and Brain Sciences, 49-64.

Pijnacker, J., Guerts, B., Van Lambalgen, M., Buitelaar, J. & Hagoort, P. (2010). Reasoning with Exceptions: An Event-related Brain Potentials Study. J. Cogn. Neuro. 23 (2), 471-480.

Pijnacker, J., Guerts, B., Van Lambalgen, M., Buitelaar, J. & Hagoort, P. (2010). Reasoning with Exceptions: An Event-related Brain Potentials Study. J. Cogn. Neuro. 23 (2), 471-480.

Potts, G. F. (2004). An ERP index of task relevance evaluation of visual stimuli.Brain and cognition, 56(1), 5-13.

Potts, G. F., Liotti, M., Tucker, D. M., & Posner, M. I. (1996). Frontal and inferior temporal cortical activity in visual target detection: Evidence from high spatially sampled event-related potentials. Brain Topography, 9(1), 3-14.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. Journal of Memory and Language, 27(4), 399-415.

Prat, C. S., Mason, R. A., & Just, M. A. (2011). Individual differences in the neural basis of causal inferencing. Brain and language, 116(1), 1-13.

Regel, S., Coulson, S. & Gunter, T. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. Brain Research, 1311, 121-135.

Rohde, D. (2005). Linger experiment presentation software.

Rohde, H., Levy, R., & Kehler, A. (2011). Anticipating explanations in relative clause processing. Cognition, 118(3), 339-358.

Ross, E. D. (2010). Cerebral localization of functions and the neurology of language: Fact versus fiction or is it something else?. The Neuroscientist, 16(3), 222-243.

Saron, C. D., & Davidson, R. J. (1989). Visual evoked potential measures of interhemispheric transfer time in humans. Behavioral Neuroscience, 103(5), 1115.

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

Siegal, M., Carrington, J., & Radel, M. (1996). Theory of mind and pragmatic understanding following right hemisphere damage. Brain and Language, 53(1), 40-50.

St. George, M., Mannes, S. & Hoffman, J. (1997). Individual Differences in Inference Generation: An ERP Analysis. J. Cogn. Neuro. 9 (6), 776-787.

Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time-course of the influence of implicit causality information: Focusing versus integration accounts. Journal of Memory and Language, 42(3), 423-443.

Swaab, T. Y., Baynes, K., & Knight, R. T. (2002). Separable effects of priming and imageability on word processing: an ERP study. Cognitive Brain Research, 15(1), 99-103.

Tabor, W., & Tanenhaus, M. K. (1999). Dynamical models of sentence processing. Cognitive Science, 23(4), 491-515.

Till, R., Mross, E. & Kintsch, W. (1988). Time-course of priming for associate and inference words in a discourse context. Memory & Cognition 16 (4), 283-298.

Tompkins, C. A., Fassbinder, W., Blake, M. L., Baumgaertner, A., & Jayaram, N. (2004). Inference generation during text comprehension by adults with right hemisphere brain damage: Activation failure versus multiple activation. Journal of Speech, Language and Hearing Research, 47(6), 1380.

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.

Tompkins, C., Scharp, V., Fassbinder, W., Meigh, K. & Armstrong, E. (2008). A different story on "theory of mind" deficit in adults with right hemisphere brain damage. Aphasiology 22 (1), 42-61.

Van Berkum, J. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In U. Sauerland, & K. Yatsushiro (Eds.), Semantics and pragmatics: From experiment to theory (pp. 276-316). Basingstoke: Palgrave Macmillan.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(3), 443.

Van Berkum, J. J., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. Journal of cognitive neuroscience, 11(6), 657-671.

Van Berkum, J. J., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. Cognitive Brain Research, 17(3), 701-718.

Van den Broek, P. (1990). The causal inference maker: Towards a process model of inference generation in text comprehension. Comprehension processes in reading, 423-445.

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. Language & Cognitive Processes 8, 485-531.

Van Petten, C. (1995). Words and sentences: Event-related brain potential measures. Psychophysiology, 32, 511-532.

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

Van den Broek, P., & Lorch Jr, R. F. (1993). Network representations of causal relations in memory for narrative texts: Evidence from primed recognition. Discourse Processes, 16(1-2), 75-98.

Virtue, S., Haberman, J., Clancy, Z., Parrish, T. & Jung-Beeman, M. (2006). Neural activity of inferences during story comprehension. Brain Res. 1084, 104-114.

Virtue, S., Parrish, T. & Jung-Beeman, M. (2008). Inferences during Story Comprehension: Cortical Recruitment Affected by Predictability of Events and Working Memory Capacity. J. Cogn. Neuro. 20 (12) 2274-2284.

Wlotko, E. & Federmeier, K. (2007). Finding the right word: Hemispheric asymmetries in the use of sentence context information. Neuropsychologia 45, 3001-3014.

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. Neuroimage, 62(1), 356-366.

Yang, C., Perfetti, C. & Schmalhofer, F. (2007). Event-Related Potential Indicators of Text Integration Across Sentence Boundaries. Journal of Experimental Psychology: Learning, Memory, and Cognition 33 (1), 55–89.