Title
Muggle or magical?: Electrophysiological investigations of variation in the language-knowledge interface during reading

Author
Troyer, Melissa

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Muggle or magical?: Electrophysiological investigations of variation in the language-knowledge interface during reading

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

in

Cognitive Science

by

Melissa Troyer

Committee in charge:

Professor Marta Kutas, Chair
Professor Seana Coulson
Professor Sarah Creel
Professor Victor Ferreira
Professor Zhuowen Tu

2019

The Dissertation is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2019

DEDICATION

To my parents and mentors for unfailing support, and to paying it forward.

EPIGRAPH

If you think about it, reading is a necessarily individual act, far more than writing. If we assume that writing manages to go beyond the limitations of the author, it will continue to have a meaning only when it is read by a single person and passes through [her] mental circuits.

Italo Calvino, *if on a winter's night a traveler…*

TABLE OF CONTENTS

LIST OF FIGURES

Chapter Five

LIST OF TABLES

Chapter Two

Chapter Three

Chapter Four

ACKNOWLEDGMENTS

I am deeply grateful to the mentors who showed me many paths to learning and gave me the space, time, and support to find my own. My advisor Marta Kutas has been a constant guide, advocate, and voice of reason. Marta invested so much in my training and always encouraged me to hone my own instincts—scientific and otherwise. Marta, thank you for allowing me to be a part of your lab! Thank you also for your guidance when it came to choosing a research topic that could form the basis for a productive career in science. I could not have asked for a more generous, thoughtful, and discerning mentor. Thanks also for leading the cognitive science department: it has been a privilege to train in the cognitive science community at UCSD.

I would also like to thank my mentor Jeff Elman, whose excitement about scientific ideas always made meetings and discussions a delight. I am so lucky to have benefited from his encouragement to tackle challenges and his glee for talking about the scientific process. Thanks also to my committee members Seana Coulson, Sarah Creel, Vic Ferreira, and Zhuowen Tu. Seana, thank you for your input on this work along the way, and for your constant encouragement to go a step further than I might have otherwise. Sarah and Vic, thank you both for your helpful feedback, especially your thoughtful input on experimental design. Zhuowen, thank you so much for your guidance in writing the Frontiers of Innovation in Science Program grant and for providing me with the context to make use of current language/semantics models.

I could not have completed this dissertation without the mentorship, guidance, and friendship that I found in members of the Kutas lab. Tom Urbach, Katherine DeLong, Wen Chan, Anna Stoermann, Reina Mizrahi—thank you for so many conversations about experiments, data, and models, and for your feedback on papers and talks. Thanks also to the Taco Friday crew for conversations ranging from linear regression to cats. Thank you to Tom Urbach for the countless hours spent thinking out loud and looking at code and data, line-by-line, with us in the Monday data analysis meetings. Thanks to former Kutas lab members, who taught me so much about data collection and analysis: Paul Krewski, Ben Amsel, Mieko Ueno, Alex Kuo, Neil Cohn, Janaina Weissheimer, Ross Metusalem, Emily Provenzano, and Alyssa Truman. Thanks also to Beverley Walton, Thanh Maxwell, and Oura Neak for their dedication to grad students.

A great part of being a grad student has been making so many interesting friends – and the bittersweet part is that graduate school is transient. Friends—near and far—thank you for being there, for your support, for conversations about life/the universe/everything, for goofy text messages, for being writing buddies, and for occasionally humoring my need for karaoke: Megan, Jeff, Laura, Josh, Vicente, Julia, Tania, Amy, Gabriela, Miren, Richard, Stephanie, Jamie, Isabelle, Drew, Ryan, Kevin, Tim, Roman, Eva, Jérémie, Morgan, Paolo, Matt, Kristina, Tristan, Sarah, Emily, Dean, Valerie, Praneeth, Zenna, Stephanie, David, Ellen, Marjorie, Tim, Becky, Eduardo, Gus, Manizeh, Liuba, Sangyu, Nuné. Megan, thank you for being a constant partner in crime—and thanks for marrying Jeff and giving me another great friend. Also to Laura, Josh, Vicente, thank you for lots of laughter over the years and especially for your support during the last few months of writing. Thanks also to friends from way back, Jessica and Keith, for your visits and friendship across the miles! Finally, thanks to my parents, Sue and Don, and sisters, Mandy and Mary, for your love and support.

Chapter Two of this dissertation has been published in an article in *Language, Cognition, and Neuroscience* for which I was the primary author and investigator (Troyer & Kutas, 2018).

Chapter Three of this dissertation comprises an article which is undergoing revision at the *Journal of Experimental Psychology: Learning, Memory, and Cognition* for which I was the primary author and investigator (Troyer, Urbach, & Kutas, under revision).

Chapter Four of this dissertation, for which I was the primary author and investigator, is being prepared for submission for publication (Troyer & Kutas, in prep).

VITA

2008	B.S. in Cognitive Science B.S. in Psychology B.A. in Linguistics, French Indiana University, Bloomington, *Summa Cum Laude*
2012	M.S. in Cognitive Science Massachusetts University of Technology
2019	Ph.D. in Cognitive Science University of California San Diego

Publications

Troyer, M., Urbach, T.P., & Kutas, M. (Under review). *Lumos!*: Electrophysiological tracking of (wizarding) world knowledge use during reading.

Troyer, M. & Kutas, M. (2018). Harry Potter and the Chamber of *What*?: the impact of what individuals know on word processing during reading. *Language, Cognition, and Neuroscience*.

Troyer, M. & Borovsky, A. (2017). Maternal socioeconomic status influences the range of expectations during language comprehension in adulthood. *Cognitive Science*, *41*, 1405-1433.

Troyer, M., Hofmeister, P., & Kutas, M. (2016). Elaboration over a discourse facilitates retrieval in sentence processing. *Frontiers in Psychology*, *7*, 374.

Wang, H., Ma, L., Wang, Y., Troyer, M., & Li, Q. (2015). An ERP study in English relative clause processing by Chinese-English bilinguals. *Chinese Journal of Applied Linguistics*, *38*(1), 3-35.

Troyer, M., Curley, L.B., Miller, L.E., Saygin, A.P., & Bergen, B.K. (2014). Action verbs are processed differently in metaphorical and literal sentences depending on the semantic match of visual primes. *Frontiers in Human Neuroscience*, *8*, 982.

DeLong, K.A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass*, *8*(12), 631-645.

Miller, B.R., Troyer, M., & Busey, T.A. (2008). Virtual EEG: A software-based electroencephalogram designed for undergraduate neuroscience-related courses. *The Journal of Undergraduate Neuroscience Education (JUNE), 7*(1), A19-A25. Software available at http://virtualeeg.psych.indiana.edu.

ABSTRACT OF THE DISSERTATION

Muggle or magical?: Electrophysiological investigations of variation
in the language-knowledge interface during reading

by

Melissa Troyer

Doctor of Philosophy in Cognitive Science

University of California San Diego, 2019

Professor Marta Kutas, Chair

Across cognitive systems, world knowledge allows individuals to organize raw sensation into meaningful experiences. Language processing is no exception—words cue world knowledge which can be rapidly brought to mind in real time. It stands to reason that how much and how well individuals know things will impact what each individual brings to mind during real-time comprehension; yet to date, models of real-time language processing have not taken this variability into account. We addressed this issue by studying a linguistically rich, yet constrained, popular domain—the fictional world of Harry Potter (HP), by J.K. Rowling. In a series of three studies, we recorded event-related brain potentials (ERPs) while young adults who varied in their knowledge of HP read sentences about HP "facts" and/or sentences describing general topics. As a measure of real-time knowledge retrieval, we focused on N400 amplitude, a brain potential with a centro-parietal maximum occurring ~250-500 ms post stimulus onset that is sensitive to factors

impacting the ease of retrieval from semantic memory, with larger reductions in N400 (i.e., more positive-going N400 potentials) associated with greater ease of retrieval.

Across all three studies, we found that individuals' domain knowledge of HP (assessed via an offline multiple-choice trivia quiz) was moderately-to-strongly correlated with average N400 brain potentials to contextually supported (i.e., accurate) words completing sentences about HP, but not to unsupported (inaccurate) endings, nor to N400 effects (i.e., difference ERPs) of more vs. less supported endings to sentences describing general topics (Experiments 1 and 3). Single-trial regression analyses pitting domain knowledge against participants' reports of knowledge about each HP "fact" (i.e., whether or not they had known each fact ahead of reading it) revealed trial-level knowledge was a strong, but not the sole, predictor of N400 amplitudes to supported words (Experiment 2). Rather, even after accounting for trial-level knowledge reports, and especially when retrieval conditions were presumably more difficult (i.e., for trials reported as unknown by an individual or those generally less likely to be known across participants), N400 amplitudes were modulated by individuals' domain knowledge of HP. We hypothesized that degree of domain knowledge might modulate real-time semantic retrieval by virtue of the differential organization of that information in semantic memory—for example, functionally organized around events and categories, as has been proposed more generally in the literature on expert knowledge.

In Experiment 3, we therefore manipulated the relationship between the final (critical) words of HP sentences and their sentence contexts according to these organizational structures. Critical words (each of which appeared across conditions) were supported continuations, unsupported/unrelated continuations, or unsupported continuations that were related via an HP-specific category to the supported ending or to the event/episode being described by the sentence context. Individuals with greater HP domain knowledge showed reduced N400 amplitudes not only to supported words, but also to unsupported but contextually related words (for both types of relationships), compared to the unsupported unrelated words. That is, domain knowledge systematically influenced the quick availability of functionally (categorically, event-based) related knowledge during written sentence comprehension.

Our results provide the first empirical demonstration that real-time retrieval of knowledge during reading is determined by how much information is known and which facts are readily available to an individual comprehender. Domain knowledge seems to influence implicit retrieval of (perhaps) partial information (Experiment 2) and the availability of relevant/related information (Experiment 3)—the very information that is needed to make sense of words in real time. We hypothesize that variation in domain knowledge leads to systematic variation in hallmark organizational structures of semantic memory, including organization of categories and events, which in turn influence the degree to which relevant information can be (pre-)activated in order to make sense of words in real time.

CHAPTER ONE: INTRODUCTION AND BACKGROUND

1.1 Knowledge and language

Undeniably, the ability of individuals to glean meaning from spoken or written language depends on their knowledge of words and of the world. Imagine a favorite hobby—e.g., basketball, music, botany. For the southern Californian plant enthusiast, comprehension of the following language from a recent *New Yorker* article might proceed with ease:

> Succulents—drought-friendly, fireproof, angular, Zen—long ago attained the status of design cliché, a living version of the shag rug, Heath mug, Eames chair. But now a particular species, Dudleya farinosa (stage name: Powdery Liveforever), a wild roseate plant with silvery, pink-tipped leaves and a spectacular yellow-flowered stalk, which thrives on California's coastal bluffs, has become the It Plant for succulent thieves. (Goodyear, 2019)

Even for the uninitiated, the descriptive text might allow for a basic understanding—plant thieves are stealing the succulents! Yet, considering the text word by word, it seems obvious that the precise information brought to mind at each moment is likely to differ when it comes to the plant connoisseur vs. a novice. However, despite the seemingly obvious impact of differences in knowledge on language understanding, current theories and models of real-time language processing do not typically take such variability into account, and little empirical work has investigated how variability in the degree of specific content knowledge influences the moment-by-moment availability of information during real-time language processing. We therefore do not yet have a precise description of how variability in knowledge shapes the nature and timing of information that is available to comprehenders as they understand language, word by word.

This thesis aims to advance our understanding of how individual differences in knowledge impact the nature of its organization in long-term memory and the quick availability of information as individuals understand words in real time. To do so, we employ multiple methods and approaches from several subfields of cognitive science, including human cognitive electrophysiology, psycholinguistics, cognitive psychology, and natural language understanding. We first provide a review of relevant background that motivates the current work before describing the precise goals of the studies in the thesis.

1.2 Understanding words in context

Poets, philosophers, and cognitive scientists alike have all questioned what it means to understand a word. Amongst psycholinguists, one predominant metaphor has been to assume that the meanings of words are stored in a "mental dictionary," or lexicon, which contains "a small chunk of phonology, a small chunk of syntax, and a small chunk of semantics" (Jackendoff, 2002, p. 131). On such views, the meanings of words are assumed to be relatively stable—presumably both across contexts and across individual speakers of a language. An alternative approach is to consider a more dynamic view of word meaning. Elman (2011, p. 16) suggests that words should be viewed "…not as elements in a data structure that must be retrieved from memory, but rather as stimuli that alter mental states… in lawful ways. In this view, words are not mental objects that reside in a mental lexicon. They are operators on mental states." On this view, an understanding of a word's meaning seems to be a natural consequence of applying new information (an incoming word) to an existing mental representation. This view invites questions about how both immediate (e.g., sentence-level) and lifelong (e.g., at the level of long-term memory) context shape how individuals understand words in real-time.

For example, Elman (2011) describes several studies focusing on the interpretation of verbs, which can provide comprehenders with rich sources of information about the actions they describe as well as the people and objects involved in the actions. Subtle differences in the sense of verbs can arise depending on the types of linguistic agents, patients, objects, and instruments that accompany them. For example, Matsuki and colleagues (2011) found that a verb like '*cut*' following an agent like '*scissors*' seemed to be interpreted differently from when it followed an agent like '*saw*'—individuals read an object like '*paper*' faster in the first case, compared to an object like '*wood*.' Similarly, Bicknell and colleagues (2010) found that combinations of agents like '*journalist*' or '*mechanic*' with a verb like '*checked*' led to both reading time differences and rapid differences in the brain's electrophysiological response to the patient of the verb. For example, in '*The journalist/mechanic checked the <u>spelling</u>…*,' differences in brain potentials were observed within about a third of a second at '*spelling*,' being facilitated when appearing after '*journalist*' compared to '*mechanic*.'

Indeed, research over the last thirty to forty years using online measures including reading times, eye-tracking (during both reading and listening), and event-related brain potentials (ERPs) has consistently shown that the ease with which words and phrases are understood is dependent on the surrounding linguistic and physical/situational context, include visual scene information and information about who the speaker is (e.g., Kutas & Hillyard, 1980; Kutas, Lindamood, & Hillyard, 1984; Tanenhaus et al., 1995; Altmann & Kamide, 1999; Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008; Brown-Schmidt, 2012; Borovsky & Creel, 2014). In particular, the N400 event-related brain potential has proved a useful tool for investigating how words are quickly understood in context. N400s are negative-going potentials occurring between ~250-500 ms after the onset of a meaningful stimulus, like a word, pseudoword (e.g., Laszlo & Federmeier, 2009) or picture (e.g., Friedman, 1990), are broadly distributed over central and parietal scalp sites, and provide a sensitive measure of semantic processing (reviewed in Kutas & Federmeier, 2011).

In early reports, Kutas and Hillyard (1980) showed that a word's N400 amplitude was sensitive to sentence context, such that anomalous words showed a large (negative-going) N400 amplitude, and contextually supported words showed a reduced (more positive-going) N400 amplitude, a finding that has been replicated numerous times (a quite comprehensive list is provided in Table 1 of Van Petten & Luka, 2012). Going beyond studies showing that N400 is large for contextually congruous words, other work has indicated that large N400s are the "default" response to words (or other meaningful stimuli), with reductions in N400 amplitude suggesting facilitated access to incoming words. Evidence supporting this interpretation includes the finding that N400 amplitudes are larger for words at the beginning of a sentence and attenuated to content words as sentential context accrues (Van Petten & Kutas, 1990). Interestingly, when words are presented in isolation, N400 amplitude is larger for words with a greater number of orthographic neighbors (words which share all but one letter; Laszlo & Federmeier, 2011) and for words with a greater number of semantic associates (Van Petten, 2014); moreover, the effect of orthographic neighborhood persists for words in sentential contexts (Laszlo & Federmeier, 2009), though context effects tend to take precedence over lexical factors like word frequency and semantic association (e.g., Van Petten & Kutas, 1990; Van Petten, 1993).

Manipulations of contextual constraint and word cloze probability suggest that a word's N400 amplitude is sensitive to the *degree* to which it is predictable from the context. In cloze tasks, individuals are provided with sentence frames and asked to provide the next (or sometimes final) word in the sentence. Cloze probability is defined as the proportion of individuals who provide a word given a specific context and has been used as a measure of a word's contextual predictability. On average (i.e., across a separate group of individuals), N400 amplitudes are reduced as a function of their predictability according to cloze tasks (Kutas & Hillyard, 1984; DeLong, Urbach, & Kutas, 2005; Federmeier et al., 2007). Graded N400 effects in certain experimental designs have been taken as evidence that people probabilistically (pre-)activate likely upcoming word information (DeLong, Urbach, & Kutas, 2005). More recently, N400 amplitudes have also been linked to *surprisal* (often defined as the negative log probability of a word given a context of length *n*), another way of operationalizing linguistic predictability (Frank, Otten, Galli, & Vigliocco, 2015; Frank & Willems, 2017). However, these studies do not provide a direct link between cloze probability (or information theoretic) estimates of a word's likelihood given a sentential context and what any given individual brings to mind in real time during reading (see discussion in Van Petten & Luka, 2012).

N400 effects are often accompanied by late positivity effects, which have also been termed P600s and post-N400 positivities (PNPs). In early studies, the functional significance of the P600 was attributed to factors relating to syntax, it being sensitive to syntactic anomalies and to syntactic difficulty (Osterhout & Holcomb, 1992; Kaan, Harris, Gibson, & Holcomb, 2000). However, modulation of late positivities can also occur in the absence of syntactic violations or difficulty and have more recently been linked to a number of high-level cognitive processes that may occur during sentence processing, including reanalysis of semantically incongruous and/or unexpected words in context (discussed in Brouwer et al., 2012; Van Petten & Luka, 2012). In a meta-analysis, Van Petten and Luka (2012) reviewed 45 studies which included sentence materials containing a congruent vs. incongruent ending. Of these studies, the majority showed a biphasic response—that is, an N400 effect (reduction in N400 amplitude for congruent vs. incongruent words) followed by a late positivity effect, with more positive-going potentials for incongruent compared to congruent words in a later time period (typically ~600-900 ms). In their review, Van Petten and Luka suggest that there may be some functional dissociation between frontally and parietally distributed late positivities, with enhanced parietal positivities more often reflecting processing of semantically anomalous words, and enhanced frontal positivities more often occurring to contextually unexpected but plausible / meaningful words (as in Federmeier, Wlotko, De Ochoa-Dewald, &

Kutas, 2007; Thornhill & Van Petten, 2012; DeLong, Quante, & Kutas, 2014). Though there is currently no consensus view on the functional significance of late positivities in language processing, one promising view put forth by Brouwer and colleagues (2012) is that they are perhaps best viewed as a "family… that reflect the word-by-word construction, reorganization, or updating of a mental representation of what is being communicated" (p. 138). In sum, late positivity context effects may reflect a set of processes involved in evaluation of a word in context beyond its initial semantic processing reflected by N400 potentials.

1.3 Semantic memory structure and real-time language processing

The N400 studies reviewed above provide a vein of evidence that sentential context has a quick influence on word understanding during reading. Words which "make sense" in context are easier to process than those which do not—whether the context is a sentence, a word, or general language experience. Moreover, N400 amplitudes are sensitive to a word's ease of access even when there is no overt "context" present. When words are presented in isolation, high-frequency words elicit reduced N400 potentials compared to low-frequency words (Kutas & Federmeier, 2000). When preceded by a category label (like '*a water sport*'), N400 amplitude to a word describing a more typical category member ('*swimming*') is reduced compared to less typical members ('*rowing*') and largest for non-members ('*pottery*'). These findings suggest that individuals' knowledge structure—including degree of experience and structure of semantic memory—are rapidly available sources of information that are used when interpreting words (reviewed in Kutas & Federmeier, 2000). In other words, N400 amplitudes seem to reflect the ease of access of information from long-term (semantic) memory. Indeed, studies using intracranial recordings have linked N400 amplitudes to regions of the brain known to support retrieval from long-term memory (McCarthy, Nobre, Bentin, & Spencer, 1995), findings consistent with a metareview comparing experimental conditions eliciting N400 effects with studies using methods (like fMRI and MEG) that can localize such effects to functionally specific brain regions (Lau, Phillips, & Poeppel, 2008).

When it comes to understanding how individuals structure and process knowledge about the world, some researchers have distinguished between semantic knowledge—rooted in theories of the semantics of words in a language and how they lawfully combine—and world knowledge, which is perhaps more flexible and is based on truth conditions of the actual world. Interestingly, Fischler and colleagues (1983) reported that brain potentials with timing and scalp distribution similar to N400s were sensitive not necessarily to the truth conditions of sentences describing simple relationships between words and categories (e.g., '*A sparrow is (not) a bird / vehicle*'), but rather were sensitive to the semantic relationship between the content words, being reduced in amplitude for '*bird*' regardless of whether the sentence was affirmative or negative.

So-called semantics and world knowledge are often conflated—for example, words which constitute inanimate objects cannot perform actions like singing or dancing, and these actions also do not often occur in the real world (though in certain fictional circumstances, they very well may). Hagoort and colleagues (2004) contrasted sentences which were felicitous, sentences which violated semantics, and sentences which violated world knowledge—based at least in part on their specific population's knowledge of the world (in this case, Dutch speakers living in the Netherlands). For example, participants would read the following sentence frame with one of the underlined critical words: '*The Dutch trains are yellow / white / sour and very crowded.*' Critically, the fact that Dutch trains are yellow (and not white) led to a world knowledge violation for '*white*.' The authors found that N400 amplitude to the world knowledge violations was similar in morphology and time course to that of the semantic violation (trains are typically not deemed to be *sour*), empirically demonstrating that individuals quickly make use of knowledge gleaned from particular experience with the world as they understand words in sentences. This general pattern of results extends to knowledge about fictional characters, like the Hulk and Scooby Doo (Filik & Leuthold, 2013) as well as to short discourses describing fictional scenarios about otherwise implausible events, such as a story about peanuts falling in love (Nieuwland & Van Berkum, 2006).

In order to more specifically probe the nature of information from long-term memory that is rapidly available in real time, many ERP studies of sentence processing have used what is often termed a *related anomaly* paradigm. For example, Kutas & Hillyard (1984) asked individuals to read sentences that were completed with either the best completion (i.e., a high-cloze and congruent word), a low-cloze/incongruent word, or a low-cloze word that was incongruent but related in some way to the best completion. They observed a three-way difference such that N400 potentials were largest for the unrelated/incongruent word, most reduced for the best completion, and intermediate for the so-called "related anomaly," suggesting that the related semantic information was activated in long-term memory. Since that study, many others have used similar related anomaly paradigms in order to probe the nature of information that is activated during moment-by-moment word understanding during sentence processing, suggesting that a variety of sources of information make a rapid contribution, including information that overlaps based on categorical relationships (Federmeier & Kutas, 1999; Federmeier et al., 2002); information related via schemata and/or event structure (Metusalem et al., 2012; Paczynski & Kuperberg, 2012; Amsel et al., 2015); perceptually-related information (Rommers, Meyer, Prammstra, & Huettig, 2013; Amsel et al., 2015); and other semantically associated information (Kutas & Hillyard, 1984; DeLong, Chan, & Kutas, 2018).

As a prime example, Federmeier and and Kutas (1999) presented individuals with sentence pairs ending either in a best completion, a categorically-related word, or a category non-member; for example, '*The gardener really impressed his wife on Valentine's Day. To surprise her, he had secretly grown some <u>roses</u>/<u>tulips</u>/<u>pines</u>.*' They hypothesized that if the structure of conceptual information stored in long-term memory influences what is brought to mind in real time, then processing of the categorically-related word '*tulips*' should be facilitated compared to '*pines*.' Indeed, this is what the authors observed: a three-way difference in N400 amplitude such that N400 amplitude was most reduced to '*roses*,' followed by '*tulips*,' and largest to '*pines*.' Moreover, the related anomaly effect (related vs. unrelated words) was most pronounced for words appearing in high (compared to low) constraint contexts.

From these results, Federmeier and Kutas concluded that, from sentential contexts, individuals can form expectations about more than simply general semantic features about upcoming words, given the observed differences between best completions and related anomalies (e.g., '*roses*' and '*tulips*'). Importantly, they also concluded that general semantic features of an upcoming word *did* seem to be able to be anticipated based on the sentence context, given the observed differences between the related and unrelated anomalies ('*tulips*' vs. '*pines*'). That the related anomaly effects were largest for high constraint contexts also led Federmeier and Kutas to conclude that when more specific word expectations were available to comprehenders, they were more likely to activate semantic features of likely upcoming words prior to encountering them. Together, their findings led them to conclude that "…the kind of perceptual and functional similarity captured by semantic categories… has an immediate impact on neural and linguistic processing" (p. 489); i.e., that long-term memory structure quickly shapes moment-by-moment word processing in sentences.

1.4 Individual differences in word-by-word reading

If long-term memory structure shapes the nature of information that is available to language comprehenders in moment-by-moment language processing, then it logically follows that differences in the contents of each individual's knowledge (i.e., the information that is stored in long-term memory) should have serious consequences for language processing and comprehension. There is, however, very little empirical work studying the link between individuals' knowledge and real-time language processing.

A handful of studies, however, have investigated real-time access to individualized "personal semantics" knowledge during real-time language comprehension (e.g., Fischler, Bloom, Childers, Arroyo, & Perry, 1984; Renoult et al., 2016). For example, Coronel and Federmeier (2016) had individuals report their personal preferences on a number of different topics (e.g., their favorite dessert). Then, in an ERP reading study, they read sentences about the same topics; for example, '*For dessert, you like to eat cheesecake*.' Endings that were consistent with each individual's actual preferences (whether the individual preferred cheesecake vs. popsicles, for example) elicited reduced N400 amplitudes compared to endings that were inconsistent, providing support for the notion that knowledge at the individual participant level is taken into account during real-time word understanding. The timecourse and scalp distribution of these effects were similar to those of consistent vs. inconsistent words which ended sentences about common preferences (e.g., that many people dislike *anchovies* vs. *pepperoni*). However, there are some limitations to generalizing the results of this study. First, participants were explicitly asked to provide answer to the exact items used in the ERP study, leading to a potential confound due to repetition of the items in the case of the consistent, compared to inconsistent, endings. Second, the study could make only limited inferences about the functional organization and use of personal semantic knowledge in real time, as there were no conditions including words that probed information semantically related to the consistent word (as in related anomaly paradigms).

A few studies have investigated the consequences of group-level affiliations which can also lead to rapid differences in behavioral and neural response to word processing. In an ERP sentence reading study, Van Berkum and colleagues (2009) asked individuals with different political affiliations to read sentences containing language about situations which were potentially morally objectionable, depending on political affiliation. They observed small but quick differences in ERPs beginning at the first word which distinguished a potential clash in values (e.g., '*I think euthanasia is an <u>acceptable / unacceptable</u>…*'). In an ERP study investigating the processing of isolated words in the context of a lexical decision task, Bechtold and colleagues (2019) showed that ERPs to abstract words about math differed for individuals with more vs. less math expertise: more knowledgeable individuals showed a trend for reduced N400 amplitude and showed a larger late positivity to words like '*multiplication*' compared to less knowledgeable individuals, but no systematic differences were observed for abstract words unrelated to math, like '*thought*' (Bechtold, Bellebaum, Egan, Tettamanti, & Ghio, 2019). In addition, Verhagen and colleagues studied individuals with different group affiliations: job recruiters, job seekers, and individuals not yet looking for jobs (Verhagen, Mos, Backus, & Schilperood, 2018). In a cued word production task, individuals read the first two or more words of phrases that were variously associated with job ads (e.g., '*40 hours per…*') or news articles (e.g., '*in the United…*') and then produced a completion that came to mind. They found that job recruiters were relatively faster to respond to the job ad compared to the news phrases; job seekers showed a similar effect that was somewhat smaller; and inexperienced individuals (those not yet looking for a job) showed the opposite pattern (faster responses for news items), suggesting a group-level sensitivity to language register/genre.

Although there is only limited work investigating the role of individual differences in content knowledge on real-time language processing, there has been a recent surge of interest in the role of more general cognitive abilities, such as working memory and cognitive control, as well as in language proficiency, on how individuals understand language in real time. Here, the focus has been mostly on grammatical processing. Variation in language proficiency in monolinguals has been linked to graded differences in the brain's response to grammatical violations (Pakulak & Neville, 2010). In addition, some studies have identified subgroups of individuals who seemed to show different neurocognitive profiles of second language processing. For example, some individuals who were at intermediate stages of learning a language showed neural patterns more similar to native language speakers, exhibiting late positivity (P600) effects to morphosyntactic violations in German, while others showed an N400 effect, and this pattern was linked with behavioral measures of grammatical processing (Tanner, McLaughlin, Herschensohn, & Osterhout, 2013). Dominance of one pattern of effects (N400 vs. P600) within an individual has also been observed in individuals as they understand grammatical violations within their native language (Tanner & Van Hell, 2014), and N400- vs. P600-dominant effects have also been linked to an individual's verbal working memory capacity[1] (e.g., Nakano, Saron, & Swaab, 2010; Kim, Oines, & Miyake, 2018).

Another general capacity which predicts certain aspects of real-time sentence processing is cognitive control, which has been linked both to N400 and late positivity ERPs. In one study,

[1] Verbal working memory is most typically measured by asking participants to complete "span" tasks which task them with reading sentences out loud while remembering a series of words (Daneman & Carpenter, 1980).

Boudewyn and colleagues (2012) asked individuals to read sentences containing lexical associates; e.g., '*In her haste, she forgot to buy the apples and oranges/bread*,' where '*oranges*' is lexically associated to '*apples*,' and '*bread*' is not. Replicating previous work (e.g., Van Petten, Weckerly, McIsaac, & Kutas, 1997), Boudewyn et al. found that such lexical associates elicited reduced N400 amplitude compared to non-associates. Further, they showed that N400 effect size was associated with individuals' performance on a Stroop task, believed to reflect the ability to suppress irrelevant information: larger effects were associated with individuals who performed the worst on the Stroop task. The authors interpreted these findings as evidence that individuals with lower suppression abilities (as measured by the Stroop task) were less able to suppress information from the lexical associate—presumed to be less important than understanding the overall content of the sentence message. Zirnstein and colleagues (2018) found that a different measure of inhibitory control (the AX-CPT) was correlated with the size of monolingual speakers' frontal positivity effects to sentence-medial words that were predictable (i.e., congruent, likely continuations) vs. unpredictable (but plausible). The authors interpreted these effects as evidence that cognitive control mediates individuals' abilities to override initial predictions about upcoming words in sentence contexts.

1.5 Domain knowledge shapes cognition

Though there have been few attempts to link individual differences in knowledge with real-time language processing mechanisms, there is a large literature on how content expertise within a domain influences cognition when it comes to processing information within the domain (reviewed in Ericsson, Charness, Feltovich, & Hoffman, 2006). Chi and Ohlsson (2005) describe multiple ways in which complex (or expertlike) knowledge may come to be organized, including via semantic networks, theories, and schemas. Semantic networks capture multiple types of associations, including categorical relationships (e.g., Rosch, Mervis, Gray, Johnson, & Boyes, Braem, 1976) and taxonomic hierarchies (e.g., Collins & Quillian, 1969). Theories capture organizing principles of complex knowledge domains (e.g., physics) and may be available only for experts (e.g., diSessa, 1993). Finally, schemas—similar to "scripts," "frames," and in some cases "chunks"—capture information about patterns of experience that give rise to predictable templates, the "slots" of which may be filled in by multiple types of elements (e.g., Schank & Abelson, 1977; Johnson-Laird, 1983; van Dijk & Kintsch, 1983; Kintsch, 1988).

As a famous example, when expert players view a chessboard with pieces from real chess game play, they seem to "chunk" the information, recognizing patterns that are not available to chess novices (de Groot, 1966; Chase & Simon, 1973). As a result, master chess players can remember more individual pieces than novices can—but only if they studied chess arrangements consistent with game play, and not with randomly shuffled pieces (de Groot, 1966; Chase & Simon, 1973; see Gobet et al., 2001, for a review). Moreover, expert chess players seem to recruit brain regions known to be involved in long-term memory tasks when they process such chess configurations, a finding which has been interpreted as indicating neural functional reorganization of information into templates and chunks that can be more flexibly retrieved from long-term memory in the case of experts (reviewed in Guida, Gobet, Tardieu, & Nicolas, 2012).

Miller described "chunking" as "… a process of organizing or grouping the input into familiar units" (1956, p. 93). By chunking information like letters into words, or like binary digits into higher-level numeric representations, individuals may increase the amount of information they are able to manipulate over a short period of time (e.g., Simon, 1974). On such a proposal, familiar and/or relevant "bits" of perceptual input differ as a function of knowledge, with experts being able to encode larger "chunks" compared to novices. In the chess example, an expert might be able to encode certain common formations consisting of multiple chess pieces whereas a novice might only encode smaller chunks, perhaps consisting of individual pieces.

In other domains, too, experts may use deeper levels of analysis compared to novices (Chi, 2006). For example, expert physics students may access deeper or higher-level organizational principles when categorizing physics problems whereas novices access more shallow or literal features of the problems (Chi, Feltovich, & Glaser, 1981). Chi & Ohlsson (2005) discuss these individual differences as variation in the availability of *theories*, in which core knowledge can form organizing principles from which other bits of information extend. Experts, but not novices, may have access to this information, allowing them to tell which bits of information are most relevant or important when understanding language or other content (e.g., images) describing physical processes.

1.6 Domain knowledge influences offline measures of text comprehension

There is also a large literature investigating how background knowledge influences text memory and offline comprehension measures. Both children and adults with greater background knowledge are more accurate when answering comprehension questions about them. For example, children who knew more about spiders were better at making inferences in a text about spiders compared to children with less spider knowledge (Pearson et al., 1979). For adults, such effects have been demonstrated in many diverse domains, including music (Arbuckle et al., 1990), cooking (Soederberg Miller, 2001, 2009), and sports (Spilich et al., 1979; Walker, 1988; Schneider et al., 1989; McNamara et al., 2004). In adults, prior experience with a domain (e.g., health or music) can also improve learning from texts about the same domain (Arbuckle et al., 1990; Beier et al., 2005; Chiesi et al., 1979), and expertise has also been shown to boost recollection in domains including biology (Long, Prat, Johns, Morris, & Jonathan, 2008) and fictional knowledge of "Star

Trek" (Long & Prat, 2002), though a demonstration of benefits of knowledge on recognition memory from texts has been less forthcoming (discussed in Long et al., 2008).

Some studies have concurrently measured knowledge in a particular domain (e.g., sports like soccer and baseball) and more general cognitive measures (e.g., IQ or working memory) in order to test whether knowledge may matter more for individuals scoring lower on general cognitive measures. Studies typically find that text comprehension in domains of knowledge does not differ as a function of cognitive ability, so long as the participants are knowledgeable in the domain (Schneider et al., 1989; McNamara et al., 2004). However, individuals who are less knowledgeable and also score lower on cognitive measures are at a disadvantage. Domain knowledge may therefore provide more of a "boost" for individuals who otherwise may have difficulty with text comprehension.

A general interpretation for the finding that knowing more can help individuals comprehend texts is that these individuals benefit from (1) pre-existing schemas (van Dijk & Kintsch, 1983; Kintsch, 1988) or scripts (Schank & Abelson, 1977) and/or (2) situation (Foss, 1982; Hess et al., 1995; Kintsch, 1998) or mental (Johnson-Laird, 1983) models that derive both from pre-existing knowledge and the unfolding situation in the discourse. A situation model can provide a framework with which to interpret a text (Soederberg Miller et al., 2006; Soederberg Miller & Stine-Morrow, 1998; Bransford & Johnson, 1972) as well as the basis for forming new long-term memories (Kintsch, 1988). On one theory, combining pre-existing knowledge structures with incoming information is best thought of as relying upon "long-term working memory," involving the retrieval of structured chunks of knowledge (Ericsson & Kintsch, 1995). Ericsson and Moxley (2012) suggest that on this view, "…skilled readers interpret new sentences in light of their general knowledge and in particular their knowledge of the previously read text. In many narrative texts the reader generates a model of the current situation and this model allows the reader to integrate the new information in memory by updating and changing the model of the current situation (Kintsch, 1998)" (p. 411).

Indeed, many memory-based discourse processing models suggest that differences in prior knowledge should affect language comprehension. Most models propose a linguistic (or surface) level, including the words themselves, from which a textbase of meaning, including propositional content, gets constructed with reference to a knowledge base (contained in memory). In most models, this is a passive memory process (Kintsch, 1988; McKoon & Ratcliff, 1998; Myers & O'Brien, 1998), though constructivist models place more emphasis on active, goal-directed inferencing during text comprehension (Graesser et al., 1994). Finally, this information is incorporated into a larger situation model. It may well be the case that knowledgeable (compared to less knowledgeable) individuals are able to make better use of more highly organized structures (such as situation models) to guide language processing in real time, but this has not been systematically assessed.

1.7 Bridging approaches to understand the language-knowledge interface

As the above review suggests, limited work has addressed the role of individual differences in knowledge on real-time comprehension. This thesis presents a tightly connected series of studies which aim to do just this. The approach we will take is to combine methods and analyses from different strands of cognitive science. As in the vast literature on expert cognition, we focus on a domain that is well-suited to our needs—the narrative world of Harry Potter. This domain is a currently popular domain with which many young adults are familiar, though they still vary in their degree of knowledge—which we can assess with background questionnaires and objective quizzes. Moreover, it is a rich narrative world, with many (novel) cross-cutting people, places, events, and so on. Because it is text based, we also have a corpus to use as “ground truth,” making it possible to apply recently-developed language/semantic models to capture meaningful relationships between words in a quantitative manner.

We combine this approach with tools from psycholinguistics and cognitive electrophysiology, using ERPs as measures of access to information as individuals read sentences word-by-word about the HP domain and about general topics. ERPs are extremely well-suited to answering questions about real-time, moment-by-moment language comprehension and reflect averaged electrical brain activity time-locked to the onset of a stimulus and have remarkable temporal resolution, making them ideal for studying real-time language processing. The specific brain potentials of interest are described in some detail above. These include the N400 (a negative-going wave occurring approximately 250-500 ms after onset of a meaningful stimulus, with a typically centro-parietal distribution over the scalp) and so-called late positive components, or LPCs (beginning ~500 ms or later after the onset of a meaningful stimulus, with various distributions over the scalp).

In addition, we employ a relatively new approach in the field of cognitive neuroscience, regression-based event-related potentials (rERPs; Smith & Kutas, 2015), which provide a framework for analyzing effects of both categorical and continuous variables (and any interactions thereof) on ERPs. In many studies of individual differences, participants are grouped categorically based on behavioral assessments. While this approach has been fruitful, a drawback is that much of the information in behavioral variability cannot be captured in order to predict continuous measures like ERPs. With rERPs, continuous scores can be used directly to estimate how much the ERP changes with a given amount of change in continuous predictor of interest.

1.7.1 Assessing the quick availability of knowledge during real-time comprehension

If individuals with a greater degree of knowledge within a domain make quick use of this knowledge, we would expect to see differences during the earliest moments of semantic processing, reflected in N400 amplitudes, compared to those with less knowledge. To test this hypothesis, we conducted an experiment which examines the influence of HP domain knowledge on the processing of words which were contextually supported vs. unsupported in both HP sentence contexts as well as in sentences that described general topics, described in **Chapter Two**. In this study, we also assessed other individual differences in areas that have been known to influence language processing and/or seemed likely to differ along with knowledge of a fictional world, including measures of reading experience and verbal working memory, among others. If domain knowledge has a rapid influence on individuals' ability to bring information to mind in real time, then we would expect domain knowledge to have an influence on N400 effects of contextual support for HP sentences, but *not* for sentences about general topics. Moreover, given the sensitivity of N400 amplitudes to ease of semantic access, we anticipated that this effect would be driven by the brain's response to *supported*, and not *unsupported*, items, as unsupported words should be similarly difficult (i.e., not facilitated) regardless of domain knowledge, and supported should be facilitated—perhaps in a graded fashion—according to degree of domain knowledge.

Given the vast N400 literature documenting its sensitivity to factors relating to ease of semantic access—including effects of sentential context/word congruity, cloze probability, and numerous factors relating to retrieval from semantic memory—it might seem obvious that we should observe the pattern of results described above. Of course individuals vary in their knowledge, and of course this variability should influence the ease with which they bring information to mind—how could it be otherwise? Yet no other studies (to our knowledge) have systematically investigated any relationship between variation in content/domain knowledge and real-time processing measures of word understanding during language comprehension. Chapter Two therefore laid the groundwork for a series of studies designed to probe not only *whether* domain knowledge has an influence, but which carefully aimed to isolate item-specific vs. more general influences of domain knowledge (Chapter Three) and to link such influences to variation in the functional organization of knowledge (Chapter Four)—i.e., to begin to understand how domain knowledge might shape the internal organization of an individual's long-term memory such that information relevant for processing words in sentences might be more readily available for more vs. less knowledgeable comprehenders.

1.7.2 Domain knowledge beyond knowledge of specific facts

In Chapter Three, we asked whether influence of domain knowledge on real-time language processing might extend beyond mediating the likelihood that any individual would know a given fact. That is, at the neural level, can we see differences in real-time semantic processing that go beyond a link between domain knowledge mediating the proportion of items/trials (i.e., HP "facts") an individual knows? To do this, we constructed a large set of sentence pairs, all of which were descriptions about the narrative world of HP, and all of which ended in a contextually supported (i.e., correct) word. Individuals read these sentences while we recorded EEG. During the EEG study, just after having read each sentence pair, individuals indicated whether they believed they had known that information prior to having read it. These individuals also completed measures of individual differences. This single-trial design allowed us to ask questions that we were unable to ask in Chapter Two. First, we could identify which specific trials were reported as having been known, thus verifying that individuals with higher HP knowledge (as measured using our offline, trivia-style HP quiz) did indeed (at least subjectively) indicate knowing a greater proportion of the items. Second, we could verify that items reported as known did indeed lead to facilitated semantic retrieval compared to those reported as unknown—that is, that reduction in N400 amplitude, at the single trial level, was related to whether individuals actually reported knowing the information. And finally, using recent advances in regression modeling (Baayen, Davidson, & Bates, 2008; Smith & Kutas, 2015), we were able to ask whether HP domain knowledge influence ERPs beyond strictly determining the proportion of items each individual knew. That is, was there an influence of domain knowledge that went beyond specific, item-/trial-level knowledge? We would expect this to be the case given the vast literature suggesting that individuals with greater expertise within a domain enjoy access to knowledge structures with differential organization—such that relevant information is organized according to theories, schemas, and/or mental models.

1.7.3 How expertise might shape the functional organization of knowledge use in real time

In Chapter Four, we more directly tested the hypothesis that domain knowledge does shape the functional organization of knowledge in long-term memory, and that this knowledge can be readily available during real-time sentence processing for knowledgeable individuals. To do so, we used a related anomaly paradigm and manipulated words that completed sentences about HP, including a contextually supported word, a contextually unsupported and unrelated word, and a contextually unsupported word that was related to the context sentence and/or supported word. More specifically, the related words (a) were either from the same semantic category as the supported word or (b) were related to the event and/or episode described by the sentence context. These two types of related anomalies roughly paralleled those used in Federmeier & Kutas (1999) and Metusalem et al. (2012), respectively. First, assuming that all or the majority of participants would have at least some knowledge of the domain of HP, we expected to see effects of both contextual support (reduced N400 for supported words compared to unsupported/unrelated words) and of the related anomaly (reduced N400 for supported words compared to unsupported/related words). If having greater knowledge (in this case, of the fictional/narrative world of HP) shapes the functional organization of information stored in long-term memory such that related elements—i.e., co-occurring semantic features that overlap between category members and/or meaningful elements that are part of an event or situation model—are more readily accessible, then we would also expect to see that each effect (contextual support, related anomaly) would be modulated by HP domain knowledge, such that as an individual's degree of knowledge increases, so would the size of each effect. In the case of effects of contextual support, such a pattern would replicate the predicted pattern observed in Chapter Two. As for the related anomaly effects, such a pattern would provide more direct evidence that, as individuals' knowledge increases, they are readily able to (quickly) bring to mind contextually-relevant (though linguistically unsupported) information as they understand words in sentences in real time.

Because the related anomalies used in the sentence materials in Chapter Four come from a fictional, narrative world—where no featural similarity or concept relatedness norms exist—we also conducted additional quantitative analyses to justify our choice of supported, related, and unrelated word endings. We made use of Google's word2vec, a state-of-the-art language semantics model. This model computes distributed representations of semantics based on the distributional properties of words in a corpus. We trained a version of this model directly on the electronic texts of books from the HP series in order to derive similarity measures between critical words from each condition in our materials (supported, related, unrelated) as well as between sentence contexts and critical words of each type. For the former comparison, we expected the greatest similarity between supported and related words, and lower similarity between supported/unrelated

related/unrelated pairs; for the latter, we expected the highest similarity between sentence contexts and supported words, followed by related and unrelated words, respectively.

We also conducted behavioral norming studies asking individuals to rate either the similarity or relatedness (in separate norming studies) between the supported word and (a) related or (b) unrelated words. We expected that supported words would be judged as more similar (and more related) to our related, compared to unrelated, words. We also expected that, if the word2vec language model captures similarity metrics that are reliably related to the psychological similarity (and/or relatedness) metrics which individuals use to make judgments, then the item-wise word2vec similarity metrics should correlate with our human subject norming judgments, and that this correlation should be strongest amongst individuals with the greater HP knowledge.

Finally, in the same study, after participants completed the EEG study, we also asked them to complete a trial-by-trial cloze task on sentences that they had just read. This allowed us to sort ERPs based on whether individuals accurately completed each item (or not). We were therefore able to test whether accuracy had an influence on each condition type—supported, related, and unrelated endings, respectively. We expected that accuracy would have a large effect on supported trials, such that N400 amplitude would be reduced for correct compared to incorrect trials (because individuals would have been more readily able to (pre-)activate that information). We also expected to see little influence of accuracy on unrelated trials. We were most interested in whether N400 amplitude would be reduced for correct vs. incorrect trials within the related condition, which would be expected if knowledge of the correct (appropriate) word ending was necessary to obtain related anomaly effects.

1.7.4 Summary of goals

Taken together, the series of experiments described in this thesis can provide a strong foundation for studying the contribution of individual differences in knowledge to real-time linguistic processing. We believe that this work should bridge an existing gap between studies of individual differences in knowledge/expertise in many domains of cognition and the cognitive neuroscience of real-time language comprehension. It may seem obvious that what an individual knows should have an impact on how they understand the world around them, including how they understand language, yet empirical demonstrations of whether (and if so, when and how) domain knowledge quickly influences word-by-word comprehension have not been forthcoming. Here, we therefore explicitly and empirically test the assumption often implied in the literature that an individual's knowledge should have an immediate impact on real-time word comprehension. We then go beyond this to begin to test the boundaries of the conditions under which domain knowledge can influence real-time semantic processing and to probe some of the types of information that domain knowledge is likely to modulate, including featural and conceptual information that is relevant for understanding words in context.

1.8 References

Altmann, G.T.M. & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*, 79-87. doi: 10.1016/S0010-0277(99)00059-1

Amsel, B.D., DeLong, K.D., & Kutas, M. (2015). Close, but no garlic: Perceptuomotor and event knowledge activation during language comprehension. *Journal of Memory and Language*, *82*, 118-132. doi: 10.1016/j.jml.2015.03.009

Arbuckle, T.Y., Vanderleck, V.F., Harsany, M., & Lapidus, S. (1990). Adult age differences in memory in relation to availability and accessibility of knowledge-based schemas. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 305-315.

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. doi; 10.1016/j.jml.2007.12.005

Bechtold, L., Bellebaum, C., Egan, S., Tettamanti, M., & Ghio, M. (2019). The role of experience for abstract concepts: Expertise modulates the electrophysiological correlates of mathematical word processing. *Brain and Language*, *188*, 1-10.

Beier, M.E. & Ackerman, P.L. (2005). Age, Ability, and the Role of Prior Knowledge on the Acquisition of New Domain Knowledge: Promising Results in a Real-World Learning Environment. *Psychology and Aging*, *20*(2), 341-355.

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

Borovsky, A. & Creel, S. (2014). Children and adults integrate talker and verb information in online processing. *Developmental Psychology*, *50*(5), 1600-1613.

Bransford, J.D. & Johnson, M.K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 11, 717-726.

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-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*, *27*(1), 62-89.

Chi, M.T.H. (2006). Laboratory methods for assisting experts' and novices' knowledge. In Ericsson et al. (eds.), *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge, UK: Cambridge University Press, pp. 167-184.

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152.

Chi, M.T.H. & Ohlsson, S. (2005). Complex declarative learning. In Holyoak, K.J. and Morrison, R.G. (Eds.), *The Cambridge Handbook of Thinking and Reasoning*. Cambridge, UK: Cambridge University Press, pp. 371-399.

Chiesi, H.L., Spilich, G.J., & Voss, J.F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*, 257-273.

Collins, A.M. & Quillian, M.R. (1969). Retrieval time from semantic memory. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240-247.

Coronel, J.C. & Federmeier, K.D. (2016). The N400 reveals how personal semantics is processed: Insights into the nature and organization of self-knowledge. *Neuropsychologia*, *84*, 36-43.

Daneman, M. & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Memory and Verbal Behavior*, *19*(4), 450-466. doi: 10.1016/S0022-5371(80)90312-6

de Groot, A.D. (1965). *Thought and choice in chess*. Mouton, The Hague. doi: 10.5117/9789053569986

DeLong, K.A., Chan, W., & Kutas, M. (2018). Similar time courses for word form and meaning pre-activation during sentence comprehension. *Psychophysiology*, e13312. doi: 10.1111/psyp.13312

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

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

diSessa, A.A. (1993). Towards an epistemology of physics. *Cognition and Instruction*, *10*(2-3), 105-225. doi: 10.1080/07370008.1985.9649008

Elman, J.L. (2011). Lexical knowledge without a lexicon? *The Mental Lexicon*, *6*(1), 1-33.

Ericsson, K.A., Charness, N., Feltovich, P.J., Hoffman, R.R. (Eds). (2006). *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge, UK: Cambridge University Press.

Ericsson, K.A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211-245.

Ericsson, K.A. & Moxley, J.H. (2014). Experts' superior memory: From accumulation of chunks to building memory skills that mediate improved performance and learning. In T. J. Perfect & D. S. Lindsay (Eds.), *SAGE handbook of applied memory* (pp. 404-420). London, UK: Sage Publishing.

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

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

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

Filik, R. & Leuthold, H. (2013). The role of character-based knowledge in online narrative comprehension: Evidence from eye movements and ERPs. *Brain Research*, *1506*, 94-104.

Fischler, I., Bloom, P.A., Childers, D.G., Arroyo, A.A., & Perry, N.W. (1984). Brain potentials during sentence verification: Late negativity and long-term memory strength. *Neuropsychologia*, *22*(5), 559-568.

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

Foss, D.J. (1982). A discourse on semantic priming. *Cognitive Psychology*, *14*, 590-607.

Frank, S.L., Otten, L.J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain & Language*, *140*, 1-11. doi: 10.1016/j.bandl.2014.10.006

Frank, S.L. & Willems, R.M. (2017). Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. *Language, Cognition, and Neuroscience*, *32*(9), 1192-1203. doi: 10.1080/23273798.2017.1323109

Friedman, D. (1990). Cognitive event-related potential components during continuous recognition memory for pictures. *Psychophysiology*, *27*(2), 136-148.

Gobet, F., Lane, P.C.R., Croker, S., Cheng, P. C-H., Jones, G., Oliver, I., & Pine, J.M. (2001). Chunking mechanisms in human learning. *TRENDS in Cognitive Sciences*, *5*(6), 236-243.

Goodyear, D. (2019, February). Succulent-smugglers descend on California. *The New Yorker*, Retrieved from https://www.newyorker.com/news/california-chronicles/succulent-smugglers-descend-on-california.

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371-395.

Guida, A., Gobet, F., Tardieu, H., Nicolas, S. (2012). How chunks, long-term working memory and templates offer a cognitive explanation for neuroimaging data on expertise acquisition: A two-stage framework. *Brain and Cognition*, *79*, 221-244.

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

Hatano, G. & Osawa, K. (1983). Digit memory of grand experts in abacus-derived mental calculation. *Cognition*, *15*, 95-110.

Hess, D.J., Foss, D.J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, *124*(1), 62-82.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, and Evolution*. Oxford: Oxford University Press.

Johnson-Laird, P. N. (1983). *Mental models: towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159-201.

Kim, A.E., Oines, L., & Miyake, A. (2018). Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language comprehension: An ERP investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(3), 406-420.

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

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

Kutas, M. & Federmeier, K.D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(14), 1-27. doi: 10.1146/annurev.psych.093008.131123

Kutas, M. & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203-205. doi: 10.1126/science.7350657

Kutas, M. & Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(12), 161-163. doi: dx.doi.org/10.1038/307161a0

Kutas, M., Lindamood, T.E., & Hillyard, S.A. (1984). Word expectancy and event-related brain potentials during sentence processing. In: S. Kornblum & J. Requin (Eds.), *Preparatory States and Processes*, 217-237. Hillsdale, New Jersey: Lawrence Erlbaum.

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

Laszlo, S. & Federmeier, K.D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology*, *48*, 176-186.

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

Long, D.L. & Prat, C.S. (2002). Memory for *Star Trek*: The role of prior knowledge in recognition revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1073-1082.

Long, D.L., Prat, C., Johns, C., Morris, P., & Jonathan, E. (2008). The importance of knowledge in vivid text memory: An individual-differences investigation of recollection and familiarity. *Psychonomic Bulletin & Review*, *15*(3), 604-609.

Matsuki, K., Chow, T., Hare, M., Elman, J.L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 913-934.

McCarthy, G., Nobre, A.C., Bentin, S., & Spencer, D.D. (1995). Languag-related field potentials in the anterior-medial temporal lobe: I. Intracranial distribution and neural generators. *The Journal of Neuroscience*, *15*(2), 1080-1089.

McKoon, G. & Ratcliff, R. (1998). Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology*, *49*, 25-42.

McNamara, D.S. & McDaniel, M.A. (2004). Suppressing irrelevant information: Knowledge activation or inhibition? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 465-482. doi: 10.1037/0278-7393.30.2.465

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. doi: 10.1016/j.jml.2012.01.001

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In *Proceedings of Workshop at ICLR*.

Mikolov, T., Sutskever, I, Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances on Neural Information Processing Systems*.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, *63*(2), 81-97.

Myers, J.L. & O'Brien, E.J. (1998). Accessing the discourse representation during reading. Memory & Cognition, 26, 1313-1322.

Nakano, H., Saron, C., & Swaab, T. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, *22*(12), 2886-2898. doi: 10.1162/jocn.2009.21400

Nieuwland, M.S. & Van Berkum, J.J.A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. Journal of Cognitive Neuroscience, 18(7), 1098-111.

Osterhout, L. & Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785-806.

Paczynski, M., & Kuperberg, G.R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, *67*, 426-448.

Pakulak, E. & Neville, H. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, *22*(12), 2728-2744. doi: 10.1162/jocn.2009.21393

Renoult, L., Tanguay, A., Beaudry, M., Tavakoli, P., Rabipour, S., Campbell, K., Moscovitch, M., Levine, B., & Davidson, P.S.R. (2016). Personal semantics: Is it distinct from episodic and semantic memory? An electrophysiological study of memory for autobiographical facts and repeated events in honor of Shlomo Bentin. *Neuropsychologia*, *83*, 242-256.

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

Rosch, E., Mervis, C.B., Gary, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). *Cognitive Psychology*, *8*, 382-439.

Schank, R.C. & Abelson, R.B. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, N.J.: Erlbaum.

Schneider, W., Körkel, J., & Weinert, F. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology*, *81*(3), 306-312.

Simon, H.A. (1974). How big is a chunk? *Science*, *183*(4124), 482-488.

Simon, H.A. & Chase, W.G. (1973). Skill in chess. *American Scientist*, *61*(4), 394-403. doi: 10.1007/978-1-4757-1968-0_18

Smith, N.J. & Kutas, M. (2015). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, *52*, 157-168.

Soederberg Miller, L.M. (2001). The effects of real-world knowledge on text processing among older adults. *Aging Neuropsychology and Cognition*, *8*(2), 137-148.

Soederberg Miller, L.M. (2009). Age differences in the effects of domain knowledge on reading efficiency. *Psychology and Aging*, *24*(1), 63-74.

Soederberg Miller, L.M., Cohen, J.A., & Wingfield, A. (2006). Contextual knowledge reduces demands on working memory during reading. *Memory & Cognition*, *34*(6), 1355-1367.

Soederberg Miller, L.M. & Stine-Morrow, E.A.L. (1998). Aging and the effects of knowledge and on-line reading strategies. *Journal of Gerontology: Psychological Sciences*, *53B*(4), 223-233.

Spilich, G.J., Vesonder, G.T., Chiesi, H.L., & Voss, J.F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*, 275-290.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632-1634.

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(2), 367-382.

Tanner, D. & Van Hell, J.G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*, 289-301.

Thornhill, D.E. & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*, 382-392. doi: 10.1016/j.ijpsycho.2011.12.007

Van Berkum, J.J.A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science*, *20*(9), 1092-1099.

Van Berkum, J.J.A, van den Brink, D., Tesink, C.M.J.Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580-591.

van Dijk, T.A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

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

Van Petten, C. (2014). Examining the N400 semantic context effect item-by-item: Relationship to corpus-based measures of word co-occurrence. *International Journal of Psychophysiology*, *94*, 407-419.

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

Van Petten, C. & Luka, B.J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*, 176-190. doi: 10.1016/j.ijpsycho.2011.09.015

Van Petten, C., Weckerly, J., McIsaac, H.K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science*, *8*(3), 238-242.

Verhagen, V., Mos, M., Backus, A., & Schilperoord, J. (2018). Predictive language processing revealing usage-based variation. *Language and Cognition*, *10*, 329-373. doi: 10.1017/langcog.2018.4

Walker, C.H. (1988). Relative importance of domain knowledge and overall aptitude on acquisition of domain-related information. *Cognition and Instruction*, *4*(1), 25-42.

Zirnstein, M., van Hell, J.G., & Kroll, J.F. (2018). Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition*, *176*, 87-106.

CHAPTER TWO: HARRY POTTER AND THE CHAMBER OF *WHAT?:* THE IMPACT OF WHAT INDIVIDUALS KNOW ON WORD PROCESSING DURING READING

2.1 Abstract

During reading, effects of contextual support indexed by N400—a brain potential sensitive to semantic activation/retrieval—amplitude are presumably mediated by comprehenders' world knowledge. Moreover, variability in knowledge may influence the contents, timing, and mechanisms of what is brought to mind during real-time sentence processing. Since it is infeasible to assess the entirety of each individual's knowledge, we investigated a limited domain—the narrative world of Harry Potter (HP). We recorded event-related brain potentials while participants read sentences ending in words more/less contextually supported. For sentences about HP, but not about general topics, contextual N400 effects were graded according to individual participants' HP knowledge. Our results not only confirm that context affects semantic processing by ~250 ms or earlier, on average, but empirically demonstrate what has until now been assumed—that N400 context effects are a function of each individual's knowledge, which here is highly correlated with their reading experience.

2.2 Introduction

Language comprehenders rapidly use many sources of contextual information to make sense of written and spoken words in sentences. For example, a word's processing is facilitated when it is preceded by a supportive (vs. unsupportive) sentence or discourse context, as indicated by electrophysiological and behavioral measures (e.g., Kutas & Hillyard, 1980; Altmann & Kamide, 1999). Word processing is also facilitated by non-linguistic context, such as a co-present visual scene (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Altmann & Kamide, 1999) or speaker identity (Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008; Borovsky & Creel, 2014; see Brown-Schmidt, Yoon, & Ryskin, 2015, for a review).

A common method for estimating the strength of contextual support for a specific word is offline cloze probability norming, wherein a group of participants provide continuations for sentence fragments. The cloze probability of a word is defined as the proportion of participants who provide that word for a given context. A word's cloze probability is inversely correlated with the amplitude of a centro-parietal, negative-going event-related brain potential (ERP) between ~300-500 ms referred to as the N400 (Kutas & Hillyard, 1980, 1984; DeLong, Urbach, & Kutas, 2005; Wlotko & Federmeier, 2012). Empirically, large N400 amplitude appears to be part of the "default" response to words, with N400 amplitude reductions occurring for words in later, vs. earlier, sentence positions, taken to reflect greater ease of access with contextual accrual (Van Petten & Kutas, 1990), and for words semantically related to a sentence context and/or a predictable (though never encountered) sentence continuation (Kutas & Hillyard, 1984; Federmeier & Kutas, 1999; Metusalem, Kutas, Urbach, Hare, McRae, & Elman, 2012; Amsel, DeLong, & Kutas, 2015). Although accounts of the functional significance of N400 effects differ (see Kutas & Federmeier, 2011 and Kuperberg, 2016 for recent reviews), most seem to agree that variation in N400 amplitude reflects aspects of semantic processing.

As individuals process a word in context, contextual information presumably activates world knowledge, which is rapidly combined to form higher-level representations, allowing individuals to incrementally update their interpretation of the sentence and to activate, or pre-activate, a current or upcoming word's meaning (e.g., DeLong et al., 2005; Boudewyn, Long, & Swaab, 2015; see Kuperberg, 2016, for a recent discussion). On such an assumption, the relationship between a word's contextual support and N400 amplitude must be mediated by an individual's world knowledge.

Consistent with this assumption, Hagoort and colleagues (Hagoort, Hald, Bastiaansen, & Petersson, 2004) found that violations of culturally-specific knowledge elicited N400 potentials of similar scalp topography, timing, and magnitude to those of more general semantic violations. When Dutch participants read sentences such as '*The Dutch trains are {yellow / white / sour}*,' words like '*yellow*' (a supported continuation) elicited reduced N400 activity (from 250-550 ms at centro-parietal electrodes) compared to '*sour*' (unsupported), replicating many studies. However, the N400 response to words like '*white*'—a word inconsistent with culturally-specific knowledge that Dutch trains are yellow—elicited an N400 indistinguishable from that of '*sour*.' These findings, among others (e.g., Hald, Steenbeek-Planting, & Hagoort, 2007; Filik & Leuthold, 2013) underscore that world knowledge, gleaned from actual experience in the world, determines aspects of processing reflected in N400 amplitudes.

Even within the same cultural context, however, different individuals know different things and to varying extents. Consequently, measures of contextual support like cloze probability computed over groups of participants do not necessarily correspond to the contents of any given individual's knowledge or the likelihood that they entertained a given word or concept during moment-by-moment processing (see Verhagen, Mos, Backus, & Schilperoord, 2018, for discussion). Moreover, ERPs themselves are typically averaged over a group of participants and are not generally linked to individual participants' knowledge (though see Coronel & Federmeier, 2016, who established a relationship between knowledge of personal preferences and N400 effects at the individual level).

Across many domains, the content of an individual's knowledge can influence performance on memory tasks (e.g., de Groot, 1965; Simon & Chase, 1973), text comprehension (e.g., Chiesi, Spilich, & Voss, 1979; Spilich, Vesonder, Chiesi, & Voss, 1979), and other cognitive tasks (see

Chi, 2006, for a review). Yet, although variability in individuals' experiences and, therefore, knowledge seems likely to influence real-time processing (as argued by Kutas, 2006), few studies have examined the link between level of content knowledge and real-time sentence processing. More typically, the experimental focus has been on how differences in more generalised cognitive processing abilities, such as verbal working memory (WM) and cognitive control (Münte, Schiltz, & Kutas, 1998; Nakano, Saron, & Swaab, 2010; Boudewyn, Long, & Swaab, 2012), or language-specific abilities, such as language proficiency (McLaughlin, Osterhout, & Kim, 2004; McLaughlin, Tanner, Pitkänen, Frenck-Mestre, Inoue, Valentine, & Osterhout, 2010; Pakulak & Neville, 2010; Tanner, Inoue, & Osterhout, 2014), relate to individual differences in aspects of sentence processing (see Boudewyn, 2015, for a review).

Clearly a major hurdle for investigating individual differences in knowledge is the infeasibility of capturing all of an individual's knowledge using standard laboratory procedures. As an approximation, measures of print exposure and general knowledge (Stanovich & West, 1989; Stanovich & Cunningham, 1993) have been tentatively linked to real-time brain potential measures of sentence processing (e.g., Metusalem et al., 2012); however, these provide only coarse approximations of general knowledge and not precise topics or domains where individuals may vary in how much they know.

A potential solution may be to restrict knowledge to a single domain. In the vast literature on expertise and expert behaviour, researchers interested in how level of knowledge impacts perception and cognition have focused on specific domains of knowledge—e.g., chess (de Groot, 1965; Simon & Chase, 1973), physics (e.g., Chi, Feltovich, & Glaser, 1981), children's knowledge of dinosaurs or spiders (e.g., Pearson, Hansen, & Gordon, 1979; Gobbo & Chi, 1986). Few, however, lend themselves easily to the study of real-time language processing, which requires a domain with a rich set of verbal descriptions. Moreover, based on a relatively low signal-to-noise ratio, ERP studies typically require a number of linguistic stimuli on the order of a hundred or more per study.

To mitigate these challenges, we zeroed in on the narrative world of Harry Potter, based on the book series by J.K. Rowling. This domain is linguistically rich, making it an excellent choice for studying the interface of knowledge and language. In addition, it includes many novel organizational structures—including new categories, like magical creatures and spells, as well as multi-faceted events. At the same time, it is a constrained domain, allowing us to estimate each individual's level of HP knowledge. Finally, as present-day college students grew up with this series, they constitute a pool of participants who are likely to naturally vary in their HP knowledge.

Using this tractable domain of knowledge, we asked whether or not, and, if so, to what extent and when individual differences in knowledge about HP would specifically influence processing of sentences about the domain of HP (and not general topics). We therefore recorded ERPs while participants read sentences about general topics (control sentences) and Harry Potter (HP sentences) that ended in contextually supported or unsupported words (Table 2.1). These participants also completed offline questionnaires that assessed their knowledge about Harry Potter.

Many studies seem to assume that real-world knowledge at the individual participant level modulates the effect of contextual support on the earliest stages of semantic processing, as captured during the N400 time period. However, no study has directly tested the assumption that specific knowledge, at the individual-subject level, modulates the effect of contextual support. Here, we explicitly tested this assumption by using a single domain of knowledge to carefully manipulate the availability of knowledge on a subject-by-subject basis. We predicted that HP knowledge would specifically influence the size of N400 context effects in HP, but not control, sentences. Such a finding would provide direct evidence that an individual's level of knowledge of a domain modulates the influence of context during the earliest stages of semantic processing.

As manipulations that affect N400 amplitude often influence the size of so-called "post-N400" positivities (PNPs) at times, we also examined a later time period. Depending on the nature of the linguistic manipulation, words which are unsupported in context may elicit parietal and/or frontal positivities (for reviews, see Van Petten & Luka, 2012; Brouwer, Fitz, & Hoeks, 2012; DeLong, Troyer & Kutas, 2014). We therefore had reason to suspect that there might be late positivities of larger amplitude for unsupported (compared to supported) critical words in the control sentences, and, for individuals knowledgeable about HP (and therefore sensitive to the contextual support manipulation), in the HP sentences.

Table 2.1. Sample experimental stimuli.

Control Sentences		
Sentence frame	Supported	Unsupported
We had been watching the blue jay for days. The bird laid her eggs in the	nest	yard
The vampire moved in. He bit his victim on the	neck	shoulder
Alicia's first client was a failure. But her second was a	success	triumph
Harry Potter Sentences		
Sentence frame	Supported	Unsupported
The character Peter Pettigrew changes his shape at times. He takes the form of a	rat	dog
There are two Beaters on every Quidditch team. Their job is to protect their team from	Bludgers	Spellotape
Wizards are able to conjure the Dark Mark. They can use a spell called	Morsmordre	Stupefy

2.3 Methods

2.3.1 Participants

41 undergraduate students / members of the UCSD community (mean age = 20, range = 18-24; 26 women, 15 men) took part in the study for partial course credit or payment of $9 / hour. Of these, one participant was excluded from data analysis due to excessive eye movements. All participants provided informed consent reviewed by the Institutional Review Board at the University of California San Diego. To ensure that some participants would have high knowledge of the Harry Potter domain, a subset (n = 12) was recruited via an announcement that specifically required having read all seven Harry Potter books and/or seen all eight Harry Potter films.

2.3.3 Materials

2.3.3.1 Sentence materials

During the EEG portion of the experiment, participants read three blocks of control sentence pairs followed by three blocks of Harry Potter (HP) sentence pairs. The final word of the second sentence was always the critical word, which was either supported or unsupported by the context. Two lists were created; each participant only read one version of each sentence pair.

2.3.3.1.1 Control sentences

108 control sentence pairs (first sentence ranging 3-18 words, mean = 7; second ranging 4-10 words, mean = 7) described everyday topics and events. All sentence pairs were highly constraining (mean cloze of best completion = 94%; range = 87-100%). For control sentences, supported words were defined as the best completion. To create unsupported words, plausible continuations were selected that were semantically related to the best completion but were never produced during cloze norming.

2.3.3.1.2 Harry Potter sentences

108 Harry Potter (HP) sentence pairs were constructed as follows. Using freely available materials (including Wikipedia and Harry Potter fan sites) along with the text of the Harry Potter books, the first author created a set of single sentences that accurately described events and entities from the series. The final, “supported” word of these sentences was designed to be 100% predictable given perfect knowledge of the book series. To verify that this was the case, a norming study was conducted on a separate group of participants. This group included some participants who were highly knowledgeable about the world of Harry Potter (determined by a trivia quiz; see “Harry Potter Quiz” section below). 32-34 participants provided a final word for each sentence. To be included in the study, a sentence needed to be completed with the supported word by a minimum of 65% of the most knowledgeable respondents (that is, those who scored in the top

quartile on the HP knowledge quiz). Across all norming participants, mean cloze for supported words was 51% (range = 26-84%).

These single sentences were then broken into two sentences so that the first could be presented all at once. On average, the first sentence was 9 words long (range = 4 to 18 words); the second was 7 words long (range = 3 to 12 words). The second sentence ended in a critical word that either was supported by the context (see above) or was unsupported. To create unsupported endings, supported words were replaced by words that defied "ground truth" from the HP stories but seemed similarly plausible (for those with little to no HP knowledge). To achieve this, we used words that were from a the same/a similar category. For example, a standard English word describing an animal, like "rat," was replaced with another animal, "dog"; a magical object specific to HP like "Bludgers" was replaced with another magical object specific to HP, "Spellotape"; and an HP-specific proper name like "Kreacher" was replaced with another HP-specific proper name like "Hermes" (see these and other examples in Table 2.1). After data collection, we discovered that two HP sentence pairs contained factual errors in both the supported/unsupported versions. These materials were dropped from subsequent analyses for a total of 106 HP items and 108 Control items.

2.3.3.2 Memory tests

Immediately after the EEG portion of the experiment, participants completed two memory quizzes, one for control sentences and one for HP sentences. The primary purpose of the memory tests was to establish that participants had paid attention when reading during the EEG experiment. Participants were asked to circle the words they remembered seeing as a final word of the second sentences from the experiment—first for control sentences and then for HP sentences. Each quiz contained a total of 90 words—30 new, 30 critical words from supported contexts, and 30 critical words from unsupported contexts.

2.3.3.3 Additional tasks and measures

2.3.3.3.1 Overview of tasks

We collected several other measures of individual differences besides HP knowledge to better understand any group differences between individuals with high vs. low HP knowledge. We developed a measure of Harry Potter experience (self-report questionnaire) and collected other measures (see Appendix, in the supplementary materials) including general print/reading experience (media and reading habits questionnaire (MRH) and author and magazine recognition tests (ART/MRT), Stanovich & West, 1989); measures of general knowledge (a general knowledge trivia quiz (GKQ) that we developed from freely available materials and cultural knowledge checklists (CLC/MCLC, Stanovich & Cunningham, 1993)); vocabulary (PPVT, Dunn & Dunn, 2007), and verbal WM (sentence span, Daneman & Carpenter, 1980). Finally, we administered a debriefing questionnaire.

2.3.3.3.2 HP self-report questionnaire

We asked participants questions about their experience with the Harry Potter book series and related materials, e.g., how many times they had read each of the HP books, seen each of the HP movies, and additional ways in which they might have engaged with Harry Potter. As an estimate of overall experience with Harry Potter, a numeric score was determined by summing the total number of times an individual had read each book, seen each movie, and so on. In addition, we report statistics on the raw number of times participants read each book, on average (e.g., if an individual read the first book 3 times and each other book just once, their score would be 1.286).

2.3.3.3.3 Harry Potter quiz

We estimated participants' knowledge of Harry Potter from their score on a trivia-style quiz containing ten multiple choice questions; for example, *To gain access to the kitchens, one must tickle the following fruit: (a) Pear, (b) Orange, (c) Grape, (d) Banana*. HP quiz score (henceforth referred to as "HP knowledge") was the number of correct answers out of ten. For regression analyses, we z-transformed these scores.

2.3.3.3.4 Aggregate measures

An aggregate measure of reading experience was based on an average of z-transformed ART scores, MRT scores, total MRH score, and number of favorite authors listed on the MRH. An aggregate measure of general knowledge of common topics was based on an average of z-transformed CLC scores, MCLC scores, and general knowledge test scores.

2.3.3.3.5 Debriefing questionnaire

On a debriefing questionnaire, many participants indicated they noticed that sentences about Harry Potter were sometimes inaccurate. This was to be expected, as half of the sentences were designed to be inaccurate portrayals of "ground truth" based on the HP book series, and all participants had at least minimal knowledge of the HP series. After observing this trend, we asked all but the first two participants to complete additional debriefing questions, estimating approximately how many sentences they thought were true/accurate, and, of these, how many they thought they had known ahead of time. Participants reported that 60% of the Harry Potter sentences had been true (range = 30-100%). Of these, participants reported that they had known an average of 64% (range = 0-100%).

2.3.4 Procedures

2.3.4.1 Ordering of tasks

During set-up for the EEG experiment, participants completed the ART and MRT. After EEG recording, participants completed the memory tests followed by other questionnaires, with the order corresponding to that of their description in the preceding section.

2.3.4.2 EEG experiment

Before the study, participants were asked to remain relaxed and still to minimize muscle artifact. They were told they would be reading short, two-sentence stories (first three blocks about general topics, then three blocks about the world of Harry Potter) for meaning and that they would be asked questions about what they read at the end of the EEG recording session. Participants then read four practice items.

During the EEG experiment, participants sat approximately 100 cm in front of a cathode-ray tube monitor. The background of the screen was black and words were presented in white type. Each trial began with a blank screen for two seconds. Then, the first sentence of each pair appeared on the screen until the participant pressed a button to advance to the next sentence. After their button press, a crosshair appeared in the center of the screen for a duration which varied randomly between 1050 and 1450 ms. Participants were instructed to focus on the crosshair and not to move their eyes or blink while it was on the screen. The second sentence was then presented one word at a time right above the crosshair. Each word was presented for 200 ms with an interstimulus interval of 300 ms. After the sentence-final word disappeared, the crosshair stayed on the screen for a duration that randomly varied between 750 and 1150 ms. Control sentences were presented across three blocks, with short breaks in between, followed by three blocks of HP sentences. Within each block of the study, sentences with supported and unsupported endings were randomly interspersed.

2.3.4.3 EEG recording

The electroencephalogram (EEG) was recorded from 26 electrode sites arranged geodesically in an Electro-cap (as described in Ganis, Kutas, & Sereno, 1996; see Fig. 9). For all cap electrodes, online recording was referenced to the left mastoid; these electrodes were re-referenced offline to an average of the left and right mastoid. Electrodes were placed lateral to the outer canthus of each eye to create a bipolar recording used to monitor eye movements. Electrodes placed under each eye were referenced to the left mastoid and were used to monitor blinks. Throughout the experiment, all electrode impedances were maintained under 5 kΩ. The signal was amplified with Grass amplifiers which were set at a bandpass of .01 to 100 Hz; the sampling rate was 250 Hz.

2.3.4.4 EEG data analysis

Trials contaminated by eye movements, blinks, muscle activity, blocking, or other artifact were removed from subsequent analysis. This resulted in an exclusion of 17% of trials: HP-Supported: 17%; HP-Unsupported: 17%; Control-Supported: 18%; Control-Unsupported: 17%. ERPs were created by averaging from 200 ms before the onset of a critical word until 900 ms post-critical word. Then, for each electrode, a baseline was computed by averaging potentials from 200 ms before the word to the start of the word; this baseline was subtracted from the waveform.

Because our study is the first to directly compare ERPs to words in sentences about general topics with those in sentences about a fictional, narrative world, it was important to characterize overall influences of contextual support and sentence type across all participants. We therefore conducted a traditional, whole-head analysis prior to examining individual differences based on HP knowledge and other covariates. Of primary interest was a time period surrounding the typical peak of the N400 brain potential (~375 ms; e.g., Federmeier & Kutas, 1999) from 250 ms to 500 ms post-stimulus; we also examined a late positivity time period from 500-750 ms post-stimulus.

We subjected mean amplitudes of the ERP waveforms in these time periods to a whole-head ANOVA to determine effects of sentence and/or ending type across all participants, including repeated measures of electrode (26 levels), ending type (2 levels: supported/unsupported), and sentence type (2 levels: HP/control) as well as a between-subjects factor of list (2 levels). For all ANOVAs, we applied the Greenhouse-Geisser epsilon correction for *F*-tests with more than one degree of freedom in the numerator and report the corrected p-value, unadjusted degrees of freedom, and value of the Greenhouse-Geisser epsilon.

Our primary research questions hinged on whether and when HP knowledge interacted with contextual support. We therefore examined the relationship between HP knowledge, sentence type, and ending type in an ROI where N400 effects are typically largest, averaging mean amplitude between 250 and 500 ms across eight centro-parietal electrodes (MiCe, LMCe, RMCe, MiPa, LDPa, RDPa, LMOc, and RMOc) for each sentence and ending type. HP knowledge was defined as z-transformed performance on the HP knowledge quiz. For these analyses, we used hierarchical mixed-effects linear regression models (Baayen, Davidson, & Bates, 2008). All categorical fixed effects (sentence type, ending type) were sum coded ([-1, 1]). All models were fit to subject-averaged ERPs, with random intercepts for subject. Mixed effects models were implemented using the `lme4` (version 1.1-12; Bates, Maechler, Bolker, & Walker, 2015) and `lmerTest` (version 2.0-30; Kuznetsova, Brockhoff, & Christensen, 2017) packages in `R`. P-values were calculated using `lmerTest`, with the Satterthwaite option for denominator degrees of freedom for *F* statistics.

2.4 Results

2.4.1 Behavioral data

2.4.1.1 Memory task—recognition accuracy

On the control recognition test, participants correctly recognised an average of 15.26 out of 60 words (~25%) and false alarmed to an average of 2.08 out of 30 words (~7%). On the HP recognition test, participants correctly recognised an average of 29.69 out of 60 words (~49%) and false alarmed to an average of 2.28 out of 30 words (~8%). Participants were therefore able to discriminate between words they had and had not seen for both the Control and HP recognition tests.

For statistical analyses, we computed a *d*-prime sensitivity index for each participant and condition based on the false alarms for each recognition test (control, HP) and the number of items correctly recognised from each condition. We used a mixed-effects model to predict *d*-prime based on sentence type (control, HP), ending type (supported, unsupported), and HP knowledge (Supplementary Table 2.1). This model revealed a significant effect of sentence type, indicating higher accuracy for HP compared to control sentences. In addition, the interaction between HP knowledge and sentence type was significant. HP knowledge was correlated with *d*-prime for HP words ($r^2 = .136$, $p < .05$), but not for control words ($r^2 = .001$, n.s.).

2.4.1.2 Additional tasks

Table 2.2 reports descriptive statistics for scores on the HP knowledge quiz and other individual differences measures completed by participants. Intercorrelations among these measures are provided in Table 2.3.

Table 2.2. Mean, standard deviation, and range are provided for behavioural measures of individual differences.

	All participants			High HP group			Low HP group		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
HP Quiz	6.12	(2.33)	[2, 10]	8.67	(1.35)	[7, 10]	4.16	(0.96)	[2, 5]
# of HP Books	2.11	(3.75)	[0, 18]	4.70	(5.22)	[0.04, 18]	0.46	(0.55)	[0, 1.71]
HP Self Score	43.54	(41.32)	[4, 181.5]	75.92	(51.10)	[14, 181.5]	19.84	(11.07)	[4, 47]
ART	0.23	0.12	[0.02, 0.5]	0.29	0.10	[.15, .5]	0.17	0.11	[.02, .42]
MRT	0.22	0.08	[0.08, 0.38]	0.26	0.06	[.15, .38]	0.18	0.08	[.08, .35]
# of Authors Listed	2.60	(1.72)	[0, 5]	3.40	(1.18)	[1, 5]	1.74	(1.69)	[0, 5]
MRH Total	6.67	(3.08)	[0, 15]	7.67	(2.94)	[4, 15]	6.00	(3.06)	[0, 12]
GKQ	19.51	(3.62)	[10, 25]	21.33	(3.09)	[15, 25]	17.74	(3.41)	[10, 23]
CLC	0.35	(0.13)	[.07, .60]	0.41	(0.08)	[0.31, 0.60]	0.29	(0.13)	[0.07, 0.57]
MCLC	0.44	(0.18)	[0, .77]	0.51	(0.14)	[0.13, 0.77]	0.38	(0.20)	[0, 0.70]
PPVT	208.32	(6.70)	[195, 219]	212.07	(5.09)	[200, 219]	205.05	(6.17)	[196, 217]
Sentence Span	2.95	(0.64)	[1.5, 5]	2.77	(0.56)	[1.5, 3.5]	3.08	(0.75)	[2, 5]

Table 2.3. Intercorrelations (Pearson's *r*) among behavioral measures of individual differences. *r* values above .31 are significant at *a* = .05; *r* values above .403 are significant at *a* = .01.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 HP Quiz	1	.67	.73	.47	.51	.45	.15	.43	.39	.38	.48	-.13	.54	.47
2 HP Books	-	1	.85	.34	.32	.38	.07	.23	.33	.37	.18	-.04	.38	.36
3 HP Self Score	-	-	1	.43	.34	.46	.12	.25	.31	.31	.21	-.12	.46	.34
4 ART	-	-	-	1	.54	.38	.45	.42	.67	.54	.30	.13	.81	.64
5 MRT	-	-	-	-	1	.25	.22	.59	.57	.55	.65	.04	.69	.67
6 Authors Listed	-	-	-	-	-	1	.42	.49	.50	.12	.23	.12	.70	.43
7 MRH Total	-	-	-	-	-	-	1	.31	.42	.13	.20	.03	.72	.33
8 GKQ	-	-	-	-	-	-	-	1	.69	.50	.72	.19	.62	.85
9 CLC	-	-	-	-	-	-	-	-	1	.63	.50	.30	.74	.90
10 MCLC	-	-	-	-	-	-	-	-	-	1	.50	.23	.46	.83
11 PPVT	-	-	-	-	-	-	-	-	-	-	1	.12	.47	.67
12 Sentence Span	-	-	-	-	-	-	-	-	-	-	-	1	.11	.28
13 Reading Experience	-	-	-	-	-	-	-	-	-	-	-	-	1	.71
14 General Knowledge	-	-	-	-	-	-	-	-	-	-	-	-	-	1

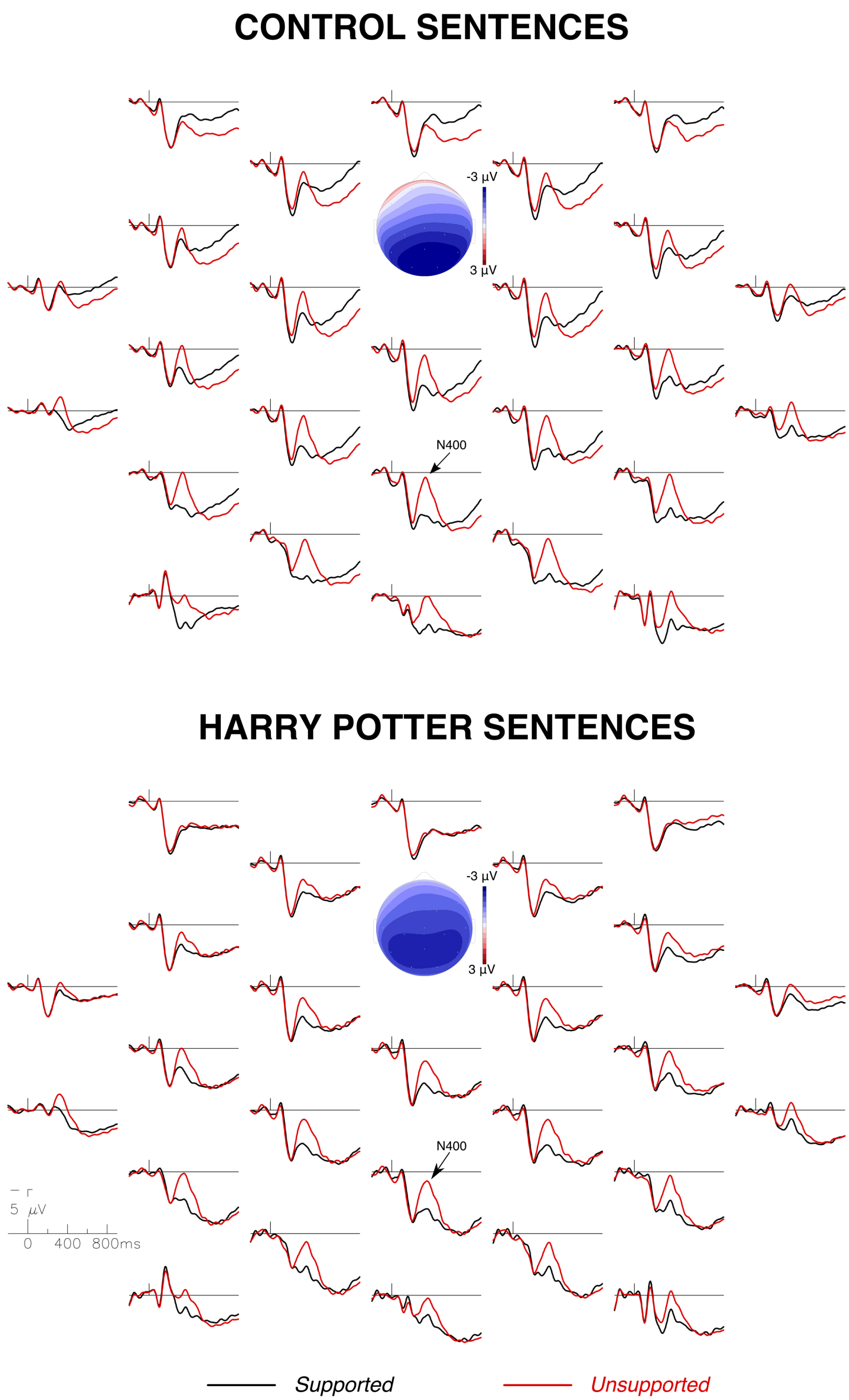


Figure 2.1. Grand average ERPs across all participants for critical words of each type (supported, unsupported) for control and HP sentences. Topographical scalp plots show the N400 effect of contextual support (unsupported minus supported) from 250 to 500 ms.

2.4.2 ERP data

Figure 1 shows the grand average ERPs for all participants across 26 scalp electrodes from 200 ms before the onset of the critical word to 900 ms post-critical word. Across most electrodes, ERPs to critical words in both control and HP sentences are characterised by two early sensory components, a negative-going peak around 100 ms (N1) and a positive-going peak around 200 ms (P2). Across all participants, for supported words, the P2 is followed by a positivity in the N400 time period (250-500 ms). For unsupported endings, the P2 is followed by a relative negativity in this same window.

Since we were specifically interested in the unique effects of HP knowledge on processing HP, compared to control, sentences, whole-head plots for the high-knowledge group (n = 15) and low-knowledge group (n = 19) for each sentence type are provided in Figure 2.2.

2.4.2.1 N400: 250-500 ms post-stimulus

2.4.2.1.1 Whole-head analyses

Results from the whole-head ANOVA for the N400 time period are provided in Table 2.4. As expected, there was a main effect of ending type, with supported endings leading to more positive-going waves (i.e., reduced negativities) than unsupported endings. Visual inspection of topographic scalp maps (Figure 1) revealed that an interaction between electrode and ending type was driven by a broadly distributed, centro-parietal N400 context effect. The distribution was roughly similar for both sentence types, though the context effect seemed somewhat more broadly distributed for HP, compared to control, sentences.[2]

---

[2] For N400 context effects, we followed up on interactions with electrode in a distribution analysis containing a subset of 16 electrodes (following the procedure in Kutas & Federmeier, 1999) which supported this interpretation. For both sentences types, N400 context effects were centro-posterior; for control, but not HP, sentences, the N400 context effect was somewhat right lateralized.

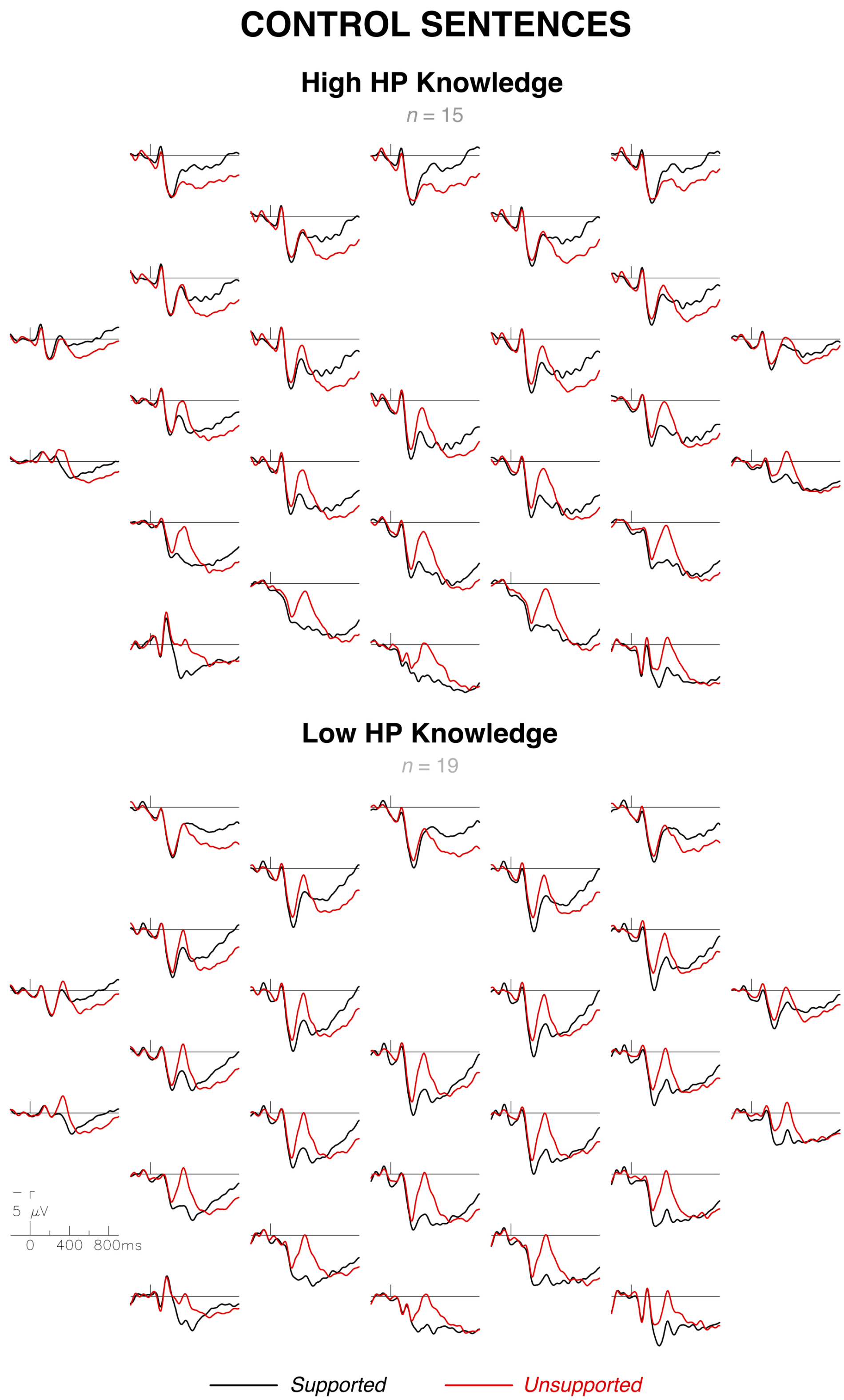


Figure 2.2a. Grand average ERPs for control sentences for high HP knowledge and low HP knowledge groups, assigned by a median (= 6) split on HP quiz score.

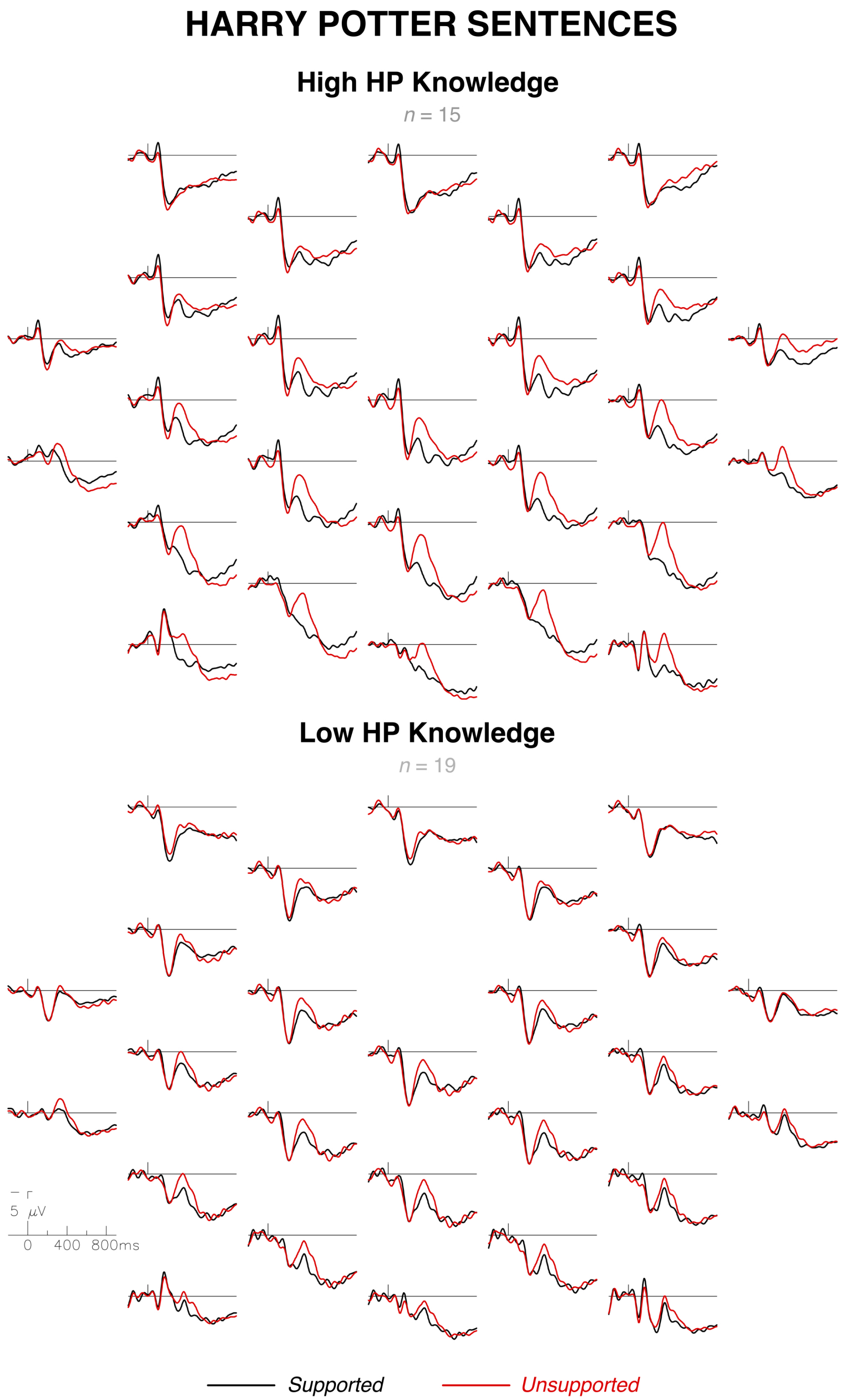


Figure 2.2b. Grand average ERPs for HP sentences for HP knowledge and low HP knowledge groups, assigned by a median (= 6) split on HP quiz score.

Table 2.4. Whole-head ANOVA results for N400 and late positivity time windows.

	DF	*F*	p-value	ε_{GG}
N400				
List	(1, 38)	0.027	.8699	
Electrode	(25, 950)	26.280	.0000	0.148
Ending Type	(1, 38)	42.625	.0000	
Sentence Type	(1, 38)	0.185	.6692	
List:Electrode	(25, 950)	0.343	.8345	0.148
List:Ending Type	(1, 38)	0.714	.4033	
List:Sentence Type	(1, 38)	0.276	.6021	
Electrode:Ending Type	(25, 950)	35.693	.0000	0.111
Electrode:Sentence Type	(25, 950)	5.347	.0004	0.168
Ending Type:Sentence Type	(1, 38)	0.307	.5828	
List:Electrode:Ending Type	(25, 950)	2.795	.0477	0.111
List:Electrode:Sentence Type	(25, 950)	0.382	.8306	0.168
List:Ending Type:Sentence Type	(1, 38)	9.470	.0039	
Electrode:Ending Type:Sentence Type	(25, 950)	7.704	.0001	0.126
List:Electrode:Ending Type:Sentence Type	(25, 950)	4.150	.0068	0.126

Table 2.4. Whole-head ANOVA results for N400 and late positivity time windows, Continued.

	DF	*F*	p-value	ε_{GG}
Late positivity				
List	(1, 38)	0.014	.9071	
Electrode	(25, 950)	27.437	.0000	0.115
Ending Type	(1, 38)	3.649	.0637	
Sentence Type	(1, 38)	4.252	.0461	
List:Electrode	(25, 950)	0.164	.9140	0.115
List:Ending Type	(1, 38)	1.819	.1854	
List:Sentence Type	(1, 38)	1.408	.2427	
Electrode:Ending Type	(25, 950)	3.463	.0237	0.106
Electrode:Sentence Type	(25, 950)	3.277	.0161	0.146
Ending Type:Sentence Type	(1, 38)	11.169	.0019	
List:Electrode:Ending Type	(25, 950)	1.456	.2343	0.106
List:Electrode:Sentence Type	(25, 950)	0.895	.4613	0.146
List:Ending Type:Sentence Type	(1, 38)	2.762	.1048	
Electrode:Ending Type:Sentence Type	(25, 950)	8.032	.0000	0.147
List:Electrode:Ending Type:Sentence Type	(25, 950)	3.085	.0211	0.147

Three- and four-way interactions with list suggest there may have been subtle differences in N400 amplitude based on which list participants saw. However, as the relevant values were in the same direction for each of the relevant comparisons (that is, unsupported values were more negative than supported values for each sentence type and for each of the two lists), we do not further pursue this point.

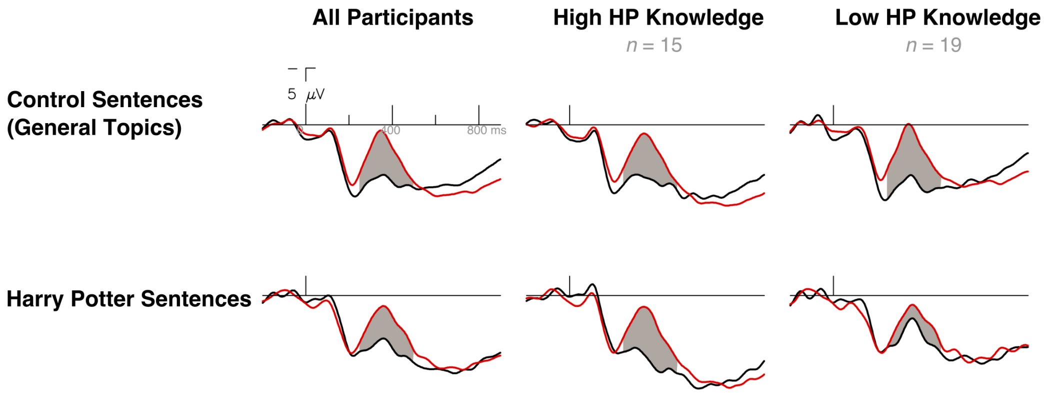


Figure 2.3. ERPs from a centro-parietal ROI for supported (black) and unsupported (red) endings to HP sentences. Shaded region from 250 to 500 ms shows N400 effect.

2.4.2.1.2 ROI analyses

Figure 2.3 shows ERPs from the centro-parietal ROI used in regression analyses. Results from a linear mixed-effects regression analysis are provided in Table 2.5. This regression confirmed the effect of ending type (i.e., contextual support) observed in the whole-head analysis. In addition, there was a significant interaction between sentence type and HP knowledge, with more knowledgeable individuals tending to have overall more positive-going N400s compared to less knowledgeable individuals. Critically, the three-way interaction between HP knowledge, sentence type, and ending type was significant. To follow up, we conducted planned analyses separately for each sentence type (Table 2.6).

Table 2.5. ROI analyses using linear mixed-effects models for N400 and late positivity time periods.

	Estimate	*SE*	*DF*	*T-value*	*Pr(>|t|)*
N400					
Intercept	3.515	0.322	38	10.914	.0000
HP knowledge	0.465	0.323	38	1.442	.1574
Sentence type	0.112	0.125	114	0.894	.3734
Ending type	1.167	0.125	114	9.309	.0000
HP knowledge:Sentence type	-0.264	0.126	114	-2.099	.0380
HP knowledge:Ending type	0.170	0.126	114	1.352	.1791
Sentence type:Ending type	0.067	0.125	114	0.531	.5964
HP knowledge:Sentence type:Ending type	-0.255	0.126	114	-2.032	.0445
Late positivity					
Intercept	5.619	0.425	38	13.214	.0000
HP knowledge	0.625	0.426	38	1.466	.1508
Sentence type	-0.393	0.147	114	-2.675	.0086
Ending type	-0.128	0.147	114	-0.869	.3864
HP knowledge:Sentence type	-0.166	0.147	114	-1.131	.2606
HP knowledge:Ending type	0.021	0.147	114	0.142	.8871
Sentence type:Ending type	-0.202	0.147	114	-1.374	.1721
HP knowledge:Sentence type:Ending type	0.021	0.147	114	0.139	.8894

Table 2.6. Follow-up ROI analyses using linear mixed-effects for HP and Control sentences during N400 time period.

	Estimate	*SE*	*DF*	*T-value*	*Pr(>|t|)*
Control sentences					
Intercept	3.627	0.358	38	10.144	.0000
HP knowledge	0.202	0.358	38	0.563	0.577
Ending type	1.233	0.175	38	7.046	.0000
HP knowledge:Ending type	-0.085	0.175	38	-0.487	0.629
HP Sentences					
Intercept	3.403	0.333	38	10.220	.0000
HP knowledge	0.729	0.334	38	2.185	.0351
Ending type	1.100	0.180	38	6.117	.0000
HP knowledge:Ending type	0.425	0.180	38	2.358	.0236

For control sentences, ending type was a significant predictor, but, critically, there were no effects of HP knowledge. For HP sentences, however, ending type, HP knowledge, and their interaction were all significant predictors. To further explore the relationship between N400 HP context effects and HP knowledge score, we conducted planned simple regressions predicting mean amplitude for supported, as well as unsupported, endings in this time period based on HP knowledge score, finding that this relationship seemed to be driven by variance in the supported (r^2 = .167, p < .01) but not unsupported (r^2 = .021, n.s.) endings (Figure 2.4).

Therefore, across the centro-parietal ROI, effects of contextual support for sentences about Harry Potter were sensitive to HP knowledge, but, critically, effects of contextual support for sentences about general topics were not. Moreover, this relationship reflected a difference in the brain's response to *supported*, not unsupported, endings.

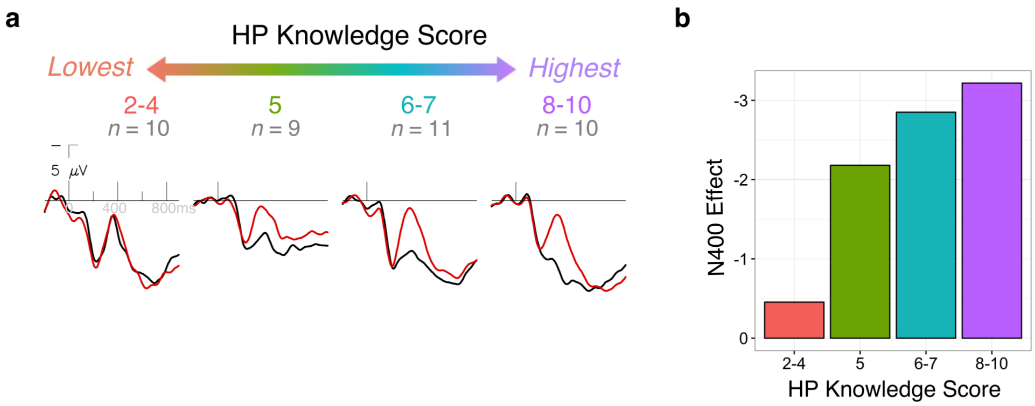


Figure 2.4. (a) ERPs from a centro-parietal ROI for supported (black) and unsupported (red) endings to HP sentences by HP knowledge subgroup. (b) N400 effect (unsupported minus supported) for each HP knowledge subgroup.

Next, we asked whether other differences between participants might modulate the influence of contextual support. We used mixed-effects linear regression to predict each mean amplitude based on four variables: HP knowledge, verbal WM, reading experience, and general knowledge, as well as their interaction with ending type (see Supplementary Table 2.2).

Confirming previous results, ending type was a significant predictor for both sentence types. For control sentences, none of the individual-differences measures nor their interaction with ending type were significant predictors. For HP sentences, however, the interaction between reading experience and ending type was a significant predictor. However, the aggregate measure of reading experience was strongly correlated with HP knowledge scores (at $r = .54$; see Table 2.3 and Supplementary Figure 2.1); this multicollinearity makes it impossible to fully dissociate effects of each predictor variable on the size of the effect of contextual support (Kutner, Nachtsheim, Neter, & Li, 2005). We return to this point in the discussion.

2.4.2.2 Late positivity: 500-750 ms post-stimulus

2.4.2.2.1 Whole-head analyses

Results from the whole-head ANOVA for the late positivity time period are provided in Table 2.4. A main effect of sentence indicated more positive-going waves for HP sentences compared to control sentence endings. An interaction of ending type × sentence type indicated a crossover effect such that the direction of the contextual support effect differed between the two sentences: whereas for control sentences, unsupported endings led to more positive-going waves compared to supported endings, the numerically reverse pattern obtained for HP sentences, with unsupported endings leading to less positive-going waves compared to supported endings. Based on visual inspection, the higher-order interactions with electrode seemed to reflect a frontal, left positivity greater for unsupported compared to supported endings for the control, but not HP, sentences.[3]

ROI analyses. A mixed-effects linear regression analysis predicting potentials from the centro-parietal ROI based on HP knowledge, sentence type, and ending type (Table 2.5) confirmed the effect of sentence type observed in the whole-head analysis, with more positive-going potentials for HP compared to control sentences. No other predictors were significant.

2.5 Discussion

2.5.1 Summary of findings

We used the narrative world of Harry Potter to ask whether level of domain knowledge could predict sentential context effects for individuals. We observed N400 context effects for sentences about Harry Potter, as well as for control sentences about general topics, but the size of the context effects—only for the HP sentences—was significantly predicted by an individual's

[3] For post-N400 context effects, we followed up on interactions with electrode in a distribution analyses containing a subset of 16 electrodes (as for our N400 context effects). These analyses confirmed that for control sentences, N400 context effects were left/frontal, while no context effects were present on the late positivity for HP sentences.

level of HP knowledge. These effects were observed during the earliest stages of semantic processing (between ~250-500 ms post-stimulus), and the relationship between context effects and knowledge seemed to be driven primarily by supported, not unsupported, words, suggesting greater ease of access to, or retrieval of, pertinent information for individuals with high knowledge scores. Indeed, differential experience in a domain can lead to differences not only in which "chunks"—i.e., "coherent patterns" (Chi, 2006, p. 181; cf. Miller, 1956; Simon, 1974)—are available to individuals but also in the depth of information that comes to mind when a cue (or chunk), like a picture or word, is processed (Chi, 2006). We speculate that our HP knowledge measure may act as a proxy for some combination of chunk size, depth of information access, or other differential organization of information.

2.5.2 N400 context effects

As expected based on the extensive N400 literature (Kutas & Federmeier, 2011), we observed N400 context effects for sentences ending in supported vs. unsupported words, whether about general topics or about Harry Potter, a well-known fictional domain with which many young adults are familiar. The control and HP N400 context effects both had similar morphology, onset (~200 ms), and timing (peaking around ~375 ms). These similarities are consistent with previous reports that contextual information combines with knowledge to almost immediately influence semantic processing (e.g., Hald et al., 2007), even for knowledge of fictional characters from popular culture (Filik & Leuthold, 2013) or for fictional descriptions that override veridical real-world knowledge (Nieuwland & Van Berkum, 2006). Our findings of N400 context effects in sentences about the people, places, objects, and ideas from the fictional world of Harry Potter affirm that experiences gleaned from many sources—including the printed word or images on the movie screen, as well as the real world—combine to form the knowledge that influences the earliest stages of semantic processing during word by word reading.

Since the original reports of N400 amplitude modulation based on contextual support (Kutas & Hillyard, 1980; Kutas & Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984), many have demonstrated that a word's N400 response is highly sensitive to factors related to semantic retrieval, including whether the word is a content word vs. function word (Kutas, Van Petten, & Besson, 1988); the word's lexical frequency (Van Petten & Kutas, 1990; Kutas & Federmeier, 2000); how much context has accrued over the course of a sentence (Van Petten & Kutas, 1990); and the word's relationship to a preceding context, whether the context be a single word (Heinze, Muente, & Kutas, 1998; Federmeier, Kutas, & Schul, 2010), a picture (Friedrich & Friederici, 2010; Knoeferle, Urbach, & Kutas, 2011), or a whole sentence (e.g., Kutas & Hillyard, 1984; DeLong et al., 2005; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007). Presumably, the relationship between a word and context is mediated by knowledge in the head of the comprehender, which has heretofore been approximated using cloze probability for sentence contexts (e.g., DeLong et al., 2005). However, offline cloze probabilities do not provide precise estimates of any one individual's knowledge nor of what an individual brings to mind in the moment.

Here, by testing participants who varied in their knowledge of Harry Potter, we were able to directly and systematically investigate the availability of relevant knowledge during real-time sentence processing. We found that N400 context effects for HP, but not for control, sentences were graded with respect to the HP knowledge measure, with minimal to no effects for individuals with the lowest scores and the largest effects for those with the highest scores. By using a constrained domain in which we could estimate individual differences in knowledge, we explicitly demonstrated what has been to date implicitly assumed—namely, that specific knowledge relevant to interpreting written sentences is brought to bear—i.e., activated, or perhaps even pre-activated—relatively early (by ~200 ms or so) as words are encountered during real-time sentence processing.

Our interpretation is based on our assumption that HP knowledge level (determined by score on a trivia quiz) was directly proportional to the likelihood that an individual knew the information described in any given HP sentence pair in real time, during the EEG study. Beyond this relationship, if an HP expert and an HP novice both know a given fact (e.g., that Harry's scar has the form of a lightning bolt), they might nonetheless process the information differently, bringing more/less information to mind or doing so with a different time course. That is, differences in the functional organization of knowledge (e.g., Federmeier & Kutas, 1999) may impact processing above and beyond *whether* an individual knows a specific fact. To investigate these possibilities, we would want to compare individuals of differing knowledge levels on trials we know that they know vs. do not know (see Brothers, Swaab, & Traxler, 2015, 2017, who utilize trial-by-trial approaches to investigate effects of contextual support and judgments of lexical prediction).

We also asked whether the effect of contextual support observed in the HP sentences was influenced by other individual differences, including measures of reading experience, general knowledge, and verbal working memory. In the presence of these additional variables, we found that reading experience modulated the effect of contextual supported observed for HP sentences; HP knowledge had an overall effect, but did not interact with ending type.

It is important to note that HP knowledge and reading experience are correlated at $r = .54$ in our sample, thereby limiting our ability to determine precisely which drives the individual variation in N400 amplitude. This relatively high correlation is not surprising, as HP knowledge comes in large part from reading. For the moment, that HP knowledge (considered separately from reading experience) is predictive of contextual support effects for HP sentences but *not* for control sentences about general topics (see Table 2.5) leads us to speculate that HP knowledge mediates the observed relationship between reading experience and HP effects of contextual support on N400 amplitude. In future studies, we aim to better dissociate HP knowledge and reading experience, either by testing a sample of participants in whom the two measures are less correlated, or, better yet, by simultaneously investigating multiple domains of knowledge, allowing individuals who are experienced in one domain but inexperienced in the other to serve as their own controls.

2.5.3 Late positivity context effects

Although late positivity effects were not a focus of our study, we did observe systematic post-N400 positivity context effects for control sentences in the whole-head analysis, with unsupported words eliciting larger left anterior positivities compared to supported words. In our study, supported words were the best completion, and unsupported words were low-cloze yet plausible endings related in meaning to the best completion. These sentences were similar in nature to a subset of the sentences in Thornhill and Van Petten (2012), who likewise observed a frontal positivity for low-cloze, yet plausible words (both related and unrelated to the best continuation), compared to the best completion. Thornhill and Van Petten suggested anterior PNPs reflected the processing of lexically unexpected words; elsewhere, anterior PNPs have been linked to the processing of low-cloze congruent/plausible words, with posterior PNPs linked to the processing of low-cloze incongruent/implausible words (Van Petten & Luka, 2012; DeLong, Quante, & Kutas, 2014). Our findings from control sentences are consistent with both sets of hypotheses.

Perhaps surprisingly, we did not observe any effects of contextual support, nor any interaction between HP knowledge and contextual support (in the ROI analysis), for the HP sentences during the late time period. Because we designed our HP sentences such that, given little to no knowledge of HP, unsupported endings would seem similarly plausible to supported endings, we did not expect effects of the contextual support manipulation for low-knowledge individuals. It is, however, unclear why we did not observe effects of contextual support on late positivities for high-knowledge participants. Future work investigating trial-by-trial variation in sentence-specific knowledge along with individual differences in domain knowledge may shed light on this.

In both the whole-head and ROI analyses, post-N400 positivities were overall larger for HP than control sentence endings. HP sentences, by contrast to control sentences which described generalized situations, often described episodes/events in HP that had occurred. Individuals thus may have entertained specific, episodic information for some of the HP sentences. In the memory literature, late parietal positivities have been associated with episodic retrieval (cf. "old/new" effects, reviewed in Rugg & Curran, 2007). In future studies, asking participants to provide more information on a trial-by-trial level (e.g., whether they had known the information ahead of time, or whether they had brought episodic information to mind during reading) might shed light on knowledge-based individual differences of this nature.

2.5.4 Limitations and future directions

A limitation of our correlational approach is that participants were not randomly assigned levels of HP exposure, and there are other individual differences besides HP knowledge. Moreover, in a model including several measures of individual differences, it was reading experience, and not HP knowledge, that interacted with ending type. As reading experience and HP knowledge were positively correlated, we were unable to tease apart pure effects of HP knowledge from overall differences in reading experience in this study, though we leaned to the former. Future studies including multiple domains of knowledge simultaneously (e.g., separate book series with which individuals have differing amounts of exposure/knowledge) may be able to better identify individual differences due to specific domain knowledge vs. those differences due to differences in general levels of reading experience.

That individuals with differing levels of domain knowledge represent and process information within that domain differentially has been substantiated across many areas of expertise (see reviews in Ericsson, Charness, Hoffman, & Feltovich, 2006). However, the precise nature of differences in the representation of knowledge as a function of expertise remains an open question. How experts chunk, store, process, retrieve, and/or otherwise use information seem likely to differ depending on the nature of the information (motoric, perceptual, procedural, declarative, etc.) and the goals of the task at hand (e.g., winning a chess game, completing a physics problem, or enjoying a narrative) (see Chi, 2006, for discussion). In future work, we aim to better understand how differences in specific domain knowledge alter the nature of information that is readily cued from linguistic input and brought to mind during real-time sentence comprehension—including differences in the presence/level of detail brought to mind when reading about events (e.g., Metusalem et al., 2012; Amsel et al., 2015) or differences in how the features/categories related to a word are conceived of and accessed during real-time processing (e.g., Federmeier & Kutas, 1999; Federmeier, McLennan, De Ochoa, & Kutas, 2002).

2.5.5 Conclusions

In this study, we went beyond group-level measures of contextual support (i.e., cloze probability) by estimating individuals' knowledge in a specific domain, the narrative world of Harry Potter. We find that such knowledge estimates seem to predict patterns of neural activity, at the earliest neural stages of semantic processing, as reflected in N400 amplitude modulations. Future work can combine measures of individual-level knowledge in a domain with trial-by-trial estimates. Such an approach could prove powerful for investigating the relative contributions of (a) the functional organization of knowledge, which is likely to substantively differ between domain experts and novices, and (b) the likelihood that an individual knows any given item, which is probabilistically related to, but not necessarily determined by, an individual's overall level of knowledge. In sum, our results lay the groundwork for investigating how inter-individual differences in organization of knowledge (amount, content, or other aspects of internal organization/connectivity) influence aspects of knowledge use (including timing and depth/level of semantic processing) in real time sentence processing.

2.6 Acknowledgments

This chapter has been published in an article in *Language, Cognition, and Neuroscience*, 2018, Troyer, Melissa & Kutas, Marta, for which I was the primary author and investigator.

2.7 References

Altmann, G.T.M. & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*, 79-87. doi: 10.1016/S0010-0277(99)00059-1

Amsel, B.D., DeLong, K.D., & Kutas, M. (2015). Close, but no garlic: Perceptuomotor and event knowledge activation during language comprehension. *Journal of Memory and Language*, *82*, 118-132. doi: 10.1016/j.jml.2015.03.009

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. doi; 10.1016/j.jml.2007.12.005

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

Borovsky, A. & Creel, S. (2014). Children and adults integrate talker and verb information in online processing. *Developmental Psychology*, *50*(5), 1600-1613. doi: 10.1037/a0035591

Boudewyn, M.A. (2015). Individual differences in language processing: electrophysiological approaches. *Language and Linguistics Compass*, *9/10*, 406-419. doi: 10.1111/lnc3.12167

Boudewyn, M.A, Long, D.L., & Swaab, T.Y. (2012). Cognitive control influences the use of meaning relations during spoken sentence comprehension. *Neuropsychologia*, *50*, 2659-2668. doi: 10.1016/j.neuropsychologia.2012.07.019

Boudewyn, M.A., Long, D.L., & Swaab, T.Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive and Affective Behavioral Neuroscience*, *15*, 607-624. doi: 10.3758/s13415-015-0340-0

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. doi: 10.1016/j.brainres.2012.01.055

Brothers, T., Swaab, T.Y., & Traxler, M.J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, *136*, 135-149. doi: j.cognition.2014.10.017

Brothers, T., Swaab, T.Y., & Traxler, M.J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, *93*, 203-216. doi: 10.1016/j.jml.2016.10.002

Brown-Schmidt, S., Yoon, S.O., & Ryskin, R.A. (2015). People as contexts in conversation. *Psychology of Learning and Motivation*, *62*, 59-99. doi: 10.1016/bs.plm.2014.09.003

Chi, M.T.H. (2006) Laboratory methods for assessing experts' and novices' knowledge. In Ericsson et al. (Eds.), The Cambridge Handbook of Expertise and Expert Performance. Cambridge University Press. doi: 10.1017/CBO9780511816796.010

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152. doi: 10.1207/s15516709cog0502_2

Chiesi, H.L., Spilich, G.J., & Voss, J.F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*, 257-273. doi: 10.1016/S0022-5371(79)90146-4

Coronel, J.S. & Federmeier, K.D. (2016). The N400 reveals how personal semantics is processed: Insights into the nature and organization of self-knowledge. *Neuropsychologia, 84*, 36-43. doi: 10.1016/j.neuropsychologia.2016.01.029

Daneman, M. & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Memory and Verbal Behavior*, *19*(4), 450-466. doi: 10.1016/S0022-5371(80)90312-6

de Groot, A.D. (1965). *Thought and choice in chess*. Mouton, The Hague. doi: 10.5117/9789053569986

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

DeLong, K.A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass*, *8*(12), 631-645. doi: 10.1111/lnc3.12093

DeLong, K.A., Urbach, T.P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from brain activity. *Nature Neuroscience*, *8*(8), 1117-1121. doi: 10.1038/nn1504

Dunn, D.M. & Dunn, L.M. (2007). *Peabody picture vocabulary test*. Minneapolis, MN: NCS Pearson.

Ericsson, K.A., Charness, N., Feltovich, P.J., Hoffman, R.R. (Eds). (2006). *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge, UK: Cambridge University Press.

Federmeier, K.D. & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469-495. doi: 10.1006/jmla.1999.2660

Federmeier, K.D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*(3), 149-161. doi: 10.1016/j.bandl.2010.07.006

Federmeier, K.D., McLennan, D.B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*, 133-146. doi: 10.1111/1469-8986.3920133

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

Filik, R. & Leuthold, H. (2013). The role of character-based knowledge in online narrative comprehension: Evidence from eye movements and ERPs. *Brain Research*, *1506*, 94-104. doi: 10.1016/j.brainres.2013.02.017

Friedrich, M. & Friederici, A.D. (2010). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of Cognitive Neuroscience*, *16*(8), 1465-1477. doi: 10.1162/0898929042304705

Ganis, G., Kutas, M., & Sereno, M.I. (1996). The search for “common sense”: An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*(2), 89-106. doi: 10.1162/jocn.1996.8.2.89

Gobbo, C. & Chi, M. (1986). How knowledge is structured and used by expert and novice children. *Cognitive Development*, *1*, 221-237. doi: 10.1016/S0885-2014(86)80002-8

Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438-441. doi: 10.1126/science.1095455

Hald, L.A., Steenback-Planting, E.G., Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Research*, *1146*, 210-218. doi: 10.1016/j.brainres.2007.02.054

Heinze, H.-J., Muente, T.-F., & Kutas, M. (1998). Context effects in a category verification task as assessed by event-related brain potential (ERP) measures. *Biological Psychology*, *47*, 121-135. doi: 10.1016/S0301-0511(97)00024-0

Knoeferle, P., Urbach, T.P., & Kutas, M. (2011). Comprehending how visual context influences incremental sentence processing: Insights from ERPs and picture-sentence verification. *Psychophysiology*, *48*, 495-506. doi: 10.1111/j.1469-8986.2010.01080.x

Kuperberg, G.R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition, and Neuroscience*, *31*(5), 602-616. doi: 10.1080/23273798.2015.1130233

Kutas, M. (2006). One lesson learned: frame language processing—literal and figurative—as a human brain function. *Metaphor and Symbol*, *21*(4), 285-325. Doi: 10.1207/s15327868ms2104_5

Kutas, M. & Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463-470. doi: 10.1016/S1364-6613(00)01560-6

Kutas, M. & Federmeier, K.D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(14), 1-27. doi: 10.1146/annurev.psych.093008.131123

Kutas, M. & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203-205. doi: 10.1126/science.7350657

Kutas, M. & Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(12), 161-163. doi: dx.doi.org/10.1038/307161a0

Kutas, M., Lindamood, T.E., & Hillyard, S.A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum and J. Requin (Eds.), *Preparatory States and Processes* (pp. 217-237). Hillsdale, New Jersey: Lawrence Erlbaum.

Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology*, *69*, 218-233. doi: 10.1016/0013-4694(88)90131-9

Kutner, M.H., Nachtsheim, C.J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models, Fifth Edition*. McGraw-Hill Irwin.

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26. doi: 10.18637/jss.v082.i13

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703-704. doi: 10.1038/nn1264

McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning, 60*, 123-150. doi: 10.1111/j.1467-9922.2010.00604.x

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. doi: 10.1016/j.jml.2012.01.001

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81-97. doi: 10.1037/h0043158

Münte, T.F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Letters to Nature*, *395*, 71-73. doi: 10.1038/25731

Nakano, H., Saron, C., & Swaab, T. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, *22*(12), 2886-2898. doi: 10.1162/jocn.2009.21400

Nieuwland, M.S. & Van Berkum, J.J.A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098-1111. doi: 10.1162/jocn.2006.18.7.1098

Pakulak, E. & Neville, H. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, *22*(12), 2728-2744. doi: 10.1162/jocn.2009.21393

Pearson, P.D., Hansen, J., & Gordon, C. (1979). The effect of background knowledge on young children's comprehension of explicit and implicit information. *Journal of Reading Behavior*, *11*(3), 201-209.

Rugg, M.D. & Curran, T. (2007). Event-related potentials and recognition memory. *TRENDS in Cognitive Science*, *11*(6), 251-257. doi: 10.1016/j.tics.2007.04.004

Simon, H.A. (1974). How big is a chunk? *Science*, *183*(4124), 428-488.

Simon, H.A. & Chase, W.G. (1973). Skill in chess. *American Scientist*, *61*(4), 394-403. doi: 10.1007/978-1-4757-1968-0_18

Simonson, R. & Walker, S. (1988). *Multi-cultural literacy*. Gray Wolf Press: Saint Paul.

Spilich, G.J., Vesonder, G.T., Chiesi, H.L., & Voss, J.F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*, 275-290. doi: 10.1016/S0022-5371(79)90155-5

Stanovich, K.E. & Cunningham, A.E. (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. *Journal of Educational Psychology*, *85*(2), 211-229. doi: 10.1037/0022-0663.85.2.211

Stanovich, K.E. & West, R.F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*(4), 402-433. doi: 10.2307/747605

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632-1634. doi: 10.1126/science.7777863

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, *17*(2), 277-293. doi: 10.1017/S1366728913000370

Thornhill, D.E. & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*, 382-392. doi: 10.1016/j.ijpsycho.2011.12.007

Van Berkum, J.J.A, van den Brink, D., Tesink, C.M.J.Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580-591. doi: 10.1162/jocn.2008.20054

Van Petten, C. & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, *18*(4), 380-393. doi: 10.3758/BF03197127

Van Petten, C. & Luka, B.J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*,176-190. doi: 10.1016/j.ijpsycho.2011.09.015

Verhagen, V., Mos, M., Backus, A., & Schilperoord, J. (2018). Predictive language processing revealing usage-based variation. *Language and Cognition*, *10*, 329-373. doi: 10.1017/langcog.2018.4

Wlotko, E.W. & Federmeier, K.D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, *62*, 356-366. doi: 10.1016/j.neuroimage.2012.04.054

2.7 Supplementary Materials

2.7.1 Appendix: Description of behavioral tasks

2.7.1.2 Author and magazine recognition tests (ART/MRT)

The ART and MRT were adapted from Stanovich and West (1989). For the ART, participants viewed a checklist of 80 names and were instructed to choose the ones they recognized as authors and to avoid guessing. The MRT was identical except for magazine titles. For each, scores were calculated by subtracting the proportion of false alarms from the proportion of correct names checked.

2.7.1.3 Media and reading habits questionnaire (MRH)

To complement the ART and MRT measures of print exposure, a media and reading habits questionnaire was adapted based on Stanovich and West (1989). As an estimate of reading experience, we report a numeric score determined by how frequently participants said they read books for fun, visited libraries, and so on. In addition, we report statistics on the number of favorite authors (out of a maximum of 5) that individuals indicated, a measure that Stanovich and West (1989) have noted is predictive of several reading-related variables.

2.7.1.4 General knowledge quiz (GKQ)

To estimate individual differences in their general ability to remember and report factual information, we asked participants 30 multiple-choice questions on topics including entertainment, literature, politics, religion, science, and history, drawn from publicly available sources including Common Core test questions, Pew Research quizzes, BuzzFeed News quizzes, and other websites.

2.7.1.5 Knowledge checklists

Participants completed the cultural (CLC) and multicultural checklists (MCLC), designed to tap into knowledge of culturally relevant information (Stanovich & Cunningham, 1993). For each, half of the names on the list were true names of famous people and half were foils; participants were instructed to choose the names of people they knew and not to guess. Scoring was identical to the ART/MRT.

2.7.1.6 Peabody Picture Vocabulary Test—4 (PPVT)

The PPVT-4 was administered as a standardized measure of receptive vocabulary (Dunn & Dunn, 2007). Reported standardized scores reflect the total number of words a participant got correct.

2.7.1.7 Sentence span

To estimate verbal WM abilities, we administered Daneman and Carpenter's (1980) sentence span task to each participant. The scores reported here are the maximum set size for which participants correctly reported all words for at least 3 out of 5 sets, plus an additional half point if they correctly reported all words for 2 out of 5 sets at the next-highest set size.

2.7.2 Supplementary Tables

Supplementary Table 2.1. Results from a mixed-effects regression model analysis of performance on the memory tasks.

	Estimate	*SE*	*DF*	*T-value*	*Pr(>|t|)*
Intercept	1.708	0.131	36	13.084	.0000
Ending type	-0.061	0.077	108	-0.795	.4283
Sentence type	-0.341	0.077	108	-4.432	.0000
HP knowledge	0.196	0.132	36	1.481	.1474
Ending type:SentenceType	-0.084	0.077	108	-1.098	.2745
Ending type:HP knowledge	-0.010	0.078	108	-0.124	.9015
Sentence type:HP knowledge	-0.237	0.078	108	-3.049	.0029
Ending type:Sentence type:HP knowledge	-0.021	0.078	108	-0.274	.7843

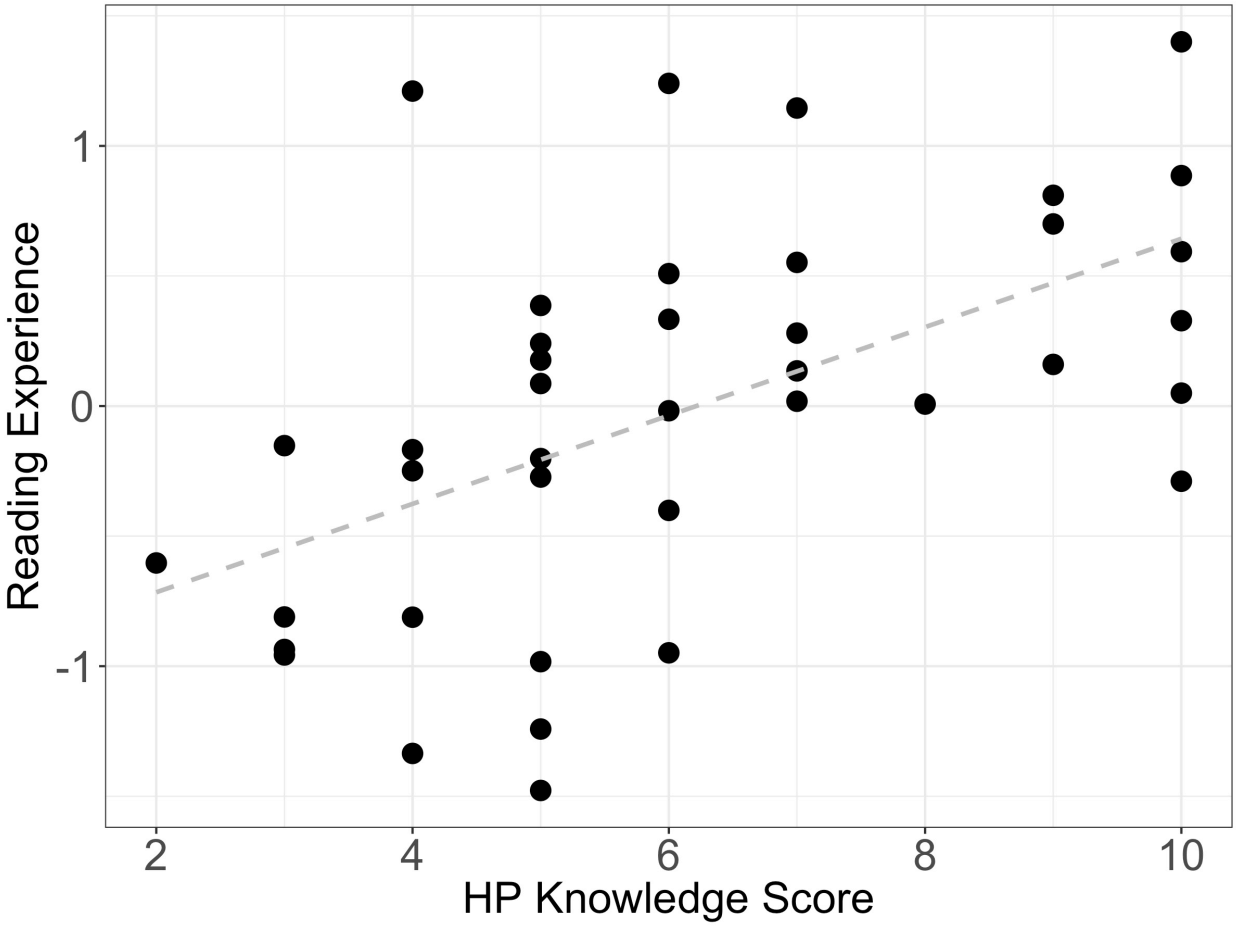
Supplementary Table 2.2. ROI analyses for each sentence type including covariates of reading experience, general knowledge, and verbal working memory and the interaction of each with ending type.

	Estimate	*SE*	*DF*	*T-value*	*Pr(>|t|)*
Control Sentences					
Intercept	3.621	0.362	34	10.006	.0000
HP knowledge	0.716	0.716	34	1.598	.1192
Reading experience	-1.121	-1.121	34	-1.504	.1417
General knowledge	-0.127	-0.127	34	-0.199	.8437
Reading span	0.191	0.192	34	0.482	.6326
Ending type	1.257	1.257	34	6.804	.0000
HP knowledge:Ending type	-0.146	-0.146	34	-0.640	.5263
Reading experience:Ending type	-0.078	0.380	34	-0.204	.8398
General knowledge:Ending type	0.222	0.327	34	0.680	.5013
Reading span:Ending type	-0.019	0.203	34	-0.095	.9248
HP Sentences					
Intercept	3.428	0.351	34	9.763	.0000
HP knowledge	0.738	0.435	34	1.696	.0989
Reading experience	-0.403	0.723	34	-0.557	.5810
General knowledge	0.428	0.622	34	0.688	.4961
Reading span	0.16	0.385	34	0.416	.6802
Ending type	1.056	.170	34	6.205	.0000
HP knowledge:Ending type	0.205	0.211	34	0.971	.3385
Reading experience:Ending type	0.903	0.351	34	2.576	.0145
General knowledge:Ending type	-0.313	0.302	34	-1.039	.3061
Reading span:Ending type	0.057	0.187	34	0.308	.7603

Supplementary Table 2.3. Post-hoc ROI analysis for HP sentences including covariates of reading habits, general knowledge, and verbal working memory and the interaction of each with ending type. The reading habits score (which came from the Media and Reading Habits questionnaire) was not strongly correlated with HP knowledge ($r^2 = .15$, n.s.). This model showed a trend for HP knowledge, but not the reading habits measure, to interact with ending type.

	Estimate	*SE*	*DF*	*T-value*	*Pr(>|t|)*
Intercept	3.425	0.353	34	9.714	.0000
HP knowledge	0.664	0.415	34	1.601	.1187
Reading habits	0.063	0.377	34	0.166	.8691
General knowledge	0.197	0.531	34	0.370	.7135
Reading span	0.172	0.388	34	0.445	.6594
Ending type	1.059	0.181	34	5.856	.0000
HP knowledge:Ending type	0.386	0.213	34	1.814	.0785
Reading habits:Ending type	0.276	0.193	34	1.426	.1630
General knowledge:Ending type	0.024	0.272	34	0.088	.9301
Reading span:Ending type	0.065	0.199	34	0.326	.7463

2.7.3 Supplementary Figures



Supplementary Figure 2.1. (Raw) HP knowledge score is plotted against the aggregate reading experience measure (*r* = .54).

CHAPTER THREE: *LUMOS!*: ELECTROPHYSIOLOGICAL TRACKING OF (WIZARDING) WORLD KNOWLEDGE USE DURING READING

3.1 Abstract

In Troyer & Kutas (2018), individual differences in knowledge of the world of Harry Potter (HP) rapidly modulated individuals' average electrical brain potentials to contextually supported words in sentence endings. Using advances in single-trial EEG analysis, we examined whether this relationship is strictly a result of domain knowledge mediating the proportion of facts each participant knew; we find it is not. Participants read sentences ending in a contextually supported word, reporting online whether they had known each fact. Participants' reports correlated with HP domain knowledge and reliably modulated ERPs to sentence-final words within 250 ms. Critically, domain knowledge had a dissociable influence in the same time window for endings which participants reported not having known and/or were less likely to be known/remembered across participants. We hypothesize that knowledge impacts written word processing primarily by affecting the neural processes of (implicit) retrieval from LTM: greater knowledge eases otherwise difficult retrieval processes.

3.2 Introduction

Understanding words in context requires quick, dynamic access to knowledge in long-term memory. Indeed, over the past fifty years, psycholinguists have convincingly demonstrated that sentence processing not only occurs incrementally but at times predictively (Kutas, DeLong, & Smith, 2011; DeLong, Troyer & Kutas, 2014). Online measures (eye-tracking, self-paced reading, and eye movements to images in scenes) indicate rapid sensitivity to word frequency (Trueswell, 1996), plausibility of thematic relationships (Trueswell, Tanenhaus, & Garnsey, 1994), discourse (Ehrlich & Rayner, 1981), and other linguistic/non-linguistic information gleaned from world knowledge (e.g., Kamide, Altmann, & Haywood, 2003; Borovsky & Creel, 2014). Scalp recordings of electrical brain activity (e.g., event-related brain potentials or ERPs) during word-by-word reading and speech track brain functioning in real time and reflect neural sensitivity to word, sentential, and pragmatic factors that impact semantic retrieval[4] within 250 ms of a word's occurrence (e.g., N400, a negativity between ~250-500 ms post-stimulus onset which is especially large for semantic anomalies but is a default response to all words; e.g., Federmeier & Kutas; reviewed in Kutas & Federmeier, 2011). Compared to ERPs elicited by a correct (true) word, lexico-semantic violations and lexically appropriate but untrue world-knowledge violations both elicit indistinguishably large N400s (e.g., for Dutch speakers: '*The Dutch trains are <u>yellow</u> / <u>white</u> / <u>sour</u> and very crowded.*'), suggesting similar time courses for retrieval of these two knowledge types (Hagoort, Hald, Bastiaansen, & Petersson, 2004). In short, understanding language requires knowledge of words and of the world, both assumed to be quickly available.

However, although individuals vary considerably in what they know and how well they know it, with documented consequences for perception, categorization, and/or memory across different domains—e.g., cooking, sports, chess, physics, medicine (Ericsson, Charness, Feltovich, & Hoffman, 2006)—the consequences for real-time sentence processing have not yet been detailed. To date, research on individual differences in language processing has focused primarily on differences in general cognitive abilities (Nakano, Saron, & Swaab, 2010; Boudewyn, Long, & Swaab, 2012; Kim, Oines, & Miyake, 2018) and/or in language-specific abilities, such as language proficiency in a first or second language (e.g., McLaughlin, Osterhout, & Kim, 2004; Pakulak & Neville, 2010; Tanner, McLaughlin, Herschensohn, & Osterhout, 2013; reviewed in Boudewyn, 2015).

[4] Here and elsewhere, by retrieval, we mean the implicit activation of semantic memory, which N400 brain potentials have been argued to reflect.

A major challenge for investigating knowledge-based individual differences in real-time sentence comprehension is allowing for the requisite variability in what people know while probing their comprehension with natural sentences. We have approached this using the narrative world of Harry Potter (HP) by J.K. Rowling—a constrained, yet rich, domain with complex intersecting trajectories of characters, objects, actions and events (Troyer & Kutas, 2017, 2018). Troyer and Kutas (2018) recorded EEG/ERP as participants varying in HP knowledge read sentences about (a) the world of HP or (b) general topics (i.e., control sentences). Control sentences ended with the highest-cloze/best completion (supported) or a plausible low cloze probability word (unsupported). HP sentences ended with a word that accurately described "facts" from the books (supported) or word that seemed plausible, but was factually incorrect given HP knowledge (unsupported). As expected, HP domain knowledge modulated N400 effects of contextual support for sentences about HP, but not for control sentences; the N400 effects were driven by variance in the *supported* (but not unsupported) HP words, consistent with the proposal that greater domain knowledge facilitates retrieval of relevant information during real-time sentence processing. We reasoned that greater HP domain knowledge likely led to participants knowing/remembering a greater number of the HP facts, and, consequently, to smaller mean N400 amplitudes to supported words as a function of the number of trials each participant knew. However, as we did not measure *which* facts each individual knew, we could not be certain that (a) those with greater HP knowledge knew a greater proportion of the critical words or that (b) this alone determined the proportion of larger versus smaller N400s in their averages. Moreover, given attested individual differences in cognition and behavior as a result of domain expertise (Ericsson et al., 2006), we reasoned that domain knowledge might influence real-time processing beyond strictly determining the proportion of items individuals knew. Such individual differences in knowledge seem likely to influence the timing and/or contents of knowledge that can be rapidly brought to mind, yet they are strikingly absent from current models of real-time sentence processing. We thus decided to investigate the influences of HP domain knowledge and participants' specific knowledge of individual HP "facts" on written word ERPs by asking participants whether or not they had known each fact.

This single-trial experimental design required analyzing ERPs as a function of categorical (response type: known/unknown) and continuous (HP domain knowledge) variables in participants who varied in the proportion of trials reported as known (vs. not). These aspects of the design pose challenges for standard ERP analyses, which rely on averages (not single trials) across participants and conditions. For statistical hypothesis tests, we therefore employed hierarchical mixed-effects regression (Baayen, Davidson, & Bates, 2008), which can model single-trial data, categorical and/or continuous predictors, and data not evenly matched across cells.

In many ERP studies, the grand mean to an event of interest—that is, the point-by-point average of participant averages—is plotted for each experimental condition. This method, however, is not well-suited for visualizing data from conditions with unbalanced cell counts or which co-vary with continuous variables. We thus turned to a relatively new, thus far little-used technique, the regression ERP (rERP), to visualize our time series data (Smith & Kutas, 2015a,b). rERPs are calculated from the same scalp potential data as conventional ERPs, time-locked to events of interest at each electrode location. Besides estimating averages, the rERP estimated regression coefficients (i.e., weights) can represent, for visualization and statistical analysis, the influence of categorical and/or continuous predictors as well as any interactions thereof on the event-related EEG signal. Moreover, these coefficients can be used to compute predicted ERPs for unobserved values of variables at the level of the participant, item (each sentence pair / HP "fact"), and/or trial.

The present study had several aims. First, we aimed to replicate the positive correlation between HP domain knowledge and average N400 amplitude to contextually supported words (Troyer & Kutas, 2018). Second, we directly tested our hypothesis that each participant's HP domain knowledge score would correlate strongly with the number of trials they reported having known during the EEG experiment. Third, we assessed our prediction of smaller N400 amplitudes for trials reported as "known" versus "unknown." Finally, with this novel design and analyses, we aimed to determine whether HP domain knowledge modulates the N400 to contextually supported words after controlling for single-trial-level reports of knowledge, indicating that effects of domain knowledge on N400 amplitude are not merely a consequence of averaging different proportions of known vs. unknown trials by participant. This, in turn, would suggest that knowledge-based differences in cognitive processes (e.g. perception, categorization, memory), as reported in other domains, may also be evident in knowledge retrieval processes during real-time language comprehension.

3.3 Method

3.3.1 Participants

41 students (mean age = 20 years; range = 18-23; 29 women, 12 men) at UCSD took part in the study for partial course credit or payment of $9 / hour. To ensure that some participants would have high knowledge of the Harry Potter domain, a subset (*N* = 18) were recruited contingent on having read all seven Harry Potter books and/or having watched all eight Harry Potter films. All participants provided informed consent reviewed by the Institutional Review Board at the University of California San Diego. We estimated that ~40 participants was

appropriate based on a previous study run in the lab, in which N400 amplitudes to contextually supported words correlated with HP domain knowledge at $r = .41$ (Troyer & Kutas, 2018); we replicate this result at $r = .37$ (Figure 3.4).

3.3.2 Materials

3.3.2.1 Sentence materials

During the EEG portion of the experiment, participants read 172 descriptions of facts / events from the narrative world of Harry Potter.[5] Using freely available materials (Wikipedia, fan sites) referring to the text of the Harry Potter books, the first author created a set of sentences that accurately described events/entities from the book series. The final word was designed to be 100% predictable given perfect knowledge of the series, verified by asking a separate group of participants to complete a cloze norming study ($N = 32$-34[6] participants, who varied in HP domain knowledge as determined by a 10-question trivia quiz, per item). Mean cloze probability across all norming participants and items was .49 (range: .03 to 1.00) (Figure 3.1b; examples provided in Table 3.1). For participants scoring in the top half on the quiz, mean cloze probability was .73 (range: .05 to 1); for those scoring in the bottom half, mean cloze probability was .24 (range: 0 to 1).

[5] 108 of these sentences were identical to the (supported) HP sentences from Troyer & Kutas (2018).

[6] For the norming study, due to time constraints, each participant provided completions for only half of the materials; half were completed by 32 participants and the remainder by 34 different participants.

Table 3.1. Sample sentence pairs.

Sentence frame	Final word	Cloze
There is one main sport in the wizarding community. It is known as	Quidditch	1.00
The character Peter Pettigrew changes his shape at times. He takes the form of a	rat	0.72
Harry eventually learns the truth about Sirius Black. Sirius is Harry's	godfather	0.56
Hermione owns a large, orange feline. Her pet is called	Crookshanks	0.44
To combat boggarts, wizards must think of something funny. They must use the spell	Riddikulus	0.38
Hogwarts students shop at Madam Malkin's. This is where they buy their	robes	0.31
Looking for Sirius, Harry and his classmates fly to the Ministry of Magic. They ride winged horses called	thestrals	0.13

The original 172 descriptions were split into sentence pairs for presentation purposes. First sentences appeared as a whole to participants (mean length = 9.5 words; range = 4-18 words); the second were presented word-by-word (mean = 6.8 words; range = 3-13 words).

3.3.2.2 Individual differences tasks and measures

For EEG participants, our primary measure of HP domain knowledge was their score on a 40-question multiple choice HP "trivia quiz[7]." Participants also completed an HP self-report questionnaire; scores were determined by summing the total number of times an individual had read each book, seen each movie, etc.

We also collected several measures of individual differences to assess other group differences, including general print/reading experience (media and reading habits questionnaire (MRH); author and magazine recognition tests (ART/MRT); Stanovich & West, 1989), general knowledge trivia quiz (GKQ); cultural literacy checklists (CLC/MCLC; Stanovich &

[7] With one exception, care was taken to avoid overlap between the HP trivia quiz and the ERP sentence materials, such that individuals would not be able to answer any of the quiz questions based on the ERP sentences.

Cunningham, 1993)); vocabulary (PPVT; Dunn & Dunn, 2007); and verbal working memory (sentence span; Daneman & Carpenter, 1980).

3.3.3 Procedures

3.3.3.1 Ordering of individual differences tasks

ART/MRT were administered during EEG set-up. After the EEG study, the HP trivia quiz, HP self-report, and all other tests were administered in a quiet room in the order described in the preceding section, followed by debriefing.

3.3.3.2 EEG experiment

Participants were asked to relax to minimize muscle artifact. They were told they would be reading two-sentence stories about the world of Harry Potter for meaning and asked to answer two questions after each pair of sentences. First, “Did you know this ahead of time?” (Q1). They were instructed to respond by button press “Yes” or “No”. Following “Yes” responses, they were asked, “How sure are you?”, responding by button press with “Certain” or “Not sure”; following “No” responses, they were asked, “After you read it, did it seem familiar?”, responding with “Yes, seemed familiar” or “No, not familiar” (Q2).

During the EEG experiment, participants sat approximately 100 cm in front of a CRT. Words flashed in white on a near-black background and subtended about 2.3 horizontal degrees of visual angle (range = 1.9-4.7°). Each trial began with a blank screen for two seconds. Then, the first sentence of each pair was presented until the participant pressed a button. Next, a small crosshair appeared just below the center of the screen for a duration which varied randomly between 900 ms and 1100 ms. Participants were instructed to focus on the crosshair and not move their eyes or blink while it was on the screen. The second sentence was then presented one word at a time just above the crosshair for 200 ms with an interstimulus interval of 300 ms. After the sentence-final word disappeared, the crosshair remained on the screen for a duration that randomly varied between 900 and 1100 ms followed by a blank screen for 1 second. Next, Q1 appeared, remaining on the screen until the participant answered. Q2 then appeared and remained on the screen until the participant answered.

3.3.3.3 EEG recording

The electroencephalogram (EEG) was recorded from 26 electrode sites arranged geodesically in an Electro-cap (Ganis, Kutas, & Sereno, 1996; Fig. 9). Online recording was to a common left mastoid reference; data were re-referenced offline to the average of the left and right mastoid (A1, A2). Electrodes located adjacent to the outer canthus of each eye with a bipolar derivation monitored lateral eye movements. Electrodes placed below each eye referenced to the left mastoid and used to monitor vertical eye-movements and blinks. Throughout the experiment, electrode impedances were maintained under 5 kΩ. The EEG was amplified with Grass Model 12 Neurodata Acquisition System amplifiers set at a bandpass of .01 to 100 Hz; the sampling rate was 250 Hz.

3.3.4 Data analysis

3.3.4.1 Behavior

We report Pearson's *r* for correlations between HP domain knowledge and the proportion of each response type, by participant, and for correlations between offline cloze and the proportion of each response type, by item.

3.3.4.2 EEG

Single-trial epochs of EEG data were extracted from the continuous recordings 200 ms before the onset of a critical word until 900 ms post-critical word. Trials containing artifacts (e.g., eye movements, blinks, muscle activity, blocking) were removed from subsequent analyses, resulting in an exclusion of 17% of trials. Because we sorted trials based on participants' reports of their own knowledge (leading to vastly different numbers of trials per bin across participants), those with relatively high artifact rates were not excluded. Due to recording error, only partial data were collected for one participant, who was included. For each trial and channel, a baseline was computed by averaging activity from 200 ms before the word to word onset; this was subtracted from the single-trial waveform.

3.3.4.2.1 Time window analyses

We analyzed a window surrounding the typical peak of the N400 brain potential from 250 to 500 ms after critical word onset and a window from 500 to 750 ms during which post-N400 positivities—hypothesized to reflect a number of high-level processing mechanisms including mental model updating (e.g., Brouwer, Fitz, & Hoeks, 2012)—sometimes appear. We focused on a centro-parietal region of interest (ROI) where N400 effects are typically most prominent, averaging across 8 electrodes (MiCe, LMCe, RMCe, MiPa, LDPa, RDPa, LMOc, RMOc; Figure 3.5).

Our experimental design hinged on participants' subjective, trial-by-trial responses to Q1, leading to different numbers of trials per cell for this variable across participants.[8] To that end, we used hierarchical mixed-effects linear regression models, which allow for different counts per cell (Baayen et al., 2008), on trial-by-trial measures of ERP data, including mean amplitude in N400 window. Because factors modulating N400 amplitude often modulate brain potentials in a post-

[8] We recognize that there may be systematic differences in how participants answered Q2 as a function of HP domain knowledge. However, we do not report ERP data based on response to Q2 because there were too few trials per cell, especially for the most knowledgeable individuals, who rarely responded that they were certain they had *not* known an item/fact (Figure 3.2). In future studies with more power, we plan to address the relative influence of domain knowledge on items never known, those seeming familiar (perhaps forgotten), and those actually remembered/known.

N400 time period (see Van Petten & Luka, 2012, for a review), we also analyzed mean amplitude in a late time window (500-750 ms). Unless explicitly indicated otherwise, time window analyses were performed on these single-trial data. Models included random intercepts[9] for item and participant and were implemented using `lme4` (1.1-12; Bates, Maechler, Bolker, & Walker, 2015) and `lmerTest` (3.0-1; Kuznetsova, Brockhoff, & Christensen, 2017) packages in `R` (3.3.2). Where relevant, we performed model comparison and report Chi-squared statistics on nested models to do significance testing on covariates of interest. To further understand covariates of interest, p-values on beta coefficients were computed using the Satterthwaite option for denominator degrees of freedom for *F* statistics. Categorical predictors were deviance-coded (i.e., Yes=1, No=-1, for Q1) and continuous predictors (e.g., HP domain knowledge, offline cloze) were z-transformed so that a value of +/- 1 reflected a single standard deviation above or below the mean (at 0). To visualize effects of and interactions between Q1 response and HP domain knowledge we computed predicted ERPs from the coefficients of the mixed-effects linear regression model.

3.4 Results

3.4.1 Behavioral data

3.4.1.1 Individual differences tasks

Supplementary Table 3.1 reports descriptive statistics for participants' scores on the HP trivia quiz and other individual difference measures. The distribution of HP domain knowledge

[9] There is some controversy among experts about when to use maximally-specified random effects structures. Whereas Barr et al. (2013) argue for using the maximal random effects structure justified by the design, Bates et al. (2015) argue for parsimonious random effects justified by the data. For our ERP data, the pattern of results for inferential statistical tests was the same regardless of the random effects structure; we therefore chose to present results from models with simpler (intercepts-only) random effects structures.

scores is plotted in Figure 3.1a. Intercorrelations among individual differences measures are provided in Supplementary Table 2.

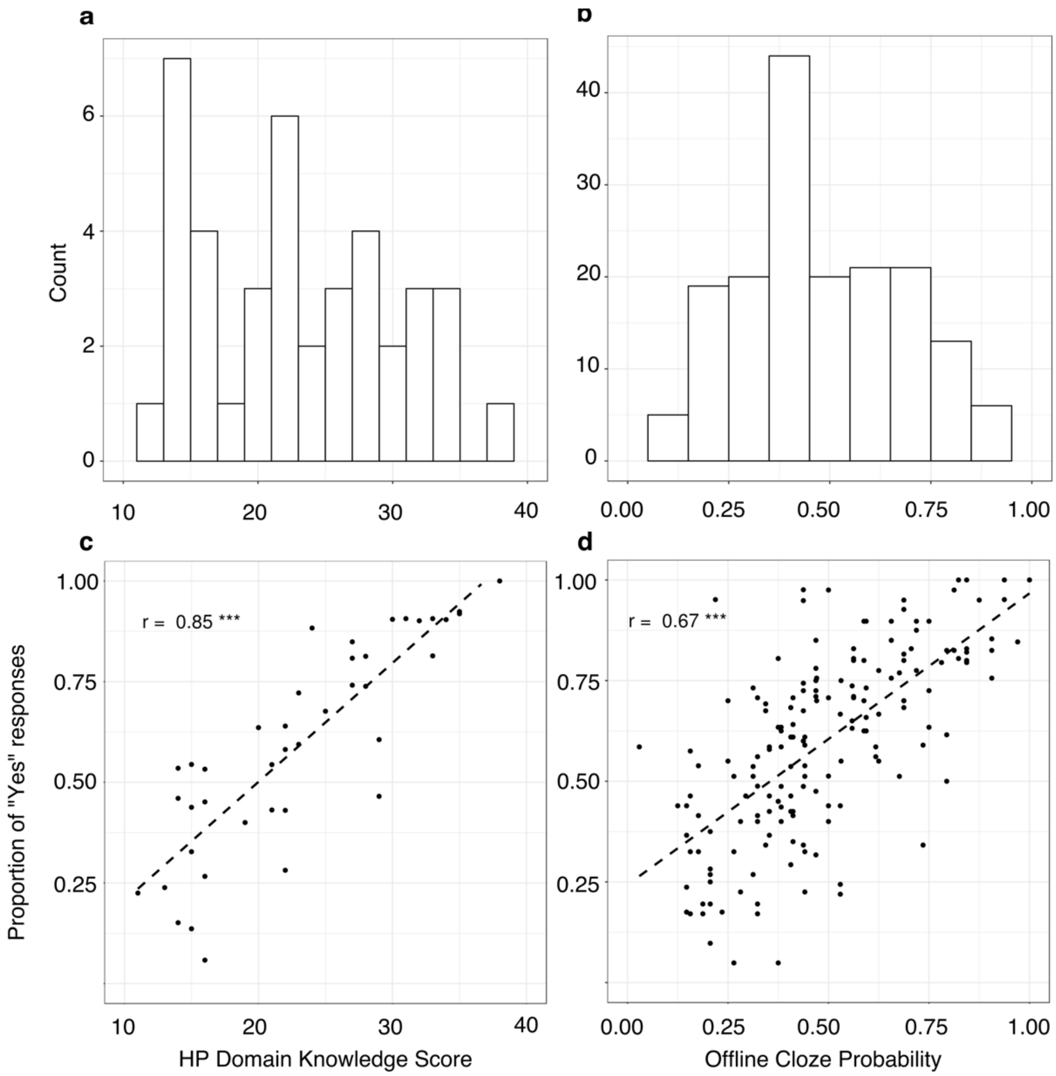


Figure 3.1. (a) A histogram shows the distribution of HP knowledge scores across participants. (b) A histogram shows the distribution of cloze probability across items. (c) Each participant's raw HP knowledge score from the offline, 40-question trivia quiz is plotted against the proportion of trials they reported having known during the ERP study (i.e., the proportion of "Yes" responses to Question 1). The two are correlated at $r = .85$, $p < .001$. (d) For each item, its offline cloze probability is plotted against the proportion of participants who reported having known it during the ERP study. The two are correlated at $r = .65$, $p < .001$.

3.4.1.2 Question responses

Participants indicated that they had known information described by each sentence pair (i.e., responded "Yes" to Q1) on an average of 59% (95% CI [52%, 67%]) of trials. As anticipated, the number of statements that participants reported having known correlated strongly with their performance on the HP domain knowledge quiz, $r = .85$, $p < .001$ (Figure 3.1c). Item-wise offline cloze probability, measured in a separate group of participants, was also strongly correlated with the proportion of participants who reported having known each item, $r = .67$, $p < .001$ (Figure 3.1d). Participants responded with "Yes—Certain" on 50% (95% CI [41%, 58%] of trials; "Yes—Not Sure" on 10% (95% CI [8%, 12%] of trials; "No—Seems Familiar" on 18% (95% CI [14%, 21%] of trials; and "No—Not Familiar" on 23% (95% CI [17%, 29%]) of trials (Figure 3.2).

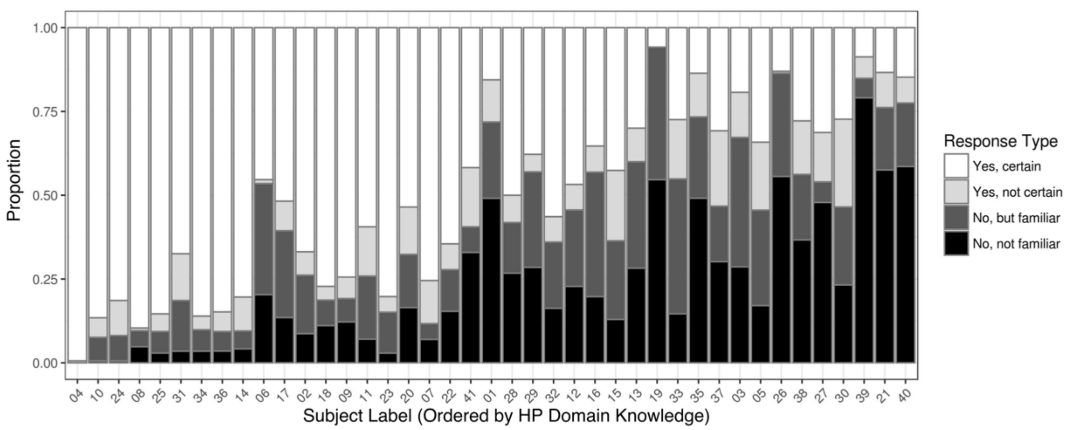


Figure 3.2. The proportions of trials of each responses type (for Q1 and Q2) are plotted by participant, ranked by HP knowledge score (highest on the left, lowest on the right).

3.4.2 ERP data

Figure 3.3 displays trial-averaged ERPs across 26 scalp electrodes from 200 ms before the critical word onset to 900 ms post-critical word, with separate waveforms computed for trials on which participants responded "Yes" vs. "No" to Q1.[10] Across most electrodes, ERPs to critical words for "Yes" and "No" responses are characterized by two early sensory components (N1 and

[10] The pattern of reduced N400 amplitude for "Yes" compared to "No" responses is similarly apparent whether data are averaged across trials (as in Figure 3.3), averaged within participants and then across participants (as is typical in ERP studies), or averaged by trial or participant for a subsample of participants ($N = 28$) with a minimum of 20 items of each response type (Yes, No).

P2). The P2 is followed by a wave which is mostly positive-going for “Yes” responses and which shows a relative negativity for “No” responses.

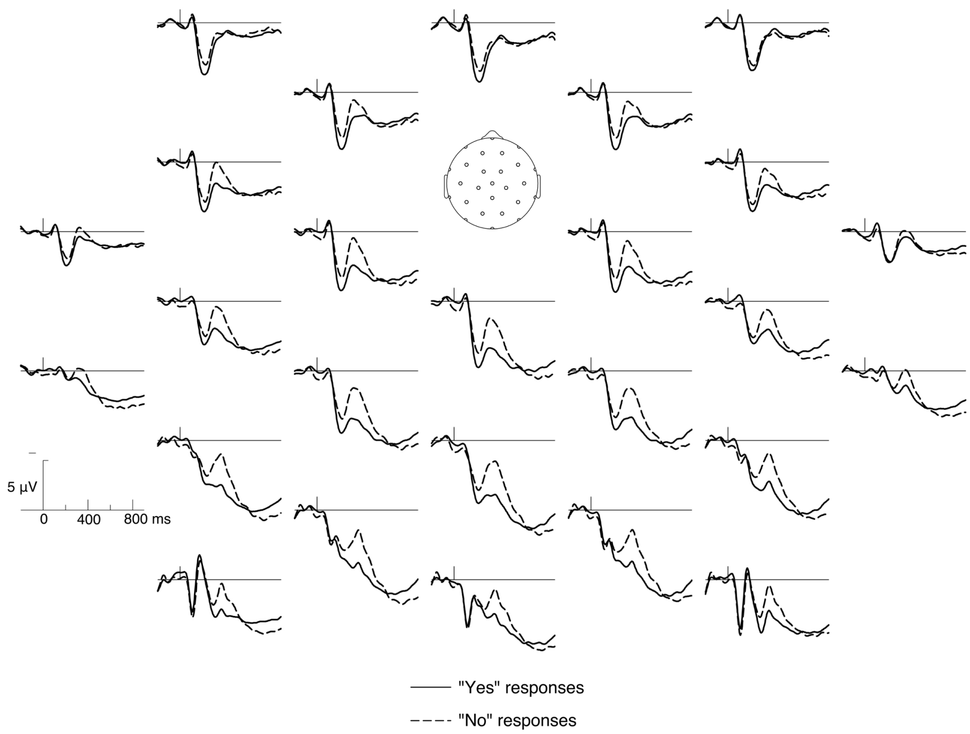


Figure 3.3. Grand average ERPs across all single trials to critical words are plotted across the whole head using a low-pass filter with a cutoff of 10 hz (electrode locations shown on the head in the center).

Before turning to single-trial analyses, we examined participants’ mean N400 amplitude to critical words in the centro-parietal ROI; it was correlated with HP domain knowledge at r = .37 (p = .018), replicating the pattern observed for HP-supported items in Troyer & Kutas (2018) (compare Figure 3.4a and 3.4b). Figure 3.4c presents the three-way relationship between participants’ HP domain knowledge, mean N400 amplitude, and the proportion of trials reported.

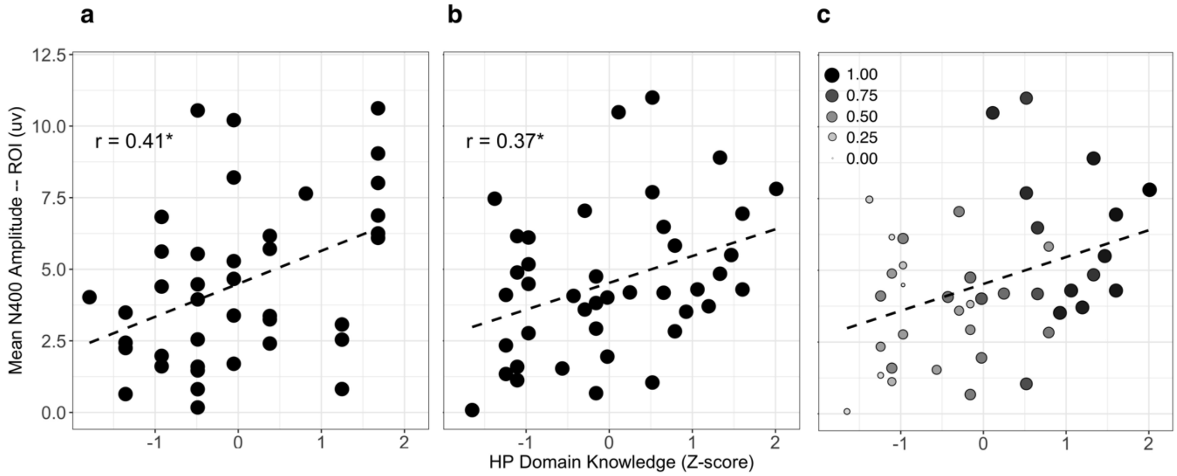


Figure 3.4. (a) In Troyer & Kutas (2018), HP domain knowledge and mean N400 amplitude to supported words in HP contexts were correlated at $r = .41$ ($p < .05$). (b) This pattern is replicated in the current study, $r = .37$ ($p < .05$). (c) The data shown in panel (b) are presented again, shaded and sized according to the proportion of items that each individual participant reported knowing during the ERP study.

3.4.2.1 Single-trial ROI analyses

3.4.2.1.1 N400 time window

Results from a linear mixed-effects model crossing Q1 response type and HP domain knowledge as fixed effects are presented in Table 2 (see Supplementary Figure 3.1 for a visualization of the beta coefficients fit to ROI data for the mixed-effects rERP model). Q1 response type was a significant predictor (Table 2); items that participants reported having known elicited reduced N400 amplitude compared to those reported as unknown. The effect of HP domain knowledge was marginal (Table 2), with higher-knowledge individuals exhibiting somewhat more positive-going N400 potentials compared to lower-knowledge individuals. Critically, the interaction term for Q1 response and HP domain knowledge was a significant predictor. This model was preferred over one incorporating only Q1 response as a fixed effect ($\chi^2(2) = 12.601$, $p = .002$), indicating that HP domain knowledge added explanatory power over and above participants' trial-by-trial reports of knowledge.

Table 3.2. Statistics for fixed-effect predictors of mean ERP amplitude in the N400 time period for ROI analyses.

Fixed effects	*β*	*SE*	*DF*	*t-value*	*p-value*
Intercept	4.65	0.42	59	11.05	.0000
Q1 Response	0.73	0.16	4512	4.66	.0000
HP Knowledge	0.72	0.39	44	1.83	.0747
Q1 Response : HP Knowledge	-0.55	0.17	4916	-3.28	.0011

Table 3.3 Statistics for fixed effects predictors of mean amplitude in the N400 time window for ROI analyses of "Yes" and "No" responses, respectively.

Fixed effects	*β*	*SE*	*DF*	*t-value*	*p-value*
"Yes" responses					
Intercept	5.55	0.46	56.99	12.12	.0000
HP Knowledge	0.03	0.42	39.54	0.07	.9470
"No" responses					
Intercept	3.92	0.49	51.33	8.00	.0000
HP Knowledge	1.11	0.50	41.55	2.24	.0306

Since the planned test found the interaction effect reliable, we conducted follow-up analyses of "Yes" and "No" responses separately using mixed-effects models with HP domain knowledge as a fixed effect (Table 3.3). For "Yes" responses, a model incorporating HP domain knowledge was not a significant predictor; this model was not preferred over an intercept-only model ($\chi^2(1) = .006$, n.s.), indicating no explanatory power of HP domain knowledge for "Yes" trials. For "No" responses, however, HP domain knowledge was a significant predictor, and this model was preferred over an intercept-only model ($\chi^2(1) = 5.003$, $p = .025$): having higher HP domain knowledge led to more positive going potentials in the N400 time window compared to having lower HP domain knowledge. We visualize this influence using both a standard approach, dividing participants into two subgroups using a median split based on HP domain knowledge and by plotting predicted ERPs based on regression modeling for hypothetical subjects (details in Figure 3.5).

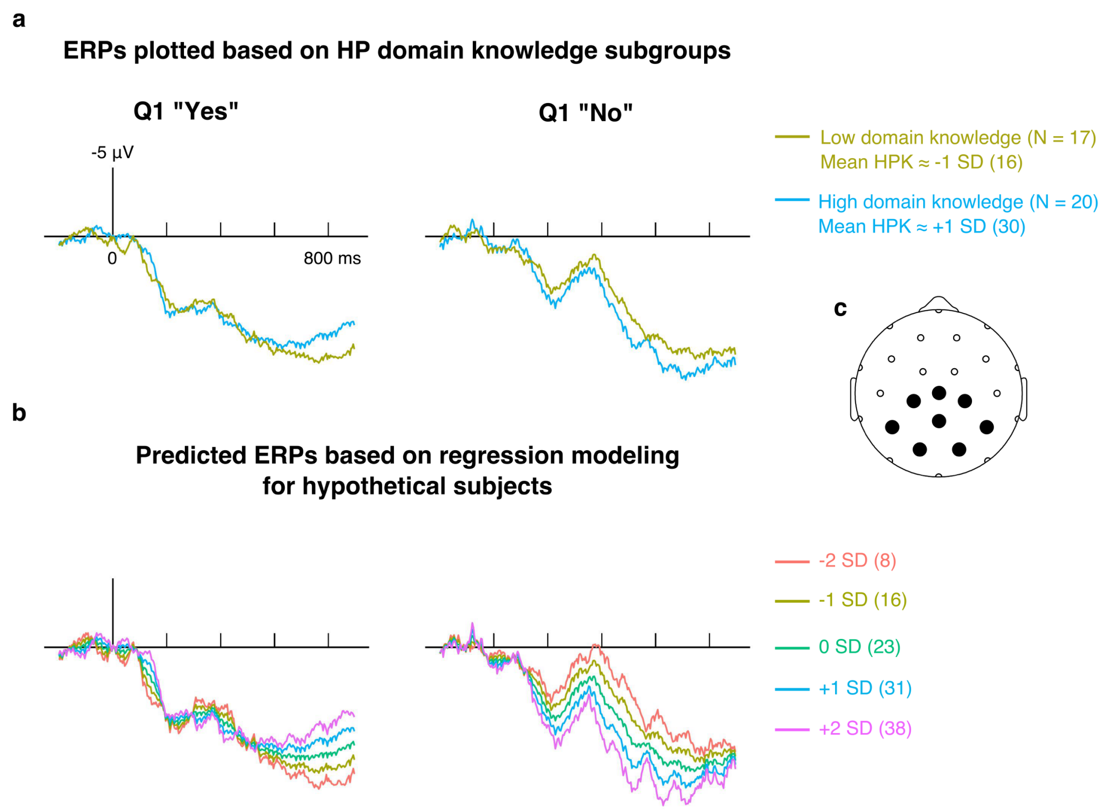


Figure 3.5. (a) Trial-averaged ERPs to critical words for a centro-parietal ROI (see text) are plotted by Q1 response type and overlapped for two subgroups of participants based on a median split on HP knowledge (HPK) scores. The high-knowledge subgroup had an HPK score of ~30 (about 1 SD above the sample mean score of 23), and the low-knowledge subgroup had an HPK score of ~16 (about 1 SD below the sample mean). (b) With standard approaches it is not possible to sort the data according to fine grains of a continuous variable when there aren't enough data points at each level of the variable. The regression ERP approach that we introduce and apply to visualize these data allows us to estimate what the data might look like, based on a generalization extracted not just over small subsamples of the data (which in this case are too limited) but over the whole sample. Predicted ERPs for hypothetical subjects and Q1 response type are therefore plotted based on regression modelling: using the estimated coefficients from linear mixed-effects models of centro-parietal ROI voltage based on Q1 response type, HPK, and their interaction, fit at each time point, we illustrate the time course of variation in ERPs as a function of HPK (in parentheses) at 0, +/- 1, and +/- 2 standard deviations from the sample mean. Because the actual values of HPK scores in (a) roughly correspond to +/- 1 SD, the median-split ERPs in (a) roughly map on to the predicted ERPs for HPK scores of +/- 1 SD in (b). (c) The centro-parietal ROI electrode locations are indicated by the filled-in circles.

Next, we asked whether individual differences apart from HP domain knowledge might modulate the influence of participants' subjective reports of knowing individual items. We therefore used a linear mixed-effects model predicting N400 amplitude based on four variables: HP domain knowledge, verbal working memory[11], reading experience, and general knowledge, as well as Q1 and its interaction with each of these variables (Supplementary Table 3.3). Consistent with previous analyses, there was a significant effect of Q1 response type, with "Yes" trials associated with more positive-going potentials compared to "No" trials, and Q1 response type interacted with HP domain knowledge, but not with any other individual difference measure. Indeed, nested model comparison of this more complex model with one incorporating HP domain knowledge and Q1 response type (described above) indicated that the simpler model was preferred ($\chi^2(6) = 3.577$, n.s.); this confirms results from our previous analysis and strongly supports the notion that individual-participant-level HP domain knowledge influenced the brain's response to critical words on trials that participants reported *not* having known.

3.4.2.1.2 Late positivity time window

Results from a mixed-effects model crossing Q1 response type and HP domain knowledge score as fixed effects are presented in Table 3.4. Both Q1 response and the interaction term crossing Q1 response and HP domain knowledge were significant predictors. This model was preferred over a model incorporating only Q1 response ($\chi^2(2) = 22.609$, $p < .001$), indicating that HP domain knowledge added explanatory power over and above participants' trial-by-trial reports of knowledge during the late positivity time window in the centro-parietal ROI.

[11] Verbal working memory scores were not collected for two participants due to time constraints; for this analysis, $N = 39$.

Table 3.4. Statistics for fixed-effect predictors of mean ERP amplitude in the late positivity time window for ROI analyses.

Fixed effects	*β*	*SE*	*DF*	*t-value*	*p-value*
Intercept	7.84	0.50	54	15.81	.0000
Q1 Response	-0.15	0.17	4419	-0.88	0.379
HP Knowledge	0.38	0.47	43	0.80	0.427
Q1 Response : HP Knowledge	-0.87	0.18	5230	-4.77	.0000

Table 3.5. Statistics for fixed-effect predictors of mean ERP amplitude in the late positivity time window for ROI analyses of "Yes" and "No" responses, respectively.

Fixed effects	*β*	*SE*	*DF*	*t-value*	*p-value*
"Yes" responses					
Intercept	7.69	0.47	64.83	16.22	.0000
HP Knowledge	-0.41	0.43	42.13	-0.97	0.339
"No" responses					
Intercept	8.13	0.62	45.45	13.04	.0000
HP Knowledge	1.27	0.65	43.07	1.94	.0585

To follow up on the Q1 by HP domain knowledge interaction, we fit separate models with HP domain knowledge as a fixed effect to subsets of the data based on Q1 response (Table 3.5). HP domain knowledge was a marginal predictor for "No" but not "Yes" responses: for "Yes" responses, a model incorporating HP domain knowledge was not preferred over an intercept-only model ($\chi^2(1) = 0.952$, n.s.) while for "No" responses, a model incorporating HP domain knowledge was marginally preferred over an intercept-only model ($\chi^2(1) = 3.809$, p = .051). To sum up, in the late positivity time window, for the centro-parietal ROI, higher (compared to lower) HP domain knowledge trended toward more positive-going amplitude on trials that participants reported *not* having known.

In an analysis incorporating individual differences (Supplementary Table 3.4) we again found an interaction between Q1 response type and HP domain knowledge. Q1 did not interact with any other individual difference measure, nor were any other terms in the model significant. Indeed, the more complex model was dispreferred compared to the simpler model reported above ($\chi^2(6) = 4.802$, n.s.).

3.5 Discussion

The view that rapid access to world knowledge is a part of real-time language comprehension is now widely held. Yet, despite considerable variability in what different people know, studies of language processing have overlooked those differences (e.g., Hagoort et al., 2004; Hald, Steenback-Planting, & Hagoort, 2007; Van Berkum, Holleman, Nieuwland, Otten, & Murre, 2009; Filik & Leuthold, 2013). We have leveraged recent statistical advances—the rERP technique and mixed-effects linear regression models—along with measurable variance in knowledge of a well-known narrative world to better delineate world knowledge influences on real time sentence (and word) processing. In Troyer & Kutas (2018), we found that each individual participant's degree of HP domain knowledge predicted N400 effects of contextual support, suggesting that it was an important determinant of real-time access to meaning during reading. We reasonably assumed that an individual's degree of domain knowledge was associated with the proportion of facts they knew, but did not know which these actually were. Hence, we could not test whether domain knowledge had additional influences on their ERPs, as might be expected from the expertise literature.

Here, we addressed these unresolved questions in an electrophysiological reading study incorporating trial-by-trial participant knowledge reports. We replicated the moderate correlation between N400 amplitudes to contextually supported words with HP domain knowledge reported in Troyer & Kutas (2018). With our single-trial design, we confirmed our hypothesis that offline HP domain knowledge scores would be highly correlated with the proportion of online sentence comprehension trials participants reported knowing. As expected, we showed that single-trial-level participant reports of knowledge were strong predictors of ERPs to supported HP sentence endings. Critically, HP domain knowledge had yet an additional influence on ERPs, even after controlling for single-trial-level participant reports of knowledge. This effect was evident at least by ~250 ms, approximately when information retrieval from long-term memory is thought to occur (i.e., beginning during the N400 time period), demonstrating a rapid impact of domain knowledge on lexical/semantic retrieval. Moreover, this additional influence was reliable only for trials individuals reported *not* knowing. That is, domain knowledge had its greatest influence when retrieval difficulty was highest. This effect persisted beyond the N400 time period into a late positivity period; during a time period that has been associated with several high-level cognitive processes including retrieval of episodic/specific information (c.f. “old/new” effects, reviewed in Rugg & Curran, 2007), semantic reanalysis (Van Petten & Luka, 2012), and/or updating of a mental model (Brouwer et al., 2012).

At the item level, real-time reports of knowledge and offline cloze probability (which correlated at $r = .67$; Figure 3.1c,d) indicated substantial variability in how likely each item was to be known/remembered in real-time and in how easy/difficult it was to produce the final word of each item. We reasoned that this item-level variability might also contribute to retrieval difficulty and be modulated by domain knowledge. We thus performed a post-hoc regression analysis incorporating Q1 response type, participant-level HP domain knowledge, and item-level offline cloze probability (Supplementary Table 3.5). In the model, each of these predictors was significant, with N400 amplitudes reduced for high- compared to low-cloze items (Supplementary Table 3.5).

Importantly, there was a three-way interaction between Q1 response type, HP domain knowledge, and offline cloze: HP domain knowledge seems to have its greatest influence on low-cloze items individuals reported having known, and on high-cloze items individuals reported *not* having known (Supplementary Figure 3.2)—both reflecting less than optimal retrieval conditions—i.e., cases in which items are least likely to be certainly known (or unknown) and which seem most likely to experience retrieval difficulty.

The present study provides, for the first time, evidence that domain knowledge influences real-time (implicit) retrieval of word information during sentence processing—beyond determining how much (i.e. proportion of facts) an individual knows. Admittedly, the explanatory mechanism(s) for these findings are yet to be determined. One set of explanations suggests that individuals with greater domain knowledge have deeper and/or differentially functionally-organized knowledge; e.g., experts in several domains have been found to organize facts according to higher-order principles and "core concepts" (Chi & Ohlsson, 2005). Experts' domain-level semantic networks thus likely differ from those of novices in both content and network connectivity (Steyvers & Tenenbaum, 2005). The real-time differences in the present study similarly may result from individuals with greater knowledge being more likely to retrieve information *related* to (and perhaps relevant for) our experimental sentences/"facts." By activating relevant information during sentence processing, knowledgeable individuals may have enjoyed facilitated retrieval upon encountering critical words—even those reported as not known. Additionally or alternatively, more knowledgeable individuals may have used different criteria or thresholds for their trial-level knowledge reports or may otherwise have perceived task demands differently, leading to a parcellation of trials ("Yes" and "No" responses) that differed systematically as a function of domain knowledge. We recognize that the N400 time window could be contaminated by overlapping P300 potentials elicited by the task-related decisions made on the critical word (see Rohrbaugh, Donchin, & Eriksen, 1974), and which may have overlapped with N400 potentials in timing and scalp distribution. However, the current results (that HP domain knowledge positively correlates with the degree of reduction in N400 amplitude to contextually supported words in HP sentences) replicate findings from Troyer & Kutas (2018, in which the task did not require participants to make any overt decisions during the ERP study, but simply to read the sentences for comprehension. That the findings look so similar leads us to infer that the N400 effects are separable from task effects based on participant report. Whatever the precise explanatory mechanism(s), we suggest that a variable mediating the influence of HP domain knowledge seems to be the ease of (implicit) information retrieval from memory.

In sum, we investigated fine-grained influences of world knowledge on real-time sentence comprehension in a novel experimental design using state-of-the-art analyses on single-trial ERP data. For the first time, we were able to dissociate individuals' reports of knowledge of specific facts from their knowledge of a rich domain of world knowledge. Our findings illustrate that domain knowledge can have a rapid influence—by less than a third of a second—on retrieval processes during reading, especially in cases where retrieval is likely to be difficult.

3.6 Acknowledgments

This chapter comprises an article which is under review at the *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Troyer, Melissa, Urbach, Thomas P., & Kutas, Marta, for which I was the primary author and investigator.

3.7 References

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. Available from arXiv:1506.04967 (stat.ME).

Borovsky, A. & Creel, S. (2014). Children and adults integrate talker and verb information in online processing. *Developmental Psychology*, *50*(5), 1600-1613.

Boudewyn, M.A. (2015). Individual differences in language processing: electrophysiological approaches. *Language and Linguistics Compass*, *9/10*, 406-419.

Boudewyn, M.A, Long, D.L., & Swaab, T.Y. (2012). Cognitive control influences the use of meaning relations during spoken sentence comprehension. *Neuropsychologia*, *50*, 2659-2668.

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.

Chi, M.T.H. & Ohlsson, S. (2005). Complex declarative learning. In Holyoak, K.J. and Morrison, R.G. (Eds.), *The Cambridge Handbook of Thinking and Reasoning*. Cambridge, UK: Cambridge University Press, pp. 371-399.

Daneman, M. & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Memory and Verbal Behavior*, *19*(4), 450-466.

DeLong, K.D., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass*, *8*(12), 631-645.

Dunn, D.M. & Dunn, L.M. (2007). *Peabody picture vocabulary test*. Minneapolis, MN: NCS Pearson.

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

Ericsson, K.A., Charness, N., Feltovich, P.J., Hoffman, R.R. (Eds.). (2006). *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge, UK: Cambridge University Press.

Filik, R. & Leuthold, H. (2013). The role of character-based knowledge in online narrative comprehension: Evidence from eye movements and ERPs. *Brain Research*, *1506*, 94-104.

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

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

Hald, L.A., Steenback-Planting, E.G., Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Research*, *1146*, 210-218.

Kamide, Y., Altmann, G.T.M., & Haywood, S.L. (2003). The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133-159.

Kim, A.E., Oines, L., & Miyake, A. (2018). Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language comprehension: An ERP investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(3), 406-420.

Kutas, M., DeLong, K.A., & Smith, N.J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the Brain*. Oxford, UK: Oxford University Press, pp. 190-207.

Kutas, M. & Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. Trends in Cognitive Sciences, 4(12), 463-470. doi: 10.1016/S1364-6613(00)01560-6

Kutas, M. & Federmeier, K.D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(14), 1-27.

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703-704.

Nakano, H., Saron, C., & Swaab, T. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, *22*(12), 2886-2898.

Pakulak, E. & Neville, H. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, *22*(12), 2728-2744.

Rohrbaugh, J.W., Donchin, E., & Eriksen, C.W. (1974). Decision making and the P300 component of the cortical evoked response. *Perception & Psychophysics*, *15*(2), 368-374.

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

Smith, N. & Kutas, M. (2015a). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, *52*, 157-168.

Smith, N. & Kutas, M. (2015b). Regression-based estimation of ERP waveform: II. Nonlinear effects, overlap correction, and practical considerations. *Psychophysiology*, *52*, 169-181.

Stanovich, K.E. & Cunningham, A.E. (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. *Journal of Educational Psychology*, *85*(2), 211-229.

Stanovich, K.E. & West, R.F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*(4), 402-433.

Steyvers, M. & Tenenbaum, J.B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*, 41-78.

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(2), 367-382.

Troyer, M. & Kutas, M. (2017). It's all in your head: Effects of expertise on real-time access to knowledge during written sentence processing. In G. Gunzelman, A. Howes, T. Tenbrink, & E.J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 1212-1217). Austin, TX: Cognitive Science Society.

Troyer, M. & Kutas, M. (2018). Harry Potter and the Chamber of *What*?: The impact of what individuals know on word processing during reading. *Language, Cognition, and Neuroscience*, 1-17.

Trueswell, J.C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, *35*, 566-585.

Trueswell, J.C., Tanenhaus, M.K., & Garnsey, S.M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285-318.

Van Berkum, J.J.A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science*, *20*(9), 1092-1099.

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

3.7 Supplemental Information

3.7.1 Supplementary Tables

Supplementary Table 3.1. Mean, standard deviation, and range are provided for behavioural measures of individual differences.

	All participants			High HP knowledge subgroup			Low HP knowledge subgroup		
	Mean	95% CI	Range	Mean	95% CI	Range	Mean	95% CI	Range
HP Quiz	23.17	[20.91, 25.43]	[11, 38]	29.55	[27.68, 31.42]	[23, 38]	15.94	[14.61, 17.27]	[11, 21]
# of HP Books	1.67	[1.02, 2.33]	[0, 8.43]	2.54	[1.44, 3.64]	[0, 8.43]	0.86	[0.21, 1.51]	[0, 5.57]
HP Self Score	52.61	[40.78, 64.44]	[7, 157]	77.80	[60.40, 95.20]	[28, 157]	27.44	[19.34, 35.54]	[7, 62]
ART	0.17	[0.14, 0.19]	[0.05, 0.45]	0.18	[0.15, 0.22]	[0.08, 0.40]	0.15	[0.11, 0.20]	[0.08, 0.45]
MRT	0.18	[0.14, 0.21]	[0.02, 0.55]	0.16	[0.11, 0.21]	[0.02, 0.45]	0.18	[0.12, 0.24]	[0.02, 0.55]
# of Authors Listed	2.49	[1.98, 3.00]	[0, 5]	3.40	[2.90, 3.90]	[1, 5]	1.29	[0.67, 1.92]	[0, 4]
MRH Total	7.44	[6.61, 8.26]	[3, 14]	7.25	[6.09, 8.41]	[3, 11]	7.65	[6.38, 8.92]	[5, 14]
GKQ	19.02	[17.87, 20.18]	[11, 27]	20.70	[18.99, 22.41]	[13, 27]	17.00	[15.68, 18.32]	[11, 20]
CLC	0.29	[0.26, 0.32]	[0.11, 0.57]	0.31	[0.26, 0.35]	[0.15, 0.57]	0.26	[0.21, 0.32]	[0.11, 0.57]
MCLC	0.40	[0.36, 0.44]	[0.17, 0.73]	0.43	[0.37, 0.49]	[0.20, 0.73]	0.37	[0.29, 0.45]	[0.17, 0.73]
PPVT	206.76	[204.85, 208.67]	[198, 218]	208.80	[205.76, 211.84]	[199, 218]	204.71	[202.41, 207.00]	[198, 216]
Sentence Span[12]	2.86	[2.61, 3.11]	[2, 4.5]	3.29	[2.99, 3.59]	[2, 4.5]	2.56	[2.21, 2.92]	[2, 4.5]
Reading Experience	0.00	[-0.22, 0.22]	[-1.19, 2.30]	0.13	[-0.13, 0.38]	[-0.86, 1.20]	-0.19	[-0.58, 0.20]	[-1.19, 2.30]
General Knowledge	0.00	[-.24, 0.24]	[-1.44, 1.90]	0.27	[-0.06, 0.61]	[-1.20, 1.90]	-0.33	[-0.69, 0.04]	[-1.44, 1.73]

[12] Note: Two participants did not complete sentence span (one from each HP knowledge subgroup).

Supplementary Table 3.2. Intercorrelations (Pearson's *r*) among behavioral measures of individual differences. *r* values above .32 are significant at *a* = .05; *r* values above .40 are significant at *a* = .01.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 HP Quiz	1	0.63	0.73	0.11	0.01	0.50	-0.07	0.36	0.21	0.22	0.39	0.40	0.19	0.33
2 HP Books	-	1	0.70	0.07	0.14	0.20	0.13	0.18	0.12	0.13	0.07	0.27	0.19	0.18
3 HP Self Score	-	-	1	0.15	0.03	0.30	-0.03	0.22	0.02	0.14	0.06	0.21	0.15	0.16
4 ART	-	-	-	1	0.55	0.37	0.40	0.15	0.57	0.61	0.27	0.29	0.80	0.56
5 MRT	-	-	-	-	1	0.16	0.41	0.10	0.68	0.72	0.38	0.12	0.73	0.63
6 Authors Listed	-	-	-	-	-	1	0.37	0.22	0.38	0.21	0.22	0.27	0.65	0.34
7 MRH Total	-	-	-	-	-	-	1	0.11	0.27	0.19	-0.01	0.35	0.75	0.24
8 GKQ	-	-	-	-	-	-	-	1	0.31	0.28	0.4	0.39	0.20	0.67
9 CLC	-	-	-	-	-	-	-	-	1	0.72	0.64	0.24	0.65	0.85
10 MCLC	-	-	-	-	-	-	-	-	-	1	0.53	0.23	0.59	0.84
11 PPVT	-	-	-	-	-	-	-	-	-	-	1	0.30	0.29	0.66
12 Sentence Span	-	-	-	-	-	-	-	-	-	-	-	1	0.36	0.36
13 Reading Experience	-	-	-	-	-	-	-	-	-	-	-	-	1	0.61
14 General Knowledge	-	-	-	-	-	-	-	-	-	-	-	-	-	1

Supplementary Table 3.3. Statistics for fixed-effects predictors of mean ERP amplitude in the N400 time period in an analysis including HP knowledge and other individual differences measures.

Fixed effects	*β*	*SE*	*DF*	*t-value*	*p-value*
Intercept	4.60	0.43	50.38	10.60	.0000
Q1	0.77	0.16	4274.53	4.78	.0000
HP Knowledge	0.58	0.45	36.96	1.27	.2113
Reading Experience	-0.70	0.71	35.67	-0.98	.3329
General Knowledge	1.04	0.65	34.86	1.60	.1193
Verbal WM	-0.38	0.46	34.88	-0.83	.4143
Q1 : HP Knowledge	-0.45	0.18	4793.49	-2.48	.0131
Q1 : Reading Experience	-0.05	0.24	5231.59	0.20	.8387
Q1 : General Knowledge	0.00	0.22	5143.38	-0.02	.9877
Q1 : Verbal WM	-0.04	0.16	5040.15	-0.25	.8042

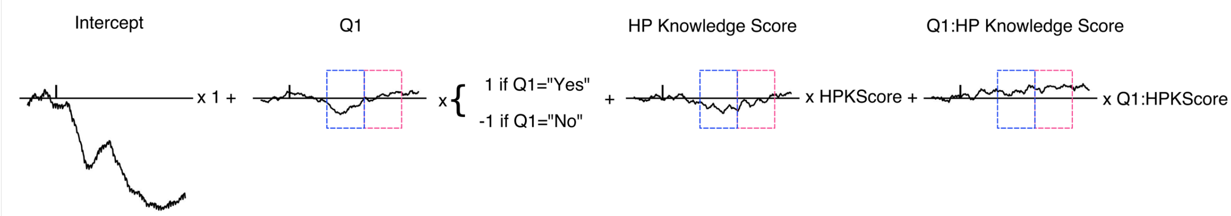
Supplementary Table 3.4. Statistics for fixed-effects predictors of mean ERP amplitude in the late positivity time period in an analysis including HP knowledge and other individual differences measures.

Fixed effects	*β*	*SE*	*DF*	*t-value*	*p-value*
Intercept	7.84	0.51	46.70	15.48	.0000
Q1	-0.13	0.17	4188.72	-0.74	.4598
HP Knowledge	-0.02	0.54	36.49	-0.04	.9661
Reading Experience	-1.17	0.84	35.49	-1.39	.1746
General Knowledge	0.68	0.77	34.80	0.87	.3879
Verbal WM	0.50	0.55	34.79	0.91	.3706
Q1 : HP Knowledge	-0.73	0.20	5008.47	-3.76	.0002
Q1 : Reading Experience	0.35	0.26	5264.89	1.36	.1728
Q1 : General Knowledge	-0.14	0.24	5214.60	-0.57	.5656
Q1 : Verbal WM	-0.18	0.17	5145.68	-1.02	.3082

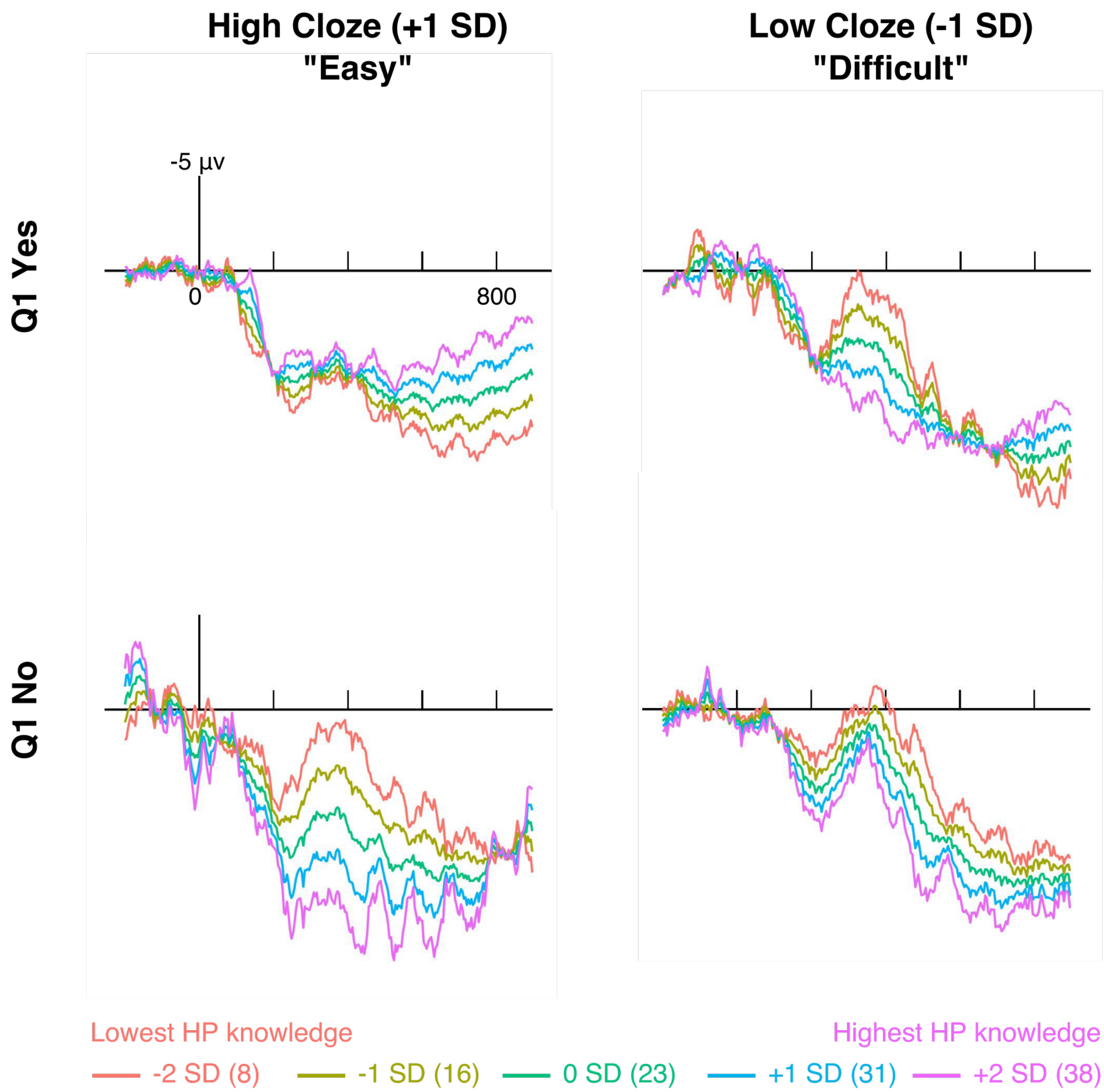
Supplementary Table 3.5. Statistics for fixed effects predictors of mean amplitude in the N400 time period for ROI analyses incorporating offline cloze probability.

Fixed effects	*β*	*SE*	*DF*	*t-value*	*p-value*
Intercept	2.38	0.64	197	3.71	.0003
Q1 Response	2.06	0.38	4581	5.45	.0000
HP Knowledge	1.34	0.51	130	2.62	.0100
Cloze	5.32	1.07	279	4.98	.0000
Q1 : HP Knowledge	0.94	0.37	5676	2.54	.0111
Q1: Cloze	-3.71	0.80	4762	-4.64	.0000
HP Knowledge : Cloze	-0.62	0.78	5697	-0.79	.4312
Q1 : HP Knowledge : Cloze	-3.27	0.78	5665	-4.21	.0000

3.7.2 Supplementary Figures



Supplementary Figure 3.1. Using the rERP framework, these waveforms show the unsmoothed, timepoint-by-timepoint fitted beta coefficients for the fixed effects of a mixed-effects linear regression model predicting centro-parietal ROI voltage based on Q1 response, HP knowledge score, and their interaction, along with an intercept term. The dashed boxes indicate the time periods that were used for N400 (blue) and late positivity (pink) analyses. A linear combination of each waveform can be used to predict the waveform for a hypothetical trial (i.e., a trial on which a hypothetical participant with some value of HP knowledge responds with either “Yes” or “No”), as in Figure 3.5 (bottom).



Supplementary Figure 3.2. rERP model-predicted ERPs are plotted for the centro-parietal ROI by the two possible Q1 response types (Yes or No), hypothetical cloze values of +/- 1 SD from the mean, and HP knowledge for five hypothetical participants with z-transformed HP knowledge scores ranging from -2 to 2 SD, where 0 represents the mean HP score of the sample (corresponding raw HP knowledge scores are provided in parentheses).

CHAPTER FOUR: TO CATCH A SNITCH: BRAIN POTENTIALS REVEAL VARIABILITY IN THE FUNCTIONAL ORGANIZATION OF (FICTIONAL) WORLD KNOWLEDGE DURING READING

4.1 Abstract

We harnessed the temporal sensitivity of event-related brain potentials alongside individual differences in Harry Potter (HP) knowledge to investigate the extent to which the availability and timing of information relevant for real-time word comprehension are influenced by variation in domain knowledge. We manipulated meaningful (category, event) relationships between sentence contexts about HP stories and critical words (HP sentence endings). During reading, N400 amplitudes to (a) linguistically supported and (b) unsupported but meaningfully related, but not to (c) unsupported and unrelated sentence endings reflected variability in domain knowledge. Moreover, single-trial analyses revealed that N400s to only words that were linguistically supported varied as a function of whether individuals knew (or could remember) the correct (supported) ending to each HP "fact." We conclude that the quick availability of information relevant for word understanding varies as a function of individuals' domain knowledge, beyond knowledge of specific facts.

4.2 Introduction

Across cognitive systems, world knowledge allows individuals to organize raw sensation into meaningful experiences. For example, chess players who become experts know a wide variety of chess positions. Accordingly, they show better performance on memory tasks involving chess pieces compared to novices, but only when the chess pieces occupy real game positions, and not when they are randomly shuffled on the board; their knowledge allows them to make sense of the meaningful arrangements (Simon & Chase, 1973). Understanding language is no exception—words cue world knowledge which can be rapidly brought to mind in real time (e.g., Hagoort, Hald, Bastiaansen, & Petersson, 2004), incrementally (Altmann & Kamide, 1999) and sometimes even predictively (reviewed in Kutas, DeLong, & Smith, 2011). The goal of the current study is to develop a more precise description of how this might occur—including which types of knowledge, their organization, and the timing of their use—by closely examining knowledge use in real time in word by word sentence reading.

Much of our understanding of the nature of information that people bring to mind during real-time, moment-by-moment sentence processing comes from online methods such as eye-tracking and event-related brain potentials (ERPs). The earliest ERP studies of word-by-word reading revealed that the N400 brain potential, a negative-going brain potential peaking between 250-500 ms with a centroparietal distribution over the scalp, was quite a sensitive measure of word- and sentence-level meaning (Kutas & Hillyard, 1980; Kutas & Hillyard, 1984; reviewed in Kutas & Federmeier, 2011). N400 amplitude is sensitive to the relationship between an incoming word and its preceding sentence (or any meaningful) context, being smallest for words which are contextually supported and largest for unsupported or anomalous words (Kutas & Hillyard, 1980; reviewed in Kutas & Federmeier, 2011). It is also reduced in amplitude (i.e., facilitated) for words which are anomalous but are related to the overall context or to a highly predictable upcoming word via a category (Federmeier & Kutas, 1999), perceptual relationship (Rommers, Meyer, Praamstra, & Huettig, 2013; Amsel, DeLong, & Kutas, 2015), event relationship (Metusalem et al., 2012; Paczynski & Kuperberg, 2012; Amsel et al., 2015), or other semantic relationship (Kutas & Hillyard, 1984; DeLong, Chan, & Kutas, 2018). Such findings using a so-called "related anomaly" paradigm have been taken as evidence that information cued by the related words is more readily accessible compared to the contents of memory cued by unrelated and anomalous words.

Federmeier and Kutas (1999) used a related anomaly paradigm to examine whether processing words in real time would be sensitive to the categorical organization of concepts in semantic memory. For example, in sentence contexts setting up an expectation for the word *pines*, categorically related words (e.g., *palms*) elicited reduced N400 amplitudes compared to words from a different category (e.g., *tulips*), but which were larger than those to the expected word. Moreover, the related anomaly effect (i.e., the difference between the N400 to related vs. unrelated anomalies) was strongest for words completing high-constraint sentences. They interpreted these findings as evidence that people pre-activate a highly predictable upcoming word's semantic features, which strongly overlap with those of the categorically related word (e.g., palms and pines both have leaves and trunks), thereby facilitating semantic retrieval for the related words.

Metusalem and colleagues (2012) used a related anomaly paradigm to investigate the extent to which individuals made use of generalized event knowledge to bring linguistically unlicensed, yet contextually relevant, information to mind. In their study, individuals read short paragraphs about common events (e.g., playing football) that set up linguistic expectations for a word like *touchdown*. They found that unexpected and linguistically unlicensed words related to the event being described (e.g., *helmet*) elicited reduced N400 amplitudes compared to unrelated words (e.g., *license*). They interpreted these findings as evidence that people make use of their event knowledge (also referred to as *schema* (van Dijk & Kintsch, 1983; Kintsch 1988), *scripts* (Schank & Abelson, 1977), and *situation models* (Foss, 1982; Hess et al., 1995; Kintsch, 1988)) as they understand words in sentences as real time, (pre-)activating concepts that are related / likely to be relevant even when they are not likely to be immediately mentioned.

In contrast to the Federmeier & Kutas study, in the Metusalem et al. experiment, there was only limited overlap in semantic features between the contextually supported word and the related anomaly. A helmet and a touchdown do not share many semantic features—they are related by virtue of being associated with the same contexts/events (football games). Nonetheless, the related anomaly ERP effects in both studies had a similar timecourse and scalp topography, maximal around 400 ms over centro-parietal recording sites, suggesting that people made quick use of both types of related information during real-time sentence processing.

These results, among others, imply that real-time language processing draws on rich, structured representations stored in long-term memory, and accessing these representations seems to be a normal part of processing words in context. When viewed with this lens, it is obvious that how much and how well individuals know things must impact what information is available for each individual to bring to mind during real-time comprehension as well as how it the information is organized semantic (i.e., long-term) memory (see Yee, Jones, & McRae, 2017). Yet to date, models of real-time language processing have not taken such knowledge variability into account, with studies of individual differences focusing more on general cognitive abilities (e.g., working memory, cognitive control; Nakano, Saron, & Swaab, 2010; Boudewyn, Long, & Swaab, 2012; Kim, Oines, & Miyake, 2018) and variation in language proficiency in a first or second language (Pakulak & Neville, 2010; Tanner & Van Hell, 2014). One reason for this disconnect may be that it is experimentally challenging to capture the specifics of an individual's world knowledge with standard laboratory procedures. A potential solution is to focus on a restricted domain of knowledge, as is often done in the extensive literature on the psychology of expertise (Ericsson, Charness, Feltovich, & Hoffman, 2006).

Troyer & Kutas (2018) focused on a domain of knowledge with the requisite properties for online language processing studies, including a large, rich set of verbal descriptions, wherein college-aged young adults differed in their degree of knowledge—the fictional world of Harry Potter (HP) by J.K. Rowling. They recorded EEG while participants with varying degrees of knowledge about HP first read sentences that described general topics, and then read sentences that described events from the HP stories; sentences ended either in contextually supported or unsupported words. Across participants, and for both sentence types, the effect of contextual support was present on N400 amplitudes, reflecting greater ease of retrieval for the supported words. But critically, participants' degree of HP knowledge influenced the size of this effect only for the sentences about HP, with more knowledgeable individuals showing large effects and less knowledgeable individuals showing small effects. These results demonstrate empirically that the rapid influence of written sentence context, known to modulate N400 brain potentials, is a function of each individual's knowledge.

In a follow-up study, Troyer, Urbach, and Kutas (under review) focused on the relationship between domain knowledge and real-time retrieval of contextually supported words, asking whether this relationship was strictly due to the proportion of items an individual knew, or whether, rather, domain knowledge had an influence beyond increasing the likelihood an individual would know any given HP "fact." To that end, they conducted single-trial analyses in a paradigm in which participants reported whether they had known each fact, just after reading it. In this study, all sentences were about HP and all ended in a supported (or "correct") word. Results showed that HP domain knowledge was a strong predictor of whether an individual trial was reported as known; but there was an interaction between HP domain knowledge and participant report ("known" vs. "unknown") such that domain knowledge had its greatest influence on N400 amplitude for trials individuals reported *not* having known. Post-hoc analyses indicated that the influence of domain knowledge on N400 amplitude was also mediated by cloze probability, such that domain knowledge had its greatest influences on trials for "low-cloze" items (i.e., those with an ending that fewer participants were able to provide in an offline cloze norming task) which participants reported knowing, as well as on "high-cloze" trials that individuals reported as *not* knowing. These instances likely reflect moments where complete retrieval of word information might be difficult, although retrieval of some word information is possible. Together, these findings converge to suggest that domain knowledge may have its greatest influence under difficult retrieval conditions, and that domain knowledge seems to have an influence on real-time retrieval beyond strictly mediating the likelihood that the fact is known to the comprehender. Troyer et al. hypothesized that domain knowledge might ease (implicit) semantic retrieval of information relevant for word understanding by virtue of its organization, presumably differing across individuals with varying degrees of knowledge.

This interpretation seems reasonable given the vast literature showing that people rapidly make use of a variety of word and world knowledge as they understand words in real time, such as orthographic neighborhood density (Laszlo & Federmeier, 2009), word frequency (Van Petten & Kutas, 1990), and non-linguistic knowledge including the organization of categories in semantic memory (Federmeier & Kutas, 1999), facts about the world (Hagoort et al., 2004), generalized event knowledge (Metusalem et al., 2012), personal preferences (Coronel & Federmeier, 2016), and fictional characters (Filik & Leuthold, 2013). It stands to reason that the structure and organization of individuals' knowledge would impact the availability, contents, and timecourse of bringing to mind these varied sources of knowledge in real time. Indeed, the literature on expert knowledge proposes that the functional organization of information around themes, events, and categories is likely to depend on individuals' degree of expertise (reviewed in Ericsson et al., 2006). We therefore designed an experiment to examine whether domain knowledge influences quick access to information that is functionally related to the sentence content via well-attested organizational structures of semantic memory, including category and event relationships.

4.2.1 The current study

We hypothesized that domain knowledge might systematically influence the functional organization of information stored in long-term memory such that individuals with greater knowledge would enjoy access to more and/or richer information relevant for processing words in sentences—for example, access to more semantic features for incoming words and/or more concepts relevant to processing the sentence's content. To probe such information, we tested individuals who varied in their knowledge of the narrative world of Harry Potter using a related anomaly paradigm applied to sentences describing HP. Using freely available materials (including Wikipedia and HP fan sites) along with the text of the HP book series by J.K. Rowling, the first author created a set of 156 sentence pairs that accurately described events and entities from the series. Each sentence context ended either in (a) a contextually *Supported* (and linguistically probable) word; (b) a word which was factually incorrect and *Unrelated* to the context and to the supported word; and (c) a word which was factually incorrect but which was *Related* to the context and/or contextually supported word in one of two ways. For half of the materials, the related words were taken from the same category as the linguistically expected word, as in Federmeier & Kutas (1999). For the remaining materials, the related words were in some way related to the episode/event being described by the preceding sentence context, as in Metusalem et al. (2012).

To establish that all participants (regardless of degree of HP knowledge) would exhibit well-attested effects of sentential context on ERPs to contextually supported vs. unsupported words in sentences about general topics, all participants first took part in a control ERP reading experiment (as in Troyer & Kutas 2018). We anticipated that individuals of both high and low HP knowledge would show similar N400 effects of contextual support for the control sentences.

If individuals' degree of domain knowledge influences how information is organized and used in real time, we would expect individuals to be more sensitive to the related anomaly manipulation as a function of their degree of HP knowledge. Specifically, we would expect a reduction in N400 amplitude for contextually related words compared to unrelated words, with larger effect sizes for individuals with greater HP knowledge. In addition, we expected that the size of the contextual support effect (i.e., N400 amplitude to supported words compared to unrelated words) would vary with HP knowledge, being largest for individuals with the greatest knowledge, in line with previous studies.

We were also interested in whether the ability to (pre-)activate contextually relevant (though linguistically unlicensed) word information (as in the related anomaly condition) would rely on knowing each HP fact—i.e., knowing the contextually supported (correct) word completing the HP sentence pair. To measure this knowledge, we conducted a post-ERP-experiment modified cloze task where participants re-read the same 156 sentence pairs, however, with the final word removed. Participants were instructed to provide the word which they thought best completed the sentence—*not* necessarily the word that they had read during the EEG study. If individuals did not know what word to provide, they were instructed to instead make a judgment about whether or not they had seen the correct word during the study. This information allowed us to sort trials from the ERP experiment according to both experimental condition (Supported, Related, Unrelated) and the participant's ability to supply the appropriate word (Correct, Incorrect). We expected that individuals would demonstrate greater overall accuracy for final words of sentences they had seen in the Supported condition (because they had just read the correct word during the ERP study). We also expected that, for the Supported words, individuals' accuracy on the modified cloze task would be reflected in ERPs, with a reduction in N400 amplitude to Supported words for Correct compared to Incorrect trials. Moreover, we expected that for the Unsupported words, individuals' accuracy on the modified cloze task would matter little. Our critical test was for the Related words. If the ability to access related (linguistically unlicensed) word information during sentence processing depends on knowing the appropriate word, then we would expect a reduction in N400 amplitude for Related trials which were Correct, compared to Incorrect. If, however, the ability to access such information does not depend on knowing a specific (correct) lexical item, then we would expect no or little difference in N400 amplitude to Related trials as a function of trial correctness on the post-ERP-experiment task.

4.3 Method

4.3.1 Participants

53 UCSD students (mean age = 20, range = 18-25; 39 women, 14 men) took part in the study for partial course credit or payment of $9 / hour. Of these participants, 5 were excluded from data analysis due to excessive artifacts in the EEG, primarily due to eye movements. Data from 48 participants were included in our analyses. All participants provided informed consent reviewed by the Institutional Review Board at the University of California San Diego. To ensure that some participants would have high knowledge of the Harry Potter domain, a subset of participants (N = 18) was recruited via an announcement that specifically required participants to have read all of the Harry Potter books and/or have seen all of the Harry Potter films. Regardless of recruitment method, all participants were informed that they would be reading some sentences about Harry Potter before the start of the study.

4.3.2 Materials

4.3.2.1 Sentence material construction

During the EEG portion of the experiment, participants read sentences pairs in two experiments, which were analyzed separately (see Appendix A for all sentence materials). We included a short control experiment to verify that participants in the study would exhibit typical N400 effects to unsupported compared to supported words in sentential contexts. In the control experiment, participants read 40 sentence pairs, half ending with a contextually Supported word and half ending with an Unsupported word. Next, participants completed five blocks of the Harry Potter (HP) experiment, with sentence pairs ending in contextually Supported, unsupported but Related, or unsupported and Unrelated endings (described below).

4.3.2.1.1 Control experiment

40 sentence pairs described everyday topics and events. Due to time constraints, these sentences were a subset of the control sentences used in Troyer & Kutas (2018). On average, the initial sentence was 6 words long (range = 3 to 13 words); the average second sentence was 8 words long (range = 5 to 11 words). All sentence pairs were highly constraining (mean Cloze of best completion = 94%; range = 89-100%). For control sentences, supported words were defined as the best completion. To create unsupported words, plausible continuations were selected that were semantically related to the best completion but were never produced during Cloze norming (see Table 4.1 for examples). Two lists were created such that each participant only read one version of each sentence (ending in a word that was either supported or unsupported).

Table 4.1. Sample experimental stimuli.

Control sentences

	Sentence frame	Supported	Unsupported
1	We had been watching the blue jay for days. The bird laid her eggs in the	nest	yard
2	The grade-schoolers stood on the corner and waited. They rode to school on the	bus	train
3	The vampire moved in. He bit his victim on the	neck	shoulder
4	For his second birthday, Leonard got a stuffed sheep. His mother baked him a big chocolate	cake	pie
5	The detective arrived at the office. Within minutes he spilled his thermos full of	coffee	tea
6	Alicia's first client was a failure. But her second was a	success	triumph

Harry Potter sentences

	Sentence frame	Supported	Related	Unrelated
7	There is one main bank in the wizarding world. It is run by	goblins	werewolves	Alohomora
8	Sybill Trelawney is a Hogwarts professor. She teaches	Divination	Transfiguration	basilisk
9	There is a branch of magic focused on changing the form of objects. It is called	Transfiguration	Divination	Alley
10	Ginny opens the Chamber of Secrets and unleashes an ancient monster. It turns out to be a	basilisk	diary	Shunpike
11	Professor McGonagall recruits Harry for the Gryffindor Quidditch team. She saw him save Neville's	remembrall	broomstick	dog
12	Before Harry's second year, he is rescued by Ron, Fred, and George Weasley. They pick him up in a flying	car	Dursleys	Draco

4.3.2.1.2 Harry Potter experiment

156 Harry Potter (HP) sentence pairs were constructed as follows. Using freely available materials (including Wikipedia and Harry Potter fan sites) along with the text of the Harry Potter books, the first author created a set of sentence pairs that accurately described events and entities from the Harry Potter book series. A subset of these sentence pairs (*N* = 114) were taken directly from the HP sentences used in Troyer & Kutas (under review). On average, the initial sentence was 10 words long (range = 4 to 18 words); the second sentence was 7 words long (range = 3 to 16 words). The final word of these sentences was designed to be 100% predictable given perfect knowledge of the book series (this was termed the "supported" word). To verify that this was the case, a norming study was conducted on a separate group of participants. This group included some participants who were highly knowledgeable about the world of Harry Potter (determined by a trivia quiz; see "Harry Potter Quiz" section under "Additional materials"). Participants provided a final word for each sentence. Across these participants, mean Cloze for supported words was 39% (range = 0-92%).

For each of the 156 sentences, two additional versions were created by replacing the final (supported) word for a total of three versions with Supported, Related, and Unrelated endings. Critical words in the Related condition were related to the context in one of two ways—either by categorical relation to the supported word (*N*=78; examples 7-9 in Table 4.1) or by association to the overall event/episode being described (*N* = 78; examples 10-12 in Table 4.1); Unrelated words were taken from the critical words used elsewhere. Three lists were then constructed so that every participant read each sentence frame and each critical word only once. That is, even though the same critical word appeared in other conditions on other lists, it never appeared in the critical position more than once in the same list. All but three words appeared as critical words in two or all three conditions.

4.3.2.2 Sentence material norming

To verify that the words we deemed related via category or event were indeed more closely related to our sentence contexts/supported words than the unrelated ending, we conducted a series of experiments to examine these relationships. First, we trained a high-dimensional semantics/language model (Google's word2vec) directly on the text of the HP book series; we then asked whether the word embeddings learned by the model reflected the manipulation in our materials (e.g., with Supported-Related word embeddings being closer in semantic space than Supported-Unrelated word embeddings). Next, we conducted two experiments asking participants of varying degrees of HP knowledge to rate critical words from our materials for their similarity and relatedness, respectively.

4.3.2.2.1 Distributed semantics model

We trained a word2vec model[13] (Mikolov, Chen, Corrado, & Dean, 2013; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) on the text from the seven books of the HP series, taken from the official electronic publication[14]—a total of 1,125,854 words, with a vocabulary size of 8,046 words (subject to the constraint of each word appearing at least 5 times in the HP books). This model uses a neural net to learn word embeddings (vectors) in high-dimensional semantic space from word co-occurrences in the input. The semantic "contents" of such embeddings can reflect various aspects of meaning, including category and event-based relationships (reviewed in Lenci, 2018). We could then use these embeddings to quantify relative similarities/differences

[13] We used the distribution by D. Yaginuma, https://github.com/dav/word2vec.
[14] Available at https://usd.shop.pottermore.com.

between word pairs (or average vectors computed over sequences of words). We used the continuous bag-of-words (CBOW) architecture, which learns to predict a word based on its context—in our case, a window of 10 words on either side. Each word from the HP books was modeled as a point (i.e., vector) in a 200-dimensional space, subject to the constraint that it appeared at least five times across all seven books.[15]

Using this model, we extracted word embeddings for critical words from each of our experimental conditions (Supported, Related, and Unrelated). For each item (156 total), and computed the cosine similarity (angular distance) between the word embeddings for Supported-Related and Supported-Unrelated pairs of critical words. As expected, we found that word embeddings for Supported words were more similar to Related words than to Unrelated words (Figure 4.1a), $t(148) = 7.19$, $p < .0001$. This pattern held for both category- and event-related item subsets, although the size of the effect (i.e., the cosine similarity for Supported-Related pairs minus the cosine similarity for the Supported-Unrelated pairs) was larger within category-related items ($t(157.52)$, $p < .005$).

We also extracted word embeddings for each word (where possible) of our sentence pair frames/contexts. To create a single embedding (i.e., vector) for each item's sentential context, we took the average of all its words' vectors. We then computed the cosine distance between this aggregate context vector and the vector for each ending type (Supported, Related, Unrelated). An ANOVA revealed a significant effect of similarity type, $F(2, 296) = 39.87$, $p < .0001$; as expected, average word embeddings for sentential contexts were most similar to Supported words, followed by Related and Unrelated words (a three-way difference, all ps < .0001). This pattern was similar

[15] Because of this constraint, for the word2vec analyses, we excluded items in which one of the critical words (i.e., any of the Supported, Related, or Unrelated words for that item) appeared fewer than five times across the HP book series (a total of five items).

for category- and event-related item subsets (Figure 4.1b). These findings show that the high-dimensional semantic space learned by the word2vec model captured systematic, meaningful differences in the relationships between the sentence context and the Supported, Related, and Unrelated endings.

4.3.2.2.1 Similarity and relatedness norming studies

To further assess the manipulation in the HP sentences, and to examine the extent to which the manipulation was dependent on HP knowledge, we conducted two behavioral studies, asking participants to rate critical word-pairs (Supported-Related and Supported-Unrelated) on similarity (N=24) or relatedness (N=23, separate group of participants). These criteria were chosen specifically to examine the two types of relationships we targeted in our HP sentence materials, namely categorical relationships (words share many similar features) and event relationships (words are related via an event/episode from the HP books). In addition, these experiments allowed us to assess the ratings of similarity/relatedness as a function of individuals' degree of HP knowledge.

For the similarity ratings experiment, participants were asked to consider word-pairs in the context of the Harry Potter stories and to judge their similarity in meaning using a scale ranging from 1 ("not similar at all") to 7 ("nearly the same meaning"). For the relatedness ratings experiment, instructions were similar, except participants were asked to judge words on how related they were using a scale ranging from 1 ("not related at all") to 7 ("very closely related)". For both tasks, participants were given specific guides as to how to judge similarity and relatedness, respectively (see Appendix B). For each norming study, participants also completed a 10-question trivia quiz assessing their HP knowledge and a questionnaire about their HP experience (see descriptions below, under "Additional tasks").

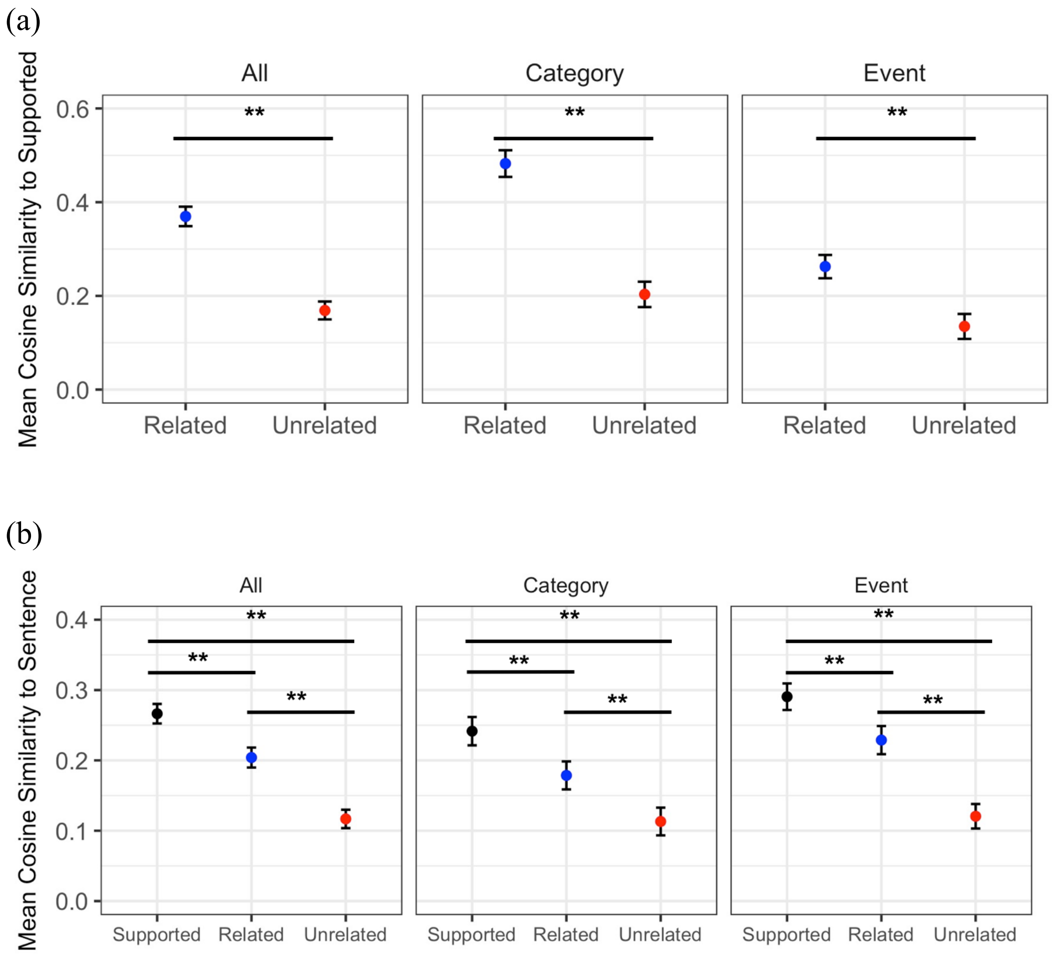
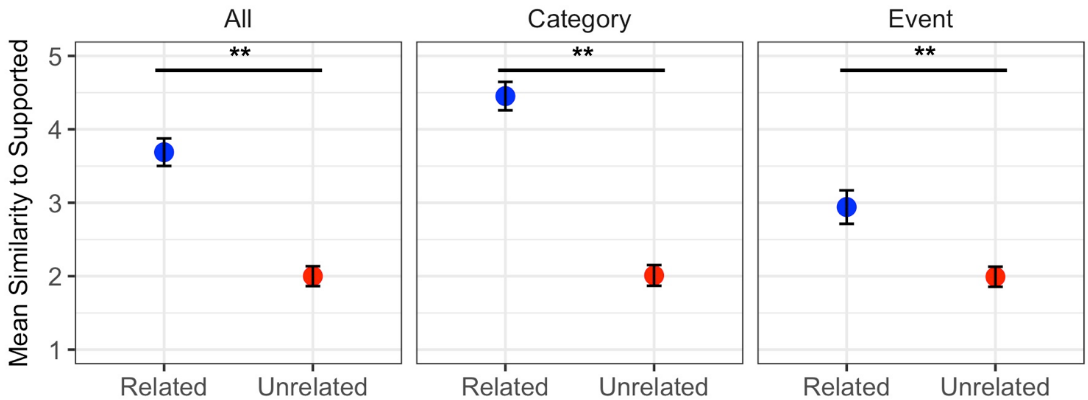


Figure 4.1. (a) Across items and within each subset (Category-related, Event-related), mean cosine similarity for Supported & Related endings is greater than for Supported & Unrelated endings. (b) For all 156 items (and within each subset) there was a significant three-way difference between cosine similarity of averaged word embeddings for sentences and Supported < Related < Unrelated endings.

As expected, mean similarity ratings for Supported-Related word pairs were greater than those for Supported-Unrelated word pairs, $t(23) = 11.19$, $p < .0001$ (Figure 4.2a). This pattern was similar for both the category- and event-related item subsets, but was larger for the category-related subset, $t(41.82) = 5.67$, $p < .0001$. Also as expected, HP knowledge was positively correlated with the size of the effect (i.e., similarity for Supported-Related word pairs minus similarity for Supported-Unrelated word pairs) at $r = .51$, $p < .05$.

(a)



(b)

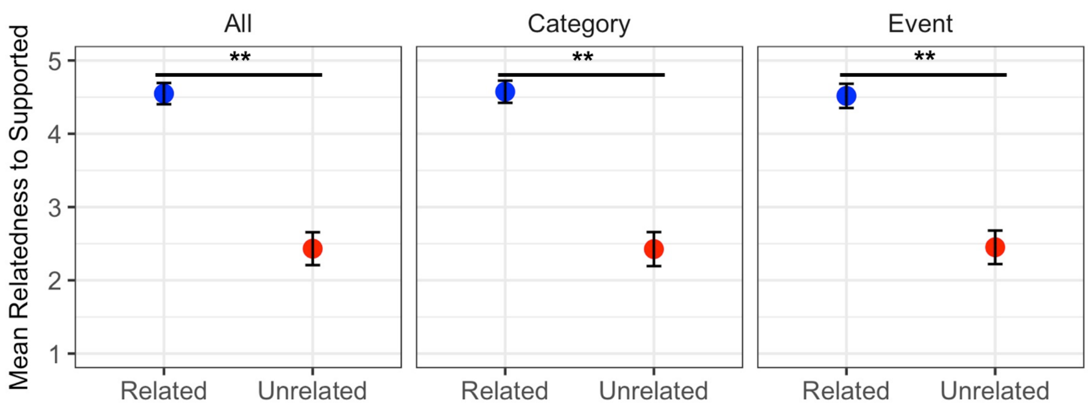


Figure 4.2. (a) Across items and within each subset (Category-related, Event-related), similarity ratings for Supported & Related endings are greater than Supported & Unrelated; this effect was larger for the category-related subset of items compared to the event-related subset. (b) Across items and within each subset, relatedness ratings for Supported & Related endings are greater than Supported & Unrelated.

In addition, mean relatedness ratings for Supported-Related word pairs were greater than those for Supported-Unrelated word pairs, $t(22) = 10.67$, $p < .0001$ (Figure 4.2b). This pattern was similar for both the category- and event-related item subsets. Also as expected, HP knowledge was positively correlated with the size of the effect (i.e., relatedness for Supported-Related word pairs minus relatedness for Supported-Unrelated word pairs) at $r = .68$, $p < .001$.

We also examined the correlation between the word2vec cosine similarity measures and the similarity and relatedness ratings for Supported-Related and Supported-Unrelated word pairs, respectively. Cosine similarity was positively correlated with both similarity ($r = .43$, $p < .0001$) and relatedness ($r = .26$, $p < .01$) ratings for the Supported-Related word pairs, but not for the Supported-Unrelated word pairs (n.s.).

These results empirically indicate that our Supported sentence endings were indeed more similar/related to our Related, compared to Unrelated, endings. Moreover, that the size of these effects was positively correlated with HP knowledge further supports the notion that sensitivity to the relatedness manipulation depends on knowledge specific to the HP book series.

4.3.2.3 Additional tasks

Our primary behavioral measure of interest was performance on a trivia quiz about HP, which served as our measure of HP knowledge (see description below). In addition, we collected several other measures of individual differences to better understand group differences among participants (see Troyer & Kutas, 2018, for more details). These included an additional measure of Harry Potter experience (the self-report questionnaire), measures of general print/reading experience (media and reading habits questionnaire (MRH), author and magazine recognition tests (ART/MRT), Stanovich & West, 1989), a measure of general knowledge (trivia quiz developed from freely available materials), and verbal working memory (sentence span, Daneman & Carpenter 1980). Finally, we administered a debriefing questionnaire.

4.3.2.3.1 Harry Potter quiz

Our primary measure of participants' knowledge about the domain of Harry Potter was computed based on a HP trivia quiz. Participants answered 10 multiple-choice questions about HP; for example, *To gain access to the kitchens, one must tickle the following fruit: (a) Pear, (b)*

Orange, (c) Grape, (d) Banana. HP quiz score (henceforth "HP knowledge") was the number of correct answers out of ten. For regression analyses, we z-transformed these scores.

4.3.2.3.2 HP self-report questionnaire

Participants answered questions about their experience with the Harry Potter book series, movies, and other media. As an estimate of overall experience with Harry Potter, a numeric score was determined by summing the total number of times an individual had read each book, seen each movie, listened to each audiobook, and so on.

4.3.2.3.3 Aggregate measure of reading experience

An aggregate measure of reading experience was based on an average of z-transformed ART scores, MRT scores, MRH score, and number of favorite authors listed on the MRH.

4.3.2.3.4 Debriefing questions

Immediately after completing the post-EEG experiment modified cloze task, participants answered two additional questions about these materials. First, they were asked to indicate, on a sliding scale of 1 to 100, what proportion of sentences during the brainwave study were *true* (i.e., accurately represented events that took place in the Harry Potter books/movies). Next, they were asked to indicate, on the same scale, what proportion of the *true* sentences they believed they had known *before* reading them during the study.

After completing all of the additional tasks, participants responded to a short questionnaire asking what had caught their attention during the study, what the experiment was trying to determine, and what parts of the experiment were particularly easy or difficult. The researcher (the first author) then offered to describe main goals of the study and to answer any questions.

4.3.3 Procedures

4.3.3.1 Ordering of tasks

During set-up for the EEG experiment, participants completed the ART and MRT. Immediately after the EEG portion of the study, participants completed the media and reading habits questionnaire. After clean-up, participants moved to a separate room to complete the remainder of the tasks, in this order: post-EEG modified cloze production task; two debriefing questions; HP self-report questionnaire; HP quiz; general knowledge quiz; reading span; remainder of debriefing questions.

4.3.3.2 EEG experiment

Before the study began, participants were asked to remain relaxed and still to minimize muscle artifact. They were told that they would be reading short, two-sentence stories (with the first block about general topics, followed by five blocks about the world of Harry Potter) for meaning and that they would be asked questions about what they read at the end of the EEG recording session.

During the EEG experiment, participants sat approximately 100 cm in front of a cathode-ray tube monitor. The background of the screen was black and words were presented in white type. Each trial began with a blank screen for two seconds. Then, the first sentence of each pair was presented until the participant pressed a button to advance to the next sentence. After their button press, a crosshair appeared in the center of the screen for a duration which varied randomly between 900 and 1100 ms. Participants were instructed to focus on the crosshair and to minimize eye movements during presentation of the second sentence. The second sentence was then presented one word at a time right above the crosshair. Each word was presented for 200 ms with an interstimulus interval of 300 ms. After the last word disappeared, the crosshair stayed on the screen for a time that varied randomly between 900 and 1100 ms. Control sentences were presented in a single block, followed by five blocks of HP sentences. Within each block of the study, sentences of different ending types (control: Supported, Unsupported; HP: Supported, Related, Unrelated) were randomly interspersed.

4.3.3.3 EEG recording

The electroencephalogram (EEG) was recorded from 26 electrode sites arranged geodesically in an Electro-cap (as described in Ganis, Kutas, & Sereno, 1996). For all cap electrodes, online recording was referenced to the left mastoid; these electrodes were re-referenced offline to an average of the left and right mastoid. Electrodes were placed lateral to the outer canthus of each eye to create a bipolar recording used to monitor eye movements. Electrodes placed under each eye were referenced to the left mastoid and were used to monitor blinks. Throughout the experiment, all electrode impedances were maintained under 5 kΩ. The signal was amplified with Grass amplifiers which were set at a bandpass of .01 to 100 Hz; the sampling rate was 250 Hz.

4.3.3.4 EEG data analysis

Trials contaminated by eye movements, blinks, muscle activity, blocking, or other artifact were removed from subsequent analysis. This resulted in an exclusion of 15% of trials: 18% control-Supported, 18% control-Unsupported, 15% HP-Supported, 15% HP-Related, 15% HP-Unrelated. ERPs were created by averaging from 200 ms before the onset of a critical word until 900 ms post-critical word. Then, for each electrode, a baseline was computed by averaging potentials from 200 ms before the word to the start of the word; this baseline was subtracted from the waveform.

We analyzed the control and HP data separately. First, we characterized effects of contextual support (for both studies) and of related anomaly effects (HP only) across all participants. We conducted a traditional, whole-head analysis for each prior to examining any individual differences. We were primarily interested in a time period surrounding the typical peak of the N400 brain potential from 250 to 500 ms post-stimulus. We also examined a late positivity time period from 500 to 750 ms post-stimulus. In each case, we subjected participants' mean amplitudes of ERP waveforms in these time periods to a whole-head ANOVA, including repeated measures of electrode (26 levels) and ending type (control: 2 levels; HP: 3 levels) as well as a between-subjects factor of list (control: 2 levels; HP: 3 levels). For all ANOVAs, we applied the Greenhouse-Geisser epsilon correction for *F*-tests with more than one degree of freedom in the numerator. We report the corrected p-value, unadjusted degrees of freedom, and value of the Greenhouse-Geisser epsilon.

Our primary research question centered on whether HP knowledge would differentially influence the neural response to words of differing functional relationships to the sentential context. We therefore examined the relationship between HP knowledge and ending type in an ROI where N400 effects are typically largest, averaging mean amplitude between 250 and 500 ms across eight centro-parietal electrodes (MiCe, LMCe, RMCe, MiPa, LDPa, RDPa, LMOc, and RMOc) for each sub-experiment (i.e., for control and HP sentences, separately). Here and elsewhere, HP knowledge was defined as the z-transformed performance on the HP knowledge quiz. For these analyses, we used hierarchical mixed-effects linear regression (Baayen, Davidson, & Bates, 2008). Except for single-trial analyses, all mixed-effects models were fit to participant-averaged data and included by-participant random intercept terms. For single-trial analyses, mixed-effects models were fit to single trials on the centro-parietal ROI data, and incorporate by-participant and by-item random intercepts. All used sum coding for categorical variables and z-transformed values for continuous variables. For statistical inferences about predictors of interest, we conducted model comparison using Chi-square tests on nested models, comparing a full model with a reduced model which omitted the term being tested, using the `anova` function in `R`. P-values for individual predictors were computed using `lmerTest`, with the Satterthwaite option for denominator degrees of freedom for *F* statistics.

4.3.3.5 Modified cloze production task

Following the EEG portion of the study, the participants completed a modified cloze production task. Participants were told that they would now see the same set of sentence pairs as in the brainwave study, but with no final word. They were reminded that, in the brainwave part of the study, some of the final words had been consistent with the Harry Potter stories, and some of them had been inconsistent. They were asked to provide a final word to fill in the blank that they believed was consistent with the HP stories—that is, the word that belonged. If they were not certain which word to provide, they were asked to provide their best guess. If they could not provide a word for a given sentence pair, they were asked to make a judgment about the sentence based on what they had initially read during the EEG study: to type "C" if they thought the original word had been consistent with the stories, or "I" if they thought it to be inconsistent.

This task allowed us to conduct additional analyses on the ERP data by sorting trials based on whether individuals were correct (i.e., had provided the appropriate word during the cloze task) or not. For these analyses, we used mixed-effects linear regression models fit to single-trial data from our centroparietal ROI in our two time periods of interest (N400, late positivity).

4.4 Results

4.4.1 Behavioral data

4.4.1.1 Modified cloze production task—accuracy

For the modified cloze production task, performed after the ERP study, mean accuracy (defined as the total proportion of trials for which each participant provided the correct, i.e., supported, word) was 49% across participants (range: 6%-95%). As expected, accuracy on this task correlated strongly with participants' HP domain knowledge scores, $r = .90$, $p < .0001$. Participants provided an incorrect word response on 14% of trials (range: 2%-44%) and provided no response whatsoever on less than 1% of trials.

Because participants had recently read all the sentence pairs containing completions during the ERP study, we expected that they might be more likely to produce the correct words which they had seen in the Supported condition compared to the Related and Unrelated conditions. This is indeed the pattern we observed, with higher mean accuracy for sentences and words viewed in the supported condition (60%) compared to the Related (43%) or Unrelated (43%) conditions ($F(2, 94) = 82.06$, $p < .001$). This pattern was only slightly different as a function of HP knowledge, with a significant interaction between HP knowledge and accuracy ($\chi^2(2) = 6.58$, $p = .037$) driven by a slightly steeper slope for the Related and Unrelated conditions (which did not differ from one other); see Figure 4.3.

We also considered the trials on which participants indicated that they did not know and could not guess the correct completion by responding with judgments about whether the word they initially read (in the ERP study) had been consistent or inconsistent with the world of HP. Participants responded with such judgments on an average of 36% of trials; they were correct on an average of 65% of these trials.

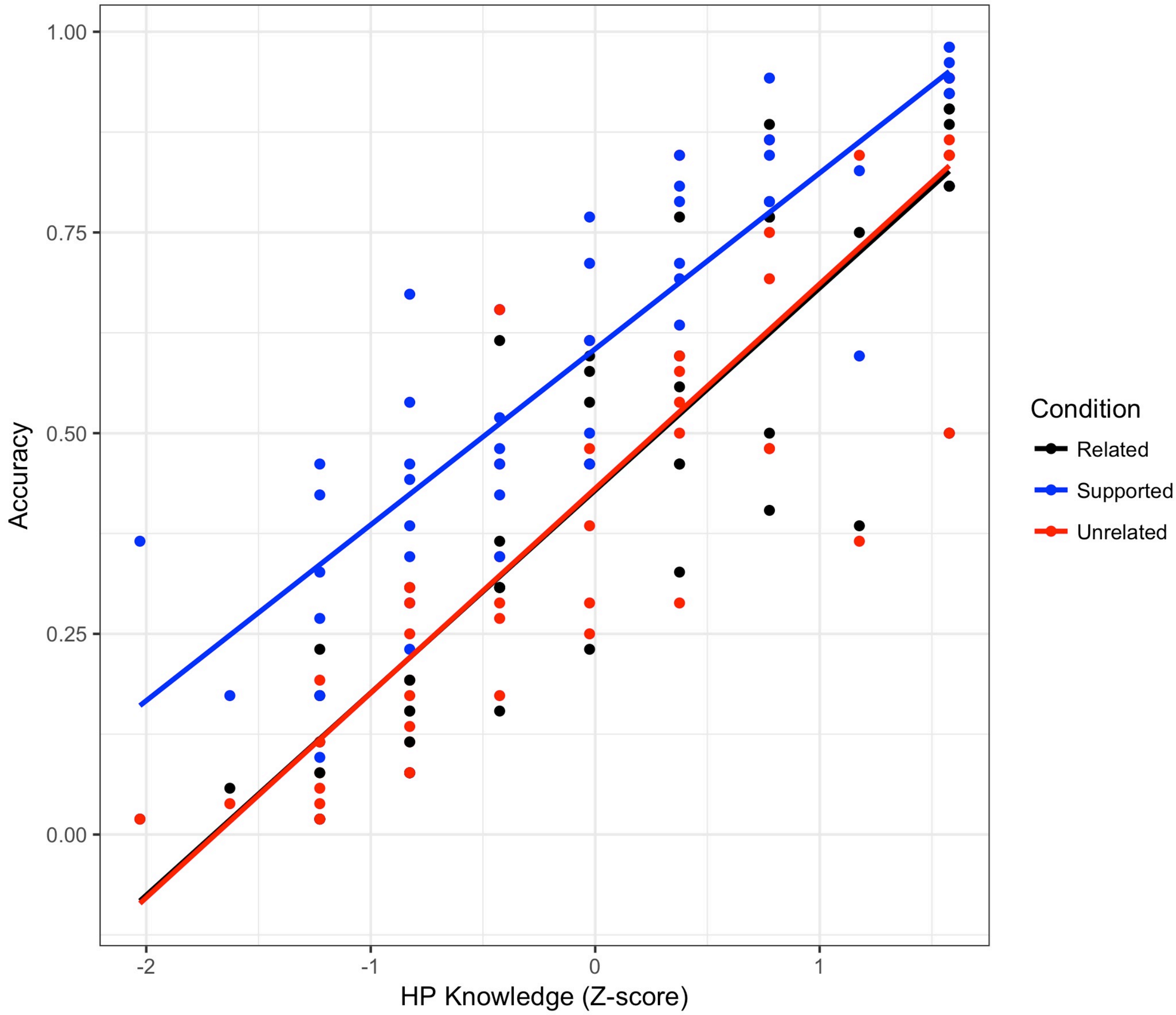


Figure 4.3. Accuracy on the post-experiment cloze task is plotted against HP knowledge (Z-scored) by condition. This relationship was slightly stronger for Related and Unrelated items compared to Supported items.

Table 4.2. Mean, standard deviation, and range are provided for behavioural measures of individual differences.

	All participants			High HP group			Low HP group		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
HP Quiz	6.06	(2.50)	[1, 10]	8.48	(1.25)	[7, 10]	3.77	(1.07)	[1, 5]
HP Self Score	34.35	(21.21)	[1, 84]	49.81	(19.66)	[18, 84]	20.09	(12.62)	[1, 43]
ART	0.17	(0.08)	[0.05, 0.40]	0.21	(0.08)	[0.08, 0.40]	0.13	(0.05)	[0.05, 0.22]
MRT	0.18	(0.08)	[0, 0.42]	0.2	(0.10)	[0.05, 0.42]	0.17	(0.07)	[0, 0.28]
MRH Total	4.81	(2.27)	[1, 12]	5.24	(2.26)	[1, 12]	4.45	(2.26)	[1, 8]
GKQ	19.31	(3.18)	[14, 27]	20.38	(3.50)	[15, 27]	18.45	(2.84)	[14, 24]
Sentence Span	2.76	(0.64)	[1, 4.5]	2.88	(0.57)	[2, 4.5]	2.64	(0.68)	[1, 4]

4.4.1.2 Additional tasks

Table 4.2 reports descriptive statistics for scores on the HP knowledge quiz and other individual difference measures. Intercorrelations among measures are provided in Table 4.3.

Table 4.3. Intercorrelations (Pearson's *r*) among behavioral measures of individual differences. *r* values above .29 are significant at *a* = .05; *r* values above .37 are significant at *a* = .01.

	1	2	3	4	5	6	7	8	9
1 HP Quiz	1	0.76	0.49	0.25	0.56	0.22	0.35	0.26	0.55
2 HP Self Score		1	0.49	0.19	0.67	0.36	0.23	0.11	0.62
3 ART			1	0.56	0.4	0.2	0.35	0.23	0.78
4 MRT				1	0.06	0.19	0.44	0.2	0.65
5 Authors Listed					1	0.41	0.26	0.16	0.68
6 MRH Total						1	-0.05	0.12	0.65
7 GKQ							1	0.21	0.36
8 Sentence Span								1	0.26
9 Reading Experience									1

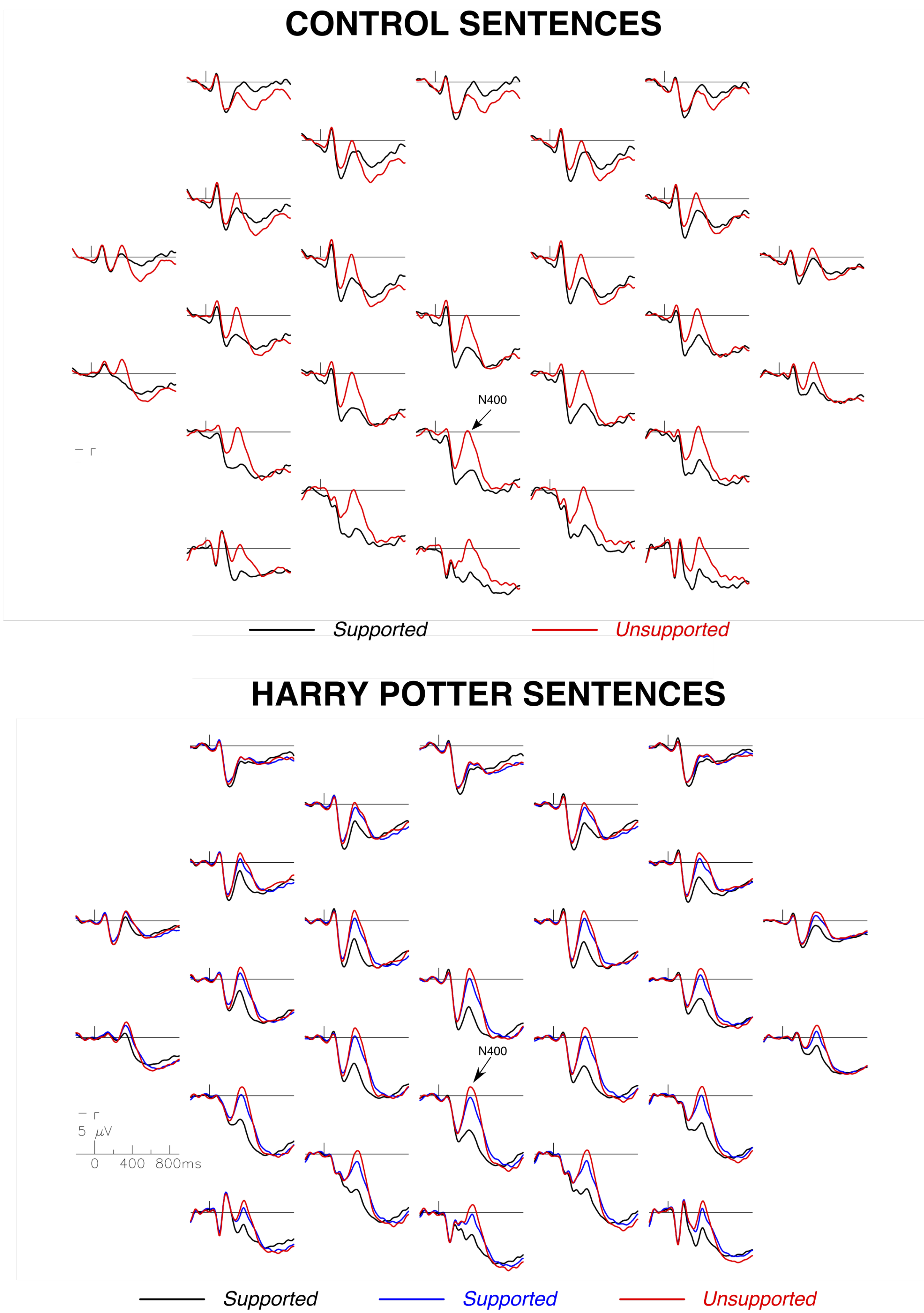


Figure 4.4. Grand average ERPs across all participants for critical words of each type for control (top) and (bottom) sentences.

4.4.2 Subject-averaged ERP data

Figure 4.4 shows the grand average ERPs for all participants across 26 scalp electrodes from 200 ms before the onset of the critical word to 900 ms post-critical word for control sentences and for HP sentences. Across most electrodes, ERPs to critical words are characterized by two early sensory components: a negative-going peak around 100 ms (N1) and a positive-going peak around 200 ms (P2). Across all participants, supported words (for both sentence types) are characterized by a positivity in the N400 time period (250-500 ms); for the unsupported ending types, the P2 is followed by a relative negativity in this time window.

Because we were specifically interested in the influence of individual differences in HP knowledge on ERPs, we provide whole-head plots for a high-HP-knowledge subgroup (n = 21) and low-HP-knowledge subgroup (n = 22) for each sentence type in Figures 4.5-4.6.

4.4.2.1 N400: 250-500 ms post-stimulus

4.4.2.1.1 Whole-head analyses

Results from the whole-head ANOVA for control and HP sentences in the N400 time window are provided in Table 4.4.

For control sentences, as expected, there was a main effect of ending type on ERPs, with Supported endings leading to more positive-going waves (i.e., reduced negativities) than Unsupported endings. Visual inspection indicates that an interaction with electrode site results from a broad distribution of the contextual support effect across central and parietal sites.

Table 4.4. Whole-head ANOVA results for N400 time window.

	DF	*F*	p-value	ε_{GG}
Control sentences				
Electrode	(25, 1175)	19.804	.0000	0.136
Ending Type	(1, 47)	22.240	.0000	
Electrode:Ending Type	(25, 1175)	21.476	.0000	0.096
HP sentences				
Electrode	(25, 1175)	19.231	.0000	0.121
Ending Type	(2, 94)	59.023	.0000	0.853
Electrode:Ending Type	(50, 2350)	26.067	.0000	0.129

Table 4.5. ROI analysis for N400 time period.

	Estimate	SE	DF	*t*-value	Pr(>|t|)
Control sentences					
Intercept	3.167	0.283	46	11.179	.0000
Ending Type	-1.526	0.245	46	-6.228	.0000
HP Knowledge	0.756	0.286	46	2.639	.0113
Ending Type:HP Knowledge	0.250	0.248	46	1.010	.3178
HP sentences					
Intercept	2.650	0.266	46	9.943	.0000
Ending Type 1	-1.503	0.172	92	-8.759	.0000
Ending Type 2	-0.714	0.172	92	-4.163	.0000
HP Knowledge	1.012	0.269	46	3.759	.0005
Ending Type 1: HP knowledge	-0.483	0.173	92	-2.783	.0065
Ending Type 2: HP knowledge	0.085	0.173	92	0.490	.6250

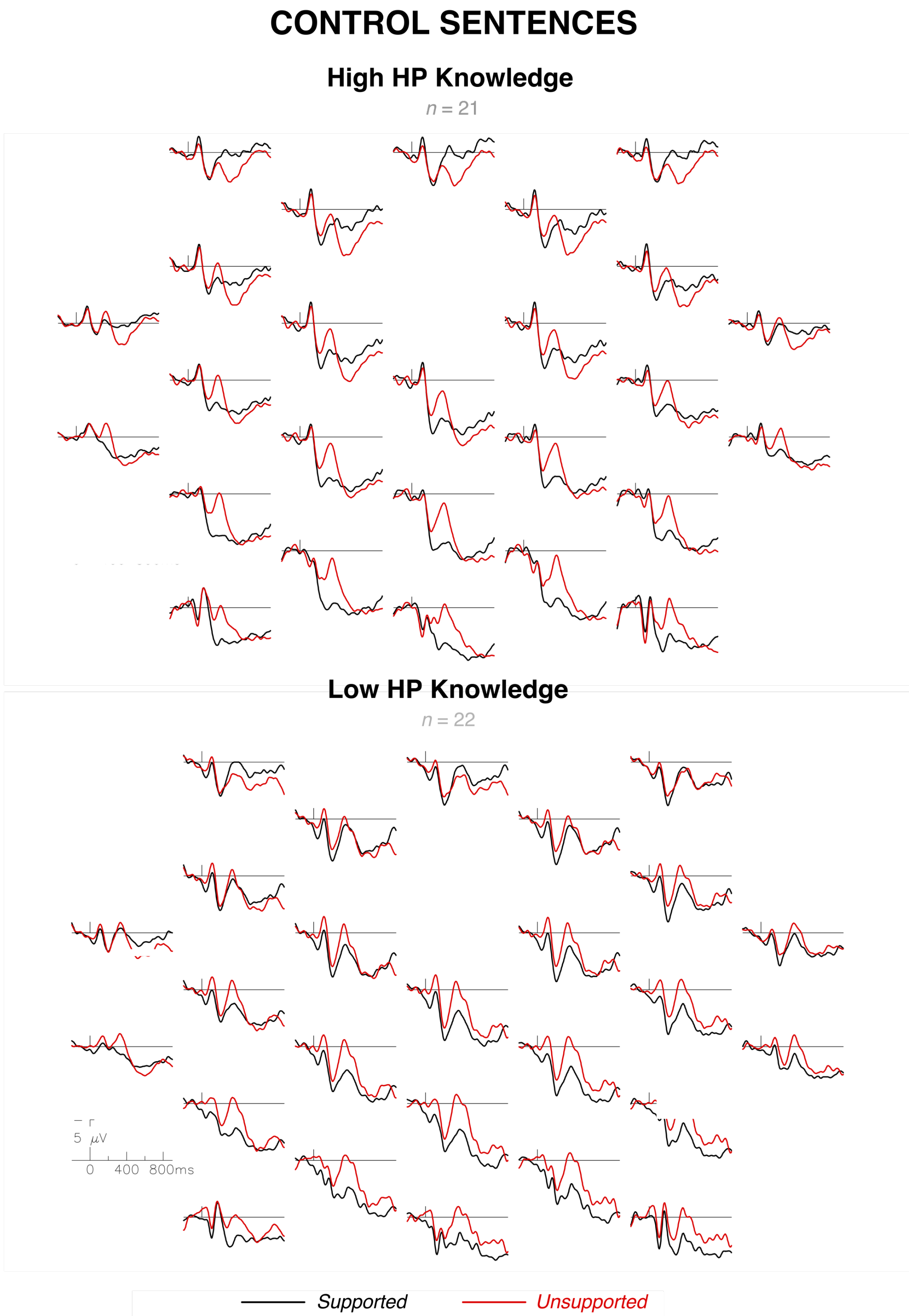


Figure 4.5. Grand average ERPs for High (top) and Low (bottom) HP knowledge subgroups (based on a median split) to critical words in the control sentences.

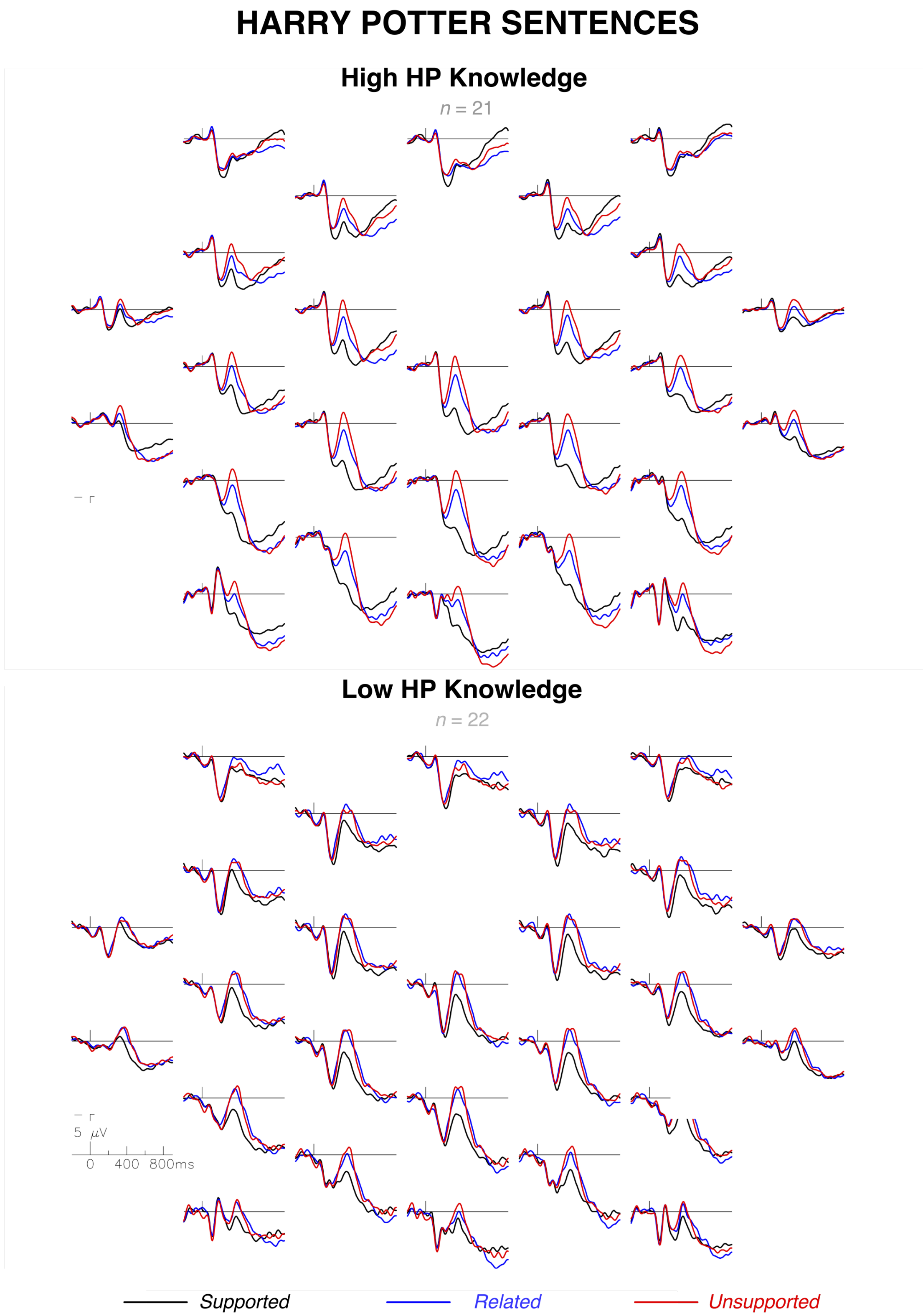


Figure 4.6. Grand average ERPs for High (top) and Low (bottom) knowledge subgroups (based on a median split) to critical words in the HP sentences.

For HP sentences, there was a main effect of ending type as well as an interaction with electrode. Planned comparisons revealed that mean N400 amplitude for the Unrelated condition was larger than the Related condition ($F(1, 47) = 7.87$, $p = .007$), which was in turn larger than the Supported condition ($F(1, 47) = 64.57$, $p < .0001$). Visual inspection suggests that the interactions between electrode and ending type result from a broad centro-parietal distribution of both the contextual support and related anomaly effects.

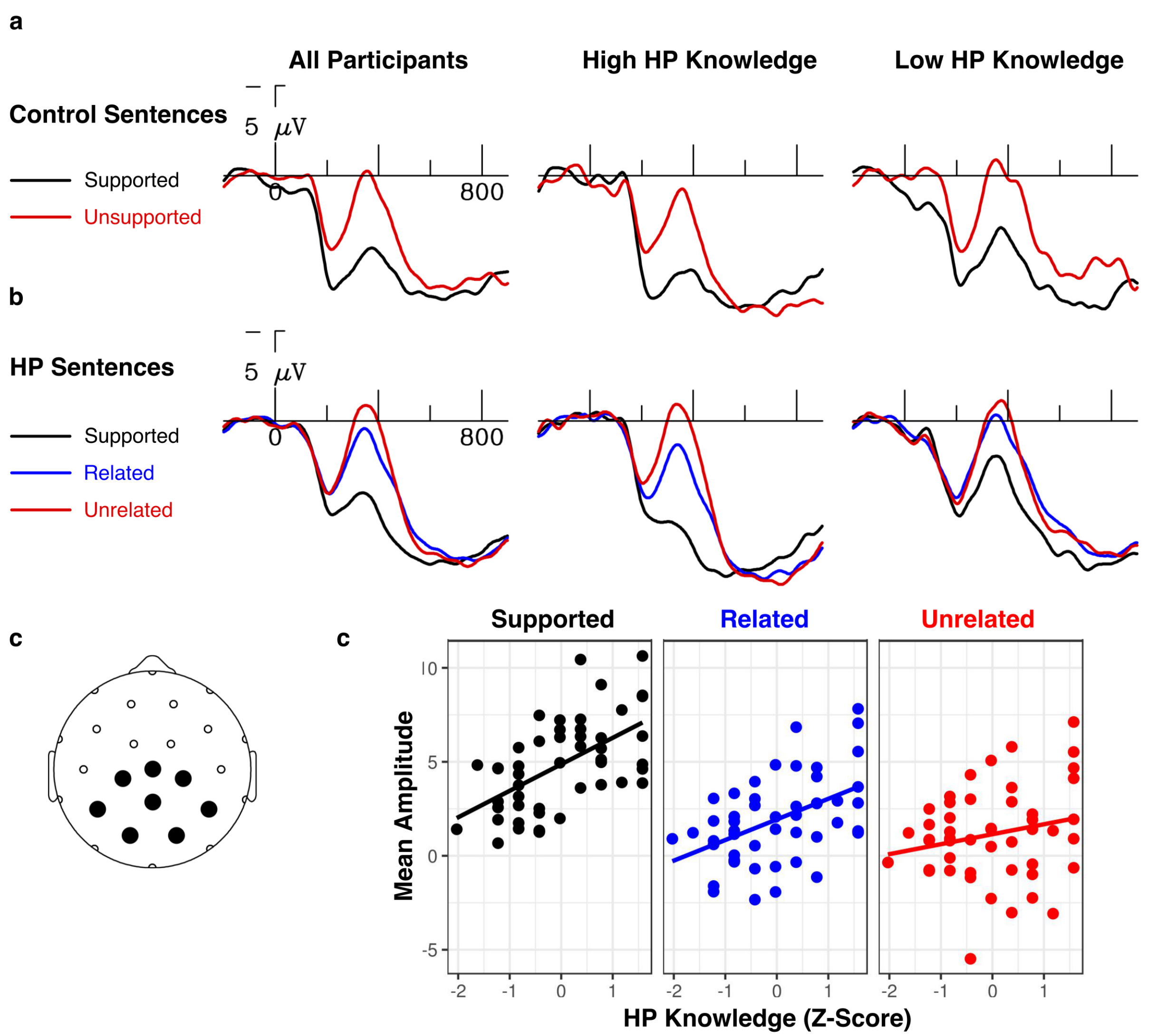


Figure 4.7. ERPs are plotted to critical words from control (a) and HP (b) sentences for an ROI based on an average of 8 centro-parietal scalp electrodes (c). (d) N400 amplitude is plotted against HP knowledge (z-scored) by condition.

4.4.2.1.2 ROI analyses of individual differences

Figure 4.7a shows ERPs for control sentences from the centro-parietal ROI used in the regression analyses. As expected, HP domain knowledge did not interact with ending type ($\chi^2(1) = 1.05$, n.s.) in the N400 time window (see Table 4.5 for full model results). That is, for control sentences, HP domain knowledge did not influence N400 effects of contextual support, replicating our previous results (Troyer & Kutas, 2018).

Figure 4.7b shows ERPs for HP sentences from the centro-parietal ROI used in regression analyses. As predicted, the interaction between HP domain knowledge and ending type was significant ($\chi^2(2) = 8.79$, $p = .012$). To unpack this result, we conducted planned pairwise comparisons for each pair of ending type conditions. Interactions between HP domain knowledge and ending type were significant for the Unrelated vs. Supported ($\chi^2(1) = 6.55$, $p = .010$) and Unrelated vs. Related ($\chi^2(1) = 5.41$, $p = .020$) comparisons, but not for the Related vs. Supported comparison ($\chi^2(1) = 1.04$, n.s.).

To confirm our prediction that these interactions were driven by HP domain knowledge explaining variance in the the Supported and Related conditions, but not in the Unrelated condition, we conducted follow-up correlational analyses between HP domain knowledge and average N400 amplitude for each condition. As predicted, we observed significant correlations between HP domain knowledge and N400 amplitude for Supported endings ($r = .574$, $p < .0001$) and for Related endings ($r = .471$, $p < .001$), but not for Unrelated endings ($r = .21$, n.s.) (Figure 4.7d).

Visual inspection of category- and event-related subsets of data suggests that the main pattern of N400 results (Supported < Related < Unrelated) is similar for each. To verify this, we confirmed that there was no three-way interaction between the three predictors ($\chi^2(2) = 0.72$, p =

.700), nor was there any interaction between ending type and related anomaly type across all participants ($\chi^2(2) = 1.81$, $p = .405$).

Finally, to rule out the possibility that other existing individual differences (namely reading experience, general knowledge, and verbal working memory scores) could better account for the observed variability in N400 ERPs, we tested a model that incorporated fixed effects of ending type, HP domain knowledge, general knowledge scores, reading span scores, and aggregate reading experience scores along with interaction terms for each individual differences measure with ending type. We compared this model and a similar model that did not incorporate interaction terms with any individual differences measures (except for the HP domain knowledge-by-ending type interaction term), and found that the more complex model did not explain additional variance ($\chi^2(3) = 9.05$, n.s).

4.4.2.2 Late positivity: 500-750 ms post-stimulus

4.4.2.2.1 Whole-head analyses

Results from the whole-head ANOVA for both control and HP sentences in the late positivity time window are provided in Table 4.6.

For control sentences, in the late positivity time window, a whole-head ANOVA revealed no main effect of ending type; however, ending type interacted with electrode. Based on visual inspection, this interaction seems to reflect a slightly greater positivity for unsupported compared to supported endings across left frontal sites, a pattern reported for similar materials in several other studies (e.g., Thornhill & Van Petten, 2012; DeLong, Quante, & Kutas, 2014; Troyer & Kutas, 2018).

For HP sentences, in the late positivity time window, a whole-head ANOVA revealed no main effect of ending type; however, ending type interacted with electrode. Based on visual inspection, this interaction seemed to reflect a tendency for ERPs to Unrelated and Related words to be somewhat more positive-going across central and parietal channels compared to Supported words in this time period.

Table 4.6. Whole-head ANOVA results for late positivity time window.

	DF	*F*	p-value	ε_{GG}
Control sentences				
Electrode	(25, 1175)	42.156	.0000	0.144
Ending Type	(1, 47)	0.527	.4716	
Electrode:Ending Type	(25, 1175)	8.479	.0000	0.144
HP sentences				
Electrode	(25, 1175)	63.871	.0000	.112
Ending Type	(2, 94)	0.361	.6978	
Electrode:Ending Type	(50, 2350)	2.728	.0000	.890

4.4.2.2.2 ROI analyses of individual differences

For control sentences, in the late positivity window, we observed a marginal interaction between HP knowledge and ending type ($\chi^2(1) = 3.16$, p = .075; Table 4.7), which seemed to be driven by the fact that individuals with higher HP domain knowledge had marginally more positive-going potentials in the late positivity time window for unsupported (Pearson's $r = .26$, p = .07), but not for supported ($r = .06$, p = .68), words.

In the late positivity time window, there was a significant interaction between HP domain knowledge and ending type ($\chi^2(2) = 15.79$, p < .001; full model provided in Table 4.7). To follow up, we conducted planned pairwise comparisons for each pair of ending type conditions. The interaction between HP domain knowledge and ending type was significant for Unrelated vs. Supported endings ($\chi^2(2) = 9.608$, $p < .002$) and for Related vs. Supported endings ($\chi^2(1) = 11.125$, $p < .001$) but not for Unrelated vs. Related endings ($\chi^2(2) = .003$, $p = .955$).

Table 4.7. ROI analysis for the late positivity time period.

	Estimate	SE	DF	*t*-value	Pr(>|t|)
Control sentences					
Intercept	5.963	0.406	46	14.706	.0000
Ending type	-0.224	0.220	46	-1.017	.3143
HP knowledge	0.565	0.410	46	1.378	.1748
Ending type:HP knowledge	0.394	0.223	46	1.770	.0834
HP sentences					
Intercept	7.193	0.394	46	18.268	.0000
Ending type 1	0.133	0.195	92	0.684	.4960
Ending type 2	-0.347	0.195	92	-1.784	.0778
HP knowledge	1.001	0.398	46	2.515	.0154
Ending type 1:HP knowledge	0.391	0.197	92	1.985	.0501
Endinge type 2:HP knowledge	0.407	0.197	92	2.070	.0413

To follow up on the interactions between HP domain knowledge and ending type, we conducted correlation analyses between HP domain knowledge and late positivity amplitude for each condition. We observed a significant correlation between HP domain knowledge and late positivity amplitude for Unrelated (r = .41, p = .004) and Related (r = .45, p = .001), but not for Supported (r = .06, n.s.), endings.

Visual inspection of category- and event-related subsets of data suggests that the pattern of late positivity results (Supported < Related < Unrelated) is similar for each. To verify this, we confirmed that there was no three-way interaction between the three predictors ($\chi^2(2) = 0.72$, p = .700), nor was there any interaction between ending type and related anomaly type across all participants ($\chi^2(2) = 1.81$, p = .405). Finally, we tested a model incorporating additional individual differences (measures of reading experience, general knowledge, and verbal working memory scores) along with HP domain knowledge along with interactions with ending type. Comparing this model to a model which was similar but did not incorporate any interaction terms with individual differences measures (except for the HP domain knowledge-by-ending type interaction term), finding that the more complex model did not explain additional variance ($\chi^2(6) = 10.69$, n.s.).

4.4.3 Single-trial ERP data

Based on participants' responses during the post-ERP modified cloze task, we sorted trials according to whether participants correctly produced the appropriate (i.e., Supported) ending to each HP sentence pair. We expected that accuracy on this task would be related to N400 amplitude for Supported endings, with a reduction for correct compared to incorrect trials, replicating previous findings (Troyer et al., under review). We were most interested in whether such a reduction would be present on the N400 for the Related endings. We also analyzed the late positivity time window.

4.4.3.1 N400: 250-500 ms post-stimulus

We first asked whether accuracy on the post-experiment cloze task had a similar influence across conditions (Figure 4.8). For the N400 time period in the centro-parietal ROI, for all participants (regardless of HP knowledge), there was an interaction between ending type and accuracy on the post-experiment cloze task ($\chi^2(2) = 79.39$, $p < .0001$). Planned comparisons by ending type revealed N400 amplitude was reduced for correct compared to incorrect items only for Supported words ($\chi^2(1) = 71.87$, $p < .0001$), but not to Related ($\chi^2(1) = 0.02$, n.s.) or Unrelated words ($\chi^2(1) = 0.26$, n.s.).

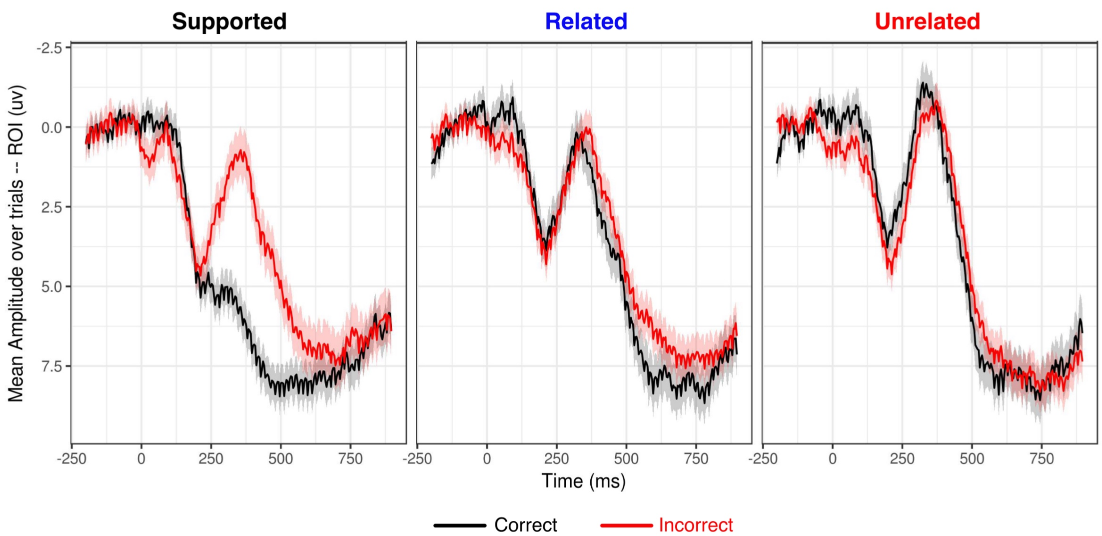


Figure 4.8. ERPs averaged over single trials in the centroparietal ROI are plotted by condition and by accuracy on the post-experiment cloze task.

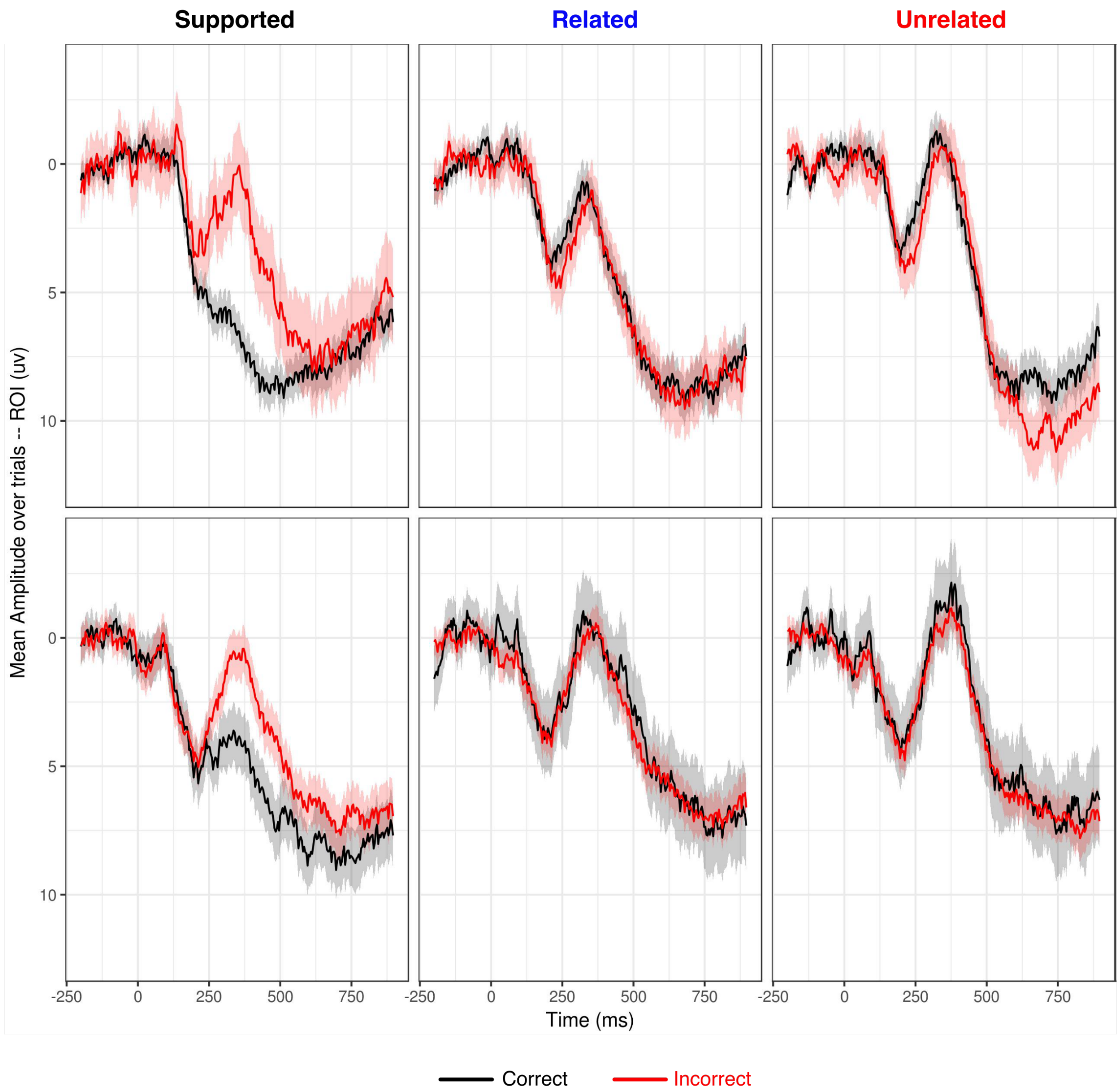


Figure 4.9. ERPs averaged over single trials in the centroparietal ROI are plotted by condition and by accuracy on the post-experiment cloze task and by HP knowledge subgroup based on a median split (top: high HP knowledge; bottom: low HP knowledge).

We also asked whether HP knowledge modulated this interaction (between accuracy and ending type). To do so, we compared a fully crossed model incorporating all three terms with a reduced model containing two two-way interactions (ending type X HP knowledge and ending type X accuracy), thereby eliminating the three-way term we were interested in testing. This analyses revealed that HP knowledge did not modulate the higher order interaction ($\chi^2(3) = 3.17$, n.s.); that is, the influence of accuracy on N400 amplitude was similar across ending type condition, regardless of participants' domain knowledge (Figure 4.9).

4.4.3.2 Late positivity: 500-750 ms post-stimulus

We also asked whether accuracy on the post-experiment cloze task would influence late positivities. Across all participants, we observed no interaction between accuracy and ending type ($\chi^2(2) = 1.72$, n.s.).

4.5 Discussion

4.5.1 Summary of findings

In the current study, we were test whether, and, if so, when and under what conditions, individuals' degree of domain knowledge influences the organization and use of information in real time. To that end, we constructed sentence materials containing related anomalies, which were designed to probe contextually (linguistically) unsupported information that was nonetheless meaningfully related to the sentence context and/or a supported (but unpresented) continuation. As expected, we found that individuals were more sensitive to the related anomaly manipulation as a function of their degree of HP knowledge.

More specifically, HP domain knowledge influenced the differential ERP response to words of different types of relations to the context in both the N400 and late positivity time windows, albeit with a different pattern in each. As expected, greater HP domain knowledge was associated with the greatest reduction in N400 amplitude for the best completion (i.e., Supported endings) compared to Unrelated words, and N400s were intermediate in amplitude for initial words related to the sentence context and/or best completion (i.e., Related endings). In the late positivity time window, HP domain knowledge did not have a influence on brain potentials to the best completion; rather, greater domain knowledge was more strongly associated with more positive-going potentials for words which were inappropriate continuations (i.e., both the Related and Unrelated).

Using our post-experiment cloze task, we found that, as expected, individuals' trial-level knowledge of facts (inferred from their cloze productions) had a strong influence on N400 brain potentials for the Supported continuations. Also as expected, for Unrelated words, there was no such relationship. However, perhaps surprisingly, there was also no relationship between trial-level knowledge of facts and N400 amplitude for Related words. We consider this finding in more detail below.

4.5.2 Contextual support effects

The current study replicates previous findings that individual differences in domain knowledge of a fictional narrative world (Harry Potter) modulate the influence of sentential support on real-time processing of words in sentences. Troyer & Kutas (2018) compared sentences about HP ending with contextually supported (i.e., "correct") and unsupported (i.e., "incorrect") words. As expected, effects of contextual support on N400 amplitude were largest for individuals with greater HP knowledge and smallest for those with the least HP knowledge. Moreover, this effect was driven by variation in the N400 response to contextually supported words (also replicated in Troyer et al., under review); there was no significant relationship between domain knowledge and the N400 response to contextually unsupported words. This pattern of results strongly suggests that information cued by the sentential contexts was available to facilitate (pre-) activation of the appropriate, supported word.

In the current study, we replicate this pattern once again, finding that the N400 response to words supported by the sentence context was reduced as a function of individuals' HP domain knowledge, and that there was no significant relationship between HP domain knowledge and the

N400 response to unsupported (and unrelated) words. These findings empirically demonstrate that knowledge at the level of the individual participant is critical for bringing to mind relevant meaning from sentence contexts, and that this knowledge is quickly used to make sense of contextually supported words in real time.

For the HP sentences, we also observed an interaction effect of HP knowledge and ending type on amplitude in the late positivity time period, such that individuals with greater HP knowledge showed increased positivities (relative to those with less HP knowledge) for Unrelated and Related words, but not for the Supported words. In a previous ERP reading study, there was no such link between individuals' HP knowledge and words which were inappropriate continuations to sentences about HP (Troyer & Kutas, 2018). The difference may be attributable to the demands of the reading experiments: in the current study, only a third of the items were congruent / contextually supported, whereas in the previous study, half of the items ended in supported words, and the other half in words that were not correct (according to the HP stories) but were designed to seem plausible, otherwise.

In the sentence processing literature, variation in late parietal positivities has been linked to the processing of implausible words / semantic anomalies (DeLong, Quante, & Kutas, 2014; Van Petten & Luka, 2012). Moreover, in the memory literature, variation in late parietal positivities has been linked to conscious recollection (reviewed in Hillyard & Kutas, 2002; Rugg & Curran, 2007). Based on these interpretations of late positivities, we speculate that individuals with greater HP knowledge were more likely to detect the anomalies (Related and Unrelated words) and/or were more likely to engage in effortful, conscious recollective processes during this time period.

4.5.3 Related anomaly effects

The critical condition in our HP sentences was the related anomaly, which was related via category to a contextually supported (i.e., correct) word or via relationship to the event/episode described by the sentence context. In previous studies, related anomalies have been shown to elicit reduced N400 amplitudes compared to unrelated/unsupported words, though larger than for words which are best completions/contextually supported (Federmeier & Kutas, 1999; Metusalem et al., 2012; Amsel et al., 2015; DeLong et al., 2018). Similar to these findings, we demonstrated related anomaly effects for sentences about the narrative world of Harry Potter. Moreover, we showed that N400 potentials to the related anomalies were a function of individuals' degree of knowledge about the HP domain. These findings strongly suggest that the quick availability of information relevant for word processing varies as a function of individuals' domain knowledge. We hypothesize that this is a result of differences in the functional organization of individuals' domain knowledge, such that related pieces or "chunks" of information are more readily activated and brought to mind for knowledgeable individuals due to their differential functional organization in semantic memory. Consistent with previous studies, related anomaly effects were similar in magnitude, timing, and distribution over the scalp, regardless of whether the related anomaly was from the same category as the supported word or whether it was related via event/episode to the content of the sentence context.

4.5.4 Knowledge of specific facts

Our post-experiment cloze task was designed as a proxy for trial-level knowledge of the specific facts about HP that participants read during the ERP study. Our primary measure of interest was whether or not individuals could provide the appropriate completion for each sentence. Of course, because participants had read this word in a third of the trials during the ERP study (i.e., during the Supported conditions), we expected that on average, individuals would be more accurate for those Supported trials compared to Related or Unrelated trials. This was indeed the case, with an approximately 17% boost for individuals who produce the correct response. This overall pattern was similar across individuals, with a slightly greater boost for individuals with relatively less knowledge, which is not surprising, as more knowledgeable individuals presumably knew more of the items overall, and to a stronger degree, meaning they likely have less to gain from having seen the appropriate word during the ERP study.

As expected, we found that N400 amplitudes to Supported words were strongly influenced by whether or not each individual could correctly report that fact during the post-experiment cloze task: words which were correctly reported elicited reduced N400 amplitude compared to words which were incorrectly reported. Also as expected, N400 amplitude to Unrelated words was unaffected by participants' post-ERP cloze task reports. However, perhaps surprisingly, we also observed that N400 amplitude to the Related words was unaffected by whether individuals reported the correct word during the post-ERP task. This pattern of results—that is, the influence of task accuracy on N400 potentials by condition—was the same regardless of individuals' degree of HP knowledge. Combined with the fact that individuals' domain knowledge *did* modulate N400 amplitudes to Related (as well as Supported) words, the absence of an influence of single-trial-level knowledge on N400 amplitude in the Related condition leads us to conclude that domain knowledge can influence implicit retrieval of word information in the absence of specific, item-level knowledge of "facts" (fictional facts, in this case).

4.5.5 Domain knowledge and variability in the functional organization of knowledge

It is well-attested that knowledge in long-term memory is organized along many cross-cutting dimensions (reviewed in Yee, Jones, & McRae, 2017), including taxonomic organization based on categories (e.g., Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976); event knowledge (e.g., knowledge based on scripts; Schank & Abelson, 1977); and perceptual / sensorimotor relationships between concepts (e.g., Barsalou, 1999). These (among other) systems of functional organization have all been shown to impact real-time information processing, suggesting that each is used to make sense of words as they are understood in sentences in real-time. Here, we investigated the engagement of two cross-cutting functional systems of organization of long-term memory, namely categorical and event relationships, during real-time language processing as a function of individuals' domain knowledge and of individuals' knowledge of specific, trial-level facts.

It is also well-established that experts within a given domain of knowledge (e.g., cooking, sports, chess, academic disciplines, professions, and so on) are likely to perceive, categorize, and otherwise process information from that domain in specialized ways (Ericsson, Charness, Feltovich, & Hoffman, 2006). For example, a novice and expert may view the same medical scan, but only the expert can attend to the relevant information, determine its functional significance, and use the information to make the appropriate diagnosis (e.g., Gilhooly, McGeorge, Hunter, Rawles, Kirby, Green, & Wynn, 1997). Similarly, in physics, expert students seem to access deeper or higher-level organizational principles when categorizing physics problems whereas novices access more shallow, surface, or literal features of the problems (Chi, Feltovich, & Glaser, 1981). Chi & Ohlsson (2005) discuss such individual differences as variation in the availability of *theories*, in which core knowledge can form organizing principles from which other bits of information extend. Experts, but not novices, may have access to this information, allowing them to determine which bits of information are most relevant and/or important when processing information (including images and text) within the domain.

Here, we have explored the extent to which variation in knowledge of a fictional domain shapes the availability of such knowledge structures in real-time. We expected that, as in other domains of knowledge, expertise in a fictional world of knowledge would result in differential functional organization of knowledge. That is, not only would HP experts know more facts, but expert knowledge of concepts and the relationships between them would result in information that was more highly structured and organized in principled ways according to the plots and concepts in the narrative world of HP. Our behavioral norming results supported this notion—individuals with greater HP knowledge were sensitive to manipulations of category and event relatedness between critical words in our materials.

Moreover, we expected that this differential organization of knowledge according to expertise in HP would be reflected in real-time language comprehension. Extending many other studies of real-time language comprehension (e.g., Federmeier & Kutas, 1999; Metusalem et al., 2012), we used a related anomaly paradigm to probe the ease with which individuals were able to retrieve word knowledge when the words themselves were not linguistically licensed, but were, rather, related either to event knowledge cued by the sentence context or via the category of the contextually supported word. We assume that the implicit activation of such knowledge results as a natural consequence of sentence comprehension in real time.

For example, consider our sample context, '*Professor McGonagall recruits Harry for the Gryffindor Quidditch team. She saw him save Neville's…*,' which is appropriately completed by the contextually supported (i.e., correct) word, '*remembrall*.' As highly-knowledgeable individuals process each incoming word, they may (implicitly) retrieve relevant information based on extensive knowledge of the stories, including knowledge about events/schema/scripts (for example, *Quidditch* is a sport that is played at the wizarding school of Hogwarts, where players fly around on broomsticks). As a result, at the point of processing the critical word (Supported = 'remembrall'; Related = 'broomstick'; Unrelated = 'dog'), individuals with a high degree of knowledge (compared to those with less knowledge) may have activated information related to the word 'broomstick' in long-term memory to a higher degree.

In addition, knowledgeable individuals may similarly retrieve information that is related via category. For example, consider the sample context, '*Sybill Trelawney is a Hogwarts professor. She teaches…*,' which is appropriately completed by the contextually supported word '*Divination*.' Professors at Hogwarts teach subjects across a range of topics, and the contextual information in the sentence may cue other concepts from the same category, like *Transfiguration* (the Related completion used in our study). For both types of materials (containing category and event related anomalies), we assume that activation of relevant information is more readily available because it has been implicitly activated (to some degree) in long-term memory as individuals make sense of each incoming word and of the content of the sentence. Our findings reveal that real-time word understanding exhibits a quick sensitivity to individuals' domain knowledge, showing that differential implicit activation happens rapidly as a function of degree of knowledge.

4.5.6 Domain knowledge and specific item-level knowledge

Interestingly, we found that related anomaly effects were *not* modulated by trial-level explicit knowledge of the contextually supported word information (assessed via cloze productions). Tentatively, we infer that in the current study, related anomaly effects were driven by the (pre-)activation of (perhaps implicit) information cued by the sentential contexts, and not necessarily by (pre-)activation of the contextually supported word, per se.

It is important to note that in the current study, trial-level knowledge was based on an offline task, and we had no direct measure of trial-level knowledge in the moment. This is in contrast to a previous study (Troyer et al., under review), in which individuals read contextually supported sentences and then immediately reported whether they had known the information (or not) prior to reading the sentence. In both studies, domain knowledge had an influence on N400 amplitude to words in context beyond mediating the availability of specific knowledge at the trial level, though this pattern played out in different ways in the two studies.

In Troyer et al., domain knowledge interacted with trial-level reports of knowledge, such that domain knowledge had its primary influence on trials which participants reported *not* knowing. In that article, the authors interpreted the results as suggesting implicit, partial access to relevant information in the absence of subjective reports of knowledge of the fact. In the current study, domain knowledge did not interact with the pattern of N400 amplitude results based on trial-level accuracy and ending type. That is, accuracy on the cloze task seemed to influence N400 amplitude similarly across ending types.

However, there were many differences between task design and materials in the two studies that could account for this apparent discrepancy. Troyer et al. used only contextually supported sentence endings; here, sentences ended in a contextually supported word only a third of the time. In the previous study (Troyer et al., under review), participants immediately provided a subjective report just after reading each sentence pair. Though, in that study, trial-level reports were strongly correlated with individual-level HP knowledge, we had no objective measure of trial-level knowledge. In contrast, in the current study, we did have an objective measure of knowledge—participants were cued with the sentence contexts, and had to provide the correct word. However, this task occurred offline, and in a third of the sentences, participants had recently read the correct continuation during the ERP study (which we think led to the ~17% boost in accuracy for the Supported condition). This may well have obliterated any relationship between HP domain knowledge and partially-known information contributing to the processing of contextually supported words in context in the absence of explicit or stated knowledge of those facts. That is, the processing of facts/words in context which are less-well-known may indeed be facilitated by domain knowledge, but we did not have the ability to detect this in the current design.

In the current study, we found that domain knowledge influenced N400 amplitude to related anomalies whether or not individuals were able to produce the correct (Supported) word for these sentences. We interpret this as complementing results from Troyer et al. (under review) that indicate that domain knowledge has a quick influence on the processing of words in context beyond mediating specific knowledge at the trial level.

4.5.7 Conclusion

In conclusion, the current study provides strong evidence that individual differences in knowledge have a rapid and systematic influence on access to relevant information during real-time sentence processing. Having relatively more knowledge, and, we argue, a knowledge system that is systematically organized around structures related to categories, events, and so on, allows for quick availability of this relevant organization during real-time reading. That is, knowledgeable individuals can quickly (pre)-activate relevant featural and/or associative information—the very knowledge that is needed to make sense of words in real time. Moreover, knowledgeable individuals seemed to access this information regardless of individuals' specific knowledge of item-level facts (inferred from our offline cloze task). These methods and findings invite new research using knowledge-based individual differences to better understand how language processing interfaces with knowledge in real time.

4.6 Acknowledgments

This chapter is being prepared for publication (Troyer & Kutas, in prep). I was the primary author and investigator.

4.7 References

Altmann, G.T.M. & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*, 79-87. doi: 10.1016/S0010-0277(99)00059-1

Amsel, B.D., DeLong, K.D., & Kutas, M. (15). Close, but no garlic: Perceptuomotor and event knowledge activation during language comprehension. *Journal of Memory and Language*, *82*, 118-132. doi: 10.1016/j.jml.2015.03.009

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. doi; 10.1016/j.jml.2007.12.005

Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577-660.

Boudewyn, M.A, Long, D.L., & Swaab, T.Y. (2012). Cognitive control influences the use of meaning relations during spoken sentence comprehension. *Neuropsychologia*, *50*, 2659-2668. doi: 10.1016/j.neuropsychologia.2012.07.019

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152. doi: 10.1207/s15516709cog0502_2

Chi., M.T.H. & Ohlsson, S. (2005). Complex declarative learning. In Holyoak, K.J. and Morrison, R.G. (Eds.), *The Cambridge Handbook of Thinking and Reasoning*. Cambridge, UK: Cambridge University Press, pp. 371-399.

Coronel, J.S. & Federmeier, K.D. (2016). The N400 reveals how personal semantics is processed: Insights into the nature and organization of self-knowledge. *Neuropsychologia*, *84*, 36-43. doi: 10.1016/j.neuropsychologia.2016.01.029

Daneman, M. & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Memory and Verbal Behavior*, *19*(4), 450-466. doi: 10.1016/S0022-5371(80)90312-6

DeLong, K.A., Chan, W., & Kutas, M. (2018). Similar time courses for word form and meaning preactivation during sentence comprehension. Psychophysiology, 1-14. DOI: 10.1111/psyp.13312

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

Ericsson, K.A., Charness, N., Feltovich, P.J., Hoffman, R.R. (Eds). (2006). *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge, UK: Cambridge University Press.

Federmeier, K.D. & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469-495. doi: 10.1006/jmla.1999.2660

Filik, R. & Leuthold, H. (2013). The role of character-based knowledge in online narrative comprehension: Evidence from eye movements and ERPs. *Brain Research*, *1506*, 94-104. doi: 10.1016/j.brainres.2013.02.017

Foss, D.J. (1982). A discourse on semantic priming. Cognitive Psychology, 14, 590-607.

Ganis, G., Kutas, M., & Sereno, M.I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*(2), 89-106. doi: 10.1162/jocn.1996.8.2.89

Gilhooly, K.J., McGeorge, P., Hunter, J., Rawles, J.M., Kirby, I.K., Green, C., & Wynn, V. (1997). Biomedical knowledge in diagnostic thinking: The case of electrocardiogram (ECG) interpretation. *European Journal of Cognitive Psychology*, *9*(2), 199-223.

Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438-441. doi: 10.1126/science.1095455

Hess, D.J., Foss, D.J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. Journal of Experimental Psychology: General, 124(1), 62-82.

Kim, A.E., Oines, L., & Miyake, A. (2018). Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language comprehension: An ERP investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(3), 406-420.

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

Kutas, M. & Federmeier, K.D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(14), 1-27. doi: 10.1146/annurev.psych.093008.131123

Kutas, M. & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203-205. doi: 10.1126/science.7350657

Kutas, M. & Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(12), 161-163. doi: dx.doi.org/10.1038/307161a0

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

Lenci, A. (2018). Distributional models of word meaning. *Annual Review of Linguistics*, *4*, 151-171.

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. doi: 10.1016/j.jml.2012.01.001

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In *Proceedings of Workshop at ICLR*.

Mikolov, T., Sutskever, I, Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances on Neural Information Processing Systems*.

Nakano, H., Saron, C., & Swaab, T. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, *22*(12), 2886-2898. doi: 10.1162/jocn.2009.21400

Paczynski, M., & Kuperberg, G.R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, *67*, 426-448.

Pakulak, E. & Neville, H. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, *22*(12), 2728-2744. doi: 10.1162/jocn.2009.21393

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

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382-439.

Rugg, M.D. & Curran, T. (2007). Event-related potentials and recognition memory. *TRENDS in Cognitive Science*, *11*(6), 251-257. doi: 10.1016/j.tics.2007.04.004

Schank, R.C. & Abelson, R.B. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, N.J.: Erlbaum.

Simon, H.A. & Chase, W.G. (1973). Skill in chess. *American Scientist*, *61*(4), 394-403. doi: 10.1007/978-1-4757-1968-0_18

Stanovich, K.E. & West, R.F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*(4), 402-433. doi: 10.2307/747605

Tanner, D. & Van Hell, J.G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*, 289-301.

Thornhill, D.E. & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*, 382-392. doi: 10.1016/j.ijpsycho.2011.12.007

Troyer, M. & Kutas, M. (2018). Harry Potter and the Chamber of What?: The impact of what individuals know on word processing during reading. *Language, Cognition, and Neuroscience*, 1-17.

Troyer, M., Urbach, T.P., & Kutas, M. (under review). Lumos!: Electrophysiological tracking of (wizarding) world knowledge use during reading.

van Dijk, T.A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Van Petten, C. & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, *18*(4), 380-393. doi: 10.3758/BF03197127

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

Van Petten, C. & Luka, B.J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*, 176-190. doi: 10.1016/j.ijpsycho.2011.09.015

Yee, E., Jones, M. N., & McRae, K. (2017). Semantic Memory. In J. T. Wixted & S. Thompson-Schill (Eds.), *The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (4th Edition, Volume 3: Language and Thought). New York: Wiley.

4.8 Supplementary Materials

4.8.1 Appendix A

Control Sentences

Critical words are second-sentence-final: { supported / unsupported }.

1. We had been watching the blue jay for days.
The bird laid her eggs in the { nest / yard }.

2. Unfortunately, people always get her confused.
She has an identical but much grumpier { twin / sister }.

3. The rescue team found signs of the lost boy down by the river.
So they ardently continued their { search / investigation }.

4. Something was cooking in the oven.
Whatever it was had burned and set off the fire { alarm / buzzer }.

5. We hired a decorator.
She positioned the painting on the { wall / panel }.

6. The old woman was very cautious.
At night she always locked the { door / entrance }.

7. I tried to stop but couldn't.
Chewing your nails is a bad { habit / tendency }.

8. Alicia's first client was a failure.
But her second was a { success / triumph }.

9. Lillian loves dogs.
She thinks they are so much friendlier than { cats / felines }.

10. The grade-schoolers stood on the corner and waited.
They rode to school on the { bus / train }.

11. The vampire moved in.
He bit his victim on the { neck / shoulder }.

12. The detective arrived at the office.
Within minutes he spilled his thermos full of { coffee / tea }.

13. The chef was very experienced.
He knew how to blend sweet and { sour / lemon }.

14. The girl had written a book.
She dedicated it to her best { friend / buddy }.

15. The turtle was startled by the sounds.
It tucked its head inside its { shell / covering }.

16. The parents feared for their daughter.
They were afraid she had joined some sort of strange { cult / religion }.

17. The couple had to be somewhere right after dinner.
They left the dirty dishes in the { sink / tub }.

18. The goaltender kicked the ball.
We watched as it went down the { field / grass }.

19. The cardiologist was prepping.
He was performing an operation on my { heart / valve }.

20. After weeks backpacking, Victor was ready for a good night's sleep.
All he wanted was a comfortable { bed / bunk }.

21. Maura heard thunder in the distance.
She knew they were in for a big { storm / squall }.

22. The bear had hibernated.
All winter it would remain in its { cave / cavern }.

23. For his second birthday, Leonard got a stuffed sheep.
His mother baked him a big chocolate { cake / pie }.

24. I was completely stuck.
I could not remember his { name / title }.

25. Paul painted the clown's suit blue.
And he gave him a big red { nose / lip }.

26. The two guys were finally getting to meet.
One of them held out his { hand / fist }.

27. I only need six hours.
But some people need more than eight hours of { sleep / rest }.

28. We stared at the tiny kitten.
It was lapping the milk with its { tongue / mouth }.

29. Nancy heard it ringing.

But she was too busy cleaning to get the { phone / call }.

30. Our puppy wasn't little anymore.
He was growing to be a huge { dog / hound }.

31. She promised she would get to it.
She would have to start first thing in the { morning / evening }.

32. You can't turn left at that intersection.
It is against the { law / statute }.

33. While touring Florida Arthur couldn't see the swamp.
He was at the back of the crowded tour { bus / van }.

34. Giraffes are known for their long necks.
Zebras are known for their { stripes / markings }.

35. The contractor was doing some work.
He was replacing the shingles on the { roof / eaves }.

36. The waiter served us.
He ladled some soup into the { bowl / dish }.

37. The billionaire was in the news lately.
He was seeking a divorce from his { wife / partner }.

38. The waiter was very rude.
We ended up not giving him a { tip / gratuity }.

39. It was time for Joan to feed her baby.
She was just preparing some warm { milk / cream }.

40. The Hell's Angel rolled up the sleeve of his leather jacket.
He showed the girls where he had gotten a { tattoo / piercing }.

<u>Harry Potter Sentences</u>

Critical words are second-sentence-final: { supported / related / unrelated }.
For items #1-78 the unsupported-related endings are category-related.
For items #79-156 the unsupported-related endings are event-related.

1. There is one main bank in the wizarding world.
It is run by { goblins / werewolves / Alohomora }.

2. Creatures skilled in healing and astrology live in the forest.
They are known as { centaurs / unicorns / Errol }.

3. The Chasers try to score in Quidditch using a special ball.
It is called a { Quaffle / snitch / Krum }.

4. Sybill Trelawney is a Hogwart professor.
She teaches { Divination / Transfiguration / basilisk }.

5. There is a branch of magic focused on changing the form of objects.
It is called { Transfiguration / Arithmancy / Alley }.

6. In their second year, Harry, Ron, and Hermione create a lab.
It is located in the bathroom haunted by Moaning { Myrtle / Peeves / gillywater }.

7. Rowena Ravenclaw is the mother of another ghost.
The ghost is known as the Grey { Lady / Baron / werewolf }.

8. In life, Helena Ravenclaw was killed by a ghost.
His name is the Bloody { Baron / Friar / Lupin }.

9. Witches and wizards can rapidly travel from one Wizarding building to another.
To do so, they can use Floo { Powder / Portkey / memories }.

10. Professor Flitwick is head of Ravenclaw.
He is also teacher of { Charms / Potions / Whinging }.

11. Ron Weasley joins the Gryffindor Quidditch team in his fifth year.
He plays in the position of { Keeper / Seeker / Animagus }.

12. There are two Beaters on every Quidditch team.
Their job is to protect their team from { Bludgers / Quaffles / Filch }.

13. In his first year at Hogwarts, Neville Longbottom excels at one subject.
He does really well at { Herbology / Astronomy / Goblet }.

14. After Harry's Nimbus 2000 is destroyed, Sirius Black gives him a present.
It is an even better broomstick called a { Firebolt / Cleansweep / Cauldron }.

15. There is a good way to knock a weapon out of a victim's hands.
It is best to use the spell { Expelliarmus / Alohomora / arm }.

16. Wizards can open locks on doors and windows.
They can use the spell { Alohomora / Lumos / butterbeer }.

17. Hermione owns a large, orange feline.

Her pet is called { Crookshanks / Trevor / Brothers }.

18. In 1945, Dumbledore defeated a Dark Wizard.
His name was { Grindelwald / Voldemort / chess }.

19. Xenophilius Lovegood puts out a magazine.
It is known as The { Quibbler / Prophet / turban }.

20. Viktor Krum is one of the participants in the Triwizard Tournament.
He is the representative from { Durmstrang / Beauxbatons / stag }.

21. The Hogwarts houses are known for having different values.
The house that most values intelligence is known as { Ravenclaw / Hufflepuff / Expelliarmus }.

22. Wizards can create light at the end of their wands.
They can use a spell called { Lumos / Accio / skin }.

23. Some witches and wizards have the rare ability to transform into a particular animal at will.
Such a person is called an { Animagus / Occlumens / Apparition }.

24. In order to preserve his life, Voldemort splits his soul into seven pieces.
This creates a series of { horcruxes / Inferi / giants }.

25. Ginny Weasley replaces Harry Potter on the Gryffindor Quidditch team.
She takes over as { Seeker / Keeper / necklace }.

26. Hogwarts students shop for robes in Diagon Alley.
They buy them at a shop run by Madam { Malkin / Ollivander / centaurs }.

27. There is a very well-known pub in Hogsmeade.
It is known as the Three { Broomsticks / Honeydukes / cards }.

28. There is one very widely read newspaper in the wizard community.
It is called the Daily { Prophet / Quibbler / George }.

29. Wizards are able to conjure the Dark Mark.
They can use a spell called { Morsmordre / Crucio / clothes }.

30. To combat boggarts, wizards must think of something funny.
They must use the spell { Riddikulus / Sectumsempra / Hagrid }.

31. There is a branch of magic involving the study of magical properties of numbers.
It is known as { Arithmancy / Divination / Hippogriffs }.

32. Earwax, dirt, and vomit.
These are some unsavory flavors of Bertie Bott's Every Flavor { Beans / Frogs / Padfoot }.

33. Fleur Delacour is one of the competitors during the Triwizard Tournament.
She is competing on behalf of the school { Beauxbatons / Durmstrang / howler }.

34. There is a drink for younger wizards which contains very little alcohol.
It is known as { butterbeer / gillywater / biography }.

35. One Horcrux is destroyed by Fiendfyre.
This is Rowena Ravenclaw's { diadem / cup / egg }.

36. Fred and George Weasley invent an eavesdropping device.
It is called an Extendable { Ear / Snackbox / Patel }.

37. The Hufflepuff house has a ghost.
He is called the Fat { Friar / Myrtle / Parselmouth }.

38. Upon arrival at Hogwarts, each student is assigned to a school house.
These are chosen by the Sorting { Hat / Goblet / mandrakes }.

39. Harry receives an owl as a gift from Hagrid.
He decides to name her { Hedwig / Hermes / tower }.

40. Percy's parents give him a gift after he is made a prefect.
He receives a pet owl named { Hermes / Errol / mischief }.

41. There is a sweet shop in Hogsmeade where you can find Cauldron Cakes and other treats.
It is called { Honeydukes / Zonko's / Nagini }.

42. In the wizarding world, parents can send their children angry letters.
These are called { Howlers / Sneakoscopes / Skrewts }.

43. There is an art of reading another person's mind.
It is known as { Legilimancy / Apparition / cave }.

44. At Hogwarts, there are tubers that look like babies when they are young.
These are called { mandrakes / gillyweed / Gringotts }.

45. Pure-blood supremacists use an offensive term for Muggle-borns.
They use the derogatory word { Mudblood / squib / godfather }.

46. As a young child, Harry grows up within the ordinary world of non-magical people.
They are called { muggles / mudbloods / Astronomy }.

47. Luna Lovegood tells her friends about magical creatures that live in mistletoe.
She says they are called { nargles / wrackspurts / Beans }.

48. Harry does not bring a rat or cat to Hogwarts.
Instead, he brings a pet { owl / toad / Charms }.

49. During a duel club meeting, Harry stops a snake attack.
Ron is surprised to learn Harry is a { Parselmouth / seer / Quirrell }.

50. There is a potion that allows the drinker to assume the appearance of someone else.
The potion is known as { Polyjuice / Veritaserum / Firebolt }.

51. There is a way to make invisible ink appear.
You can use a bright-red eraser called a { revealer / Howler / Firenze }.

52. Some wizards and witches are clairvoyant.
A wizard or witch who can predict an event is called a { seer / Legilimans / motorcycle }.

53. In Harry's fourth year, Hagrid has his class look after some creatures.
They are called Blast-Ended { Skrewts / Flobberworms / Baron }.

54. Hogwarts has a house that is associated with the dark arts.
It is called { Slytherin / Ravenclaw / umbrella }.

55. Harry receives a gift from Ron on his thirteenth birthday.
It is a Pocket { Sneakoscope / Remembrall / Trevor }.

56. In Quidditch, games are usually won in one way.
This is when the seeker catches the { snitch / Bludger / dragon }.

57. Some children are born to magical parents but develop no magical abilities.
These people are called { squibs / muggles / Hedwig }.

58. Fleur Delacour and her sister Gabrielle are very beautiful.
They are both part { veela / house-elf / Powder }.

59. The main story arc concerns Harry and a villain.
Harry must defeat the Dark wizard Lord { Voldemort / Grindelwald / Griphook }.

60. Ron and Hermione's friendship suffers after a misunderstanding.
Ron thinks Hermione's cat ate his pet rat called { Scabbers / Crookshanks / prophecy }.

61. Hagrid loves animals and takes care of a dragon.
He receives Norbert from Professor { Quirrell / Trelawney / Patronus }.

62. Firenze tells Harry that mortally injured individuals can be saved.
They must drink the blood of a { unicorn / centaur / Cleansweep }.

63. Professor Binns teaches History of Magic.

He is the only Hogwarts teacher who is a { ghost / dwarf / Fleur }.

64. At Hogwarts, there is a Keeper of Keys and Grounds.
He is called { Hagrid / Filch / Quidditch }.

65. Hogwarts has a hospital wing.
It is run by Madam { Pomfrey / Hooch / garden }.

66. Hogwarts has a Quidditch referee.
Her name is Madam { Hooch / Pomfrey / boggarts }.

67. A poltergeist at Hogwarts causes a lot of chaos.
He is known as { Peeves / Nick / pudding }.

68. In their first year at school, Harry, Ron, and Hermione discover that Hagrid has been doing something illegal.
He has been raising a baby { dragon / fairy / Flitwick }.

69. Hagrid and Madame Maxime are similar.
They are proof that humans are capable of breeding with { giants / goblins / Hallows }.

70. For a time, Gilderoy Lockhart teaches Defense Against the Dark Arts.
In his first class, he releases a bunch of Cornish { pixies / gnomes / Pettigrew }.

71. There is a tree that attacks anyone who approaches it on the grounds of Hogwarts.
It is called the Whomping { Willow / Shack / knight }.

72. Argus Filch is the caretaker at Hogwarts.
He has a cat called Mrs. { Norris / Crookshanks / parents }.

73. The Hogwarts staff member who heads up the Inquisitorial Squad is detested by students.
Her name is Professor { Umbridge / Quirrell / Flobberworms }.

74. There are shape-shifters that can take the form of an intended victim's worst fear.
These are called { boggarts / thestrals / password }.

75. The Malfoy family owns a house elf.
His name is { Dobby / Kreacher / Hat }.

76. Harry gets sorted when he arrives at Hogwarts.
He is assigned to the house of { Gryffindor / Slytherin / wings }.

77. Each Hogwarts house has a headmaster or headmistress.
The head of Gryffindor is Professor { McGonagall / Sprout / squib }.

78. In his second year, Harry dislikes the new Defense Against the Dark Arts teacher.

His name is Professor { Lockhart / Flitwick / wrackspurts }.

79. Harry Potter's parents were killed by Voldemort.
Harry was left with a lightning { scar / Lily / Fudge }.

80. Harry gets a ride to school in the summer before his third year.
He is picked up by the Knight { Bus / Shunpike / Lily }.

81. Harry overhears some teachers talking about Sirius Black.
This happens on an illegal trip to the village { Hogsmeade / McGonagall / car }.

82. Sirius Black was sentenced to prison.
He spent time in { Azkaban / dementors / diadem }.

83. When Vernon Dursley insults Dumbledore, Hagrid responds.
He gives Dudley a pig's { tail / umbrella / Glory }.

84. In Harry's third year, Hagrid is promoted.
He begins to teach Care of Magical { Creatures / Hippogriffs / scar }.

85. Ron uses Wingardium Leviosa just after learning the spell.
He knocks out a { troll / Hermione / sweater }.

86. Professor McGonagall recruits Harry for the Gryffindor Quidditch team.
She saw him save Neville's { remembrall / broomstick / dog }.

87. Harry's parents left him a large inheritance.
It's located at the wizarding bank called { Gringotts / Galleons / Lockhart }.

88. Harry travels to the Quidditch World Cup with the Weasleys.
For transportation, they use a { Portkey / George / Knockturn }.

89. Ginny opens the Chamber of Secrets and unleashes an ancient monster.
It turns out to be a { basilisk / diary / Shunpike }.

90. Draco Malfoy threatens to kill Dumbledore.
However, Dumbledore is ultimately killed by { Snape / tower / nargles }.

91. Neville Longbottom destroys the final horcrux.
It is Voldemort's snake { Nagini / sword / Polyjuice }.

92. Along with three other Champions, Harry takes part in the Triwizard Tournament.
At one point, he has to retrieve a golden { egg / Krum / beetle }.

93. Harry is saved from a basilisk in the Chamber of Secrets.
His savior is Fawkes, Dumbledore's pet { phoenix / tears / Hufflepuff }.

94. Ron attempts to protect the Sorcerer's Stone.
However, he must sacrifice himself in a life-size game of wizard's { chess / knight / Mysteries }.

95. In his sixth year, Harry and Dumbledore travel to retrieve a horcrux.
They journey to the inside of a { cave / locket / badger }.

96. In order to save the sorcerer's stone, Harry must go through a trapdoor.
It is guarded by a three-headed { dog / flute / Life }.

97. On a visit to the zoo, Harry sets a boa constrictor free.
This horrifies the { Dursleys / Parseltongue / Norris }.

98. Ron and Harry wake up to gifts during their first Christmas at Hogwarts.
Hagrid gives Harry a { flute / sweater / snake }.

99. A goblin takes Harry and Hagrid into the vaults at Gringotts.
His name is { Griphook / cart / Potions }.

100. During Harry's first year, Professor Quirrell is possessed.
The back of his head begins to sprout the face of { Voldemort / turban / Pomfrey }.

101. During a Quidditch game, Harry is chased by a rogue Bludger.
It eventually breaks his { arm / Dobby / tears }.

102. Before Harry's second year, he is rescued by Ron, Fred, and George Weasley.
They pick him up in a flying { car / Dursleys / Draco }.

103. At the Dursleys', Dobby tries to get Harry in trouble.
He destroys a fancy dessert made by { Petunia / pudding / Hooch }.

104. The final task of the Triwizard Tournament involves a hedge maze.
At the end of the maze is the Triwizard { Cup / Cedric / Malkin }.

105. During Harry's fourth year, Defense Against the Dark Arts is believed to be taught by Mad-Eye Moody.
It is later revealed that the real Moody was placed under the Imperius { curse / Crouch / unicorn }.

106. Before the Yule Ball, Ron becomes jealous.
He hears that Victor Krum is attending the ball with { Hermione / Triwizard / deluminator }.

107. During Harry and Ron's sixth year, Ron is surprised by a Christmas present.
He wakes up to find a { necklace / Lavender / Snackbox }.

108. In his fourth year, Harry needs to find a date for a fancy ball.

He winds up going with Parvati { Patel / Yule / Riddikulus }.

109. During the Triwizard Tournament, Harry must swim to the bottom of the Black Lake.
His task is to rescue { Ron / gillyweed / seer }.

110. After the Triwizard Tournament, Harry duels with Voldemort.
Voldemort's wand releases its most recently performed { spells / parents / Veritaserum }.

111. During a private lesson with Snape, Harry is left alone with some of Snape's memories.
He sees Snape being bullied by his { father / Occlumency / Flamel }.

112. Professor Trelawney is fired during Harry's fifth year.
She is replaced by { Firenze / Divination / cart }.

113. Harry studies with Snape to protect his mind against invasions by Voldemort.
He takes lessons in { Occlumency / pensieve / house-elf }.

114. Before his fifth year at Hogwarts, Harry is taken to Number 12, Grimmauld Place.
It serves as headquarters for the Order of the { Phoenix / Sirius / Snap }.

115. In his fifth year, Harry and his classmates form their own Defense Against the Dark Arts group.
They meet in the Room of { Requirement / Army / toad }.

116. During his fifth year at Hogwarts, Harry is punished for rebelling.
He is forced to write with a cursed quill that carves the words into his { skin / Umbridge / Snare }.

117. In Harry's fifth year, he witnesses an attack during a vision.
He sees this through the eyes of Voldemort's { snake / Arthur / Trelawney }.

118. Looking for Sirius, Harry and his classmates fly to the Ministry of Magic.
They ride winged horses called { thestrals / Luna / Parseltongue }.

119. Harry and Dumbledore find a suspected horcrux in an underground lake.
The lake is filled with { Inferi / potion / pixies }.

120. Toward the end of Harry's fifth year, he attempts to rescue Sirius.
Voldemort lures him to the Department of { Mysteries / prophecy / Willow }.

121. Before Harry's sixth year, Snape meets with Narcissa Malfoy and Bellatrix Lestrange.
Snape shocks Bellatrix when he makes the Unbreakable { Vow / Draco / Hogsmeade }.

122. Harry, Ron, and Hermione ask Xenophilius Lovegood about his necklace.
He responds by suggesting they read The Tale of the Three { Brothers / Hallows / Crouch }.

123. Dumbledore bequeathes objects to Harry and his friends.
He leaves Ron a { deluminator / snitch / Triwizard }.

124. Ultimately, Voldemort kills Snape.
Before he dies, Snape gives Harry some of his { memories / pensieve / Moony }.

125. In the Forest of Dean, Harry and Hermione encounter a mysterious silver doe Patronus.
It guides them to a pond containing Godric Gryffindor's { sword / horcrux / spells }.

126. Hogwarts students play a game that involves spontaneous explosions.
It is called Exploding { Snap / cards / revealer }.

127. Before his third year, Harry accidentally causes his aunt to blow up like a balloon.
He winds up receiving a visit from the Minister of Magic in Diagon { Alley / Fudge / Arithmancy }.

128. In his third year, dementors cause Harry to faint and fall off his broomstick.
Afterwards, he begins to work with professor { Lupin / Patronus / arm }.

129. When Ron Weasley attempts to chase down his pet rat, he is attacked by a dog.
The dog drags him through a tunnel to a place called the Shrieking { Shack / Pettigrew / Accio }.

130. Sirius Black recognizes Peter Pettigrew in the Daily Prophet.
They confront each other in the Shrieking { Shack / Scabbers / Bus }.

131. At Honeydukes Sweetshop, you can find lots of treats.
A collectible card accompanies every chocolate { frog / biography / Map }.

132. When Harry is one year old, Hagrid brings him to the Dursleys'.
For transportation, he uses a borrowed { motorcycle / Sirius / Vow }.

133. Harry eventually learns the truth about Sirius Black.
Sirius is Harry's { godfather / Padfoot / Luna }.

134. Snape eventually reveals something about Remus Lupin.
Lupin is a { werewolf / Moony / diary }.

135. Harry's aunt and uncle treated him poorly.
However, they spoiled their own child { Dudley / Privet / Astronomy }.

136. Only some people can see Thestrals.
These people have witnessed { death / wings / Ear }.

137. The Sorcerer's Stone can be used to brew a potion.
It is called the Elixir of { Life / gold / Arthur }.

138. After the Triwizard Tournament, Hermione discovers that Rita Skeeter is an unregistered Animagus.
She can take the form of a { beetle / quill / Morsmordre }.

139. One of the houses at Hogwarts is called Hufflepuff.
Its symbol is a { badger / Sprout / Zonko's }.

140. Professor Sprout provides one protectant of the Sorcerer's Stone.
She supplies Devil's { Snare / Herbology / Army }.

141. Harry has a patronus.
It takes the form of a { stag / dementor / Sectumsempra }.

142. The Sorcerer's Stone can change the form of metal.
The metal turns to { gold / Flamel / Yule }.

143. House-elves can be set free.
Their masters must present them with { clothes / Dobby / Lavender }.

144. A patronus can defend against the soulless creatures that gradually deprive the mind of happiness.
They are known as { Dementors / Azkaban / Ollivander }.

145. Harry is able to track the movements of other wizards.
He uses something called the Marauder's { Map / mischief / Petunia }.

146. Harry often visits the Weasleys.
Their home is known as the { Burrow / garden / Accio }.

147. A prize award is given for the Triwizard Tournament.
It is set at one thousand gold { Galleons / Fleur / veela }.

148. Rita Skeeter is a journalist who often fabricates information to write an appealing story.
She is aided by devices like the Quick-Quotes { Quill / beetle / fairy }.

149. There is a gateway between the non-wizarding world and Diagon Alley.
It is located in a pub called the Leaky { Cauldron / London / dwarf }.

150. The Weasley family home has a garden.
It is infested with { gnomes / Burrow / umbrella }.

151. Only some students may participate in the Triwizard Tournament.
They are chosen by the Goblet of { Fire / Cedric / portrait }.

152. There is a way to make a candle give light only to the person holding it.
You can place it in the Hand of { Glory / Knockturn / Gryffindor }.

153. At one point, a horcrux is hidden at 12 Grimmauld Place.
It was taken there by a house-elf called { Kreacher / locket / owl }.

154. Members of Gryffindor house enter by speaking with the Fat Lady.
She appears in a { portrait / password / tail }.

155. There is one main sport in the wizarding community.
It is known as { Quidditch / broomstick / Nick }.

156. Harry still stays with the Dursleys during the summers.
They live in Little { Whinging / Dudley / troll }.

4.8.2 Appendix B

<u>Instructions for the similarity task</u>

(1) Do the two word meanings behave similarly (e.g., do they perform the same actions)?

(2) Do the two word meanings share physical / sensory properties (e.g., do they look, taste, smell, sound or feel similarly)?

(3) Do the two word meanings share many functional properties (e.g., are they used in similar ways, or do they serve a similar purpose)?

(4) Do the two word meanings share any other properties and/or features in common?

<u>Instructions for the relatedness task</u>

(1) How likely are the words to show up within the same context (that is, in/around the same part of the HP stories)?

(2) How important does one word seem to be for understanding the meaning of the other?

(3) Are the two words related via some theme, topic, event, or episode/scenario in the HP stories?

(4) Are the two words related via any other relationship?

CHAPTER FIVE: GENERAL DISCUSSION AND CONCLUSIONS

5.1 Summary of goals

The purpose of this thesis was to investigate the contributions of individual differences in knowledge to moment-by-moment access to information during sentence processing. To this end, we explored individual differences in knowledge of a fictional domain, the narrative world of Harry Potter, and their influence on word processing during sentence comprehension, asking a series of inter-related questions:

1. Does domain knowledge influence the quick availability of contextually-supported information?
2. Given that it does: is the link between domain knowledge and real-time effects of contextual support strictly due to more knowledgeable individuals knowing more facts? Or is there an influence of domain knowledge beyond predicting the likelihood that an individual will know a given fact?
3. Does knowledge go beyond mediating effects of contextual support to influence the functional organization of information in semantic memory, thereby modulating the quick availability of sentence-related information, e.g., category- and event-related information—the sorts of information that can support word-by-word sentence processing? And if so, is this relationship due specifically to more knowledgeable individuals knowing (more) specific facts, or, rather, might it reflect implicit, "partial" information (e.g., at the level of semantic features)?

Next, we consider the answers to each of these questions in light of results from the thesis taken as a whole, pointing out consistencies (and any inconsistencies) across the three studies. We conclude by describing the implications of our findings for a conceptualization of moment-by-moment use of knowledge during real-time sentence processing and comprehension.

5.2 Knowledge and context

In Chapter Two, we asked whether such domain knowledge would interact with context, such that effects of sentential context would be most pronounced for those with the greatest domain knowledge. To do so, we compared ERPs to critical words which completed sentences about general topics (control sentences) or Harry Potter (HP sentences) and were either contextually supported or unsupported. As expected based on the real-time sentence processing literature (reviewed in the introduction), we found that domain knowledge did indeed have a rapid influence on ERPs to the critical word, modulating effects of contextual support in sentences about HP within a third of a second. Moreover, these N400 context effects were specific: no such influences of domain knowledge were observed on control sentences about general topics, and effects of contextual support in HP sentences were driven by the relationship between HP domain knowledge and N400 amplitudes to supported, but not unsupported words.

In Chapters Three and Four, we replicated the finding that HP domain knowledge (measured via a trivia quiz) was positively related to N400 amplitudes to supported HP sentence endings, such that greater knowledge was associated with more positive-going (i.e., reduced) N400 amplitudes. Visualizations of these moderate-to-strong correlations are provided in Figure 5.1, and we plot ERPs to contextually supported words for high- vs. low-HP-knowledge subgroups (based on a median split, as discussed in Chapters Two, Three, and Four) across the three studies in Figure 5.2.

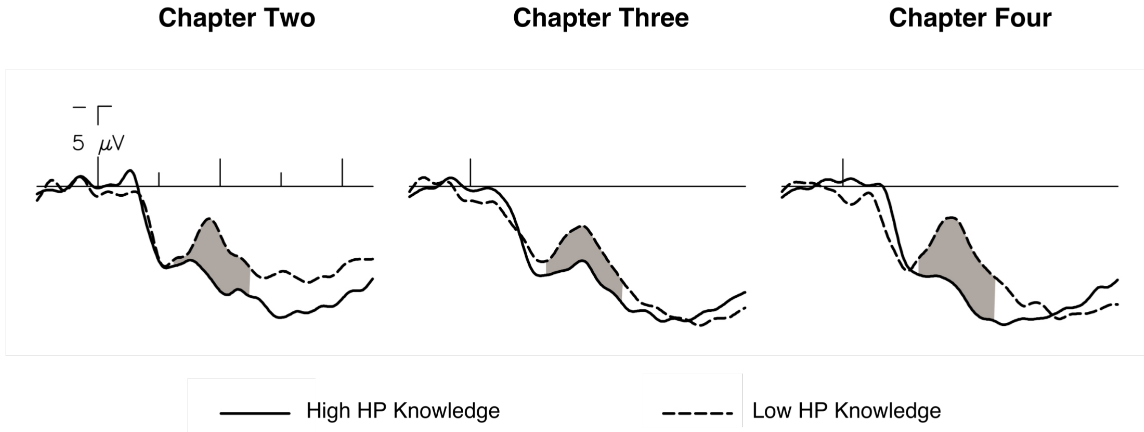


Figure 5.1. ERPs to words in the supported condition are plotted for high and low HP knowledge subgroups for experiments in Chapters Two, Three, and Four. Data are from a centro-parietal ROI, and shaded regions correspond to the N400 time period between 250 and 500 ms.

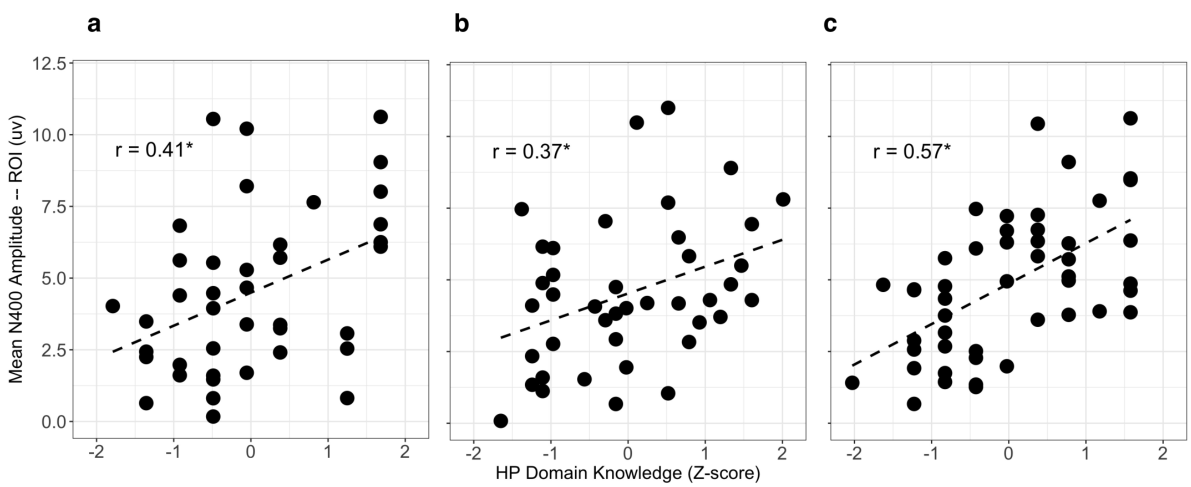


Figure 5.2. Mean N400 amplitude averaged across all items within the HP supported condition, by participant, is plotted against participants' HP domain knowledge scores for (a) Chapter Two, (b) Chapter Three, and (c) Chapter Four.

These findings support the notion that individual-level knowledge can have a quick influence on individuals' ability to bring contextual information to mind in real time. The sentence frames in our studies were all designed to end in a single contextually-appropriate "correct" word—for example: '*The character Peter Pettigrew changes his shape at times. He takes the form of a…*'. In this particular example, all of the words (save for the proper name '*Peter Pettigrew*') would be familiar to an English speaker, regardless of their knowledge of the domain of Harry Potter. However, our empirical results support the (rather obvious) intuition that the knowledge cued by this language must depend on specific knowledge about the events and characters of the HP book series.

In the studies presented here, we only analyzed the final (critical) word, showing that N400 amplitudes were reduced as a function of degree of domain knowledge only to contextually supported (or, in the case of the study presented in Chapter Four, related) words. However, if our interpretation is correct, we would expect that domain knowledge should modulate what information is accessed not only at the critical word, but over the course of the sentence, as context accrues. Indeed, ERP studies of contextual accrual have indicated that as a sentence unfolds, processing the semantics of incoming content words increases in ease (as inferred from N400 potentials; Van Petten & Kutas, 1990). As the materials we used in our studies are quite variable in terms of sentence structure and length, we cannot easily assess such influences of contextual accrual as a function of domain knowledge. In future work, however, tighter control of sentence materials would afford an analysis of how domain knowledge might influence semantic processing across sentential contexts.

Though the focus across our studies was on N400 amplitudes (and primarily in a single centro-parietal spatial region of interest), we also examined the late positivity window across studies. In the literature, late positive complexes have also been known to be modulated by contextual support, with effects showing different scalp topographies according to whether "unsupported" words are plausible continuations (linked to posterior positivities) and/or are unexpected but plausible and/or semantically related to the best completion (linked to frontal positivities; Thornhill & Van Petten, 2012; Van Petten & Luka, 2012; DeLong, Quante, & Kutas, 2014). Although we had no critical hypotheses about how domain knowledge might interact with context in this time window, we speculated that, to the extent that our unsupported words constituted contextual violations *more so* for individuals with high vs. low HP knowledge, we might expect to see effects of contextual support in the late positivity time window which were largest for individuals with the greatest knowledge.

The studies presented in Chapters Two and Four allowed us to examine this possibility, as both included HP sentences ending in contextually supported vs. unsupported words. However, the results from the two studies seem inconsistent with one another. In Chapter Two, there was no influence of contextual support for HP sentences during this time period, and no interaction between contextual support and HP knowledge. In Chapter Four, however, we did observe an interaction between HP domain knowledge and context, such that individuals showed increased late (parietal) positivities to both types of unsupported (i.e., related and unrelated) words compared to supported words as a function of their degree of domain knowledge. One difference between the studies presented in Chapter Two and Four is that the sentence materials in Chapter Four were more likely to contain true "anomalies" rather than words which were simply incorrect. Indeed, in Chapter Two, the unsupported words were designed to seem plausible to the HP-uninitiated, e.g., '*Harry has a patronus. It takes the form of a stag / lion*.' In Chapter Four, however, sentences in both the unrelated and related conditions were not designed with this constraint in mind, and, though we did not norm these sentences for plausibility, it seems likely that the degree of anomaly/plausibility was more variable in Chapter Four. For example, consider a sentence pair like '*In order to save the sorcerer's stone, Harry must go through a trapdoor. It is guarded by a three-headed* *dog / flute / Life*.' Here, the unsupported/related ending '*flute*' is not something which typically has heads, and the unsupported/unrelated ending '*Life*' is perhaps an even less plausible continuation (dogs have heads, flutes are physical objects which do not have heads, and life is an abstract concept). By contrast, in '*One Horcrux is destroyed by Fiendfyre. This is Rowena Ravenclaw's* *diadem / cup / egg*,' there is no obvious difference in plausibility among the three ending types. Late parietal positivities have been linked specifically to instances when a word is implausible given the context (e.g., DeLong et al., 2014a). It stands to reason that, to the extent that individuals with greater HP knowledge had a fuller understanding of the sentence contexts, they may have been more sensitive to the implausibility of critical words, leading to the observed variation in late positivities in Chapter Four. This pattern of results—domain-knowledge-driven variation in context effects in Chapter Four, in which many sentence endings were truly anomalous, but not Chapter Two, in which endings were designed to seem plausible—hints at a novel way to distinguish between the functional significance of late positivities with different scalp distributions. In the future, one way to shed more light on this issue might be to conduct a study to assess the influence of domain knowledge on processing words that are contextually supported vs. "incorrect but seemingly plausible" vs. "incorrect and outright anomalous." This would allow for a more fine-grained look at how brain's sensitivity to plausibility (contrasting words which are plausible at a surface level but with those which are more obviously anomalous) during a time period (~500-900 ms post-stimulus) when processes related to mental model updating and reanalysis of sentence content are occurring (Brouwer et al., 2012).

In sum, results from three ERP studies show that degree of knowledge (here, of a fictional domain) can rapidly modulate effects of sentential context. Next we discuss findings relevant to determining whether differential organization of knowledge in semantic memory due to greater domain knowledge indeed leads to this pattern of results, compared to an account on which these findings could be explained by individuals with greater domain knowledge simply knowing *more* of the facts that we presented to our participants.

5.3 Domain vs. specific knowledge

In Chapter Three, we employed a single-trial paradigm in order to test whether the influence of domain knowledge on the N400 context effects observed in Chapter Two were strictly a result of the proportion of trials each individual knew, or whether domain knowledge had an influence beyond this. We were able to determine empirically that individuals with greater HP domain knowledge (measured by our trivia quiz, as in Chapter Two and elsewhere) did indeed report knowing more facts in real time. We could then ask whether individuals' domain knowledge had an influence beyond strictly mediating the overall proportion of trials that a given individual reported knowing. We found that this was indeed the case and seemed to be driven by items which were most likely to incur relatively high retrieval difficulty—that is, trials which participants reported *not* knowing, and, in an analysis incorporating item-level offline cloze probability, items which were generally more difficult to know / remember across individuals.

There are multiple (non-mutually exclusive) interpretations for the finding that domain knowledge quickly influenced the brain's response to contextually supported words beyond mediating the proportion of trials reported as known. One set of possibilities relates to the idea that individuals with different degrees of domain knowledge may have used systematically different criteria when making a decision to respond affirmatively that they had known a trial. It is possible, for example, that more knowledgeable individuals were more conservative in making responses of “yes.” Framed another way, a response of “yes” might have meant something different to an individual with a thorough knowledge of main facts within the HP domain, compared to an individual with a more cursory knowledge—the task demands may have seemed subjectively different to such individuals.

Another set of possibilities relates to putative differences in the functional organization of domain knowledge stored in long-term memory. As reviewed in the introduction, experts often seem to have access to highly structured/organized systems of semantic memory within the domain of expertise, with concepts organized around high-level schemas, and with the potential to access large and consolidated “chunks” of information (compared to information that is represented more piecemeal and/or is less structured, as has been proposed for those with less domain expertise). We propose that the observed results (that domain knowledge influences N400 amplitudes to contextually supported words beyond mediating the proportion of trials known) may indeed reflect such differential functional organization of semantic memory. We explored this hypothesis further in Chapter Four by manipulating sentence ending types using a related anomaly paradigm (discussed in section 5.4).

The influence of domain knowledge that we observed in Chapter Three was most prominent on trials likely to incur retrieval difficulty, i.e., trials which participants reported not having known and/or which were generally less likely to be known / remembered (as determined by offline cloze probability judgments by a separate group of participants of varying degrees of HP knowledge). What could account for this pattern of results? One possibility is that even though individuals with greater domain knowledge report not knowing a given trial, they still make use of access to deeper / richer information cued by the sentential context which is relevant for processing the critical word. Taking this a step further, it is possible that such knowledgeable individuals might actually know the correct word at some level, even though they did not feel as though they were able to bring it to mind in real time; another possibility is that the factual knowledge was not (and perhaps had never been) present, but other "partial" knowledge may have allowed them to more readily process the supported word.

As a concrete example, consider a hypothetical reader with moderate HP knowledge. She reads, '*To combat boggarts, wizards must think of something funny. They must use the spell <u>Riddikulus</u>.*' Perhaps she knows that such a spell exists, but until processing the word '*Riddikulus*,' she did not recall what the word was. Therefore, she responds that no, she had not known the information prior to having read the sentence, even though *after* reading the sentence the linking information was available. Later in the experiment, she reads, '*Wizards are able to conjure the Dark Mark. They can use a spell called <u>Morsmordre</u>.*' Perhaps in this case, the reader knows what the Dark Mark is, knows there is a spell that can conjure it, and has heard of something called '*Morsmordre*,' but has no recollection that this is the name of the spell in question. In this case, she also responds no, she had not known the information prior to having read the sentence. In both examples, it is possible that the sentence contexts cue information that may be relevant for processing the final word—but in the first case, a link already existed in the mind of the reader. It is possible that this type of link is more likely to exist for high-knowledge individuals, thereby leading to the observed semantic facilitation (inferred from reduced N400 amplitudes) at the critical word on trials which high-knowledge individuals report not having known.

On such an account, so-called "partial" (and likely implicit) information may be more readily available as a function of expertise. First, the "chunks" of information available to readers, cued by the words in the sentence context, are likely to differ in number, degree of depth, and/or degree of availability as a function of the individual's degree of expertise. Moreover, these "chunks" are themselves likely to be associated (to different degrees) with additional "chunks" of information, also as a function of the individual's degree of expertise. Elman and McRae (2019) propose a model of event knowledge (though not of linguistic processing) in which such "chaining" of events (or sub-events) occurs as activation cascades from one event to related events. In the case of our materials, it may be that the "chunks" of information relevant for processing the sentences behave similarly—especially for individuals with greater HP expertise.

The argument in Chapter Three that domain knowledge might have its most prominent influence in cases of the greatest retrieval difficulty was also based on the patterning of the interaction of trial-level knowledge reports from individuals and offline cloze probability measures from a separate group of participants, which we took as a proxy for overall difficulty of the item.[16] In particular, we observed that domain knowledge had its greatest influence on N400 amplitudes to critical words for trials which were reported as unknown, but which were high cloze items (based on offline norming studies), and on trials which were reported as known, but which were low cloze. In other words, these were trials which were not the easiest (as in the case of trials participants reported knowing *and* were high cloze, more generally) and which might incur retrieval difficulty, but on which retrieval was not outright impossible. Considering all three studies from Chapters Two, Three, and Four, we propose that having greater domain knowledge may result in (a) increasing the availability of chunks of information available in long-term memory that can be cued by the sentence context, (b) increasing the number of links (and/or

[16] Here, we recognize that we are taking ERPs to the critical (final) word as a proxy for a real-time processing measure of knowing each fact; we think this is reasonable because we constructed our sentences such that the critical words indeed supplied a critical concept in completing the propositional content of our sentence pairs.

strength of links) stored in long-term memory between these chunks, such that relevant chunks may more readily / rapidly cue related chunks, thereby (c) increasing the availability of relevant cues for processing critical “factual” (in this case, fictional) words.

We also explored specific, trial-level knowledge using an offline cloze task in Chapter Four. There were numerous differences between the design of the ERP experiments in Chapters Three and Four, making it hard to directly compare results of specific knowledge in each. In Chapter Four, our sentence materials included not only contextually supported endings (as in Chapter Three) but also two types of contextually unsupported endings (related, unrelated). Across the two studies, we did replicate the main effect of single-trial knowledge on contextually supported words (via our post-experiment cloze task in Chapter Four and reports of knowledge during the ERP experiment in Chapter Three): N400 amplitude was reduced to contextually supported words that participants knew, vs. those which they did not. However, in Chapter Four, we did not replicate the finding that domain knowledge had a specific influence on items that were *not* known (i.e., to which participants incorrectly responded on the offline cloze task), finding instead that HP domain knowledge did not modulate the overall pattern of the influence of accuracy and ending type (i.e., that accurate trials led to reduced N400 amplitude only for the supported condition). However, one major difference between the studies in Chapters Three and Four which may have contributed to this apparent discrepancy is that in Chapter Four, for the supported condition, participants had already seen the appropriate (correct) ending during the ERP study when they made their post-experiment report on the offline cloze task. As such, items which might have incurred some retrieval difficulty—i.e., those items which might have been known, but to a lesser degree or with less confidence, and the very items which seemed to be most modulated by domain knowledge in Chapter Three—may have received a “boost” such that they were accurately recalled and reported during the cloze task. Future studies might use a pre-experiment offline cloze task to assess a less contaminated measure of trial-level knowledge in order to determine if this account is indeed accurate.

5.4 Expertise and the functional organization of knowledge

In Chapter Four, we further investigated the extent to which domain knowledge might modulate real-time sentence processing beyond strictly determining the ability to quickly make use of knowledge to process contextually supported words. To do so, we used a related anomaly paradigm, in which participants read HP sentences ending in a contextually supported word, an unsupported but related word, or an unsupported and unrelated word. Across all participants, we observed the predicted three-way difference in N400 amplitude according to ending type (supported endings < related endings < unrelated endings). In support of the hypothesis that domain knowledge mediates availability of related/relevant information for understanding words in sentential contexts (beyond linguistically supported words themselves), we observed that not only the N400 effects of contextual support but also N400 related anomaly effects were influenced by an individuals' degree of domain knowledge. That is, individuals with greater domain knowledge showed a greater reduction in amplitude of related, compared to unrelated, sentence endings.

We interpreted these findings as consistent with the notion that individuals with greater domain knowledge have access to knowledge structures that are tightly organized around important principles. Whereas in a domain like physics, the organizing principles might relate to theories about physical entities (like force and motion), in the case of the narrative world of Harry Potter, organizational structures might include novel categories (like types of spells, magical creatures, magical school subjects), typical events (e.g., what happens during a Quidditch match—

a magical sport), specific episodes within the story narrative (e.g., the time a flying car hit a scary magical tree), main characters (Harry, Hermione), and so on. We used two types of related anomalies, category- and event-related words, finding the same pattern of results across each. This said, it is likely that the nature of the (pre-)activated information leading to the related anomaly effect was different for each. For category-related anomalies, it is likely that the (pre-)activation of semantic features shared by category members (e.g., the fact that spells are conducted by wizards, using wands, involving the utterance of Latinate phrases, etc.) are what leads to related anomaly effects. For event-related anomalies, it seems more likely that the (pre-)activated information is not limited to semantic features shared with a best completion, but rather that (pre-)activated information would include concepts related via the general event (or, in some cases, a specific episode). Though measures from a language model (word2vec) and human judgments of our sentence materials hinted at a potential dissociation between the two sentence types, any differences between the two were not discernible from our N400 results using the current analyses. In the future, more fine-grained analyses at the level of item-specific semantic properties (for example, the degree of relatedness/similarity between critical words and sentential contexts derived from language models and/or human norming studies) might shed light on any systematic differences between category and event relationships.

Interestingly, although N400 amplitudes to words in the related anomaly condition were reduced as a function of domain knowledge, they did not show a sensitivity to whether individuals could correctly produce the correct word in a post-ERP experiment offline cloze task. This suggests that knowing the correct completion of the sentence was not necessary to lead to a reduction in N400 amplitude for these words. Of course, it is difficult to make arguments based on a single null result—it is possible that with a greater number of items and/or participants we would be able to detect an effect of item-specific knowledge on the N400 response to related anomalies. However, given that, as expected, we saw large effects of accuracy on the N400 to supported words (and, also as expected, no effects of accuracy on the N400 to unrelated words), such an effect (if it exists) is likely to be quite small. A second interpretation of this pattern of results is that, as individuals with greater knowledge read our sentence contexts, they activated / accessed relevant (though not necessarily linguistically licensed) information for understanding the meaning of the sentence. This information could have been sufficient to (pre-)activate information relevant for making sense of the related anomalies, even in the absence of (pre-)activation of the contextually supported (correct) word.

5.5 Knowledge and reading experience

Our primary individual-difference measure of interest across studies was HP knowledge, measured by performance on a trivia quiz about Harry Potter. In Chapters Two and Four, this quiz consisted of 10 multiple choice questions; in Chapter Three, we used a longer quiz of 40 questions. In addition, all participants responded to a subjective questionnaire asking about their experience with the HP books, movies, and other HP-related media (including listening to podcasts, visiting theme parks, and so on). A numeric, standardized score was assigned to participants based on their responses to this questionnaire. Unsurprisingly, this score was strongly correlated with performance on the HP trivia quiz across all three studies (range of *r*s = [.73 - .76]). Also unsurprisingly, HP knowledge—gleaned from *reading* among other sources of information—was positively correlated with an aggregate measure of reading experience; this correlation was significant in Chapters Two and Four (range of *r*s = [.54 - .55]), but did not reach significance in Chapter Three (*r* = .19).

In Chapter Two, in a regression model fit to N400 data from HP sentences incorporating not only HP knowledge and sentence ending type, but also measures of individual differences (including the aggregate measure of reading experience along with a measure of verbal working memory and an aggregate measure of general knowledge), we found that reading experience, rather than HP knowledge, interacted with ending type. These findings seem to suggest that differences in overall reading experience (and perhaps ability), rather than HP knowledge, was the primary reason for observed effects. However, we argued that, since HP knowledge and reading experience were moderately correlated (and therefore partially collinear in our regression model), it was impossible to tease the effects apart with our design and analyses. And, since HP knowledge specifically influenced the HP (and not control) sentences in our simpler regression analyses (i.e., those not incorporating individual differences), we concluded it was more parsimonious to interpret our findings as being driven by HP-specific knowledge, which was, in our group of participants, partially confounded with reading experience.

In further support of this interpretation, in Chapters Three and Four, we did not observe the same pattern of results in regression analyses fit to ERP data incorporating influences of individual differences. Instead, we found that models incorporating individual differences in reading experience, verbal working memory, and general knowledge did not explain variance beyond models which did not incorporate these differences. Future studies may more systematically assess the extent to which reading experience matters (or does not) during word-by-word comprehension by more carefully controlling the population (such that reading experience is matched across high- and low-knowledge subgroups) and/or by investigating multiple domains of knowledge simultaneously, such that individuals who are "experts" in one domain can serve as their own controls as "novices" in another domain.

5.6 Knowledge variability and models of sentence processing

In the introduction, we summarized findings from the text and discourse processing literatures showing that domain knowledge can influence offline measures of text memory and comprehension. However, only limited work has addressed the role of individual differences in domain knowledge during real-time sentence processing. As such, current models and descriptive theories of moment-by-moment sentence processing do not typically take variability in domain / background knowledge into account. Here, we briefly review several current veins of models of sentence processing and discuss how the work presented in this thesis might make contact with each.

It is no longer contested that sentence processing proceeds incrementally, with linguistic information often being used as soon as it is made available (e.g., Tanenhaus et al., 1995; Altmann & Kamide, 1999) and, often, anticipatorily, with many theories positing that individuals make (explicit or implicit) predictions about upcoming linguistic features (reviewed in Kutas, DeLong, & Smith, 2011; DeLong, Troyer, & Kutas, 2014; Kuperberg & Jaeger, 2016). One line of theories describing incremental sentence processing has placed the emphasis on syntactic (word order) processes and has focused on explaining behavioral effects of syntactic complexity. For example, there is a well-documented difference in difficulty in processing object vs. subject relative clauses, which are syntactic structures that differ in word order. Subject relative clauses use the canonical word order in English (Subject-Verb-Object) and are more frequent in the input; object relative clauses use an atypical word order (Object-Subject-Verb) and are less common. Some studies suggest that attested differences in processing object vs. subject relative clauses can be explained by differences in language experience and frequency of usage (MacDonald & Christiansen, 2002), while others have proposed theories which place the crux of the difficulty on memory processes

(Gibson, 1998, 2000; McElree, 2000; Lewis & Vasishth, 2005). The latter class of theories tends to predict that individual differences in (verbal) working memory ability—in brief, the ability to hold and manipulate information for the short-term—should be correlated with behavioral syntactic complexity effects, though the evidence for this has been mixed (Just & Carpenter, 1992; Waters & Caplan, 1996). The former class of theories, however, predicts that an individual's experience with syntactic structures (and other linguistic structures) will more strongly influence sentence processing—which is empirically supported (e.g., Wells, Christiansen, Race, Acheson, & MacDonald, 2009; Montag & MacDonald, 2015).

The current data could potentially inform both types of accounts. As for experience-based theories of syntactic complexity effects, it is straightforward to see how individual differences in knowledge (gleaned from experience) might shape not only the availability of syntactic structures but also the nature of the "chunks" of meaningful information (i.e., the semantics) that syntactic structures must make contact with. And, as for memory-based accounts of similarity-based retrieval interference (e.g., Gordon, Hendrick, & Levine, 2002), an incoming word can be difficult to process if a (syntactic) connection needs to be made to a preceding word but an intervening word is similar in nature to the one that needs to be retrieved. For example, in the object relative clause '*The banker that the barber praised…*,' *banker* and *barber* are both descriptions of occupations; changing the phrase '*the barber*' to a pronoun (e.g., '*The banker that you praised…*') leads to easier processing (as inferred from reaction time studies), which has been interpreted as evidence that reducing the similarity of items leads to less interference and, therefore, easier retrieval (Gordon, Hendrick, & Johnson, 2001). It stands to reason that if individual differences in knowledge do indeed shape the nature of semantic/linguistic representations stored in long-term memory, then they may also shape the extent to which words are similar. That is: the nature of "chunks" stored in long-term memory seems likely to have consequences for such a view of sentence processing.

In computational linguistics, many researchers have drawn from information theoretic approaches to language, modeling word-by-word difficulty in sentences using mathematical concepts like *surprisal* (e.g., Hale, 2001; Levy, 2008a) and *entropy* (Hale, 2016). The surprisal of an upcoming word is its relative probability of occurring given some recent context (i.e., preceding words). Surprisal is typically estimated using a language model that is trained on a large corpus of text. Similarly, a word's entropy can also be derived from text corpora and is a measure related to the distribution/equiprobability of likely upcoming words—entropy reduction has to do with how much the current word reduces the relative degree of ambiguity of upcoming words. These measures may be viewed as roughly analogous to behaviorally-derived measures of cloze probability and sentential constraint and have been linked to N400 and late positive ERPs, respectively (Frank et al., 2015; Frank & Willems, 2017). In addition, distributional semantic models of word meaning (discussed briefly in Chapter Four) have recently (re-)gained popularity, with models like Google's word2vec (Mikolov, Chen, Corrado, & Dean, 2013; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) and GloVe (Pennington, Socher, & Manning, 2014) being readily available for researchers to use.

Because these models rely on the statistics of large text corpora to make quantitative predictions, it is not immediately clear how they could incorporate individual differences in knowledge without having to build models from entirely different input. Though potentially computationally intensive, such an approach could provide an interesting way to model word and meaning representations for individuals with different degrees of knowledge within a domain and, indeed, with different profiles of background knowledge, more generally. In Chapter Four, we took a step in this direction by training Google's word2vec with CBOW architecture directly on the text of the HP book series in order to quantify meaning for our sentence materials. In the future, researchers could exploit this approach by comparing how models trained on corpora containing different degrees of information about a given topic relate to empirical knowledge-based differences in word processing and/or word meaning representation.

Noisy channel models are a branch of probabilistic inference models which aim to explore the notion that comprehenders must understand language in a noisy environment which is prone to uncertainty and to comprehension error (Shannon, 1948; Levy, 2008b). Gibson and colleagues (2013) describe these models as making several key predictions about comprehension, including the notion that, in the face of "noisy" input, individuals should be biased towards plausible interpretations and should also be sensitive to the statistics of the noise in the environment (Gibson, Bergen, & Piantadosi, 2013). Gibson et al. propose a Bayesian model of how comprehenders decode intended meanings, which is based on the individual comprehender's priors as well as a noise term (which involves knowledge of how sentences are likely to be "corrupted" or miscommunicated). Gibson et al. describe these priors as "…all of the comprehender's relevant linguistic and world knowledge, including for instance the base-rate frequencies of different grammatical constructions and the plausibility of different meanings" (p. 8051). However, in their empirical tests of some of the predictions of their theory, this linguistic and world knowledge is left implicit, and is assumed to give rise to differences in plausibility of the sentences they test. It stands to reason that individual differences in knowledge are quite likely to give rise to individual differences in the perceived plausibility of sentences, a hypothesis that could be empirically tested and potentially incorporated in such a model.

More broadly, a range of models from computational linguistics, psycholinguistics, and the neurobiology of language converge to suggest that comprehenders understand language in an anticipatory and probabilistic manner (Kuperberg & Jaeger, 2016). Recent neurocomputational approaches have aimed to model language-relevant event-related brain potentials—most notably N400 effects and, to a lesser degree, late positive complexes (e.g., Brouwer & Hoeks, 2013). One line of proposals posits that N400 amplitudes reflect implicit "prediction error" – i.e., that at the point of processing a word, it is evaluated with respect to the previous sentence or discourse context, and the amplitude of the resulting N400 reflects the additional semantic information needed to make sense of the word (Rabovsky & McRae, 2014; Rabovsky, Hansen, & McClelland, 2018). This line of work subtly contrasts with a neurocomputational model put forth by Brouwer and colleagues, which views N400s as reflecting retrieval of word-level information and late positive complexes as reflecting the integration of this information into the ongoing representation of the meaning of the sentence (Brouwer, Crocker, Venhuizen, & Hoeks, 2017). As such, both models are consistent with the notion that language processing is anticipatory in nature, but seem to propose a different timecourse for the locus of "updating" or "integration" of information into an existing mental model.

These models echo debates from descriptive theories which have viewed N400 effects as being primarily driven by semantic integration (e.g., Brown & Hagoort, 1993) vs. retrieval of semantic information (Kutas & Federmeier, 2000; see Lau, Phillips, & Poeppel, 2008 for a discussion). On the former view, N400 amplitudes are primarily viewed as reflecting comparison of an incoming word to an ongoing construction of meaning of the sentence. On the latter view, N400 amplitudes are believed to reflect ease of access to semantic features of an incoming word, which may be influenced by the ongoing construction of meaning of the sentence (or any other relevant/meaningful contextual information). The studies presented in this thesis were not designed to distinguish between these proposals. However, our results seem to provide support for theories in which the N400 reflects semantic retrieval, and less for theories in which it reflects prediction error / sentential integration. For example, in Chapter Three, we observed differences in N400 amplitudes to words in sentential contexts describing facts that individuals (of differing degrees of HP knowledge) reported *not* knowing. Even so, we observed reduced N400 potentials for individuals with higher compared to lower knowledge. The simplest explanation for this seems to be that individuals with greater knowledge differentially accessed semantic information from long-term memory, leading to facilitated retrieval. Indeed, on an integration account, we might have predicted the opposite pattern of results: individuals who know more about the world of HP should have a better understanding of the sentential context and might have found it *more* difficult to integrate an unknown incoming word, with unknown words thereby eliciting *greater* N400 amplitudes compared to those for individuals with less knowledge. However, as current computational models do not take individual differences in knowledge into account, is it difficult to know exactly what they would predict.

A separate inter-connected series of proposals has provided a neurobiologically plausible model of N400 potentials by implementing the appropriate excitatory dynamics of pyramidal neurons (believed to give rise to scalp-recorded brain potentials; Luck, 2014) and inhibitory properties of interneurons (Laszlo & Plaut, 2012; Laszlo & Armstrong, 2014; Cheyette & Plaut, 2017). These models attribute observed N400 activity to over-excitation of semantic information at the level of the excitatory dynamics, which is then modulated by the interneuron activity over time. This approach can account for a range of empirical phenomena, primarily at the level of isolated words (as opposed to words occurring in sentential or discourse contexts), for example by (correctly) predicting that words with greater numbers of semantic features and/or orthographic neighbors should elicit large-amplitude N400s (as opposed to those with fewer neighbors) because of the additional semantic features that can be activated. The putative differences in functional organization of information we have proposed in this thesis may have consequences for the degree to which such neural excitation can occur—for example, if individuals with greater knowledge have more links between semantic content stored in long-term memory, and more / deeper "chunks" of information available, then perhaps there should be *more* excitation and therefore overall greater N400 amplitudes to single words from the HP domain (in isolated contexts). However, a recent ERP study of single words in the domain of math suggested the opposite pattern (Bechtold, Bellebaum, Egan, Tettamanti, & Ghio, 2019). Incorporating individual-level (domain) knowledge within a framework like that of the model by Plaut, Laszlo, and colleagues might therefore yield interesting predictions which could be tested empirically.

In sum, most descriptive and computational models of language seem to assume that knowledge stored in long-term memory is fundamental to how people process language, and many current theories espouse views in which language content is probabilistically pre-activated according to context and knowledge. Yet most of these models seem to implicitly assume that this knowledge is similar across individuals—or at least they do not explicitly incorporate individual-level variation in knowledge. Results from the current thesis show empirically that differences in knowledge (when they exist) have a quick influence on the neural processing of words in sentence contexts. Should current models incorporate such individual differences, they might afford more precise predictions about the nature and timing of information access during sentence processing, and ultimately to a better understanding of how individuals use knowledge in real time to make sense of language.

5.7 Conclusions

Everyone knows that we use our knowledge to make sense of sentences. But here, we have demonstrated this empirically, and we have learned that the influence of domain knowledge in real time is quick and specific. Our results show that domain knowledge can influence retrieval of word information even when that information is unlikely to be wholly retrieved, and even when the words we encounter only partially overlap in meaning with a word licensed by the linguistic context – indicating that what individuals know about a domain rapidly influences access to partial and implicit word information as they attempt to comprehend words in written sentences in real time.

Our studies have employed a range of methods from cognitive science along with state-of-the-art regression modeling analyses to investigate individual differences in knowledge and how these impact moment-by-moment linguistic and neural processing as individuals read. It is our hope that these studies can spur new research in individual differences in language, with more of a focus on how variation in knowledge can influence the moment-by-moment dynamics of language processing. We also think these findings can connect with the literature on pedagogy, in that they begin to show how the same language can be processed (and presumably interpreted) quite differently depending on what the individual knows. What is brought to mind by the knowledgeable individual may spur them to draw connections and make inferences in ways that are unavailable to a less knowledgeable individual. These results underscore the importance of providing individuals with the tools they need to build knowledge—and think like experts.

5.8 References

Bechtold, L., Bellebaum, C., Egan, S., Tettamanti, M., & Ghio, M. (2019). The role of experience for abstract concepts: Expertise modulates the electrophysiological correlates of mathematical word processing. *Brain and Language*, *188*, 1-10.

Brown, C. & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*(1), 34-44.

Brouwer, H., Crocker, M.W., Venhuizen, N.J., & Hoeks, J.C.J. (2017). A neurocomputational model of the N400 and P600 in language processing. *Cognitive Science*, *Suppl. 6*, 1318-1352.

Brouwer, H. & Hoeks, J.C.J. (2013). A time and place for language comprehension: mapping the N400 and P600 to a minimal cortical network. *Frontiers in Human Neuroscience*, *7*(758), 1-12.

Cheyette, S.J. & Plaut, D.C. (2017). Modeling the N400 ERP component as transient semantic over-activation within a neural network model of word comprehension. *Cognition*, *162*, 153-166.

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

DeLong, K.A., Troyer, M., & Kutas, M. (2014b). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass*, *8/12*, 631-645.

Elman, J.L. & McRae, K. (2019). A model of event knowledge. *Psychological Review*, *126*(2), 252-291.

Frank, S.L., Otten, L.J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain & Language*, *140*, 1-11. doi: 10.1016/j.bandl.2014.10.006

Frank, S.L. & Willems, R.M. (2017). Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. *Language, Cognition, and Neuroscience*, *32*(9), 1192-1203. doi: 10.1080/23273798.2017.1323109

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1-76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, Language, Brain*, 95-126. MIT Press.

Gibson, E., Bergen, L., & Piantadosi, S.T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Science*, *110*(20), 8051-8056.

Gordon, P.C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1411-1423.

Gordon, P.C., Hendrick, R., & Levine, W.H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*(5), 425-430.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *NAACL '01 Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, 1-8.

Hale, J. (2016). Information-theoretical complexity metrics. *Language and Linguistics Compass*, *10/9*, 397-412.

Just, M.A. & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122-149.

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

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

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

Laszlo, S. & Armstrong, B.C. (2014). PSPs and ERPs: Applying the dynamics of post-synaptic potentials to individual units in simulation of temporally extended Event-Related Potential reading data. *Brain & Language*, *132*, 22-27.

Laszlo, S. & Plaut, D.C. (2012). A neurally plausible Parallel Distributed Processing model of Event-Related Potential word reading data. *Brain & Language*, *120*, 271-281.

Lau, E.F., Phillips, C., & Poeppel, D. (2008). *Nature Reviews Neuroscience*, *9*, 920-933.

Levy, R. (2008a). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126-1177.

Levy, R. (2008b). A noisy-channel model of rational human sentence comprehension under uncertain input. *Proceedings of the 13th Conference on Empirical Methods in Natural Language Processing*, 234–243.

Lewis, R.L. & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 375-419.

Luck, S.J. (2014). *An Introduction to the Event-Related Potential Technique*. MIT Press.

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

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, *29*(2), 111-123.

Montag, J.L. & MacDonald, M.C. (2015). Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. *Journal of Experimental Psychology: General*, *144*(2), 447-468.

Pennington, J., Socher, R., & Manning, C.D. (2014). Glove: Global vectors for word representation. *Proceedings of the Empiricial Methods in Natural Language Processing* (EMNLP), 12.

Rabovsky, M., Hansen, S.S., & McClelland, J.L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behavior*, *2*, 693-705.

Rabovsky, M. & McRae, K. (2014). Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition*, *132*, 68-89.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 623–656.

Thornhill, D.E. & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*, 382-392. doi: 10.1016/j.ijpsycho.2011.12.007

Van Petten, C. & Luka, B.J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*, 176-190. doi: 10.1016/j.ijpsycho.2011.09.015

Waters, G.S. & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, *103*(4), 761-772.

Wells, J.B., Christiansen, M.H., Race, D.S., Acheson, D.J., & MacDonald, M.C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*, 250-271.