Journal of Memory and Language 113 (2020) 104111

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

To catch a Snitch: Brain potentials reveal variability in the functional organization of (fictional) world knowledge during reading

Melissa Troyer[a,1,*], Marta Kutas[a,b]

[a] Department of Cognitive Science, University of California, San Diego, United States
[b] Department of Neuroscience, University of California, San Diego, United States

ARTICLE INFO

Keywords:
Sentence processing
Event-related brain potentials
Knowledge
Semantic memory
Individual differences

ABSTRACT

We harnessed the temporal sensitivity of event-related brain potentials (ERPs) alongside individual differences in Harry Potter (HP) knowledge to investigate the extent to which the availability and timing of information relevant for real-time written word processing are influenced by variation in domain knowledge. We manipulated meaningful (category, event) relationships between sentence fragments about HP stories and their sentence final words. During word-by-word reading, N400 amplitudes to (a) linguistically supported and (b) unsupported but meaningfully related, but not to (c) unsupported, unrelated sentence endings varied with HP domain knowledge. Single-trial analyses revealed that only the N400s to linguistically supported (but not to either type of unsupported) sentence-final words varied as a function of whether individuals knew (or could remember) the correct (supported) ending for each HP "fact." We conclude that the quick availability of information relevant for word understanding in sentences is a function of individuals' knowledge of both specific facts and the domain to which the facts belong. During written sentence processing, as domain knowledge increases, it is clearly evident that individuals can make use of the relevant knowledge systematically organized around themes, events, and categories in that domain, to the extent they have it.

Introduction

Across cognitive systems, world knowledge allows individuals to organize raw sensation into meaningful experiences. For example, expert chess players know a wide variety of chess positions. Accordingly, they show better memory performance on tasks involving chess pieces compared to novices, but only when the pieces occupy veridical game positions, and not when they are randomly shuffled on the board. It is assumed that their knowledge allows them to capitalize on their familiarity with the meaningful arrangements (Simon & Chase, 1973). Understanding language is no exception—words cue world knowledge which can be brought to mind in real time rapidly, (e.g., Hagoort, Hald, Bastiaansen, & Petersson, 2004), incrementally (Altmann & Kamide, 1999) and sometimes even predictively (reviewed in Kutas, DeLong, & Smith, 2011).

It is incontrovertible that without knowledge (comprising long-term memory), people could not comprehend language. And yet, to date, variation in what people know (how much, how deeply, in what ways) has not played a significant role in any of the theoretical or computational accounts of real-time language processing (reading or listening). This is a glaring gap. We believe that variation in individuals' knowledge is likely to have consequences for multiple aspects of language processing, such as the speed and ease with which word-form information is retrieved, the availability of stored information from long-term memory, the mechanisms of (rapid) retrieval of the stored information, the ability to comprehend information from a sentence or discourse, and the mindsets and goals of comprehenders moment by moment.

The goal of the current study is to begin to develop a more precise description of how variation in knowledge—including which types of knowledge, their organization, and the timing of their use—might influence language processing by closely examining knowledge use in real time in word by word sentence reading. To this end, we integrate several approaches from cognitive science. Following traditions from the psychology literature on expertise and our own studies (Troyer & Kutas, 2018; Troyer, Urbach, & Kutas, 2019), we employ a specific domain for study, the narrative world of Harry Potter, which provides a semantically rich, large set of verbal descriptions (necessary for studying language processing) in a domain where individuals naturally vary in their knowledge. To go beyond our previous work, we adapt a

* Corresponding author at: Western Interdisciplinary Research Building, Room 5148, University of Western Ontario, London, Ontario N6A 5C2, Canada.
E-mail address: mtroyer2@uwo.ca (M. Troyer).
[1] Current address: University of Western Ontario in London, Ontario, Canada.

https://doi.org/10.1016/j.jml.2020.104111
Received 16 August 2019; Received in revised form 15 January 2020; Accepted 12 February 2020

related-anomaly paradigm (e.g., Federmeier & Kutas, 1999), combining a reading study employing event-related brain potentials with carefully constructed sentence materials in which we manipulate sentence-final words that are either contextually supported, unsupported and unrelated, or, critically, unsupported but semantically related to the sentential context. This allows us, for the first time, to subtly probe multiple types of knowledge activation during real-time reading in individuals who vary in their knowledge of the content.

Knowledge and real-time sentence processing

Historically, theories of real-time sentence processing focused on questions related to syntax (i.e., rules and/or patterns governing word order in a language). A long-standing debate centered on whether non-syntactic factors could influence processing prior to commitment to a single syntactic parse (see reviews in Kutas et al., 2011; Ferreira & Lowder, 2016). Over time, views have shifted, with an abundance of evidence that language processing is probabilistic, such that multiple parses can be considered concurrently (e.g., Hale, 2001; Levy, 2008), and incremental, such that multiple (syntactic and non-syntactic) sources of information have rapid and cascading influences as people understand words in sentences in real time. Beyond purely linguistic knowledge, these information sources have been proposed to include pragmatic knowledge of how people, things, places, and so on behave in the world (e.g., Hagoort et al., 2004; Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008; Filik & Leuthold, 2013) along with knowledge about common organizational structures like events (e.g., Matsuki et al., 2011; Metusalem et al., 2012); semantic and pragmatic information gleaned from local linguistic contexts (e.g., Kutas & Hillyard, 1980; DeLong, Urbach, & Kutas, 2005; Nieuwland & Van Berkum, 2006; Otten & Van Berkum, 2007), and information from visual and/or physical situations (e.g., Kamide, Altmann, & Haywood, 2003; Knoeferle, Carminati, Abashidze, & Essig, 2011). On some accounts, these information sources (often termed “constraints”) can be weighted probabilistically. Such constraint-based accounts propose that more informative constraints are preferentially used to guide processing as soon as they are made available (reviewed in McRae & Matsuki, 2013). Going beyond incremental processing, many accounts now argue for predictive processing and/or pre-activation of semantic information (prior to its being encountered), such that the activation of meaningful (though as-yet-unmentioned), relevant information is a normal part of sentence comprehension (see Federmeier, 2007; DeLong, Troyer, & Kutas, 2014; Kuperberg & Jaeger, 2016, for reviews). However, while many accounts of incremental and predictive language comprehension make reference to the use of stored knowledge about words and the world, to our knowledge, none of them appears to have seriously considered how variation in the functional organization of this knowledge— across or within individuals—might influence language processing, moment-by-moment.

In the field of computational linguistics, linguistic processing difficulty is often modeled on measures derived from large text corpora. For example, linguistic *surprisal* (Hale, 2001; Levy, 2008), a measure from information theory, is defined as the relative probability of encountering a word given some recent context (e.g., preceding words). Because this measure is based on linguistic information alone, it does not directly encode world knowledge. As one exception to this, a computational model by Venhuizen and colleagues developed an online surprisal metric which derives from a simplified measure of world knowledge in addition to linguistic experience (Venhuizen, Crocker, & Brouwer, 2019). In a microworld made up of a vocabulary of 21 words and a limited number of propositions, they independently manipulated strings of text (a proxy for linguistic experience) and the co-occurrence of propositions (a proxy for world knowledge). As the model encounters each word in a sentence, it uses both sources of information to navigate a situation-state space to arrive at a point representing the sentence meaning, such that “…intermediate points in state space are effectively accumulations of evidence about the sentence-final utterance meaning” (p. 240). Online surprisal can be viewed as the change from one point in state-space to the next, and is based on separate measures of both linguistic surprisal (based on the linguistic text input) and situation surprisal (based on the world knowledge). In its current instantiation, this model does not consider how variability in world knowledge (or linguistic experience) across individuals might influence processing. Indeed, no current processing model has taken on the challenge of modeling the computational implications of individual differences in world knowledge. However, such computational models could be extended to systematically vary the amount and/or structure of the linguistic and world knowledge input to examine systematic differences in processing as a function of variability in (linguistic and world) knowledge. The present study aims to empirically elucidate some of the processing differences that may occur as a function of such variability.

Event-related brain potentials and real-time sentence processing

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, a negative-going brain potential peaking between 250 and 500 ms with a centroparietal maximum over the scalp, is a good measure of word- and sentence-level meaning activation (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). Moreover, N400 amplitude to words is sensitive to the amount of preceding context, more generally, and a large N400 can be viewed as the “default” response to a meaningful stimulus like a word, with progressively greater reductions in N400 amplitudes to content words as context accrues (Van Petten & Kutas, 1990).

The N400 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 categorical (Federmeier & Kutas, 1999) or perceptual relationship (Amsel, DeLong, & Kutas, 2015; Rommers, Meyer, Praamstra, & Huettig, 2013), event-based relationship (Amsel et al., 2015; Metusalem et al., 2012; Paczynski & Kuperberg, 2012), or other semantic relationship (DeLong, Chan, & Kutas, 2018; Kutas & Hillyard, 1984). Such findings stemming from a so-called “related anomaly” paradigm have been taken as evidence that information cued by the related words is more readily accessible compared to information cued by unrelated and anomalous words.

Manipulations that influence N400 amplitude in sentence processing studies also often influence the amplitude of later-occurring, post-N400 positivities (PNPs), which have been variously referred to as late positive complexes (LPCs) or P600s. In early studies, the functional significance of the P600 was related to syntactic factors, being sensitive to ungrammatical and/or challenging syntactic structure (Kaan, Harris, Gibson, & Holcomb, 2000; Osterhout & Holcomb, 1992). However, modulation of late positivities often occurs in the absence of syntactic violations or difficulty and has been linked to a number of cognitive processes, including reanalysis of semantically incongruous and/or unexpected words in context (discussed in Brouwer, Fitz, & Hoeks, 2012; Van Petten & Luka, 2012). In a meta-analysis, Van Petten and Luka (2012) reviewed 45 studies which included sentences containing congruent vs. incongruent endings. Of these, the majority showed a biphasic response: an N400 effect (reduction in N400 amplitude for congruent vs. incongruent word) followed by a late positivity effect, with more positive-going potentials for incongruent compared to congruent words in a later time period. Van Petten and Luka also suggested a functional dissociation between frontally and parietally distributed

LPCs, with the former being sensitive to expectancy (amongst plausible words) and the latter reflecting differences between plausible/implausible words (see Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Thornhill & Van Petten, 2012; DeLong, Quante, & Kutas, 2014). There is currently no consensus regarding the functional significance of LPCs in language processing. One promising view put forth by Brouwer et al. (2012) is that they are perhaps best viewed as a rather broad "family [of potentials] that reflects 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 indexed by N400 potentials.

Several descriptive and computational accounts of N400 effects have been presented which construe them as reflecting aspects of semantic processing (reviewed in Kutas & Federmeier, 2011). In some, N400 effects reflect differences in the ease of *integration* of incoming semantic information (e.g., Brown & Hagoort, 1993); in others N400 effects reflect differences in the ease of *retrieval* of semantic information from long-term memory, influenced by context but not indexing integration per se (Kutas & Federmeier, 2000; also discussed in Lau, Phillips, & Poeppel, 2008).

Several different computational models of N400 amplitudes also have been put forward (e.g., Laszlo & Plaut, 2012; Laszlo & Armstrong, 2014; Cheyette & Plaut, 2017). On one line of computational models, N400 amplitude reflects implicit prediction error. As a word is being evaluated with respect to its sentence or discourse context, its N400 amplitude reflects the additional semantic information needed to make sense of it (Rabovsky, Hansen, & McClelland, 2018; see also Rabovsky & McRae, 2014). This approach can be contrasted with a model put forth by Brouwer and colleagues, in which N400s reflect retrieval of word-level information whereas late positive complexes reflect the integration of this information into the evolving representation of sentence meaning (Brouwer, Crocker, Venhuizen, & Hoeks, 2017). Both models are consistent with language processing as anticipatory in nature, but differ in the proposed timecourse for updating or integration of information in an existing mental model. These models thus echo debates from the descriptive theories of sentence processing reviewed above.

Despite the undeniable import of knowledge to descriptive and computational models or theories of language processing, none of the existing accounts are sufficiently fleshed out to support predictions about individual differences. The present study thus was not designed to adjudicate among these accounts, but to begin gathering data that any complete account of written language processing would need to explain.

N400 as a subtle probe of semantic processing in real time

As in many of the studies mentioned above, brain potentials can provide relatively unobtrusive probes of what semantic/perceptuo-semantic information is activated during real-time sentence processing. Within the appropriate experimental design, measures of N400 amplitude are well-suited to asking questions about the nature of information available to comprehenders at any given moment as they process written sentences in real time.

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. In a two-sentence context setting up an expectation for the word *pines*, another member (e.g., *palms*) of the same category (tree) elicited a smaller N400 amplitude compared to a word from a different category (e.g., *tulips*) but a larger N400 amplitude compared to the expected word. Moreover, the related anomaly effect (i.e., the difference between the N400 to related vs. unrelated anomaly) was larger for words completing high-constraint than low constraint sentences. This pattern of findings was used to argue that sentence contexts can pre-activate relatively specific semantic features of likely upcoming words and thereby facilitate the retrieval of words which share many semantic features (i.e., are members of the same category), even if they render the sentence anomalous.

Metusalem et al. (2012) likewise used a related anomaly paradigm to investigate the extent to which generalized event knowledge brought linguistically unlicensed, yet contextually relevant, information to mind as individuals read. Individuals read short paragraphs about common events (e.g., playing football) that set up linguistic expectations for a word like *touchdown*. Contextually unexpected, linguistically unlicensed words related to the event being described (e.g., *helmet*) elicited reduced N400 amplitudes compared to unlicensed, event-unrelated words (e.g., *license*). They proposed that people use their event knowledge (also referred to as *schemata* (van Dijk & Kintsch, 1983; Kintsch, 1988), *scripts* (Schank & Abelson, 1977), and *situation models* (Foss, 1982; Hess, Foss, & Carroll, 1995; Kintsch, 1988)) as they understand words in sentences in real time, perhaps (pre-)activating concepts that are related/likely to be relevant even if they are linguistically unlicensed.

In contrast to the Federmeier and Kutas study, in Metusalem et al. (2012), 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. However, 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 sites, suggesting that comprehenders seem to make use of both types of information in real time.

These results, among others, imply that real-time language processing draws on rich, structured representations in long-term memory, and that accessing these representations is part and parcel of the normal processing of words in sentences. Viewed with this lens, it is obvious that how much and how well an individual knows (and knows about) things must impact what information is available for that individual to bring to mind during real-time comprehension and, presumably, the nature of its organization in 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. Extant studies of individual differences have been focused in large part on general cognitive abilities (e.g., working memory and 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 may be that it is experimentally challenging to capture the specifics of an individual's world knowledge with standard laboratory procedures. A potential solution may be to experiment within a restricted domain of knowledge, as is often done in the literature on the psychology of expertise (Ericsson, Charness, Feltovich, & Hoffman, 2006).

We have taken this approach. In Troyer and Kutas (2018), we took advantage of a domain of knowledge with the requisite properties for online language processing studies, including a large, rich set of verbal descriptions, wherein otherwise similar college-aged young adults differed in their degree of knowledge—the fictional world of Harry Potter (HP) by J.K. Rowling. We 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 evident on N400 amplitudes, reflecting greater ease of retrieval for the supported words. Critically, however, participants' degree of HP knowledge influenced the size of this effect only for the HP sentences, with more knowledgeable individuals showing relatively larger effects and less knowledgeable individuals showing smaller effects. These results empirically demonstrate that the rapid influences of written sentence contexts, which modulate N400 brain potentials, are a function of each individual's knowledge.

In a follow-up study (Troyer et al., 2019), we examined the relationship between domain knowledge and real-time retrieval of contextually supported words, asking whether the pattern observed in Troyer and Kutas (2018) was strictly due to the proportion of items an individual knew (or not), or whether domain knowledge had an influence beyond increasing the likelihood an individual would know any given HP "fact." To that end, we conducted single-trial analyses in a paradigm in which participants reported whether they had known each fact, immediately after reading about it. All sentences were about HP and ended in a supported (or "correct") word. HP domain knowledge was a strong predictor of whether an individual trial was reported as known, but there was also 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 also was mediated by each item's offline cloze probability, which we argued was a proxy for the difficulty of reporting the final word of each sentence pair across a group of individuals. In this analysis, domain knowledge had its greatest influences on trials for "low-cloze" items (those with an ending that few participants were able to provide in an offline cloze norming task—i.e., challenging trials) which participants reported knowing, as well as on "high-cloze" (i.e., easier) trials that individuals reported as *not* knowing. These instances likely reflect moments in which complete retrieval of a specific word might be difficult, although retrieval of some word information is nevertheless possible. Together, these findings suggest that domain knowledge may be especially impactful under difficult retrieval conditions, influencing real-time retrieval beyond strictly mediating the likelihood that a fact is known to the comprehender. In Troyer et al., we hypothesized that domain knowledge might ease (implicit) retrieval of (perceptual, semantic, and/or conceptual) information relevant for word understanding by virtue of its organization, which presumably differs across individuals with varying degrees of knowledge of that domain.

This interpretation seems reasonable given the vast literature showing that people rapidly make use of a variety of word and world knowledge as they process words in real time, including information about orthographic neighborhood density (Laszlo & Federmeier, 2009), word frequency (Van Petten & Kutas, 1990), and non-linguistic knowledge such as 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 might impact the availability, contents, and timecourse of bringing these sources of knowledge to mind in real time. Indeed, the literature on expert knowledge suggests 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). Accordingly, we 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, such as categorical and event relationships.

The current study

We hypothesized that domain knowledge systematically influences 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, knowledgeable individuals may access more perceptual and/or conceptual semantic features for incoming words and/or more concepts relevant to processing the sentence's meaning. To probe such information, we tested individuals who varied in their knowledge of the narrative world of Harry Potter using a related anomaly paradigm with 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 or 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 relationships. For half of the sentence materials, the related words were taken from the same category as the linguistically expected word, as in Federmeier and Kutas (1999). For the other half, the related words were 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 versus unsupported words in sentences about general topics, they all 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 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 the literature. Though our primary ERP measure of interest was amplitude during the N400 time window, we also examined a later, post-N400 interval potential sometimes observed to vary with semantic and/or contextual manipulations (reviewed in Van Petten & Luka, 2012). Given that such late positivities have been reported to be sensitive to word plausibility (DeLong, Quante, et al., 2014) as well as to task manipulations (reviewed in Brouwer & Crocker, 2017), we thought it was likely to vary with domain knowledge. For example, domain knowledge might influence the likelihood that individuals notice anomalies and how they process "implausible" or incorrect HP sentence endings.

In a set of exploratory analyses, we also examined whether the ability to (pre-) activate contextually relevant (though linguistically unlicensed) word information (as in the related anomaly condition) would rely on actually knowing the (correct) contextually supported word completing the HP sentence pair. To measure this knowledge, we conducted a modified cloze task after the ERP experiment, in which participants re-read the same 156 sentence pairs minus the sentence final words. Participants were asked to provide the word which they thought best completed each sentence fragment—which was *not* necessarily the word which they had read during the EEG study. If individuals did not know what word to provide, they were instructed to render a judgment about whether or not they had seen the correct word during the EEG recording portion of the study. This information was used 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). Because participants had recently read these same sentences, sometimes with the correct but more often with an incorrect word in the sentence-final position, we expected that they would demonstrate greater overall accuracy for final words of sentences seen in the Supported condition. 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 categorically- or event-related (albeit contextually unsupported) word information during sentence processing depends on knowing the appropriate word, then we would expect to see 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 to see no or little difference in N400 amplitude to Related trials as a function of trial correctness on the post-ERP-experiment task.

Method

Participants

53 UCSD students (mean age = 20, range = 18–25; 39 women, 14 men) took part in the EEG 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.

Materials

Sentence material construction

During the EEG portion of the experiment, participants read sentence pairs in two experiments, which were analyzed separately (see Appendix A for all sentence materials). We included a short control experiment to verify that these participants 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).

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 and Kutas (2018). On average, the initial sentence was 6 words long (range = 3–13 words), and the second sentence was 8 words long (range = 5–11 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 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).

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 = 112) were taken directly from the HP sentences used in Troyer and Kutas (2018). On average, the initial sentence was 10 words long (range = 4–18 words), and the second sentence was 7 words long (range = 3–16 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 117 participants, with between 35 and 43 participants responding to each sentence pair by providing a final word.[2] 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 "Behavioral tasks and measures"). Performance on this 10-question quiz ranged from 1 to 10, with a median score of 5. 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 1) or by association to the overall event/episode being described (N = 78; examples 10–12 in Table 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.

Evaluating the Harry Potter sentence materials

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.

Distributed semantics model. We trained a word2vec model[3] (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[4]. The corpus consisted of a total of 1,125,854 words, with a vocabulary size of 8046 words (subject to the constraint that each word appear at least 5 times in the HP books[5]). 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 then used these embeddings to quantify relative similarities/differences between word pairs (or average vectors computed over sequences of words). We used the continuous bag-of-words (CBOW) architecture that 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.

Using this model, we extracted word embeddings for critical words from each of our experimental conditions (Supported, Related, and

[2] Due to experimenter error, cloze probability measures were not collected for two items.

[3] We used the distribution by D. Yaginuma, https://github.com/dav/word2vec.

[4] Available at https://usd.shop.pottermore.com.

[5] 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).

Table 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

Unrelated). For each item (156 total), we computed the cosine similarity (angular distance) between the word embeddings for Supported-Related and Supported-Unrelated pairs of critical words, respectively. As expected, word embeddings for Supported words were more similar to Related words (*Mean* = 0.370, *CI* = [0.329, 0.409]) than to Unrelated words (*Mean* = 0.169, CI = [0.132, 0.206]; Fig. 1a), t (1 4 8) = 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 for 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 *(Mean* = 0.266, *CI* = [0.239, 0.293]) words, followed by

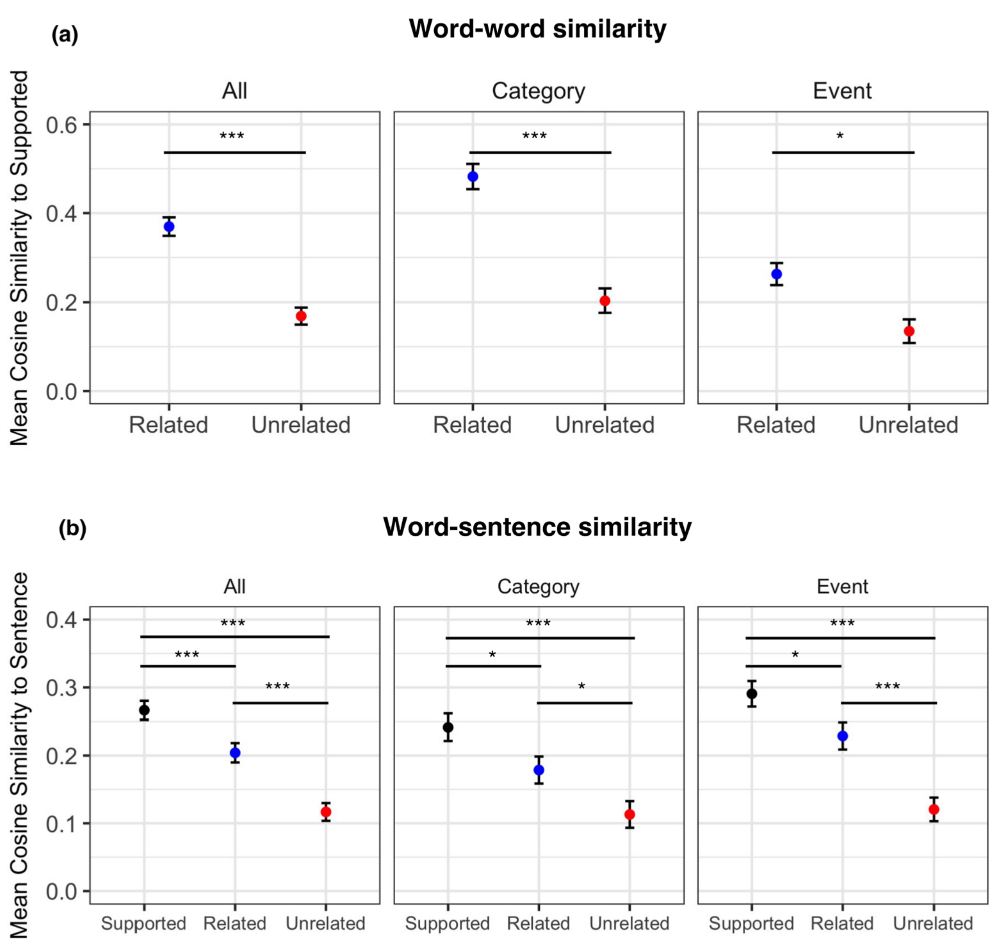

Fig. 1. (a) Across items and within each subset (Category-related, Event-related), mean cosine similarity for Supported & Related endings is greater than for Supported & Unrelated endings. (b) For all 156 items (and within each subset) there was a significant three-way difference between cosine similarity of averaged word embeddings for sentences and Supported < Related < Unrelated endings. *** *p < .0001;* ** *p < .001;* * *p < .05*; error bars represent standard error of the mean.

Related (*Mean* = 0.204, *CI* = [0.176, 0.232]) and Unrelated (*Mean* = 0.117, *CI* = [0.091, 0.142]) words (a three-way difference, all ps < .0001). This pattern was similar for category- and event-related item subsets (Fig. 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.

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").

As expected, mean similarity ratings for Supported-Related word pairs (*Mean* = 3.689, *CI* = [3.321, 4.056]) were greater than those for Supported-Unrelated word pairs (Mean = 2.001, CI = [1.735, 2.266]), $t(23)$ = 11.19, p < .0001 (Fig. 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.

In addition, mean relatedness ratings for Supported-Related word pairs (*Mean* = 4.548, *CI* = [4.264, 4.833]) were greater than those for Supported-Unrelated word pairs (*Mean* = 2.432, *CI* = [1.993, 2.871]), $t(22)$ = 10.67, p < .0001 (Fig. 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.

Behavioral tasks and measures

Our primary behavioral measure of interest was performance on a trivia quiz about HP, which served as our measure of HP knowledge

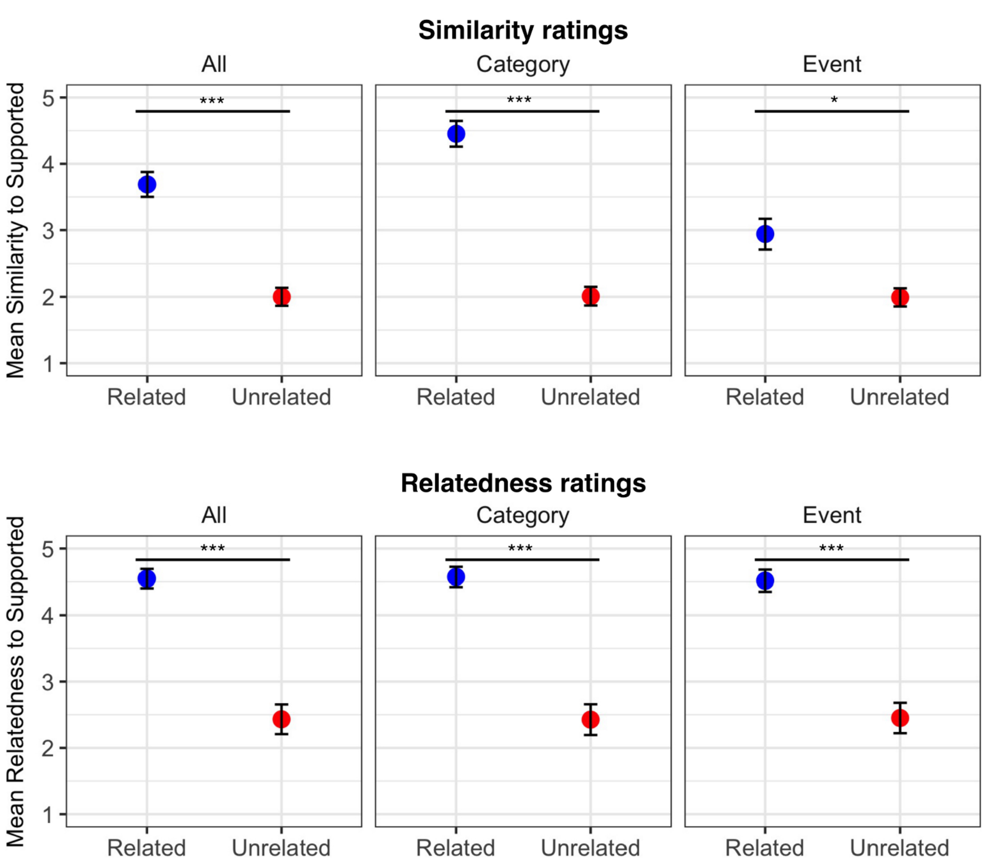

Fig. 2. (a) Across items and within each subset (Category-related, Event-related), similarity ratings for Supported & Related endings are greater than Supported & Unrelated; this effect was larger for the category-related subset of items compared to the event-related subset. (b) Across items and within each subset, relatedness ratings for Supported & Related endings are greater than Supported & Unrelated. *** p < .0001; ** p < .001; * p < .05; error bars represent standard error of the mean.

(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.

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.

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.

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.

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 1–100, 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 recording portion of study.

After completing all of the additional tasks, participants responded to a short questionnaire asking what, if anything, had caught their attention during the study, what they thought 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.

Procedures

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.

EEG experiment

Before the EEG recording 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.

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 0.01–100 Hz; the sampling rate was 250 Hz.

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 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. Because sentence-level factors modulating N400 amplitudes also often influence later, post-N400 potentials, we also examined a 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 study (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 incorporated 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.

Modified "cloze" production task

Following the EEG portion of the study, 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 not. They were asked to fill in the blank with the final word that they believed was consistent with the HP stories—that is, the word that belonged. If they were unsure which word to provide, they were asked to provide their best guess. If they could not provide a word for any given sentence pair, they were asked to type "C" if they thought the original sentence-final word they had read had been consistent with the HP stories, or "I" if they thought it had been 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).

Results

Behavioral data

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%); opted to provide a response of "Correct" or "Incorrect" on 36% of trials (see below); 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 (*Mean* = 60%, *CI* = [53%, 67%]) compared to the Related (*Mean* = 43%, *CI* = [35%, 51%]) or Unrelated (*Mean* = 43%, *CI* = [35%, 51%]) 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 Fig. 3.

We also considered the trials on which participants indicated that they did not know and could not guess the correct completion by indicating whether the word that they initially read (in the ERP study)

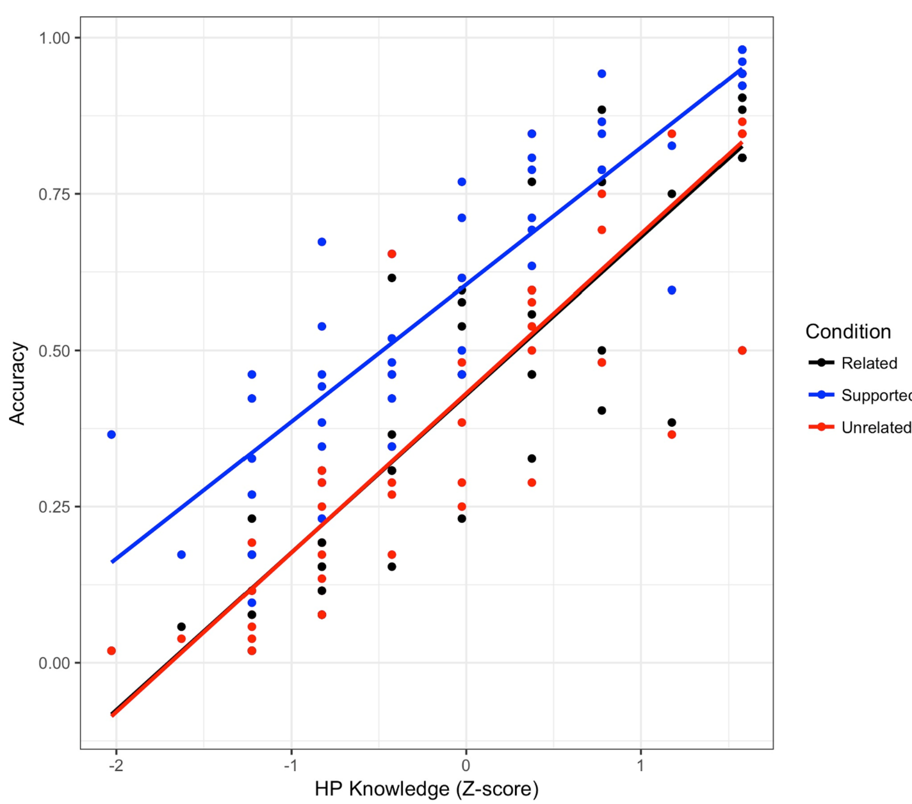

Fig. 3. Accuracy on the post-experiment cloze task is plotted against HP knowledge (Z-scored) by condition. This relationship was slightly stronger for Related and Unrelated items compared to Supported items.

Table 2
Mean, standard deviation, and range are provided for behavioural measures of individual differences for all participants and by HP knowledge subgroup based on a median split on the HP trivia quiz. Measures on which the subgroups differed significantly are indicated as follows: *** $p < .0001$; ** $p < .001$; * $p < .05$.

	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]

ending that context 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.

Additional tasks

Table 2 reports descriptive statistics for scores on the HP knowledge quiz and other individual difference measures. Intercorrelations among measures are provided in Table 3.

Subject-averaged ERP data

Fig. 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 Figs. 5–6.

N400: 250–500 ms post-stimulus

Whole-head analyses. Results from the whole-head ANOVA for control and HP sentences in the N400 time window are provided in Table 4.

For control sentences, as expected, there was a main effect of ending type on ERPs, with more positive-going waves (i.e., reduced negativities) for Supported (*Mean* = 2.96 μv, *CI* = [2.809, 3.120]) compared to Unsupported (*Mean* = 1.386 μv, *CI* = [1.232, 1.540]) 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 3
Intercorrelations (Pearson's r) among behavioral measures of individual differences. r values above .29 are significant at $\alpha = .05$; r values above .37 are significant at $\alpha = .01$.

	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

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 (*Mean* = 3.353 μv, *CI* = [3.205, 3.501]) condition was larger than the Related (*Mean* = 1.423 μv, *CI* = [1.294, 1.552]) condition ($F(1, 47) = 7.87$, p = .007), which was in turn larger than for the Supported (*Mean* = .0891 μv, *CI* = [0.754, 1.029]) 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.

ROI analyses of individual differences. Fig. 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 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).

Fig. 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 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.) (Fig. 7d).

Visual inspection of category- and event-related subsets of data suggested that the main pattern of N400 results (Supported < Related < Unrelated) was similar for each (Fig. 8). To verify this and examine whether HP domain knowledge might have a larger influence on related anomaly effects in either subset of the materials (i.e., the items with category- or event-related anomalies), we conducted analyses incorporating ending type (Supported, Related, Unrelated), the type of related anomaly (category-related, event-related) and HP domain. This analysis confirmed that there was no three-way interaction ($\chi^2(2) = 0.72$, n.s.). In addition, we confirmed that there was no two-way interaction between ending type and related anomaly type across all participants ($\chi^2(2) = 1.81$, n.s.). This suggests that the three-way effect of ending type was similar for both category- and event-related anomaly subsets of our materials.

Finally, to rule out the possibility that other existing individual

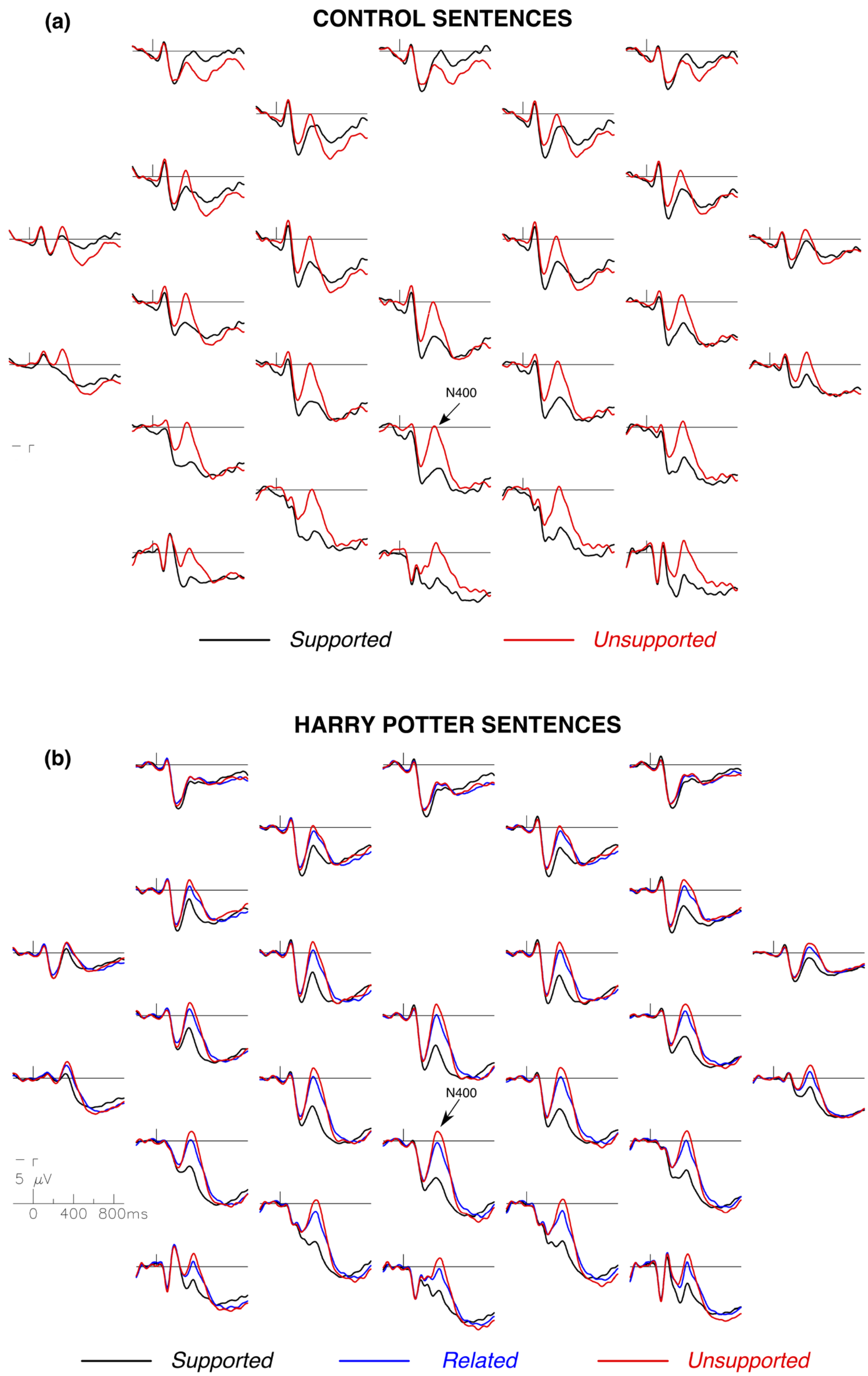

Fig. 4. Grand average ERPs across all participants for critical words of each type for (a) control and (b) HP sentences.

differences 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, verbal working memory (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 nested 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), suggesting that the interaction between HP domain knowledge and N400 potentials was not driven by other individual differences that we measured.

Late positivity: 500–750 ms post-stimulus

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 6.

For control sentences, in the late positivity time window, a whole-

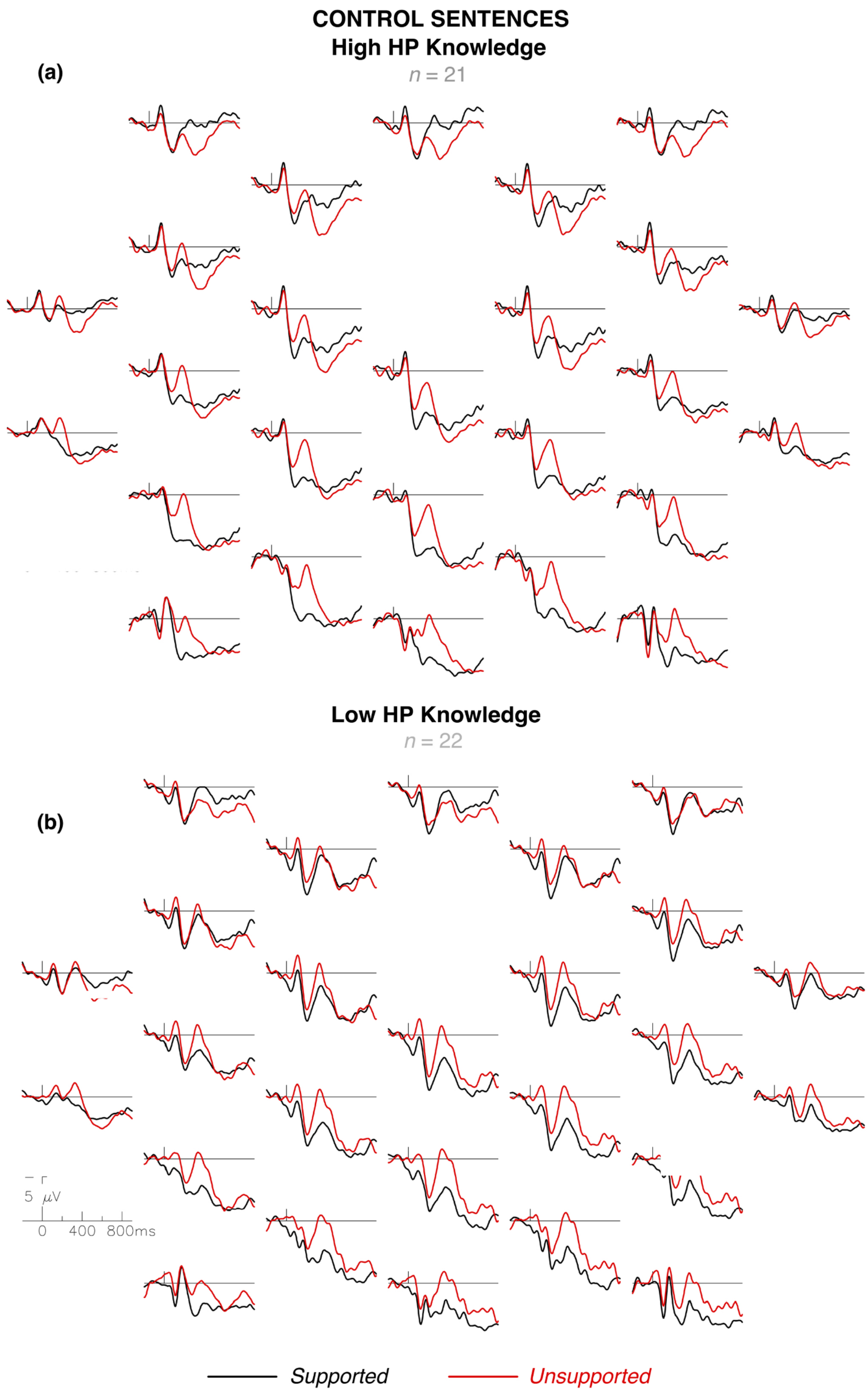

Fig. 5. Grand average ERPs for (a) High and (b) Low HP knowledge subgroups (based on a median split) to critical words in the control sentences.

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, et al., 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.

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 7), which seemed to be driven by the fact that individuals with greater 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 ($r = .06$, n.s.), words.

For HP sentences, in the late positivity time window, there was a significant interaction between HP domain knowledge and ending type

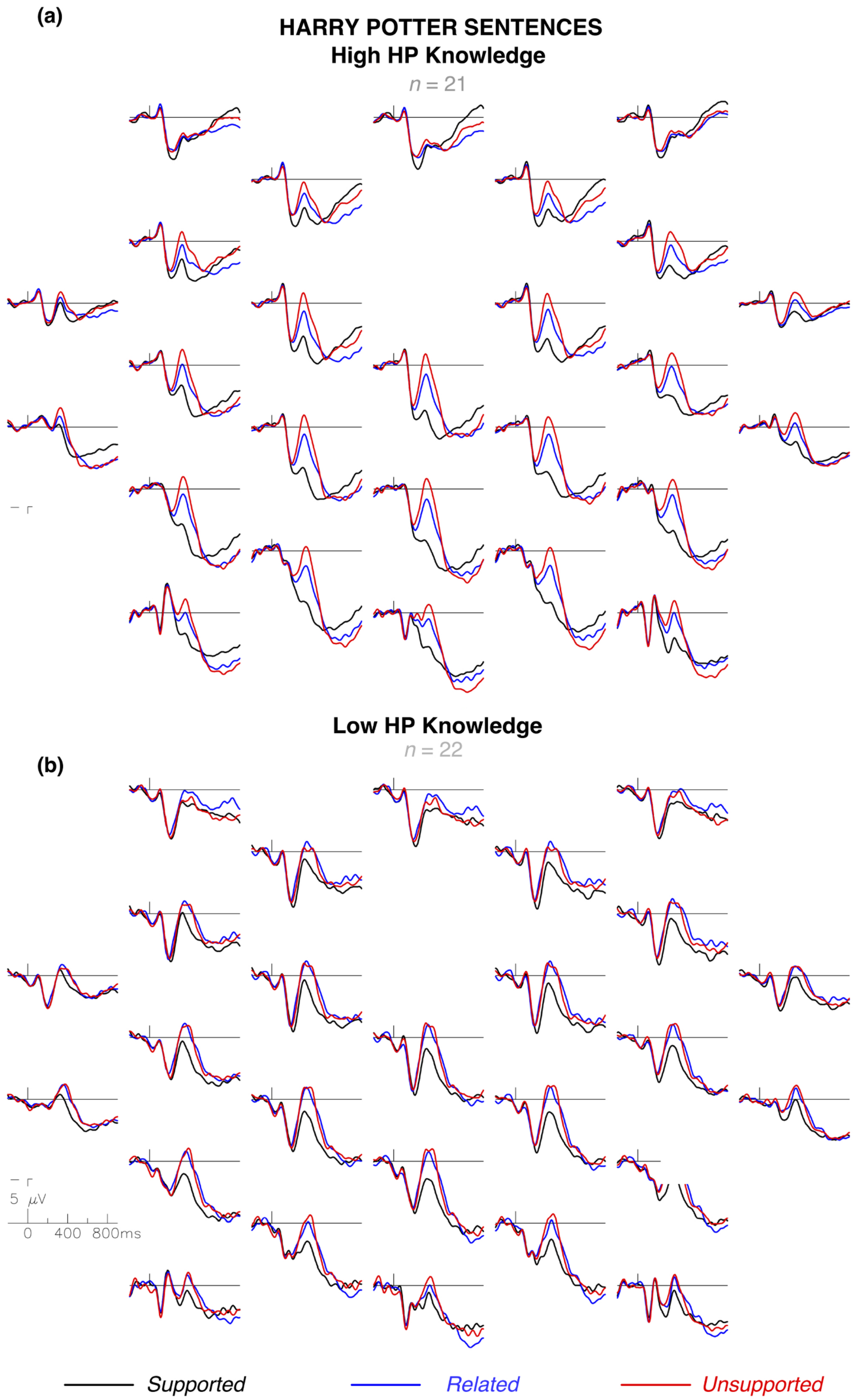

Fig. 6. Grand average ERPs for (a) High and (b) Low HP knowledge subgroups (based on a median split) to critical words in the HP sentences.

($\chi^2(2)$ = 15.79, p < .001; full model provided in Table 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, n.s.).

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 positive 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.

In the late positivity time window, there was no three-way interaction of related anomaly type (i.e., category- vs. event-related anomalies), ending type, and HP domain knowledge ($\chi^2(2)$ = 1.00, n.s.). However, across all participants, there was a two-way interaction

Table 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

between ending type and related anomaly type ($\chi^2(2)$ = 9.16, p = .01). Follow-up tests suggested this interaction resulted from modulation of late positivities by ending type for the category-related subset of items ($\chi^2(2)$ = 11.40, p = .003), with greater positivities for the Unrelated ($\chi^2(1)$ = 7.69, p = .006) and Supported ($\chi^2(1)$ = 9.70, p = .002) conditions relative to the Related condition, but no statistical differences in ending type for the event-related subset of items ($\chi^2(2)$ = 0.63, n.s.).

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 nested model that 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.).

Table 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

Single-trial ERP data

Based on participants' responses during the post-ERP modified "cloze" task, we sorted trials according to whether or not participants

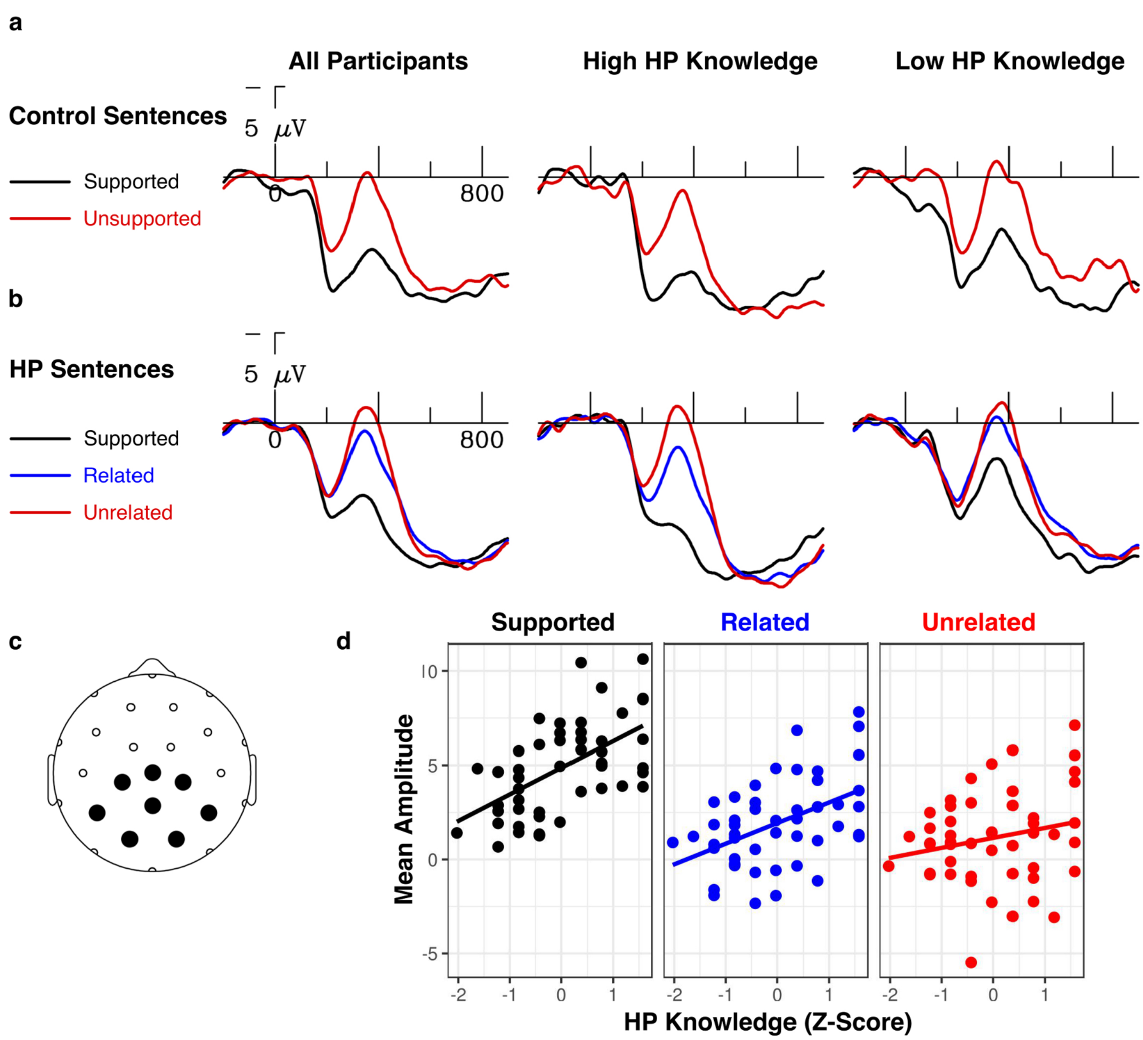

Fig. 7. ERPs are plotted to critical words from control (a) and HP (b) sentences for an ROI based on an average of 8 centro-parietal scalp electrodes (c). (d) N400 amplitude is plotted against HP knowledge (z-scored) by condition.

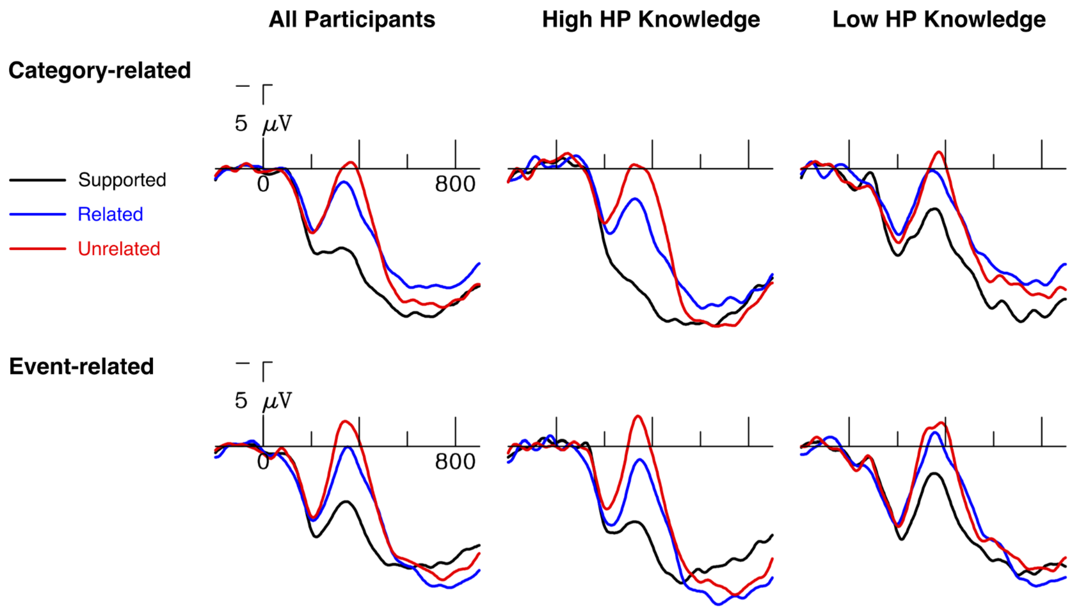

Fig. 8. ERPs are plotted to critical words from the two subsets of the HP sentences, those containing category-related anomalies (top) and those containing event-related anomalies (bottom) for an ROI based on an average of 8 centro-parietal scalp electrodes.

Table 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	0.112
Ending Type	(2, 94)	0.361	.6978	
Electrode:Ending Type	(50, 2350)	2.728	.0000	0.890

Table 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

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., 2019). We were most interested in whether such a reduction would also be present on the N400 for the Related endings. We also analyzed the late positivity time window.

N400: 250–500 ms post-stimulus

We first asked whether accuracy on the post-experiment "cloze" task had a similar influence across conditions (Fig. 9). 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.).

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 (Fig. 10).

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.).

Discussion

Summary of findings

The current study is the third in a series examining the influence of domain knowledge on the real-time processing of words in written sentence contexts. Troyer and Kutas (2018) found that domain knowledge of the narrative world of Harry Potter had a rapid and specific influence on processing words in sentences. ERP effects of contextual support were modulated by individuals' HP domain knowledge for sentences about Harry Potter, but not for sentences about general topics. These effects were driven by domain knowledge modulating N400 amplitudes to the contextually supported (but not unsupported) final words of HP sentences describing fictional "facts." Troyer et al. (2019) replicated the latter finding, and further showed that HP domain knowledge influences on N400 amplitude to contextually supported words were not strictly a result of the total proportion of HP facts that

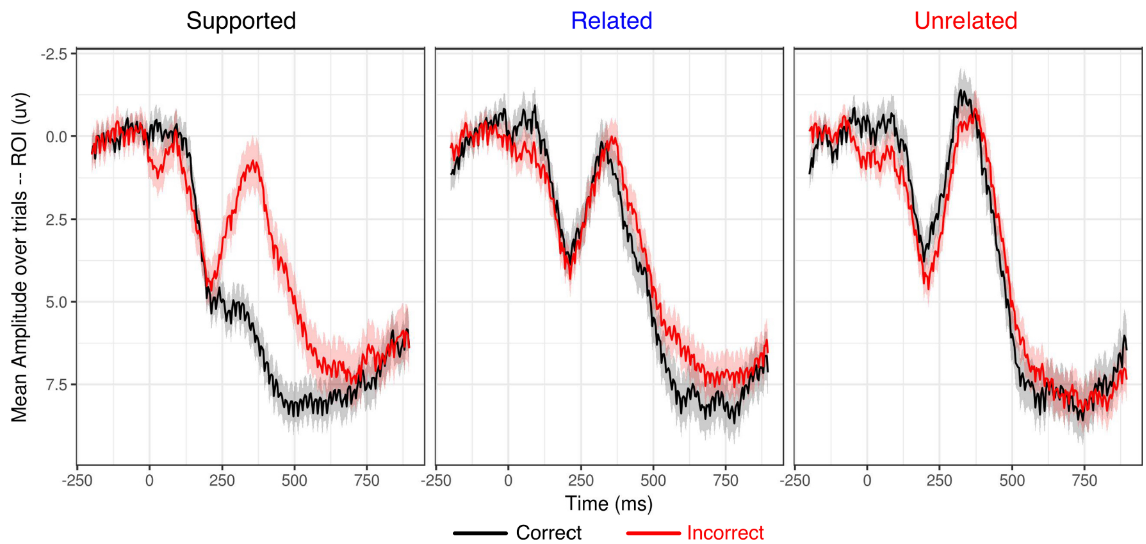

Fig. 9. ERPs averaged over single trials in the centroparietal ROI are plotted by condition and by accuracy on the post-experiment cloze task.

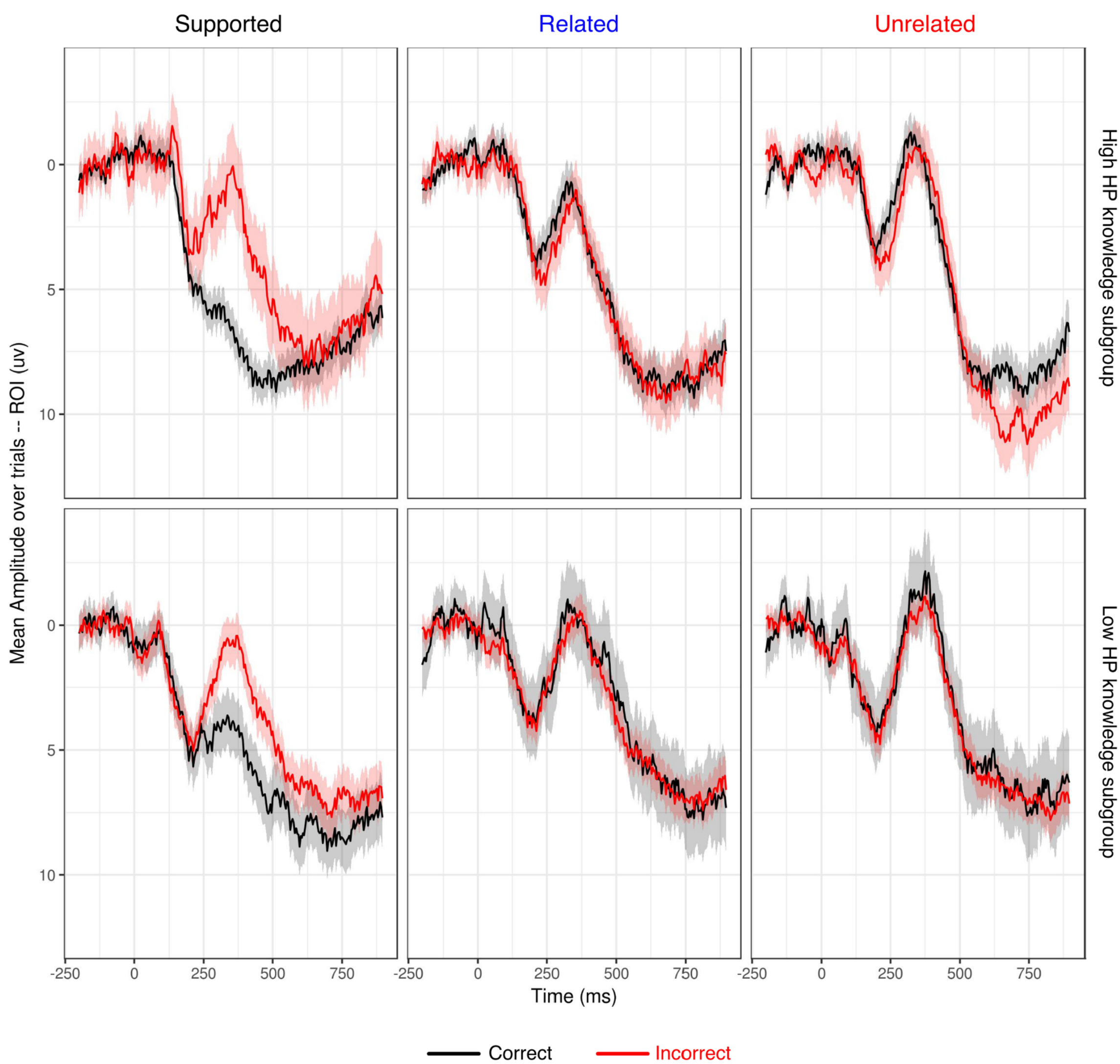

Fig. 10. ERPs averaged over single trials in the centroparietal ROI are plotted by condition and by accuracy on the post-experiment cloze task and by HP knowledge subgroup (based on a median split).

individuals knew. For individuals with greater HP knowledge, N400 amplitudes were also reduced (compared to individuals with less HP knowledge) for contextually supported words that they reported not knowing. We took this pattern of results to reflect implicit and/or partial access to meaningful information in the absence of explicit awareness of that knowledge for individuals with greater HP domain knowledge. We speculated that this might be a product of a coherent, schema-based, and multi-faceted functional organization of relevant HP information in (more) knowledgeable individuals.

In the present study, we followed up on this conjecture more systematically. We were interested in whether, and, if so, when and under what conditions, individuals' degree of domain knowledge influences the functional organization and use of that information in real time. To that end, we constructed sentence materials containing related anomalies, designed to probe individuals' access to contextually unsupported information that was nonetheless meaningfully related to the sentence context and/or to a contextually supported (albeit not presented) continuation. As expected, we observed a three-way difference in N400 amplitude, with the greatest N400 reduction observed for Supported endings; the largest N400 to Unrelated endings; and an intermediate N400 for Related endings. We also replicated our finding in Troyer and Kutas (2018) that domain knowledge modulated the influence of contextual support. In addition, for the first time, we found that individuals' ERPs were sensitive to the related anomaly manipulation as a function of their degree of HP knowledge.

More specifically, HP domain knowledge influenced the ERP to related anomalies in both the N400 and post-N400, late positivity time windows, albeit with a different pattern in each. As expected, greater HP domain knowledge was associated with a greater reduction in N400 amplitude for the best completion (i.e., Supported endings) compared to Unrelated sentence final words as well as to Related sentence endings, although to a somewhat lesser degree. In the late positivity (post-N400) time window, HP domain knowledge did not have an influence on brain potentials to the best completion; rather, greater domain knowledge was more strongly associated with more positive-going potentials for inappropriate continuations (i.e., sentence-final words in both the Related and Unrelated conditions).

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, there was no such relationship for Unrelated words. However, perhaps surprisingly, there also was no relationship between trial-level knowledge of facts and N400 amplitude for Related words. We consider this finding in more detail below.

Contextual support effects

The current study replicates Troyer and Kutas (2018) to show 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 written sentences about that world. N400s to words supported by the sentence context were reduced as a function of individuals' HP domain knowledge whereas those to unsupported (and unrelated) words were not. These findings empirically demonstrate that knowledge at the level of the individual participant is critical for bringing to mind relevant semantic/conceptual information from sentence contexts, and that this knowledge is quickly used to process contextually supported words in real time. This pattern of results strongly suggests that information cued by the sentential contexts was available to facilitate (pre-)activation/processing of the appropriate, contextually-supported word.

For the HP sentences, we also observed an interaction effect of HP knowledge and ending type on amplitude in the post-N400 late positivity time period, such that individuals with greater HP knowledge showed greater positivities (relative to those with less HP knowledge) for Unrelated and Related words, but not for the Supported words. In a previous ERP sentence 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 in ERP patterns across the two experiments may be attributable to differences in their design and/or task demands: in the current study, for example, 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, et al., 2014; Van Petten & Luka, 2012). 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 construals of various 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 processing during this time period. We consider this point in more detail in Section 'Domain knowledge as a window onto the temporal dynamics of word processing'.

Related anomaly effects

The critical condition in our HP sentences was the related anomaly, related via category to a contextually supported (i.e., correct) word or via relationship to the event/episode described by the sentence context. Across a number of studies, related anomalies have been shown to elicit reduced N400 amplitudes compared to unrelated/unsupported words, though larger than for words which are highly probable, best completions/contextually supported (Amsel et al., 2015; DeLong et al., 2018; Federmeier & Kutas, 1999; Metusalem et al., 2012). Similarly, we observed related anomaly effects for sentences about the narrative world of Harry Potter. Consistent with the literature, related anomaly effects were similar in magnitude, timing, and distribution over the scalp, whether the related anomaly was from the same category as the supported word or was related via event/episode to the sentential content.

In addition, for the first time, we showed that the amplitudes of the N400 potentials to the related anomalies were a function of individuals' degree of knowledge about the (fictional) domain. We hypothesize that this stems from differences in the functional organization of individuals' domain knowledge arising from one or more of the following: (1) Experts (compared to novices) possess more and/or different "chunks" (related pieces) of information; (2) these "chunks" are more readily activated and brought to mind for knowledgeable individuals; and (3) the "chunks" are differentially organized such that more and/or different links between the "chunks" serve to guide information retrieval and processing for those who are more (vs. less) knowledgeable. For example, 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 other words, individuals with greater knowledge may implicitly take advantage of organizing principles unavailable to those with less knowledge to access relevant information and to guide their language comprehension. Our behavioral norming results support this notion: individuals with greater HP knowledge are more sensitive to manipulations of category and event relatedness between critical words in our experimental sentence materials.

It is well-attested 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 et al., 1997). Similarly, in physics, student experts seem to access deeper or higher-level organizational principles (e.g., problems dealing with the conservation of energy) when categorizing physics problems whereas novices access more shallow, surface, or literal features of the problems (e.g., problems dealing with inclined planes; Chi, Feltovich, & Glaser, 1981). Chi and 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.

Although few studies have investigated the influence of such expertise on real-time language processing, there are a significant number of reports detailing the influence of domain knowledge on offline text comprehension and memory measures, across different domains (music: Arbuckle, Vanderleck, Harsany, & Lapidus, 1990; cooking: Soederberg Miller, 2001, 2009; sports: Spilich, Vesonder, Chiesi, & Voss, 1979; Walker, 1988; Schneider, Körkel, & Weinert, 1989; McNamara & McDaniel, 2004; biology: Long, Prat, Johns, Morris, & Jonathan, 2008; fictional knowledge: Long & Prat, 2002). A general view of these findings is that knowledgeable individuals benefit from pre-existing event knowledge (reviewed in Elman & McRae, 2019), such as schemas (Kintsch, 1988; van Dijk & Kintsch, 1983) or scripts (Schank & Abelson, 1977), which serve to interpret text. That is, depending on the content of the text, readers can use their existing knowledge of related entities, the relationships between entities, typical sequences of actions that may occur, and other event knowledge to make sense of incoming text. Indeed, comprehenders are assumed to routinely use their knowledge of common events in everyday language processing (e.g., Metusalem et al., 2012), at times generating predictive inferences based on the combination of their pre-existing knowledge and the incoming information (reviewed in Kuperberg, 2016). By investigating individuals with varying degrees of knowledge about a fictional world containing a rich array of scenarios and events, we were able to empirically demonstrate the activation of domain knowledge in the form of ad hoc categories and event schemas in real time during word by word sentence reading, to the extent that the individual had the requisite knowledge to interpret sentences about the HP world.

For example, given the context, '*Professor McGonagall recruits Harry for the Gryffindor Quidditch team. She saw him save Neville's*…,' the contextually supported (i.e., correct) word is '*remembrall*.' Even if individuals with relatively greater versus less knowledge all knew this word, we suggest that as highly-knowledgeable individuals processed this sentence, they also were likely to (implicitly) retrieve more HP domain knowledge than this word, such as relevant events, and/or make inferences based on their extensive knowledge of the HP stories. This information retrieval could happen anywhere during the course of the sentence or be triggered by the related anomaly. *Quidditch*, for example, is a sport played at the wizarding school of Hogwarts, where players fly around on broomsticks. When (if not before) processing the critical word (Supported = '*remembrall*'; Related = '*broomstick*'; Unrelated = '*dog*'), individuals with a higher degree of HP knowledge (compared to those with less knowledge) may have activated '*broomstick*', or done so to a greater degree, as a part of their event knowledge. Something akin to this is what is hypothesized to be indexed by the reduced N400 amplitude to the event-related anomaly.

A similar line of reasoning could account for the reduced N400 to categorically-related anomalies. '*Sybill Trelawney is a Hogwarts professor. She teaches*…,' for example, is appropriately completed by the contextually supported word '*Divination*.' As both *Divination* and *Transfiguration* (the Related completion in our study) are school classes, taught by professors at the wizards' school (Hogwarts), attended by young wizards, involving magic, they are categorically related. To the extent that a comprehender has this knowledge (explicitly or implicitly) we would expect ERPs sensitive to domain knowledge to be modulated.

Across both types of sentence materials (containing category and event-related anomalies), we assume that 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 sentence frame. That is, to the extent that it is available (which is dependent on each comprehender's knowledge), we assume that cuing such information is a natural and integral part of understanding sentences. A more careful examination of such item-level information alongside individual differences in participants' domain knowledge could shed light on the precise nature of such "chunks" of information.

One implication of our findings is that the same string of words in a sentence can lead to a different pattern of semantic processing as a function of variability in individual-level-knowledge. This lens on semantic processing fits well with some aspects of "good-enough" accounts of language comprehension (Ferreira & Patson, 2007; Ferreira, Bailey, & Ferraro, 2002), in which "…language processing is sometimes only partial and… semantic representations are often incomplete" (Ferreira et al., 2002, p. 11). On such theories, comprehenders do not always process linguistic information to achieve completely accurate or detailed syntactic and/or semantic representations; rather, they use processing strategies that lead to representations that are "good enough" for the goal at hand. Proposals of good-enough language processing have primarily been applied to explain processing of temporary *syntactic* ambiguities, where multiple different meaningful parses of sentences seem to linger simultaneously, even in the face of disambiguating information (e.g., Christianson, Hollingworth, Halliwell, & Ferreira, 2001; reviewed in Ferreira & Lowder, 2016). However, these theories also seem to suggest that *semantic* representations during sentence processing can be quite nuanced, with variance in the nature of processing of lexico-semantic content.

A recent expansion of good-enough processing accounts claims that information structure is a determinant of whether sentence processing is merely "good-enough" or complete. Specifically, Ferreira and Lowder (2016) maintain that when information is given (i.e., can be linked to the prior discourse) individuals are likely to engage in superficial processing of content that is "good-enough" whereas when information is considered new (i.e., adds information to the discourse) then processing is likely to be based on more detailed, predictive, and integrative processes - that is, to be complete. Viewed from this perspective, variability in knowledge/expertise seems likely to have a major impact on what is considered "good-enough" for any given comprehender. Perhaps more (vs. less) knowledgeable individuals—who by virtue of their knowledge have at their disposal more cues or routes to relevant information—may not need to be as careful, thorough, or complete in their processing. As individuals who have a great deal of background knowledge about a sentence's content process each word, they may have access to a "bigger picture" or more coherent knowledge (e.g., schemas) and, as discussed above, may rely on "chunks" of information which themselves are flexible—possibly leading to probabilistic activation of semantic features and/or other, related meaningful "chunks." Framed another way, by virtue of the rapid availability of higher-order knowledge structures like schema, high-knowledge individuals may be able to utilize compressed information structures to access and manipulate knowledge at the rapid rates necessary. As such, they may be able to encode word information that is at once (flexibly) linked with rich swaths of information, and yet is not processed thoroughly, or (in the terms of Ferreira and colleagues) is just "good enough."

On a loosely related note, Rommers and Federmeier (2018) also have promoted the idea that word processing may be more or less superficial (or deep) as a functional of sentential constraint. They maintain that words subject to high contextual constraint are highly predictable and are not processed as thoroughly as less predictable words. They found that N400 repetition effects were larger for unpredictable than predictable words, and that only unpredictable words elicited post-N400 positivities presumed to reflect episodic processing. They took their results as support for a "top-down verification account" of sentence processing (Van Berkum, 2010). On such an account, linguistic processing is subject to (general) neural predictive mechanisms which serve to (rapidly) evaluate "bottom-up" input with respect to extant expectations about what is likely to come next, such that "predictability would decrease the quality of word representations in memory" (Rommers & Federmeier, 2018, p. 17). All else being equal, it seems that individuals with relatively greater knowledge about a sentence's content would be more likely to use top-down predictive mechanisms, compared to less knowledgeable peers. Based on the account by Rommers and Federmeier, more knowledgeable individuals might therefore be less likely to process information episodically, extracting sentence gist rather than committing specific word information to memory. Though Rommers and Federmeier contrast less-thorough processing of "gist" (for predictable words) with deeper and/or more episodic processing (for unpredictable words), there is another perspective on which processing "gist" can be considered deeper (e.g., in a levels-of-processing framework; Craik & Lockhart, 1972) than processing surface-level word-form information (see Potter & Lombardi, 1990). The present results indeed suggest that information supporting the "gist" of a sentence—that is, content from the event- and category-related words inherent in the related anomaly stimulus materials—is readily available as a function of an individual's knowledge. Moreover, results from single-trial analyses based on a post-ERP-experiment cloze task (described in more detail in Section 'Knowledge of specific facts') suggest that the contextually-activated event- and category-related information was available to high-knowledge individuals even absent their explicit memory or knowledge of the contextually appropriate (supported) word.

The prediction-based top-down verification account and the good-enough processing account both imply that semantic processing during sentence comprehension is nuanced, with factors such as information structure and word predictability influencing the depth or completeness with which linguistic information is processed. Though neither mentions variation in individuals' knowledge, it is reasonable to assume that this too will critically shape the nature and depth of the swaths of information that become active as individuals attempt to comprehend words in real time.

Thus far, our use of terms like "retrieval," "access," and "activation" to refer to an individual's ability to bring relevant semantic content to mind during real-time sentence processing reflects a dominant metaphor of semantic memory as a library, where word and concept meanings can be looked up. On an alternate view, however, words serve not as cues to (relatively stable) chunks of meaning, but rather as operators over mental states (Elman, 2011; see also Onnis & Spivey, 2012). Venhuizen et al. (2019), for example, conceptualize real-time sentence processing as navigation through situation state-space, with each incoming word leading to an updating of the current mental model. In this framework, the functional organization of knowledge (i.e., the organization of semantic memory as it is used during processing) could refer to information that guides movement through a high-dimensional space of different aspects of meaningful information. Our results cannot distinguish between these conceptualizations but do demonstrate that variability in knowledge shapes the way words guide semantic processing, in real time. It may prove fruitful to conceptualize semantic memory and/or mental models as state spaces that can be navigated, with each individual's knowledge (semantic memory) serving to select and shape the relevant dimensions and trajectories.

In sum, our results suggest differences in the extent to which meaningful relations are available to individuals as a function of their degree of domain knowledge (relevant to sentence content). To our knowledge, no extant account of language comprehension incorporates knowledge-based individual differences as a determinant of the semantic contents of real-time language comprehension, but will need to in order to be complete.

Domain knowledge as a window onto the temporal dynamics of word processing

In our study, we found that domain knowledge had a different profile of influences on ERPs in different time windows post-word onset. As expected, in the N400 time window, domain knowledge was positively correlated with the brain's response to contextually supported words as well as words related to the sentence context via the category of the contextually appropriate word or the general event or episode being described. However, there was no reliable relationship between domain knowledge and words which were contextually unsupported and unrelated. These findings are consistent with accounts of N400 amplitude as reflecting rapid access to long-term knowledge/semantic memory (Kutas & Federmeier, 2000; Lau et al., 2008; reviewed in Kutas & Federmeier, 2011). By contrast, in the time window of the post N400 positivity, domain knowledge was positively correlated with the brain's response to the two anomalies, related and unrelated, but not to contextually supported words. We tentatively speculate that this pattern of results reflects knowledge-based individual differences in (a) the ability to detect the anomalies and/or (b) a set of "post-N400" semantic processes that individuals engage upon anomaly detection. We speculate that knowledge-based individual differences influence processes related to semantic retrieval (indexed by N400 contextual support and related anomaly effects) and also may influence cascading processes involved in making sense of a sentence (reflected in larger late positivities for anomalous compared to supported words).

Though our experimental design was not intended to adjudicate among the various theoretical and computational models of N400s and/or late positivities, our results seem most consistent with the Retrieval-Integration account proffered by Brouwer et al. (2012, 2017). On this account, N400 reflects retrieval from semantic memory, based on a comparison of the state of the semantic memory prior to and upon receipt of an incoming word, and late positivities reflect a family of semantic integration processes engaged as individuals attempt to construct the intended meanings of sentences. Our data suggest that domain knowledge first has an influence on contextually meaningful words (supported and related, during the N400 time period) and subsequently on contextually implausible words (related and unrelated anomalies, during the late positivity time window). This pattern parallels the Retrieval-Integration stages proposed by Brouwer and colleagues.

To our knowledge, however, the Retrieval-Integration theory does not make any explicit predictions about how individual differences in knowledge might influence each or any stage of processing. Presumably, semantic retrieval processes would be sensitive to individual differences in knowledge and to variation in the semantic relationship between an incoming word and the preceding sentential context; we find support for both. In addition, semantic integration processes also seem equally likely to be sensitive to individual differences in HP knowledge, as well as to variation in the fit between a word and the ongoing construction of sentence meaning; again, we have outlined support for both.

In any case, we are of the opinion that the Retrieval-Integration model, like any model of semantic processing during language comprehension, would be augmented by consideration of how variation in knowledge, and (putative) accompanying differences in the functional organization of semantic memory, might influence each aspect of semantic processing. Rabovsky et al. (2018), for example, describe a computational model in which the locus of semantic integration effects is during the N400 time period. We believe that modelling and empirically testing variation in (domain) knowledge may ultimately allow for a more careful delineation between such theories and, ultimately, a more precise account of at least written language comprehension.

Knowledge of specific facts

In our post-experiment, offline cloze task, the primary measure of interest was whether given time individuals could provide the appropriate completion for each sentence or not. During the EEG recording study, participants were presented with the appropriate final word on one third of the trials (i.e., the Supported condition). Unsurprisingly, individuals were thus more accurate (by about 17%) on items taken from the Supported compared to the Related and Unrelated conditions. This pattern was similar across individuals, with a slightly greater boost observed for participants with relatively less domain knowledge.

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: correctly-reported final words elicited noticeably reduced N400 amplitudes compared to incorrectly-reported words. By contrast, also as expected, N400 amplitudes to Unrelated words were unaffected by participants' post-ERP cloze task reports. However, perhaps surprisingly, N400 amplitudes to the Related words were similarly unaffected by participants' post-ERP cloze task reports. Moreover, this pattern of results—i.e., the influence of task accuracy on N400 potentials by condition—was not influenced by individuals' degree of HP knowledge. Combined with our finding that individuals' HP 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 influences implicit retrieval of word information in the absence of explicit or conscious knowledge of the "correct" (fictional) lexical item. That is, we take these findings as an indication that domain knowledge has an impact that goes *beyond* its ability to bring the contextually supported (correct) word to mind; it also shapes the information that is brought to mind during sentence reading in real time more generally. Based on the lack of modulation of related anomaly effects by trial-level/explicit knowledge of contextually supported words, we propose that at least in this study, related anomaly effects may have been driven by the (pre-)activation of (perhaps implicit) information cued by the sentential contexts, and not by (pre-)activation of the contextually supported word, *per se*.

There are, however, limitations to these inferences due to the task design, in which participants performed a cloze task after having read sentence pairs in different conditions. The difference in accuracy for Supported compared to Related/Unrelated trials poses some interpretative issues. We cannot tell if this difference is due solely to a boost for Supported trials based on participants having previously read the correct word during the ERP study, or whether, rather, individuals also experience an interference effect for the Related and Unrelated (i.e., the anomalous) trials based on having previously read an incorrect word. If there is interference, it is possible that for trials participants knew, but knew less well, they may have been less likely to produce the correct word. It is also possible that such trials contribute strongly to related anomaly effects. Splitting such trials between "Correct" and "Incorrect" bins based on our sorting may therefore have eliminated the ability to observe an influence of item-level knowledge on related anomaly effects. Future work using less contaminated measures of item-level knowledge may address this concern and offer a more conclusive test of whether specific lexical knowledge is necessary for related anomaly effects to occur.

In the current study, trial-level knowledge was based on an offline task; we had no direct measure of trial-level knowledge in the moment. This is in contrast to Troyer et al. (2019), in which individuals read contextually supported sentences and then immediately reported whether they had known the information (or not) before reading it. In both studies, HP domain knowledge had an influence on N400 amplitude to words in context beyond mediating the availability of specific knowledge at the trial level, though it seemed to play out in different ways in the two studies.

In Troyer et al., domain knowledge interacted with trial-level reports of knowledge, having its primary influence on trials which participants reported *not* knowing. In that article, we 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, by contrast, domain knowledge did not interact with the pattern of N400 amplitude results based on trial-level accuracy and ending type.

There are, however, 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, just a third of the sentences ended in a contextually supported word. In Troyer et al., participants provided a subjective report immediately after reading each sentence pair, though we had no trial-level measure of objective knowledge.[6] In contrast, in the current study, our trial-level measure of knowledge *was* objective—participants were cued with the sentence contexts, and were asked to provide the correct word. However, this task occurred offline, and in a third of the sentences, participants had recently read the correct continuation (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 may have lacked the ability to detect this in the current design.

In sum, in the current study, we found that domain knowledge influenced N400 amplitude to related anomalies whether or not individuals were able to provide the correct (Supported) word for these sentences. We take this to complement findings from Troyer et al. (2019) indicating that domain knowledge has a quick influence on the processing of words in context beyond mediating specific (lexical) knowledge at the trial level.

Conclusion

In conclusion, the present study provides strong evidence that individual differences in (domain) knowledge have a rapid and systematic influence on access to relevant information during real-time word by word written sentence processing. Having relatively more knowledge, and, we argue, a knowledge system that is organized around structures related to categories and events, among other relations, allows for quick availability (access/retrieval) of this information during reading. That is, knowledgeable individuals can quickly (pre)-activate relevant featural and/or event-related 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.

CRediT authorship contribution statement

Melissa Troyer: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Marta Kutas:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision.

[6] However, by-participant HP domain knowledge (based on the offline HP trivia quiz) was strongly correlated with the by-participant subjective reports of knowledge averaged over all trials.

Acknowledgements

This research was supported by HD022614 to MK and Frontiers of Innovation in Science Program grant to MT. We thank Thomas Urbach, Jeffrey Elman, Katherine DeLong, Seana Coulson, Sarah Creel, Zhuowen Tu, Wen-Hsuan Chan, Vic Ferreira and Ken McRae for input and feedback on this work and write-up.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2020.104111.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 79–87. https://doi.org/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. https://doi.org/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. https://doi.org/10.1037//0278-7393.16.2.305.

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. https://doi.org/10.1016/j.jml.2007.12.005.

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. https://doi.org/10.1016/j.neuropsychologia.2012.07.019.

Brouwer, H., & Crocker, M. W. (2017). On the proper treatment of the N400 and P600 in language comprehension. *Frontiers in Psychology, 8*(1327), https://doi.org/10.3389/fpsyg.2017.01327.

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, 41*(Suppl. 6), 1318–1352. https://doi.org/10.3389/fpsyg.2017.01327.

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. https://doi.org/10.1016/j.brainres.2012.01.055.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5*(1), 34–44. https://doi.org/10.1162/jocn.1993.5.1.34.

Cheyette, S. J., & Plaut, D. C. (2017). Modeling the N400 ERP component as transient semantic over-activation within a neural network of word comprehension. *Cognition, 162*, 153–166. https://doi.org/10.1016/j.cognition.2016.10.016.

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. https://doi.org/10.1207/s15516709cog0502_2.

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

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*, 368–407. https://doi.org/10.1006/cogp.2001.0752.

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. https://doi.org/10.1016/j.neuropsychologia.2016.01.029.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Memory and Verbal Behavior, 19*(4), 450–466. https://doi.org/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*. https://doi.org/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. https://doi.org/10.1016/j.neuropsychologia.2014.06.016.

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. https://doi.org/10.1111/lnc3.12093.

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

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

Elman, J. L., & McRae, K. (2019). A model of event knowledge. *Psychological Review, 126*(2), 252–291. https://doi.org/10.1037/rev0000133.

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

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. https://doi.org/10.1006/jmla.1999.2660.

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. https://doi.org/10.1016/j.brainres.2006.06.101.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158.

Ferreira, F., & Lowder, M. W. (2016). Prediction, information structure, and good-enough language processing. *Psychology of Learning and Motivation, 65*, 217–247. https://doi.org/10.1016/bs.plm.2016.04.002.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass, 1*(1-2), 71–83. https://doi.org/10.1111/j.1749-818x.2007.00007.x.

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. https://doi.org/10.1016/j.brainres.2013.02.017.

Foss, D. J. (1982). A discourse on semantic priming. *Cognitive Psychology, 14*, 590–607. https://doi.org/10.1016/0010-0285(82)90020-2.

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. https://doi.org/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. https://doi.org/10.1080/713752555.

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

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In Proceedings of NAACL (Vol. 2, pp. 159–166).

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. https://doi.org/10.1037/0096-3445.124.1.62.

Hillyard, S. A., & Kutas, M. (2002). *Event-related potentials and magnetic fields in the human brain. Neuropsychopharmacology: The Fifth Generation of Progress*. Baltimore: Lippincott, Williams and Wilkins427–439.

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. https://doi.org/10.1080/016909600386084.

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*, 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8.

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. https://doi.org/10.1037/xlm0000457.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163.

Knoeferle, P., Carminati, M. N., Abashidze, D., & Essig, K. (2011). Preferential inspection of recent real-world events over future events: Evidence from eye tracking during spoken sentence comprehension. *Frontiers in Psychology, 2*(376), 1–12. https://doi.org/10.3389/fpsyg.2011.00376.

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

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition, and Neuroscience, 31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299.

Kutas, M., DeLong, K.D., & 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 (pp. 190–207).

Kutas, M., & Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*, 463–470. https://doi.org/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. https://doi.org/10.1146/annurev.psych.093008.131123.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203–205. https://doi.org/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. https://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. https://doi.org/10.1016/j.jml.2009.06.004.
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. https://doi.org/10.1016/j.bandl.2014.03.002.
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. https://doi.org/10.1016/j.bandl.2011.09.001.
Lau, E., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience, 9*, 920–933. https://doi.org/10.1038/nrn2532.
Lenci, A. (2018). Distributional models of word meaning. *Annual Review of Linguistics, 4*, 151–171. https://doi.org/10.1146/annurev-linguistics-030514-125254.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006.
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. https://doi.org/10.1037//0278-7393.28.6.1073.
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. https://doi.org/10.3758/PBR.15.3.604.
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. https://doi.org/10.1037/a0022964.
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. https://doi.org/10.1037/0278-7393.30.2.465.
McRae, K., & Matsuki, K. (2013). Constraint-based models of sentence processing. In R. Van Gompel (Ed.). *Sentence Processing* (pp. 51–77). New York, NY: Psychology Press.
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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1162/jocn.2006.18.7.1098.
Onnis, L., & Spivey, M. J. (2012). Toward a new scientific visualization for the language sciences. *Information, 3*, 124–150. https://doi.org/10.3390/info3010124.
Otten, M., & Van Berkum, J. J. A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research, 1153*, 166–177. https://doi.org/10.1016/j.brainres.2007.03.058.
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. https://doi.org/10.1016/j.jml.2012.07.003.
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. https://doi.org/10.1162/jocn.2009.21393.
Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language, 29*, 633–654. https://doi.org/10.1016/0749-596X(90)90042-X.
Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as a change in probabilistic representation of meaning. *Nature Human Behavior, 2*, 693–705. https://doi.org/10.1038/s41562-018-0406-4.
Rabovsky, M., & McRae, K. (2014). Simulation the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition, 132*, 68–89. https://doi.org/10.1016/j.cognition.2014.03.010.
Rommers, J., & Federmeier, K. D. (2018). Predictability's aftermath: Downstream consequences of word predictability as revealed by repetition effects. *Cortex, 101*, 16–30. https://doi.org/10.1016/j.cortex.2017.12.018.
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. https://doi.org/10.1016/j.neuropsychologia.2012.12.002.
Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *TRENDS in Cognitive Science, 11*(6), 251–257. https://doi.org/10.1016/j.tics.2007.04.004.
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., & Chase, W. G. (1973). Skill in chess. *American Scientist, 61*(4), 394–403. https://doi.org/10.1007/978-1-4757-1968-0_18.
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. https://doi.org/10.1076/anec.8.2.137.843.
Soederberg Miller, L. M. (2009). Age differences in the effects of domain knowledge on reading efficiency. *Psychology and Aging, 24*(1), 63–74. https://doi.org/10.1037/a0014586.
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. https://doi.org/10.1016/S0022-5371(79)90155-5.
Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24*(4), 402–433. https://doi.org/10.2307/747605.
Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia, 56*, 289–301. https://doi.org/10.1016/j.neuropsychologia.2014.02.002.
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. https://doi.org/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*. https://doi.org/10.1080/23273798.2018.1503309.
Troyer, M., Urbach, T. P., & Kutas, M. (2019). *Lumos*!: Electrophysiological tracking of (wizarding) world knowledge use during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*. https://doi.org/10.1037/xlm0000737.
Van Berkum, J. J. A. (2010). The brain is a prediction machine that cares about good and bad – Any implications for neuropragmatics? *Italian Journal of Linguistics, 22*(1), 181–208.
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. https://doi.org/10.1162/jocn.2008.20054.
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. https://doi.org/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. https://doi.org/10.1016/j.ijpsycho.2011.09.015.
Venhuizen, N. J., Crocker, M. W., & Brouwer, H. (2019). Expectation-based comprehension: Modeling the interaction of world knowledge and linguistic experience. *Discourse Processes, 56*(3), 229–255. https://doi.org/10.1080/0163853X.2018.1448677.
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. https://doi.org/10.1207/s1532690xci0401_2.
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.