Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access
Author manuscript
J Exp Psychol Learn Mem Cogn. Author manuscript; available in PMC 2021 March 01.

Published in final edited form as:
J Exp Psychol Learn Mem Cogn. 2020 March ; 46(3): 476–486. doi:10.1037/xlm0000737.

Lumos!: Electrophysiological tracking of (wizarding) world knowledge use during reading

Melissa Troyer, **Thomas P. Urbach**, **Marta Kutas**
Department of Cognitive Science, University of California, San Diego, United States of America

Abstract

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

Keywords

sentence processing; event-related brain potentials; individual differences; knowledge; rERPs

1. Introduction

Understanding words in context requires quick, dynamic access to knowledge in long-term memory. Indeed, over the past fifty years, psycholinguists have convincingly demonstrated that sentence processing not only occurs incrementally but at times predictively (Kutas, DeLong, & Smith, 2011; DeLong, Troyer & Kutas, 2014). Online measures (eye-tracking, self-paced reading, and eye movements to images in scenes) indicate rapid sensitivity to word frequency (Trueswell, 1996), plausibility of thematic relationships (Trueswell, Tanenhaus, & Garnsey, 1994), discourse (Ehrlich & Rayner, 1981), and other linguistic/non-linguistic information gleaned from world knowledge (e.g., Kamide, Altmann, & Haywood, 2003; Borovsky & Creel, 2014). Scalp recordings of electrical brain activity (e.g., event-related brain potentials or ERPs) during word-by-word reading and speech track brain

Corresponding author: Melissa Troyer, 9500 Gilman Drive, La Jolla, CA 92039, mtroyer@ucsd.edu.
Author Contributions
M. Troyer and M. Kutas developed the study concept. All authors contributed to the study design. M. Troyer conducted data collection, analysis, and interpretation under the supervision of M. Kutas and with guidance and input from T.P. Urbach. M. Troyer drafted the manuscript, and all authors provided critical revisions. All authors approved the final version of the manuscript for submission.

Author Manuscript
Author Manuscript
Author Manuscript

functioning in real time and reflect neural sensitivity to word, sentential, and pragmatic factors that impact semantic retrieval[1] within 250 ms of a word's occurrence (e.g., N400, a negativity between ~250-500 ms post-stimulus onset which is especially large for semantic anomalies but is a default response to all words; e.g., Federmeier & Kutas; reviewed in Kutas & Federmeier, 2011). Compared to ERPs elicited by a correct (true) word, lexico-semantic violations and lexically appropriate but untrue world-knowledge violations both elicit indistinguishably large N400s (e.g., for Dutch speakers: ' *The Dutch trains are yellow / white / sour and very crowded.*'), suggesting similar time courses for retrieval of these two knowledge types (Hagoort, Hald, Bastiaansen, & Petersson, 2004). In short, understanding language requires knowledge of words and of the world, both assumed to be quickly available.

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

A major challenge for investigating knowledge-based individual differences in real-time sentence comprehension is allowing for the requisite variability in what people know while probing their comprehension with natural sentences. We have approached this using the narrative world of Harry Potter (HP) by J.K. Rowling—a constrained, yet rich, domain with complex intersecting trajectories of characters, objects, actions and events (Troyer & Kutas, 2017, 2018). Troyer and Kutas (2018) recorded EEG/ERP as participants varying in HP knowledge read sentences about (a) the world of HP or (b) general topics (i.e., control sentences). Control sentences ended with the highest-cloze/best completion (supported) or a plausible low cloze probability word (unsupported). HP sentences ended with a word that accurately described "facts" from the books (supported) or word that seemed plausible, but was factually incorrect given HP knowledge (unsupported). As expected, HP domain knowledge modulated N400 effects of contextual support for sentences about HP, but not for control sentences; the N400 effects were driven by variance in the *supported* (but not unsupported) HP words, consistent with the proposal that greater domain knowledge facilitates retrieval of relevant information during real-time sentence processing. We reasoned that greater HP domain knowledge likely led to participants knowing/remembering a greater number of the HP facts, and, consequently, to smaller mean N400 amplitudes to supported words as a function of the number of trials each participant knew. However, as we did not measure *which* facts each individual knew, we could not be certain that (a) those with greater HP knowledge knew a greater proportion of the critical words or that (b) this

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

alone determined the proportion of larger versus smaller N400s in their averages. Moreover, given attested individual differences in cognition and behavior as a result of domain expertise (Ericsson et al., 2006), we reasoned that domain knowledge might influence real-time processing beyond strictly determining the proportion of items individuals knew. Such individual differences in knowledge seem likely to influence the timing and/or contents of knowledge that can be rapidly brought to mind, yet they are strikingly absent from current models of real-time sentence processing. We thus decided to investigate the influences of HP domain knowledge and participants' specific knowledge of individual HP "facts" on written word ERPs by asking participants whether or not they had known each fact.

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

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

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

2. Method

2.1. Participants

41 students (mean age = 20 years; range = 18-23; 29 women, 12 men) at UCSD took part in the study for partial course credit or payment of $9 / hour. To ensure that some participants would have high knowledge of the Harry Potter domain, a subset ($N = 18$) were recruited contingent on having read all seven Harry Potter books and/or having watched all eight Harry Potter films. All participants provided informed consent reviewed by the Institutional Review Board at the University of California, San Diego. We estimated that ~40 participants was appropriate based on a previous study run in the lab, in which N400 amplitudes to contextually supported words correlated with HP domain knowledge at $r = .41$ (Troyer & Kutas, 2018); we replicate this result at $r = .37$ (Figure 4).

2.2. Materials

2.2.1. Sentence materials—During the EEG portion of the experiment, participants read 172 descriptions of facts / events from the narrative world of Harry Potter.[2] Using freely available materials (Wikipedia, fan sites) referring to the text of the Harry Potter books, the first author created a set of sentences that accurately described events/entities from the book series. The final word was designed to be 100% predictable given perfect knowledge of the series, verified by asking a separate group of participants to complete a cloze norming study ($N = 32$-34[3] participants, who varied in HP domain knowledge as determined by a 10-question trivia quiz, per item). Mean cloze probability across all norming participants and items was .49 (range: .03 to 1.00) (Figure 1b; examples provided in Table 1). For participants scoring in the top half on the quiz, mean cloze probability was .73 (range: .05 to 1); for those scoring in the bottom half, mean cloze probability was .24 (range: 0 to 1).

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

2.2.2. Individual differences tasks and measures—For EEG participants, our primary measure of HP domain knowledge was their score on a 40-question multiple choice HP "trivia quiz[4]." Participants also completed an HP self-report questionnaire; scores were determined by summing the total number of times an individual had read each book, seen each movie, etc.

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

[2]108 of these sentences were identical to the (supported) HP sentences from Troyer & Kutas (2018).

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

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

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

2.3. Procedures

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

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

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

2.4. EEG recording

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

2.5. Data analysis

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

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

2.5.2.1. Time window analyses.: We analyzed a window surrounding the typical peak of the N400 brain potential from 250 to 500 ms after critical word onset and a window from 500 to 750 ms during which post-N400 positivities—hypothesized to reflect a number of high-level processing mechanisms including mental model updating (e.g., Brouwer, Fitz, & Hoeks, 2012)—sometimes appear. We focused on a centro-parietal region of interest (ROI) where N400 effects are typically most prominent, averaging across 8 electrodes (MiCe, LMCe, RMCe, MiPa, LDPa, RDPa, LMOc, RMOc; Figure 5).

Our experimental design hinged on participants' subjective, trial-by-trial responses to Q1, leading to different numbers of trials per cell for this variable across participants.[5] To that end, we used hierarchical mixed-effects linear regression models, which allow for different counts per cell (Baayen et al., 2008), on trial-by-trial measures of ERP data, including mean amplitude in N400 window. Because factors modulating N400 amplitude often modulate brain potentials in a post-N400 time period (see Van Petten & Luka, 2012, for a review), we also analyzed mean amplitude in a late time window (500-750 ms). Unless explicitly indicated otherwise, time window analyses were performed on these single-trial data. Models included random intercepts[6] for item and participant and were implemented using lme4 (1.1-12; Bates, Maechler, Bolker, & Walker, 2015) and lmerTest (3.0-1; Kuznetsova, Brockhoff, & Christensen, 2017) packages in R (3.3.2). Where relevant, we performed model comparison and report Chi-squared statistics on nested models to do significance testing on covariates of interest. To further understand covariates of interest, p-values on beta coefficients were computed using the Satterthwaite option for denominator degrees of

[5]We recognize that there may be systematic differences in how participants answered Q2 as a function of HP domain knowledge. However, we do not report ERP data based on response to Q2 because there were too few trials per cell, especially for the most knowledgeable individuals, who rarely responded that they were certain they had *not* known an item/fact (Figure 2). In future studies with more power, we plan to address the relative influence of domain knowledge on items never known, those seeming familiar (perhaps forgotten), and those actually remembered/known.

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

freedom for *F* statistics. Categorical predictors were deviance-coded (i.e., Yes=1, No=−1, for Q1) and continuous predictors (e.g., HP domain knowledge, offline cloze) were z-transformed so that a value of +/− 1 reflected a single standard deviation above or below the mean (at 0). To visualize effects of and interactions between Q1 response and HP domain knowledge we computed predicted ERPs from the coefficients of the mixed-effects linear regression model.

3. Results

3.1. Behavioral data

3.1.1. Individual differences tasks—SI Table 1 reports descriptive statistics for participants' scores on the HP trivia quiz and other individual difference measures. The distribution of HP domain knowledge scores is plotted in Figure 1a. Intercorrelations among individual differences measures are provided in SI Table 2.

3.1.2. Question responses—Participants indicated that they had known information described by each sentence pair (i.e., responded "Yes" to Q1) on an average of 59% (95% CI [52%, 67%]) of trials. As anticipated, the number of statements that participants reported having known correlated strongly with their performance on the HP domain knowledge quiz, $r = .85$, $p < .001$ (Figure 1c). Item-wise offline cloze probability, measured in a separate group of participants, was also strongly correlated with the proportion of participants who reported having known each item, $r = .67$, $p < .001$ (Figure 1d). Participants responded with "Yes—Certain" on 50% (95% CI [41%, 58%] of trials; "Yes—Not Sure" on 10% (95% CI [8%, 12%] of trials; "No—Seems Familiar" on 18% (95% CI [14%, 21%] of trials; and "No—Not Familiar" on 23% (95% CI [17%, 29%]) of trials (Figure 2).

3.2. ERP data

Figure 3 displays trial-averaged ERPs across 26 scalp electrodes from 200 ms before the critical word onset to 900 ms post-critical word, with separate waveforms computed for trials on which participants responded "Yes" vs. "No" to Q1.[7] Across most electrodes, ERPs to critical words for "Yes" and "No" responses are characterized by two early sensory components (N1 and P2). The P2 is followed by a wave which is mostly positive-going for "Yes" responses and which shows a relative negativity for "No" responses.

Before turning to single-trial analyses, we examined participants' mean N400 amplitude to critical words in the centro-parietal ROI; it was correlated with HP domain knowledge at $r = .37$ ($p = .018$), replicating the pattern observed for HP-supported items in Troyer & Kutas (2018) (compare Figure 4a and 4b). Figure 4c presents the three-way relationship between participants' HP domain knowledge, mean N400 amplitude, and the proportion of trials reported.

[7]The pattern of reduced N400 amplitude for "Yes" compared to "No" responses is similarly apparent whether data are averaged across trials (as in Figure 3), averaged within participants and then across participants (as is typical in ERP studies), or averaged by trial or participant for a subsample of participants ($N = 28$) with a minimum of 20 items of each response type (Yes, No).

3.2.1. Single-trial ROI analyses

3.2.1.1. N400 time window: Results from a linear mixed-effects model crossing Q1 response type and HP domain knowledge as fixed effects are presented in Table 2 (see SI Figure 1 for a visualization of the beta coefficients fit to ROI data for the mixed-effects rERP model). Q1 response type was a significant predictor (Table 2); items that participants reported having known elicited reduced N400 amplitude compared to those reported as unknown. The effect of HP domain knowledge was marginal (Table 2), with higher-knowledge individuals exhibiting somewhat more positive-going N400 potentials compared to lower-knowledge individuals. Critically, the interaction term for Q1 response and HP domain knowledge was a significant predictor. This model was preferred over one incorporating only Q1 response as a fixed effect ($\chi^2(2) = 12.601$, $p = .002$), indicating that HP domain knowledge added explanatory power over and above participants' trial-by-trial reports of knowledge.

Since the planned test found the interaction effect reliable, we conducted follow-up analyses of "Yes" and "No" responses separately using mixed-effects models with HP domain knowledge as a fixed effect (Table 3). For "Yes" responses, a model incorporating HP domain knowledge was not a significant predictor; this model was not preferred over an intercept-only model ($\chi^2(1) = .006$, n.s.), indicating no explanatory power of HP domain knowledge for "Yes" trials. For "No" responses, however, HP domain knowledge was a significant predictor, and this model was preferred over an intercept-only model $\chi^2(1) = 5.003$, $p = .025$): having higher HP domain knowledge led to more positive going potentials in the N400 time window compared to having lower HP domain knowledge. We visualize this influence using both a standard approach, dividing participants into two subgroups using a median split based on HP domain knowledge and by plotting predicted ERPs based on regression modeling for hypothetical subjects (details in Figure 5).

Next, we asked whether individual differences apart from HP domain knowledge might modulate the influence of participants' subjective reports of knowing individual items. We therefore used a linear mixed-effects model predicting N400 amplitude based on four variables: HP domain knowledge, verbal working memory[8], reading experience, and general knowledge, as well as Q1 and its interaction with each of these variables (SI Table 3). Consistent with previous analyses, there was a significant effect of Q1 response type, with "Yes" trials associated with more positive-going potentials compared to "No" trials, and Q1 response type interacted with HP domain knowledge, but not with any other individual difference measure. Indeed, nested model comparison of this more complex model with one incorporating HP domain knowledge and Q1 response type (described above) indicated that the simpler model was preferred ($\chi^2(6) = 3.577$, n.s.); this confirms results from our previous analysis and strongly supports the notion that individual-participant-level HP domain knowledge influenced the brain's response to critical words on trials that participants reported *not* having known.

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

3.2.1.2. Late positivity time window.: Results from a mixed-effects model crossing Q1 response type and HP domain knowledge score as fixed effects are presented in Table 4. Both Q1 response and the interaction term crossing Q1 response and HP domain knowledge were significant predictors. This model was preferred over a model incorporating only Q1 response ($\chi^2(2) = 22.609$, $p < .001$), indicating that HP domain knowledge added explanatory power over and above participants' trial-by-trial reports of knowledge during the late positivity time window in the centro-parietal ROI.

To follow up on the Q1 by HP domain knowledge interaction, we fit separate models with HP domain knowledge as a fixed effect to subsets of the data based on Q1 response (Table 5). HP domain knowledge was a marginal predictor for "No" but not "Yes" responses: for "Yes" responses, a model incorporating HP domain knowledge was not preferred over an intercept-only model ($\chi^2(1) = 0.952$, n.s.) while for "No" responses, a model incorporating HP domain knowledge was marginally preferred over an intercept-only model ($\chi^2(1) = 3.809$, $p = .051$). To sum up, in the late positivity time window, for the centro-parietal ROI, higher (compared to lower) HP domain knowledge trended toward more positive-going amplitude on trials that participants reported *not* having known.

In an analysis incorporating individual differences (SI Table 4) we again found an interaction between Q1 response type and HP domain knowledge. Q1 did not interact with any other individual difference measure, nor were any other terms in the model significant. Indeed, the more complex model was dispreferred compared to the simpler model reported above ($\chi^2(6) = 4.802$, n.s.).

4. Discussion

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

Here, we addressed these unresolved questions in an electrophysiological reading study incorporating trial-by-trial participant knowledge reports. We replicated the moderate correlation between N400 amplitudes to contextually supported words with HP domain knowledge reported in Troyer & Kutas (2018). With our single-trial design, we confirmed our hypothesis that offline HP domain knowledge scores would be highly correlated with the proportion of online sentence comprehension trials participants reported knowing. As expected, we showed that single-trial-level participant reports of knowledge were strong predictors of ERPs to supported HP sentence endings. Critically, HP domain knowledge had yet an additional influence on ERPs, even after controlling for single-trial-level participant reports of knowledge. This effect was evident at least by ~250 ms, approximately when information retrieval from long-term memory is thought to occur (i.e., beginning during the N400 time period), demonstrating a rapid impact of domain knowledge on lexical/semantic retrieval. Moreover, this additional influence was reliable only for trials individuals reported *not* knowing. That is, domain knowledge had its greatest influence when retrieval difficulty was highest. This effect persisted beyond the N400 time period into a late positivity period, which has been associated with several high-level cognitive processes in different tasks. In sentence processing tasks, late positivities have been functionally attributed to semantic reanalysis (Van Petten & Luka, 2012) and/or updating of a mental model (Brouwer et al., 2012). In memory tasks, late positivity effects with similar timing and scalp distribution have been associated with retrieval of episodic/specific information (c.f. “old/new” effects, reviewed in Rugg & Curran, 2007). For example, Voss and Paller (2007) used a study-test “know/remember” paradigm and found that items judged as remembered/recollected during a test phase elicited greater parietal late positivities (beginning ~500 ms post-stimulus) compared to items judged as known/familiar (but not remembered) or items correctly judged as new. We do not see a direct link between our pattern of results and functional attributions of late positivity effects in sentence processing studies. Based on the memory literature, however, we speculate that our knowledge-based late positivity effects may result from variation in the degree to which individuals bring specific and/or episodic information to mind while processing critical words.

At the item level, real-time reports of knowledge and offline cloze probability (which correlated at $r = .67$; Figure 1c,d) indicated substantial variability in how likely each item was to be known/remembered in real-time and in how easy/difficult it was to produce the final word of each item. We reasoned that this item-level variability might also contribute to retrieval difficulty and be modulated by domain knowledge. We thus performed a post-hoc regression analysis incorporating Q1 response type, participant-level HP domain knowledge, and item-level offline cloze probability (SI Table 5). In the model, each of these predictors was significant, with N400 amplitudes reduced for high-compared to low-cloze items (SI Table 5). Importantly, there was a three-way interaction between Q1 response type, HP domain knowledge, and offline cloze: HP domain knowledge seems to have its greatest influence on low-cloze items individuals reported having known, and on high-cloze items individuals reported *not* having known (SI Figure 2)—both reflecting less than optimal retrieval conditions—i.e., cases in which items are least likely to be certainly known (or unknown) and which seem most likely to experience retrieval difficulty.

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

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

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported by HD022614 to MK and Frontiers of Innovation in Science Program grant to MT. We thank Jeffrey Elman, Katherine DeLong, Seana Coulson, Wen-Hsuan Chan, Ben Bergen, Vic Ferreira, and Kara Federmeier for input and feedback on this work and write-up.

References

Baayen RH, Davidson DJ, & Bates DM (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language, 59, 390–412.

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

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

Borovsky A & Creel S (2014). Children and adults integrate talker and verb information in online processing. Developmental Psychology, 50(5), 1600–1613. [PubMed: 24611671]

Boudewyn MA (2015). Individual differences in language processing: electrophysiological approaches. Language and Linguistics Compass, 9/10, 406–419.

Boudewyn MA, Long DL, & Swaab TY (2012). Cognitive control influences the use of meaning relations during spoken sentence comprehension. Neuropsychologia, 50, 2659–2668. [PubMed: 22842106]

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. [PubMed: 22361114]

Chi MTH & Ohlsson S (2005). Complex declarative learning In Holyoak KJ and Morrison RG (Eds.), The Cambridge Handbook of Thinking and Reasoning. Cambridge, UK: Cambridge University Press, pp. 371–399.

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

DeLong KD, Troyer M, & Kutas M (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. Language and Linguistics Compass, 5(12), 631–645.

Dunn DM & Dunn LM (2007). Peabody picture vocabulary test. Minneapolis, MN: NCS Pearson.

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

Ericsson KA, Charness N, Feltovich PJ, Hoffman RR (Eds.). (2006). The Cambridge Handbook of Expertise and Expert Performance. Cambridge, UK: Cambridge University Press.

Filik R & Leuthold H (2013). The role of character-based knowledge in online narrative comprehension: Evidence from eye movements and ERPs. Brain Research, 1506, 94–104. [PubMed: 23419895]

Ganis G, Kutas M, & Sereno MI (1996). The search for “common sense”: An electrophysiological study of the comprehension of words and pictures in reading. Journal of Cognitive Neuroscience, (2), 89–106. [PubMed: 23971417]

Hagoort P, Hald L, Bastiaansen M, Petersson KM (2004). Integration of word meaning and world knowledge in language comprehension. Science, 304, 438–441. [PubMed: 15031438]

Hald LA, Steenback-Planting EG, Hagoort P (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. Brain Research, 1146, 210–218. [PubMed: 17433893]

Kamide Y, Altmann GTM, & Haywood SL (2003). The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. Journal of Memory and Language, 49, 133–159.

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

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

Kutas M & Federmeier KD (2000). Electrophysiology reveals semantic memory use in language comprehension. Trends in Cognitive Sciences, 4(12), 463–470. doi: 10.1016/S1364-6613(00)01560-6 [PubMed: 11115760]

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

Kuznetsova A, Brockhoff PB, & Christensen RHB (2017). lmerTest Package: Tests in linear mixed effects models. Journal of Statistical Software, 82(13), 1–26.

McLaughlin J, Osterhout L, & Kim A (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. Nature Neuroscience, 7(7), 703–704. [PubMed: 15195094]

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. [PubMed: 19929760]

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. [PubMed: 19925188]

Rohrbaugh JW, Donchin E, & Eriksen CW (1974). Decision making and the P300 component of the cortical evoked response. Perception & Psychophysics, 15(2), 368–374.

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

Smith N & Kutas M (2015a). Regression-based estimation of ERP waveforms: I. The rERP framework. Psychophysiology, 52, 157–168. [PubMed: 25141770]

Smith N & Kutas M (2015b). Regression-based estimation of ERP waveform: II. Nonlinear effects, overlap correction, and practical considerations. Psychophysiology, 52, 169–181. [PubMed: 25195691]

Stanovich KE & Cunningham AE (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. Journal of Educational Psychology, 55(2), 211–229.

Stanovich KE & West RF (1989). Exposure to print and orthographic processing. Reading Research Quarterly, 24(4), 402–433.

Steyvers M & Tenenbaum JB (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. Cognitive Science, 29, 41–78. [PubMed: 21702767]

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

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

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

Trueswell JC (1996). The role of lexical frequency in syntactic ambiguity resolution. Journal of Memory and Language, 35, 566–585.

Trueswell JC, Tanenhaus MK, & Garnsey SM (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. Journal of Memory and Language, 33, 285–318.

Van Berkum JJA, Holleman B, Nieuwland M, Otten M, & Murre J (2009). Right or wrong? The brain’s fast response to morally objectionable statements. Psychological Science, 20(9), 1092–1099. [PubMed: 19656340]

Van Petten C & Luka BJ (2012). Prediction during language comprehension: Benefits, costs, and ERP components. International Journal of Psychophysiology, 83, 176–190. [PubMed: 22019481]

Voss JL & Paller KA (2007). Neural correlates of conceptual implicit memory and their contamination of putative neural correlates of explicit memory. Learning & Memory, 14, 259–267. [PubMed: 17412965]

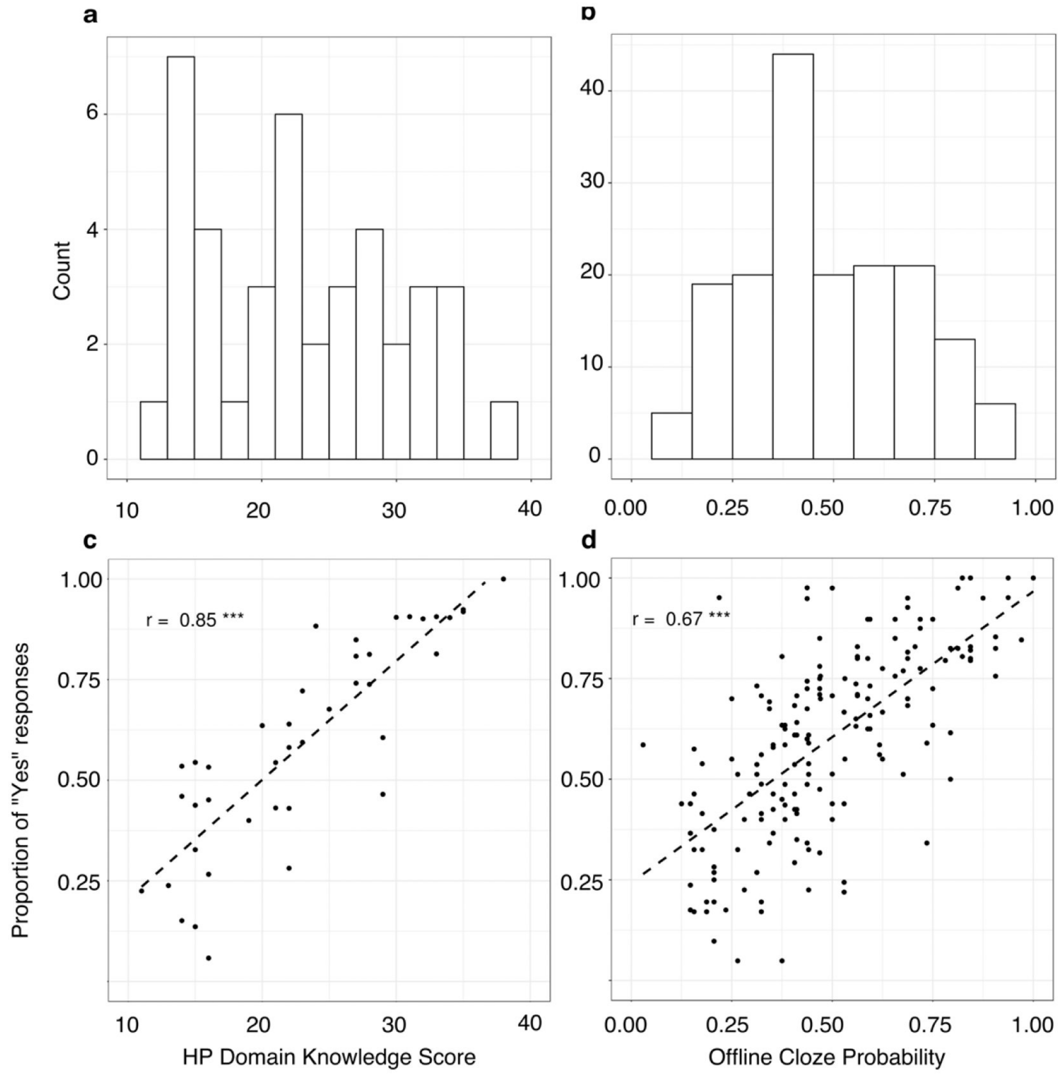

Figure 1.

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

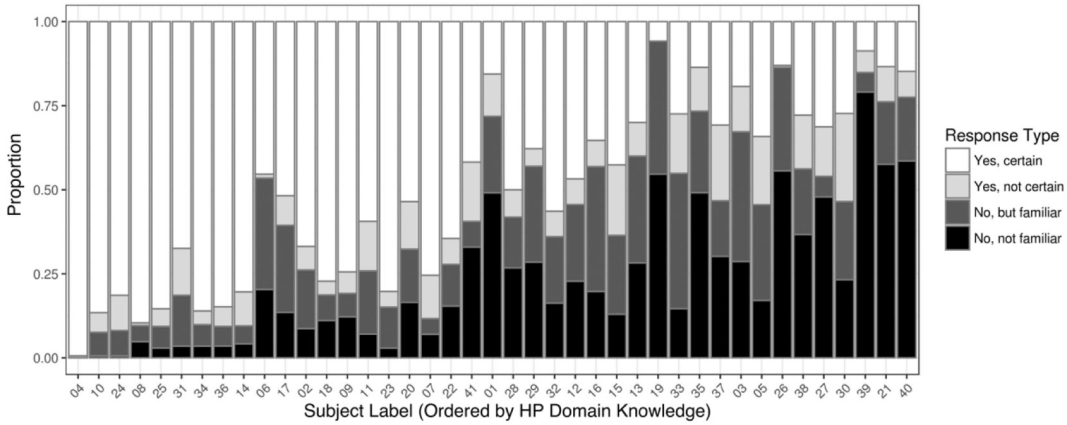

Figure 2.
The proportions of trials of each responses type (for Q1 and Q2) are plotted by participant, ranked by HP knowledge score (highest on the left, lowest on the right).

Author Manuscript
Author Manuscript
Author Manuscript

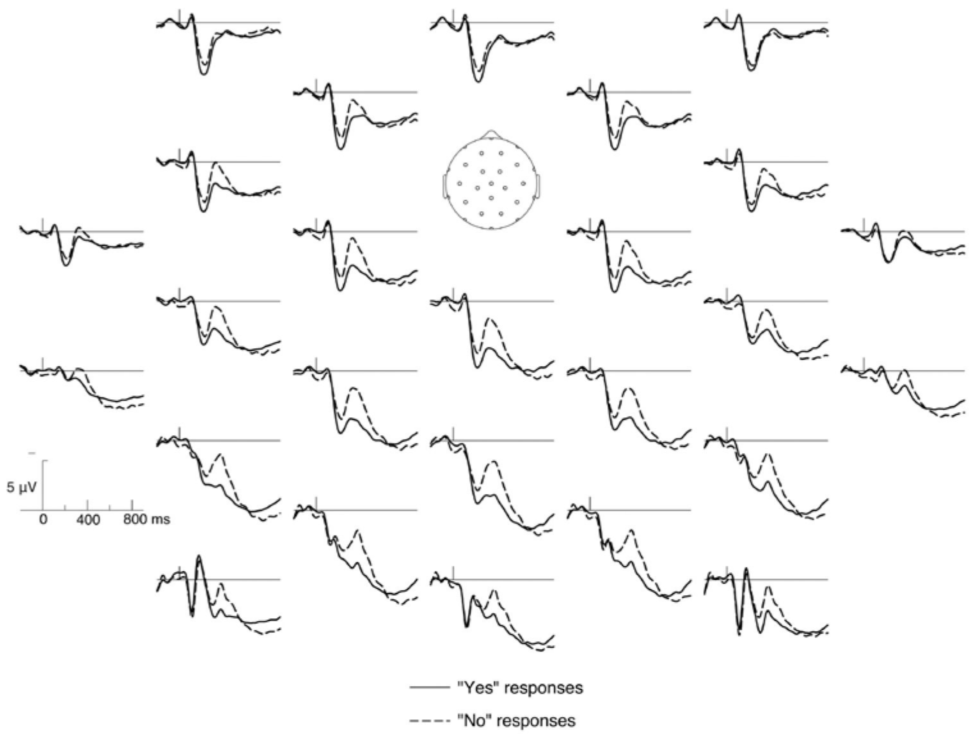

Figure 3.
Grand average ERPs across all single trials to critical words are plotted across the whole head using a low-pass filter with a cutoff of 10 hz (electrode locations shown on the head in the center).

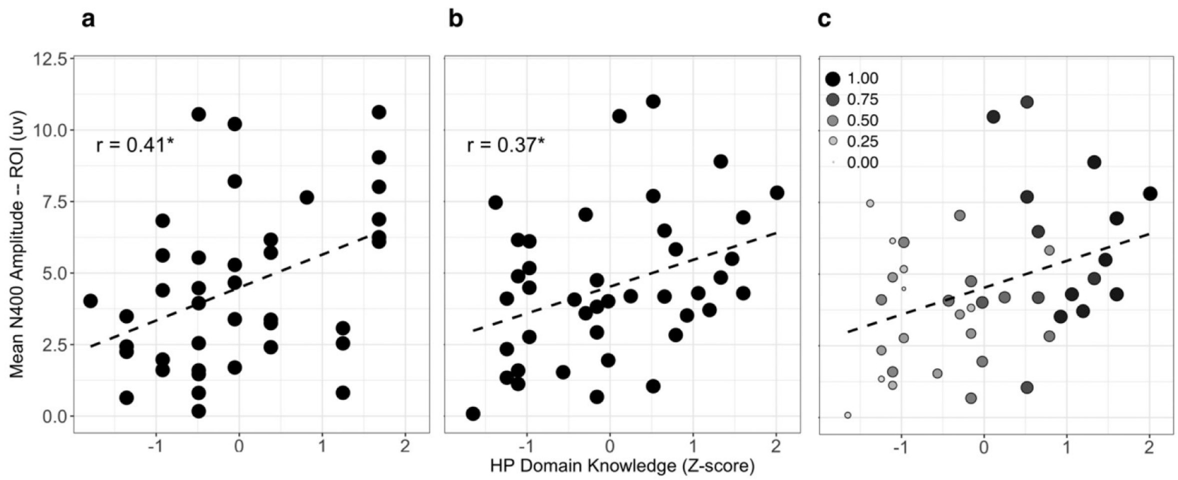

Figure 4.

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

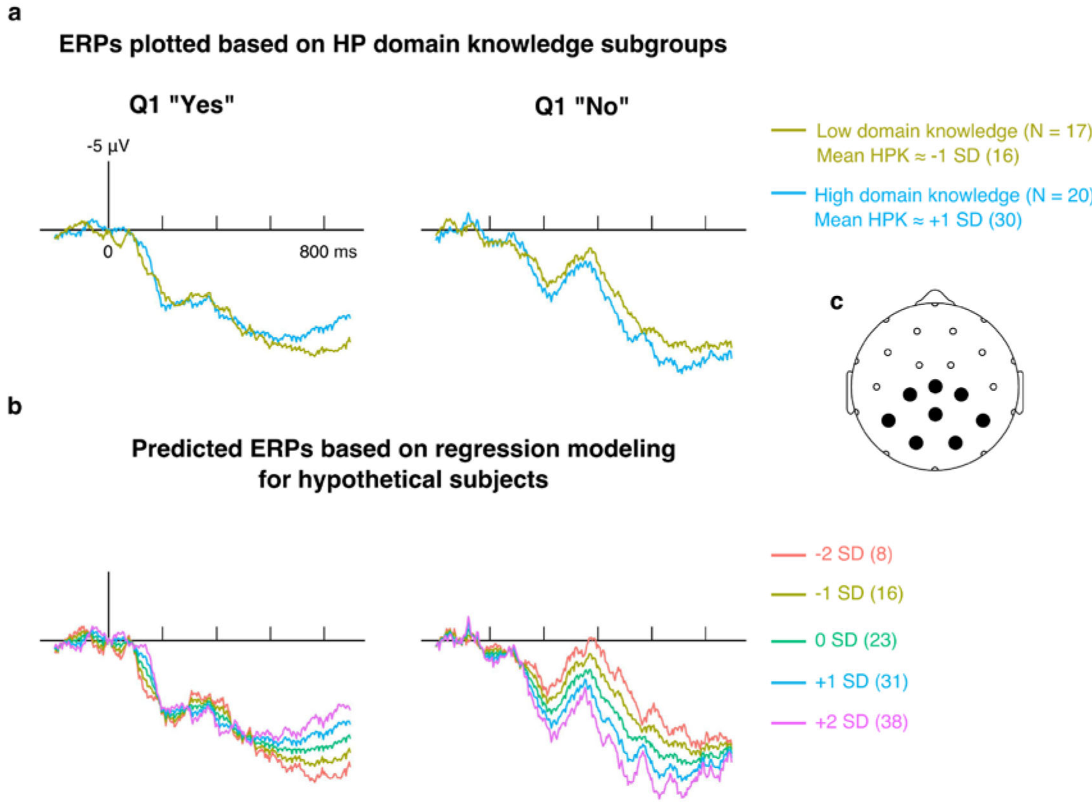

Figure 5.

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

Table 1.

Sample sentence pairs.

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

Table 2.

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

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

Table 3.

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

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

Table 4.

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

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

Table 5.

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

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