It's all in your head: Effects of expertise on real-time access to knowledge during written sentence processing

Melissa Troyer[1] (mtroyer@ucsd.edu) & Marta Kutas[1,2] (mkutas@ucsd.edu)
Department of Cognitive Science[1], Neurosciences[2], UCSD
9500 Gilman Dr., La Jolla, CA 92093

Abstract

Real-time sentence processing involves connecting linguistic input with knowledge. Here, we ask how variability in semantic memory (specific domain knowledge) may influence semantic access in real-time sentence processing. We recorded EEG while participants more/less knowledgeable about the narrative world of Harry Potter (HP) read sentences. In Experiment 1, all participants showed N400 predictability effects for general-knowledge sentences, but only those with high HP knowledge showed predictability effects for sentences about Harry Potter. This effect was driven by graded brain responses to predictable endings as a function of knowledge. Experiment 2 revealed greater semantic activation (inferred from N400 effects) for HP items participants reported knowing. High-knowledge participants also showed greater semantic activation for items they reported not knowing/remembering. These findings suggest that amount and/or functional organization of knowledge has real-time consequences on written sentence processing and implicate implicit/partial access to domain knowledge for experts when information is not explicitly recalled.

Keywords: sentence processing; knowledge; ERPs; individual differences

Introduction

Depending on your background, the question, "What's your patronus?" might leave you bewildered. But if you've spent a sizable chunk of your life obsessing over the fictional world of Harry Potter created by J.K. Rowling, you might have a response quickly at hand (e.g., a dolphin or a cat).

Variability in individual experiences helps determine an individual's knowledge, whether the domain is a fictional narrative world like Harry Potter, a game like chess, or an academic discipline, like physics. Moreover, knowledge differences in many domains have been shown to systematically influence various aspects of the organization of knowledge, including depth, breadth, and hierarchical information structure (Chi, 2006). Such differences in knowledge seem likely to impact real-time semantic access, including perceiving an utterance (or text), relating it to prior knowledge, and forming expectations about upcoming content. Yet despite the inevitable link between an individual's knowledge and semantic access, the specific role(s) of knowledge variability has received relatively little attention in models of real-time language processing.

Decades of psycholinguistic research have revealed that language processing is incremental: we update our mental representations word-by-word (e.g., Tanenhaus et al., 1995; Kamide, Altmann, & Haywood, 2003). Upon encountering an incoming word, world knowledge is used as soon as possible (e.g., Kutas & Hillyard, 1980). Real-time access to such knowledge is influenced by a host of contextual factors, both linguistic and nonlinguistic, including sentence and discourse context (Kutas & Hillyard, 1980), discourse context (Nieuwland & Van Berkum, 2006), and who the speaker is (Van Berkum et al., 2008).

These (and many other) studies have used event-related brain potentials to investigate the activation and organization of the semantic system during real-time language processing. A well-known ERP signature called N400 (a broad centro-parietally distributed, negative-going potential peaking approximately 400 ms after stimulus onset) shows fine-grained sensitivity to semantic relationships, with stronger relations between context and input yielding less negative- or more positive-going potentials between 200-500 ms post-input onset (Kutas & Federmeier, 2000).

The content and organization of long-term memory (i.e., knowledge) influence semantic access as reflected in N400 modulation both within sentences and simpler (or even no) context. For words presented in isolation, N400 amplitude is reduced for high-, compared to low-frequency, words (Kutas & Federmeier, 2000). Moreover, N400 amplitude is sensitive to category membership. Following a category label (e.g., 'a type of bird'), typical category exemplars ('robin') yield reduced N400 amplitude compared to atypical exemplars ('turkey'), and both are reduced compared to unrelated words ('broom') (Federmeier, Kutas, & Schul, 2010).

Such effects rely on long-term knowledge likely available due to years of experience with concepts like birds. N400 studies, however, have also revealed sensitivities to culturally-specific information (e.g., the fact that Dutch trains are yellow, not white; Hagoort et al., 2004) and fictional information (Nieuwland & Van Berkum, 2006; Filik & Leuthold, 2013). Taken together, these N400 findings offer a window into the relationship between language input and structured, flexible knowledge use. Moreover, N400 amplitude provides an excellent proxy measure of the ease of access to semantic information.

Here, we use the N400 to explore the notion that systematic variability in the content and organization of individuals' knowledge, *as a function of their expertise*, will have systematic influences on real-time semantic access. To that end, we conducted two ERP studies with individuals varying in their knowledge of the narrative world of Harry Potter. We first asked whether domain knowledge will systematically influence N400 effects, possibly reflecting ease of semantic access and/or availability of information in long-term memory (Experiment 1). Next, we dissociated knowledge of individual facts from domain knowledge, allowing us to ask whether or not, and if so how, domain knowledge influences ease of semantic access when individuals think they know, or don't know/remember, the information (Experiment 2).

Experiment 1

In Experiment 1, participants ranging in their knowledge of Harry Potter (based on an objective offline measure) read Control sentences about general, real-world topics as well as sentences about the narrative world of Harry Potter (HP) while we recorded EEG. Sentences of both types ended with either a Predictable or an Unpredictable word.

Based on the literature (Kutas & Federmeier, 2000), we expected predictability effects for Control sentences, with Predictable items eliciting reduced N400 amplitudes compared to Unpredictable items. Moreover, we expected that for HP sentences, specific knowledge of Harry Potter would have its effect during this N400 time window, with knowledgeable individuals showing a reliable predictability effect and less knowledgeable individuals showing a smaller (or no) difference between Predictable and Unpredictable words.

Though our study focused on predictions during the N400 time window, we also anticipated later positive effects. Late positive complexes, often occurring post-N400, have been related to attention-driven processing, including integration, revision, or updating of ongoing interpretations in the presence of unexpected items (e.g., Van Petten & Luka, 2012; Brouwer et al., 2012). We suspected that we might observe effects of Predictability on late positivities to words ending Control sentences, and possibly (for knowledgeable individuals) to words ending HP sentences, to the extent that individuals revise their interpretations.

Methods

Participants 40 right-handed students at UCSD participated for partial course credit and some monetary compensation.

Sentence Materials During the ERP portion of the study, participants read sentence pairs of two types. Control sentences described commonplace scenarios and ended in a Predictable (Offline Cloze > 87%) or an Unpredictable (Offline Cloze = 0%), albeit plausible, word, determined by offline norming studies. Harry Potter (HP) sentences described situations and entities from the Harry Potter book series and ended in a Predictable (book-consistent) or an Unpredictable (book-inconsistent) word. Unpredictable words were matched to Predictable words for the broad classes of words they belonged to (common noun, proper noun, Harry Potter-specific noun) and in many cases belonged to the same, more specific category. For example:

(1) Control: *We had been watching the blue jay for days. The bird laid her eggs in the* *nest.* (Predictable)
yard. (Unpredictable)
HP: *The character Peter Pettigrew changes his shape at times. He takes the form of a rat.* (Predictable)
dog. (Unpredictable)

The Predictable Harry Potter sentence endings were only predictable assuming perfect knowledge of the books. Unpredictable endings were inconsistent with the books but were designed to be similarly plausible endings, assuming no knowledge of the books. A total of 216 sentence frames (108 Control, 108 HP) were constructed, each with two ending types (Predictable, Unpredictable). Two lists were constructed such that each sentence frame appeared with only one ending type. Participants therefore saw a total of 216 sentences (54 sentences of each type).

Experimental Procedures Participants were told they would be reading sentences for comprehension and that they would be asked questions about the materials at the end of the study. After a practice session, blocks of Control sentences were presented first, followed by blocks of HP sentences. For each sentence pair, the first sentence appeared all at once in the center of the screen. When ready, participants pressed a button to move on to the second sentence, presented one word at a time in the center of the screen with a 500 ms SOA (200 ms on, 300 ms off). Immediately following the ERP study, participants were given a Control memory quiz followed by an HP memory quiz. For each, participants saw a list of 90 words, 60 of which had appeared in sentence-final position (half Predictable, half Unpredictable). They were instructed to circle the words they remembered as an ending to the second sentence of each pair in the study. After clean-up, participants completed 10 multiple-choice questions about the Harry Potter books. Raw scores are henceforth referred to as HP Knowledge. A median split on these scores determined High- and Low-Knowledge Groups.

ERP Recording and Data Analysis The electro-encephalogram (EEG) was recorded from 26 tin electrodes geodesically arranged in an ElectroCap, with impedances kept below 5 KΩ. Recordings were referenced online to the left mastoid and re-referenced offline to an average of the left and right mastoids. EEG was recorded by Grass bio-amplifiers with a bandpass of .01-100 Hz at a sampling rate of 250 Hz. Trials contaminated by artifacts (e.g., eye movements or blinks) were not included in analyses.

Grand average ERPs to sentence-final words were computed across all 26 recording sites by Sentence Type (Control/HP) and Ending Type (Predictable/Unpredictable). We performed statistical analyses on mean amplitudes of these waveforms in two time periods: a canonical N400 time period (250-500 ms) and a post-N400 period (500-750 ms) relative to a 500 ms pre-stimulus baseline. For each time period, we subjected data to an omnibus ANOVA including Channel[1] (26 levels), Sentence Type (Control, HP), Ending Type (Predictable, Unpredictable) as within-subjects factors and Knowledge Group as a between-subjects factor. Subsequently we focused on a region of interest (ROI) including 8 centro-parietally distributed channels (MiCe, LMCe, RMCe, MiPa, LDPa, RDPa, LMOc, and RMOc).

[1] Due to limited space, we do not report main effects of or interactions with Channel. Main effects of and interactions with Channel in the N400 region reflect the fact that N400 amplitude (and N400 effects) are largest over the middle and back of the head.

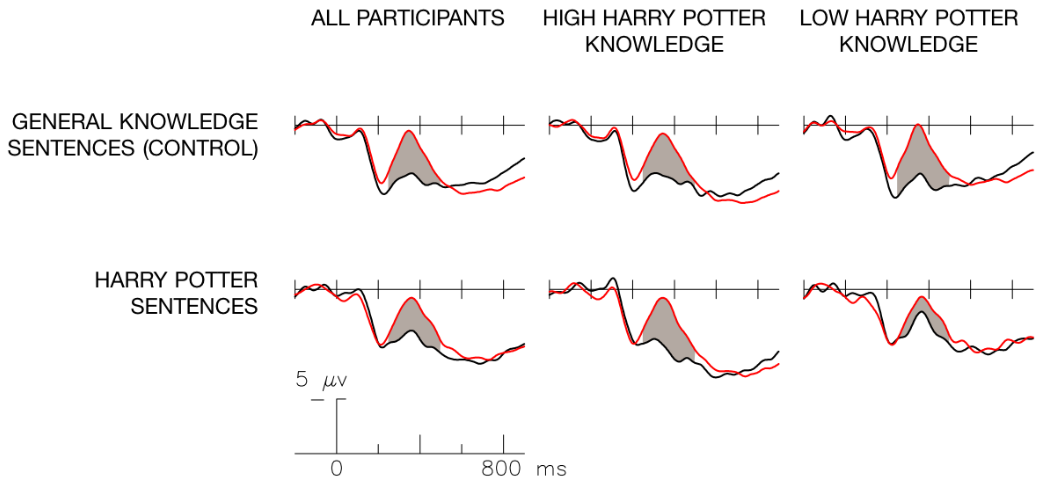

Figure 1: ERPs from a central-parietal ROI to sentence-final critical words from Experiment 1 are plotted for Predictable (black lines) and Unpredictable (red lines) endings relative to a 200 ms baseline for illustrative purposes. Shaded regions depict the area between 250 and 500 ms (N400 time window). All participants showed Predictability effects for Control sentences while Predictability effects for HP sentences were driven by the High-Knowledge group.

Results

Memory task Participants correctly recognized an average of 15 out of 60 Control words (25%) and false alarmed to an average of 2 words (7%). On the HP recognition test, participants correctly recognized an average of 30 out of 60 HP words (50%) and false alarmed to an average of 2 words (7%). Participants were therefore able to discriminate between words they had and had not seen for both the Control and HP memory tests.

To control for false alarms, we subtracted the number of false alarms for each memory test (Control, HP) from the number of items correctly recognized. We subjected these to a repeated measures ANOVA with Sentence Type (Control, HP) and Ending Type (Predictable, Unpredictable) as factors. There was a main effect of Sentence Type ($p < .0001$), with higher accuracy for HP compared to Control sentences. There was also an interaction between Sentence Type and Ending Type ($p < .001$); while memory for HP words was similar irrespective of the Ending Type (corrected accuracy for Predictable = 44%; corrected accuracy for Unpredictable = 40%), memory for Control words was better for Unpredictable words (22%) compared to Predictable words (15%).

As predicted, HP knowledge was not correlated with accuracy for Control words, but HP knowledge was correlated with accuracy for HP words (Predictable: $r = .471$, $p < .005$; Unpredictable: $r = .478$, $p < .005$).

ERPs ERPs from our centro-parietal ROI are shown in Fig. 1. ERPs for both Control and HP sentences 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, for Predictable endings, the P2 is followed by a positivity in the N400 time window (~250-500 ms). For Unpredictable endings, the P2 is followed by a relative negativity in this window.

Effects of knowledge during the N400 time window. Our primary hypothesis was that specific domain knowledge would influence semantic access, reflected by interactions between Knowledge Group, Sentence Type, and Ending Type during the N400 time window (250 to 500 ms). In the omnibus ANOVA, we observed a main effect of Ending Type ($p < .005$) but no effect of Sentence Type, reflecting the pattern observed in Fig. 1: for both sentence types, N400 amplitude is reduced for Predictable items. Of note, Knowledge Group interacted with Sentence Type ($p < .005$), and a three-way interaction was observed between Knowledge Group, Sentence Type, and Ending Type ($p < .05$).

To follow up on the effects of Knowledge Group, Sentence Type, and Ending Type on N400 amplitude, we examined Control and HP sentences separately at our centro-parietal ROI. Within Control sentences, there was an effect of Ending Type on N400 amplitude ($p < .0001$) but no main effect of Knowledge Group or interaction between Knowledge Group and Ending Type. Conversely, within HP sentences, there was a main effect of Ending Type ($p < .0001$), and an interaction between Knowledge Group and Ending Type ($p < .005$). Follow-up t-tests revealed a larger reduction in N400 amplitude for Predictable versus Unpredictable endings for the High-Knowledge Group compared to the Low-Knowledge Group ($p < .01$), supporting the notion that specific knowledge at the level of the individual reduces N400 amplitude (see Fig. 2). Furthermore, the differential knowledge had its effect primarily for Predictable endings, as High-Knowledge and Low-Knowledge individuals showed differences in N400 activity to Predictable endings ($p < .05$) but similar N400 activity to Unpredictable endings ($p = .630$).

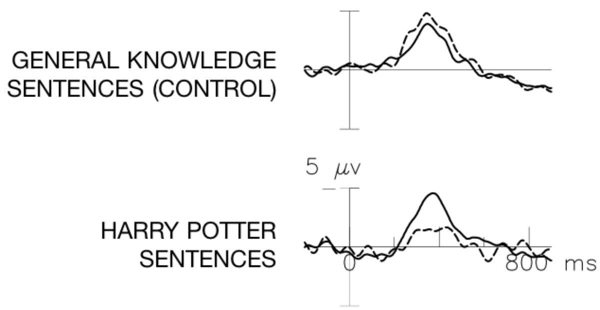

Figure 2. Difference ERPs for Unpredictable minus Predictable endings from Experiment 1 are plotted for the High HP Knowledge Group (solid lines) and the Low HP Knowledge Group (dashed lines) relative to a 200 ms baseline. Predictability effects for Control sentences were similar for both groups but only the High HP Knowledge Group showed sizable Predictability effects for HP sentences.

Analysis involving our continuous measure of HP Knowledge coincided with this pattern of results. We observed a graded relationship between the N400 effect (mean amplitude to Unpredictable minus Predictable endings) and knowledge scores, r = .457, p < .005; this relationship was driven by the correlation between knowledge and mean amplitude to Predictable endings (r = .473, p < .005) whereas no correlation obtained between knowledge and mean amplitude to Unpredictable endings (r = .171, p = .293).

Effects of knowledge post-N400 (500-700 ms). In the omnibus ANOVA, we observed an interaction effect of Sentence Type and Ending Type (p < .01), which resulted from a significant difference between Predictable and Unpredictable endings to Control sentences (p < .0001), with Unpredictable endings associated with greater positivities, but only a marginal difference between Predictable and Unpredictable endings for HP sentences, with the reverse pattern (p = .08). Apart from effects involving Channel, there were no other main effects or interactions in this analysis.[2]

Experiment 2

Experiment 1 demonstrated that specific domain knowledge influences real-time semantic access, inferred from N400 predictability effects. This pattern could obtain for multiple reasons. By definition, experts know more information, but expert knowledge also may be functionally organized differently, with greater structure and/or depth than that of less-knowledgeable individuals (Chi, 2006). We expect that this organization, in whichever form it may take, may influence semantic access above and beyond the successful retrieval of any given known item.

To tease apart contributions of (1) *knowledge of individual items* and (2) *knowledge of the domain* (Harry Potter) to semantic access, we asked participants to read sentences, all of which were consistent with the world of Harry Potter, and to respond with judgments of their knowledge along with their confidence in them. We were particularly interested in whether domain knowledge might have independent effect on semantic activation (inferred from N400 amplitude) for (a) items people say they know, (b) items people say they don't know, or (c) both.

Methods

Participants 41 right-handed students at UCSD participated for partial course credit and some monetary compensation.

Sentence Materials Materials consisted of 172 sentence pairs describing the world of Harry Potter, including the 108 from Experiment 1 plus an additional 64. All sentences ended in a word consistent with the Harry Potter books (i.e., the Predictable endings from Experiment 1).

Experimental procedure Sentence presentation was as in Experiment 1. After each sentence pair was presented, participants were first asked to make a non-speeded judgment about whether they knew the information in the sentences ahead of time, followed by a judgment of their certainty (we report only on responses to the first question in this paper).

After clean-up, participants completed 40 multiple-choice questions about the Harry Potter books (including the 10 questions used in Experiment 1). Raw scores are henceforth referred to as HP Knowledge. A median split on these scores was used to determine High- and Low-knowledge groups.

ERP recording and data analysis ERPs were recorded and processed as in Experiment 1. Because the design of Experiment 2 involves binning data based on subject responses, we used mixed-effects models (Baayen et al., 2008), which allow for the analysis of unbalanced data (e.g., Tibon & Levy, 2015). For both N400 and post-N400 time windows, we start by employing models that include fixed effects of (1) Judgment of Knowledge (two levels: "Yes," "No") and (2) HP Knowledge (continuous measure) along with random by-items and by-subjects intercepts. To unpack interactions, we follow up with similar mixed-effects models designed to isolate the root(s) of the interactions. These models were applied to data from our centro-parietal ROI (see Experiment 1). As in Experiment 1, we examined a window centered around the N400 (250-500 ms) and a post-N400 window (500-750 ms). For illustrative purposes, when plotting these data, we weight trials equally (rather than plotting grand averages as is typical, where each subject is weighted equally).

[2] More fine-grained analyses of HP sentences, however, do suggest a relationship between knowledge and late positivities that is mediated by offline, knowledge-based Cloze measures. For lower-Cloze (and, by inference, less-frequently-accessed) compared to higher-Cloze items, post-N400 activity was more positive-going, but only for high-knowledge individuals.

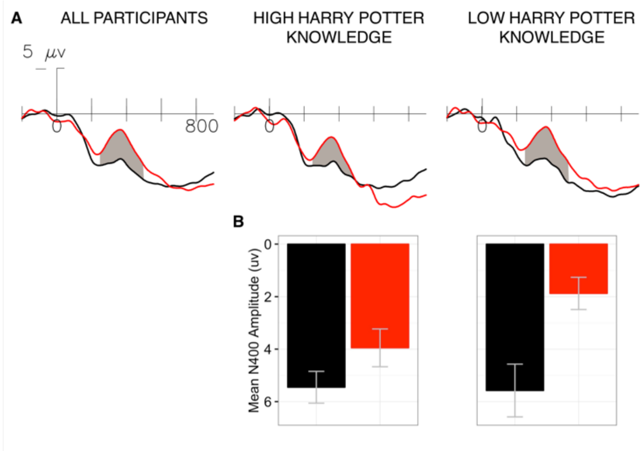

Figure 3: (A) ERPs from a central-parietal ROI to sentence-final critical words from Experiment 2 are plotted for words judged as known (black lines) and unknown (red lines) relative to a 200 ms baseline. Shaded regions depict the area between 250 and 500 ms (N400 time window). Across all participants, words judged as known led to more positive-going waves during this time. (B) During the N400 time window, HP Knowledge influenced mean amplitude only for Unknown (red), but not known (black) items.

Results

Behavior On average, participants responded that they knew 102 out of 172 items (60%). As expected, high-knowledge participants reported knowing more items (80%) than low-knowledge participants (38%), with a strong correlation between HP Knowledge and number of items judged as known, $r = .85$, $p < .0001$.

We trimmed response times three standard deviations greater than the mean for all responses. Response times for judgments of knowledge were overall slower for "No" responses (1015 ms) than "Yes" responses (851 ms), $p < .0001$. Moreover, HP knowledge interacted with Judgment of Knowledge ($p < .0001$): high-knowledge individuals responded faster for "Yes" responses (809 ms) than "No" responses (1193 ms) ($p < .0001$), but Low-Knowledge individuals showed only slightly faster RTs for "Yes" (943 ms) than "No" (964 ms) (n.s.). Pair-wise differences between High- and Low-Knowledge Groups were significant for "No" responses ($p < .005$) but not for "Yes" responses. Individuals therefore responded with similar speed when they judged items as known, but those with greater HP knowledge took longer to judge an item as unknown.

ERPs Grand average ERPs to sentence-final words were computed across all 26 recording sites grouped by participants' responses. See Fig. 3 for plots from the centro-parietal ROI.

Effects of knowledge during the N400 time window. As expected, we observed overall more positive-going waveforms during the N400 time window for high-knowledge compared to low-knowledge individuals ($p < .005$). In addition, positive Judgments of Knowledge (i.e., "Yes" responses) resulted in reduced N400 amplitudes ($p < .0001$; see Fig. 3). We also observed an interaction of Judgment and HP Knowledge ($p < .005$). Follow-up comparisons revealed that this interaction was driven by effects of HP Knowledge on N400 amplitude for "No" responses ($p < .05$) but not for "Yes" Responses, demonstrating that specific domain knowledge has its primary influence on items which participants say they did not know (recollect) at the time.

Effects of knowledge post-N400 (500-750 ms). Mean amplitude during the post-N400 window was influenced both by HP Knowledge ($p < .05$) and Judgments of Knowledge ($p < .0001$), with the two terms also interacting ($p < .0001$). Overall, "Yes" responses yielded more positive-going waves than "No" responses, and greater HP Knowledge was also related to more positive-going potentials. Follow-up analyses revealed that this interaction was driven by an effect of HP Knowledge on "No" responses ($p < .05$), with no relationship between HP Knowledge and positive-going potentials for "Yes" responses ($p = .869$). For "No" responses, individuals with more knowledge had more post-N400 positive-going activity.

General Discussion

We set out to investigate the relationship between specific domain knowledge and semantic access during real-time written sentence processing. Experiment 1 provided a strong indication that knowing about a domain (in our case, the narrative world of Harry Potter) influences semantic access, but only within that domain. As predicted, we observed no effects of Harry Potter-specific knowledge on processing of sentences about general topics. However, Harry Potter-specific knowledge did mediate N400 effects for Harry Potter sentences. We found that the size of the N400 predictability effect was correlated with HP knowledge score, with the correlation being driven by a graded relationship between knowledge and the neural response to predictable words.

In many ERP studies of sentence processing, predictability is defined using offline Cloze norming measures (that is, how likely an individual is to provide a word given a sentence context). It is worth noting that offline Cloze measures for our Harry Potter sentences provide a different type of metric than for our Control sentences. That is, predictable endings for Harry Potter sentences are predictable by virtue of being factual (within the narrative world); predictable endings for Control sentences are predictable based on world knowledge, but have no "correct" ending. Even so, our analyses revealed no main effect of sentence type (general/control vs. Harry Potter sentences) in Experiment 1. That is, across the whole group, we observed similar N400 effects of predictability for both Control and Harry Potter sentences (see Fig. 1). Our findings concur with many reports that N400 amplitude is sensitive to a word's predictability and/or contextual fit, in factual and non-factual scenarios (e.g., Hagoort et al., 2004).

Our finding that semantic access (inferred from N400 effects) is driven by knowledge is not surprising. In order to access information, the information must exist in the first place. However, there are at least two reasons why HP knowledge might relate to N400 predictability effects in Experiment 1: (1) low-knowledge individuals know fewer facts than high-knowledge individuals *on average*; and (2) there are potentially additional contributions of domain knowledge on semantic processing when individuals know (or don't know) items, respectively. In Experiment 2, we examined these possibilities by asking participants whether they knew each item (i.e., each Harry Potter fact) ahead of time. While both high- and low-knowledge groups showed large differences in N400 activity based on their own (meta-cognitive) judgments of whether they knew specific items, high-knowledge (compared to low-knowledge) individuals showed greater positivities during the N400 window even for information they reported *not* knowing at the time.

There are multiple reasons why domain knowledge might modulate N400 amplitude for items that are not immediately recognized. We cannot currently rule out the possibility that high- vs. low-knowledge individuals perceive different task demands or use different criteria when making judgments.

We believe a more likely explanation is that enhanced N400 reduction for high- compared to low-knowledge individuals suggests some level of implicit activation of information outside of conscious awareness. This activation may be restricted to a specific word and its semantic features or it may extend to related words / concepts. The precise nature of such implicitly activated information, and precisely how it is modulated by variation in level or amount of knowledge, have yet to be determined. Some possibilities include information that is taxonomically / categorically related to a predictable word (e.g., Federmeier & Kutas, 1999) and information related to the scenario / event being described (e.g., Metusalem et al., 2012).

As for post-N400 activity, we observed systematic effects of both HP Knowledge and judgments of knowledge on late positivities in Experiment 2, with high-knowledge individuals showing greater positivities than low-knowledge individuals for items they did not know. One way of interpreting this interaction is that when high-knowledge individuals do not know an item, they continue to search for it (perhaps because they believe they may know, but not currently be able to retrieve, the knowledge).

Our findings build on work showing that the functional organization of long-term memory plays an important role in the real-time construction of meaning (e.g., Federmeier & Kutas, 1999). We have demonstrated that variability among individuals in their knowledge of a domain is an important contributor to real-time access to meaning. More specifically, our data suggest that the amount and/or organization of domain knowledge appear to influence access to knowledge above and beyond explicit knowledge of individual items: expert-like knowledge organization in a domain may lead to implicit or partial activation of domain-related information, even when individuals do not explicitly recall a given piece of information.

Acknowledgments

This work was supported by NICHD grant HD22614 to Marta Kutas and UCSD Kroner Fellowship and FISP grant to Melissa Troyer. The authors would like to thank Seana Coulson, Katherine DeLong, Jeff Elman, Vic Ferreira, Zhuowen Tu, and Thomas Urbach for helpful feedback.

References

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

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

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

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

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

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.

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

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.

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

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

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.

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

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

Tibon, R. & Levy, D.A. (2015). Striking a balance: analyzing unbalanced event-related potential data. *Frontiers in Psychology*, *6*, 555.

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

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