



Published in final edited form as:

Neurobiol Lang (Camb). 2020 ; 1(1): 135–160. doi:10.1162/nol_a_00006.

Going the Extra Mile: Effects of Discourse Context on Two Late Positivities During Language Comprehension

Trevor Brothers*,

Tufts University, Medford, MA 02155

Eddie W. Wlotko,

Moss Rehabilitation Research Institute, Elkins Park, Pennsylvania 19027

Lena Warnke,

Tufts University, Medford, MA 02155

Gina R. Kuperberg

Tufts University, Medford, MA 02155

Abstract

During language comprehension, online neural processing is strongly influenced by the constraints of the prior context. While the N400 ERP response (300–500ms) is known to be sensitive to a word's semantic predictability, less is known about a set of late positive-going ERP responses (600–1000ms) that can be elicited when an incoming word violates strong predictions about upcoming content (*late frontal positivity*) or about what is possible given the prior context (*late posterior positivity/P600*). Across three experiments, we systematically manipulated the length of the prior context and the source of lexical constraint to determine their influence on comprehenders' online neural responses to these two types of prediction violations. In Experiment 1, within minimal contexts, both lexical prediction violations and semantically anomalous words produced a larger N400 than expected continuations (*James unlocked the door/laptop/gardener*), but no late positive effects were observed. Critically, the *late posterior positivity/P600* to semantic anomalies appeared when these same sentences were embedded within longer discourse contexts (Experiment 2a), and the *late frontal positivity* appeared to lexical prediction violations when the preceding context was rich and globally constraining (Experiment 2b). We interpret these findings within a hierarchical generative framework of language comprehension. This framework highlights the role of comprehension goals and broader linguistic context, and how these factors influence both top-down prediction and the decision to update or reanalyze the prior context when these predictions are violated.

General Introduction

It is well established that, during online language comprehension, readers are able to extract relevant features of the prior context to facilitate the processing of new inputs. However, what constitutes “relevant context” can vary considerably – from single words in a semantic

*Corresponding Author: trevor.brothers@tufts.edu.

priming task, to rich, multi-sentence discourse scenarios. In the present study, we investigated online neural responses when readers encountered predictable and unpredictable information, while systematically varying the length and semantic richness of the context that came before. By examining these responses in both early (300–500ms) and late (600–1000ms) time windows, we hoped to provide a clearer picture of how and when these aspects of the broader linguistic context influence the neural mechanisms engaged in language comprehension.

One important tool for examining the influence of context on language comprehension is the event-related potential (ERP) technique. This method has revealed several distinct neural components that are differentially sensitive to manipulations of linguistic context. One of the most frequently investigated ERP components is the N400, which is a negative-going waveform observed between 300–500ms after stimulus onset that is maximal over central-parietal electrode sites. During sentence comprehension, words that are semantically predictable in relation to their preceding context generate a smaller amplitude N400 than words that are semantically unpredictable (Kutas & Hillyard, 1984) or anomalous (Kutas & Hillyard, 1980). The amplitude of the N400 evoked by an incoming word is thought to reflect the retrieval or access to its semantic features that have not already been predicted (Kuperberg, 2016; Kutas & Federmeier, 2011).

The N400 component has played an important role in understanding the relative influences of local and global contexts in facilitating online comprehension. While some early, two-stage models of language comprehension argued for a delayed influence of global context on word processing (Rayner, Garrod & Perfetti, 1992; Till, Mross & Kintsch, 1988), a number of ERP studies have shown that discourse-based constraints can have an immediate influence on the amplitude of the N400 (Kuperberg, Paczynski & Ditman, 2011; Van Berkum, Hagoort & Brown 1999; Van Berkum, Zwitserlood, Hagoort & Brown, 2003). Moreover, these N400 effects are indistinguishable (in terms of latency and scalp topography) from N400 facilitation effects arising from local sentence contexts. These findings, in combination with similar results from eye-tracking while reading (Fitzsimmons & Drieghe, 2013), show that comprehenders can, in principle, rapidly combine information from both local and global contexts to help facilitate lexico-semantic processing (see Ledoux, Camblin, Swaab & Gordon, 2006 for a review).

In contrast to the N400, much less is known about the context sensitivity of language-related ERP responses that peak after the N400 time window. Here we consider two such ERP components — the *late frontal positivity* and the *late posterior positivity/P600* — both positive-going waveforms that are detected on the scalp surface between 600–1000ms following word onset.

The *late frontal positivity* is maximal over frontal electrode sites and was first characterized as a response to plausible but unexpected words appearing in highly constraining sentence contexts (Federmeier, Wlotko, Ochoa-Dewald, & Kutas, 2007; Kutas, 1993). For example, after a constraining context, “*After proposing, he put the ring on her...*”, an unexpected continuation (*dresser*) would not only elicit a larger N400 response, but also a larger *late frontal positivity*, relative to the expected completion (*finger*). It has been variously proposed

that this *late frontal positivity* response reflects the detection of a lexical prediction violation, the inhibition of an incorrectly predicted word (Federmeier et al., 2007; Kutas, 1993), and/or the incorporation of new unexpected information into a higher-level representation of meaning (Brothers, Swaab & Traxler, 2015; DeLong, Quante & Kutas, 2014; Federmeier, Kutas & Schul, 2010; Kuperberg, Brothers & Wlotko, 2019).

The *late posterior positivity/P600*, in contrast, is maximal over parietal and occipital sites, and was first characterized as an ERP response to words that are syntactically anomalous or structurally dispreferred (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992). This initial association between the P600 and syntactic processing was later revised when it became clear that syntactically correct sentences with anomalous semantic interpretations can also elicit the effect. For example, Kuperberg, Sitnikova, Caplan & Holcomb (2003) showed that semantically anomalous sentences (e.g. *Every morning for breakfast the eggs would eat...*) produced a robust *late posterior positivity/P600* effect relative to plausible control sentences (*Every morning for breakfast the boys would eat...*). Importantly, unlike the *late frontal positivity* effect, the *late posterior positivity/P600* effect is typically produced by continuations that are perceived as semantically impossible given the constraints of the prior context (Kuperberg, 2007; van de Meerendonk, Kolk, Vissers & Chwilla, 2009; see Van Petten & Luka, 2012 for a review of the early literature). This component has been linked to the detection of conflict between competing interpretations during language comprehension (Kim & Osterhout, 2005; Kuperberg, 2007; Paczynski & Kuperberg, 2012; van de Meerendonk, et al., 2008), second-pass reanalysis (van de Meerendonk, et al., 2009), and/or prolonged attempts to make sense of the input (Kuperberg, 2007).

We have recently proposed that these two late positivities and the N400 can be understood within a single theoretical framework, by appealing to a hierarchical generative model of language comprehension (Kuperberg & Jaeger, 2016; Kuperberg, 2016; Kuperberg, Brothers & Wlotko, 2020). Within this framework, the comprehender draws upon a *generative model* — a network of stored linguistic and non-linguistic representations that she believes are relevant to achieve her current goal. Information within this network is organized across multiple representational levels, which include semantic and syntactic features, as well as *event structures* which describe the semantic-thematic roles of individual propositions (see Dowty, 1989; Jackendoff 1987). If the comprehender's overarching goal is deep comprehension, then, at the highest level of the network, she will establish a *situation model*. Rather than simply encoding a surface-level representation of the text, this situation model constitutes a high-level representation of the full set of events being conveyed within the prior discourse, including the referential, spatial, temporal, motivational and causal coherence relationships that link them together (Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998).

Unlike earlier models of text/discourse comprehension (e.g. Kintsch 1988), this hierarchical generative framework is both generative and anticipatory. At any particular time, the comprehender may have high-level 'hypotheses' at the level of the situation model, and, to test these hypotheses, she propagates probabilistic predictions down to lower levels of the hierarchy ("top-down prediction"). As new bottom-up information flows up the network, any

information that has already been predicted is ‘explained away’, and any unpredicted information (prediction error) is passed up the hierarchy to update hypotheses at higher levels. In the brain, this type of message-passing scheme is known as *predictive coding* (Mumford, 1992; Friston, 2005; Rao & Ballard, 1999).

We refer to the full generative model that is deployed during deep comprehension (the situation model and levels of the network below it) as the *communication model*. Because this communication model specifies the comprehender’s assumptions about the current speaker and communicative environment (cf. Frank & Goodman, 2012), it places constraints over what sorts of inputs can (or cannot) be successfully incorporated at different levels of the hierarchy. For example, a comprehender’s communication model may incorporate the assumption that the communicator will describe events that are possible in the real world, placing constraints over the inputs that can be incorporated into the situation model.

As discussed by Kuperberg, et al. (2020), within this hierarchical generative framework, the N400 reflects the process of accessing semantic features associated with the incoming word which have not already been predicted from the prior context (semantic prediction error). The *late frontal positivity* is hypothesized to reflect shifts in belief at the level of the situation model in order to successfully incorporate new unanticipated information. Finally, the *late posterior positivity/P600* is triggered when the bottom-up input is incompatible with the communication model, and therefore cannot be successfully incorporated at a particular level of the hierarchy. In the case of the P600 evoked by semantic anomalies, this component is triggered by a failure to incorporate the new input at the level of the situation model. This results in prolonged attempts to make sense of the input through second-pass reanalysis or repair, and possibly through modifying the constraints of the communication model itself (adaptation).

The role of context and the present study

Within the hierarchical generative framework described above, both the *late frontal positivity* and the ‘semantic’ *late posterior positivity/P600* are thought to be generated by mismatches between the bottom-up input and the content/constraints present at the level of the situation model. Thus, according to this account, if no situation model has been established — that is, if the goal is not deep comprehension — then these two ERP responses should not be elicited, regardless of the predictability or plausibility of the bottom-up input. Clearly, many factors can determine whether or not we engage in deep comprehension and establish a situation model. These include the explicit task given to the reader, her internal motivation, and, of most relevance to the present study, the nature of the preceding context (see Zwaan & Radvansky, 1998, for discussion). For example, contexts that introduce characters, goals, and coherence relationships between events will encourage comprehenders to establish a situation model -- that is, to engage in deep comprehension. In contrast, following a short, semantically impoverished context, comprehenders may fail to establish a situation model, with potential consequences for the elicitation of these late positive components.

Although this issue has not been investigated in detail, there is some preliminary evidence supporting a link between the two late ERP responses and the nature of the prior context.

For example, while the *late frontal positivity* is typically observed in response to unexpected words within extended single-sentence (DeLong, Urbach, Groppe & Kutas, 2011) and multi-sentence discourse scenarios (Kuperberg, et al., 2020), this effect has not been observed in simple semantic priming paradigms. While semantically associated primes produce clear N400 reductions on expected target words (*hot – cold*) (Bentin, McCarthy & Wood, 1985, Rugg, 1985), there is no enhancement of the *late frontal positivity* to unexpected targets (e.g. *hot – brown*), even when the broader environment encourages participants to predict (Holcomb, 1988; Lau, Holcomb & Kuperberg, 2013).

The nature of the prior context has also been linked to the presence of the *late posterior positivity/P600* following semantic anomalies (see Kuperberg, 2007, for discussion). In a review of the early literature, Szewczyk and Schriefers (2011) noted that semantic anomalies presented in impoverished sentence contexts often fail to show a *late posterior positivity/P600* (e.g. Ainsworth-Darnell et al., 1998; Friederici et al., 1993, 1999; Gunter et al., 1999; Osterhout & Nicol, 1999; see also Kos, Vosse, van den Brink & Hagoort, 2010; Li, Shu, Liu, & Li, 2006; Lau, Namyst, Fogel & Delgado, 2016), while studies with longer introductory contexts (three or more content words preceding a violation) often *do* show a *posterior late positivity/P600*, at least based on based on an informal inspection of the ERP waveforms (for early examples, see Ganis et al., 1996; Kutas & Hillyard, 1980, 1983; Münte et al., 1998). Perhaps the most direct illustration of this relationship between context and the *late posterior positivity/P600* comes from an auditory comprehension study by Nieuwland and Van Berkum (2005). In this study, when participants encountered a semantic anomaly at the beginning of a discourse (*Next, the woman told the suitcase...*), this anomaly produced a robust N400 effect. However, when the same anomaly was placed at the end of an otherwise coherent narrative, it produced no N400 and a greatly enhanced *late posterior positivity/P600* effect instead.

Although these previous studies provide some evidence that the length and/or constraint of a prior context can play a role in modulating both the *late frontal positivity* and the *late posterior positivity/P600*, the hypothesized link between context and these two late ERP responses has not been systematically investigated. Specifically, it remains unclear what contextual factors may play a role in modulating these effects, including: 1) the amount of prior context, 2) its degree of constraint for an upcoming word, and/or 3) the source of lexical constraint (originating from a local context vs. a rich global context). By examining these questions, we hope to shed light on the functional significance of these two late positive effects, as well as the neurocognitive processes that are engaged when readers encounter unpredicted linguistic inputs.

Our starting point was a recent study by Kuperberg et al., (2020), who dissociated the N400, the *late frontal positivity*, and the *late posterior positivity/P600* effects in a single study during multi-sentence discourse comprehension. In that study, participants read and monitored the coherence of three-sentence scenarios, some of which were highly predictive of a particular direct object noun.

The lifeguards received a report of sharks right near the beach. Their immediate concern was to prevent any incidents in the sea. Hence, they cautioned the...

- 1a) ...swimmers (Expected)
- 1b) ...trainees (Unexpected)
- 1c) ...drawer (Anomalous)

This study showed a clear dissociation between the amplitude of the N400, which was primarily sensitive to lexical probability, and the two late positivities that were produced by words that violated prior predictions. Critically, the *late frontal positivity* was observed primarily to plausible but unexpected words (*trainees*), while semantic anomalies (*drawer*) that yielded an impossible interpretation, instead evoked the *late posterior positivity/P600*.

In the present set of experiments, we used a parallel experimental design that systematically manipulated the length of the prior context and the source of lexical constraint to investigate their role in eliciting these three ERP components. In Experiment 1, participants read minimal sentence contexts in which the constraint of the context was provided by a single verb in the local context (*Susan swept the...*), and ERPs were recorded to expected, unexpected, and anomalous continuations (Exp. 1, *single sentence*). In Experiment 2a, ERP responses were recorded to these same *locally constraining* sentences when presented as part of a longer three-sentence scenario. In Experiment 2b, ERP responses were recorded to a separate set of *globally constraining* three-sentence scenarios (a subset of those used in Kuperberg, et al., 2020), in which lexical constraint at the critical word depended on the entire preceding discourse. In this way, we separately examined how the length of the preceding context (Exp. 1. vs. 2a) and its source of lexical constraint (Exp. 2a vs. 2b) might influence how readers process contextually unexpected information during online comprehension.

Experiment 1

Introduction

In Experiment 1, we contrasted ERP responses to expected, unexpected, and anomalous critical words, following prior studies which have dissociated the late frontal and posterior positivities. However, rather than using multi-sentence discourse contexts (DeLong & Kutas, 2014; Kuperberg et al., 2020), we used very simple three-word contexts in which constraint was determined purely by the lexical properties of a single verb (“unlocked” in the examples below):

James unlocked the...

- 2a) ...door (Expected)
- 2b) ...laptop (Unexpected)
- 2c) ...gardener (Anomalous)

A large body of previous work in sentence comprehension suggests that words are recognized more rapidly and produce smaller N400 amplitudes when they are predictable given the preceding context (Kutas & Hillyard, 1984; Rayner & Well, 1996). We therefore predicted that, relative to the two other conditions, the amplitude of the N400 would be reduced for critical words that were highly expected given the constraints of the prior verb.

Our main question was whether the late frontal and posterior positivities would be produced by words that violated these short, lexically constraining sentence contexts. As noted in the Introduction, while the *late frontal positivity* is reliably elicited to unexpected words in longer sentence and discourse contexts, this effect is not typically observed in single word priming paradigms. In addition, there is evidence that the *late posterior positivity/P600* evoked by semantic anomalies is attenuated or even absent when the preceding context is short (Kuperberg, 2007; Lau, Namyst, Fogel & Delgado, 2016; Nieuwland & Van Berkum, 2005; Szewczyk & Schriefers; 2011). In the current experiment, we asked whether these two late positivities would be observed in short, relatively impoverished sentence contexts, which, nonetheless, still constrained for a particular upcoming word.

As discussed in the Introduction, there has been some debate about the neurocognitive mechanisms underlying these two positivities, with some accounts emphasizing their role in detecting violations at lower levels of representations, and others, including the generative framework outlined above, emphasizing their role in higher-level comprehension. The results of this experiment can help arbitrate between these broad accounts. Specifically, if a robust *late frontal positivity* is elicited by unexpected continuations following short but lexically constraining contexts (*James unlocked the laptop*), this would suggest that violations of lexical constraint are sufficient for triggering this ERP response. Similarly, if a robust *late posterior positivity* is elicited by anomalous continuations (*James unlocked the gardener*), this would suggest that outright violations of a verb's selectional restrictions are sufficient for triggering this response. If, however, no late positivities are observed in Experiment 1, this would suggest that, rather than simply reflecting the violation of lexical predictions these ERP effects may, at least in part, reflect activity at a higher levels of representation.

Materials

To generate sentence materials for this ERP study, we first needed to identify a set of verbs that were predictive of a particular upcoming noun (e.g. *unlocked the door*, *dimmed the lights*, *raked the leaves*). To accomplish this, we first collected a set of 617 transitively biased verbs from a variety of sources (Levin, 1993; Paczynski & Kuperberg, 2011, 2012). We then carried out a behavioral cloze norming task to determine which of these verbs were the most lexically constraining. Participants, recruited from Amazon's Mechanical Turk, were asked to read a set of short active sentences containing a proper name, the verb, and a determiner (e.g. *James unlocked the...*). For each sentence, participants filled in the first continuation that came to mind. After excluding participants who failed to follow instructions, a minimum of 90 cloze responses were obtained for all verbs. The lexical constraint of each verb was determined by calculating the proportion of participants providing the most common completion for each context.

Based on these cloze norms, we selected 87 verbs that constrained for a particular noun. The average constraint of these verbs was 63% (SD: 15%, range: 99%–36%). For each verb, we selected three nouns: an expected noun that was always the most common completion provided in the cloze test, an unexpected noun that was still plausible, and a semantically anomalous noun that violated the selection constraints of the preceding verb (*James*

unlocked the door/laptop/gardener...). The unexpected and anomalous nouns were closely matched in cloze probability (both less than 1%). Moreover, unexpected and anomalous nouns were lexically identical, with the same words being counterbalanced across sentence contexts in a Latin square design (*Judy alerted the police/gardener/laptop...*; see Appendix for additional examples). On average, the expected nouns were slightly shorter and more frequent than the unexpected/anomalous nouns (*expected*: 4.9 characters, *unexpected*: 6.6, *anomalous*: 6.6; *expected*: 3.7 log per million, *unexpected*: 2.6, *anomalous*: 2.6, SUBTLEX-US frequency; Brysbaert & New, 2009).

Using these noun-verb pairings, we generated a set of short, simple active sentences. Each sentence included a proper name, a verb, and a determiner, with the critical noun always appearing as the fourth word of the sentence (*Beth raked the leaves*). We then asked a new group of 36 participants (12 per list) to rate the plausibility of each item on a seven-point scale (ranging from “makes perfect sense” to “makes no sense at all”). As expected, there were clear differences in plausibility between the three conditions (*expected*: $M = 6.8$, $SD = 0.3$, *unexpected*: $M = 5.8$, $SD = 1.1$, *anomalous*: $M = 1.8$, $SD = 0.6$; all pairwise $p < .001$).

For the main ERP experiment, we added an additional three to seven words to each sentence, which were held constant across conditions, ensuring that the critical noun was never sentence-final (e.g. *Beth raked the leaves out back in the yard*). The three high constraint conditions described above (*James unlocked the door/laptop/gardener...*) and an additional low constraint filler condition (*Emma confiscated the laptop...*) were counterbalanced across four lists. Items in this fourth filler condition contained verbs that were not highly predictive of an upcoming noun (constraint = 21%, $SD = 9\%$) and which could also continue with a plausible or anomalous continuation. These fillers ensured that not all of the sentences were highly constraining, as in Kuperberg, et al. (2020). Because of this counterbalancing scheme, participants saw 21 to 22 items in each of our three conditions of interest, which were randomly interspersed with 154 filler sentences with a similar structure. Thus, participants saw 220 trials in total, with an equal number of sentences with high and low constraint verbs and an equal number of sentences with plausible and anomalous object nouns.

Participants

We report data from thirty participants (16 females, mean age = 22 years). Data from four additional participants were excluded - one due to excessive EEG artifact and three due to accuracies below 75% in the acceptability judgment task (see below). All participants were recruited from Tufts University and the surrounding community. They were screened on the basis of the following exclusion criteria: significant exposure to any language other than English before the age of five, history of psychiatric or neurological diagnoses or injury, and the use of psychoactive medication within the preceding six months. Participants were right-handed as assessed by the Edinburgh Handedness Inventory. They provided written informed consent, were paid for their time, and all protocols were approved by Tufts University Social, Behavioral, and Educational Research Institutional Review Board.

Procedure and Preprocessing

Participants sat comfortably in a dimly lit room, approximately 100 cm from the computer screen. After EEG preparation, they performed a short practice session, followed by four experimental blocks. On each trial, participants first saw a central fixation cross, and a sentence was then presented, one word at a time in the center of the screen. Each word appeared for 450ms with a blank 100ms inter-stimulus interval between words. Participants were asked to read each sentence silently for comprehension, and, at the end of each trial, to judge whether or not the preceding sentence “made sense”, indicating their decision by pressing one of two keys. The entire experiment lasted approximately one hour.

Throughout the experiment, EEG was recorded from the scalp using 32 active electrodes (with a modified 10/20 system montage), using a Biosemi Active-Two acquisition system. Signals were digitized at 512 Hz with a passband of DC to 104 Hz. The EEG was referenced offline to the average of the right and left mastoids, and a 0.1 – 30 Hz bandpass filter was applied. The EEG was then segmented into epochs (–100ms to 1000ms) time-locked to the onset of the critical noun. Epochs containing artifacts such as blinks, eye-movements, or muscle artifacts were rejected prior to analysis. On average, there were 18 trials per condition following artifact rejection (range: 10–22), with no significant differences in artifact rejection rates across conditions ($F(2,58) = 0.8, p = .48$).

Statistical Analysis Strategy

EEG data was averaged for each subject and condition using a –100ms pre-stimulus baseline. EEG amplitudes were then averaged across electrode sites within specific spatiotemporal regions of interest that were selected *a priori* based on a previous study using a similar design (Kuperberg, et al., 2020). The N400 was operationalized as the average voltage across electrode sites within a central region (Cz, CPz, C3/4, CP1/2) between 300–500ms. The *late frontal positivity* was operationalized as the average voltage across electrode sites within a prefrontal region (FPz, AFz, FP1/2, AF3/4) between 600–1000ms, and the *late posterior positivity/P600* as the average voltage within a posterior region (Pz, Oz, P3/4, O1/2), also between 600–1000ms. Within each spatiotemporal region, we first conducted an omnibus ANOVA (three levels of Condition), which was then followed up by pairwise comparisons of interest.

Results

Behavioral Results

In the final sample of participants, behavioral accuracy was generally high (89%). Participants were more accurate in judging the plausibility of sentences with expected critical words (95%), than those with unexpected (85%) or anomalous (86%) critical words.

ERP Results

N400 (300–500ms, Central region)—In the N400 spatiotemporal region of interest, we observed a main effect of Condition, $F(1,29) = 17.3, p < .001$, which was driven by a reduction in N400 amplitudes for expected vs. unexpected critical words, $t(29) = -4.84, p$

< .001. N400 amplitudes to unexpected and anomalous critical words did not differ significantly from one another, $t(29) = 0.90, p = .35$.

Late frontal positivity (600–1000ms, Prefrontal region)—Unlike for the N400, there was no main effect of Condition on the *late frontal positivity*, $F(1,29) = 0.64, p = .53$. Consistent with this null effect, a planned pair-wise comparison revealed no significant differences between expected and unexpected critical words, $t(29) = 0.97, p = .34$.

Late posterior positivity/P600 (600–1000ms, Posterior region)—Similar null effects were observed on the *late posterior positivity/P600*. Again, we observed no significant main effect of Condition, $F(1,29) = 0.2, p = .81$, and a planned pair-wise comparison confirmed that there was no significant difference between expected and anomalous critical words, $t(29) = 0.22, p = .83$.

Discussion

In Experiment 1, participants read short active sentences and judged whether or not they made sense while ERPs were recorded from the scalp. Our goal was to assess the effects of minimal sentential contexts on neural responses to expected, unexpected, and semantically anomalous words. As expected, the amplitude of the N400 evoked by predictable nouns (*James unlocked the door*) was smaller than that evoked by unpredictable but plausible continuations (*laptop*) or anomalous continuations (*gardener*). This finding is consistent with many previous findings relating cloze probability and the N400 (Kutas & Hillyard, 1984), as well as with studies demonstrating facilitated semantic access to words preceded by semantically associated contexts (Camblin, Gordon & Swaab, 2007; Li, Shu, Liu & Li, 2006).

Despite these clear differences on the N400, we saw no significant effects following the N400 time window (600–1000ms), either at frontal or posterior electrode sites (see Figure 1). This finding differs from the results of a recent ERP study using rich discourse contexts (Kuperberg, et al., 2020), which demonstrated both late positivities in sentences with a similar syntactic structure and similar levels of lexical constraint. In terms of the *late frontal positivity*, this result suggests that the presence of a lexical prediction violation alone is not sufficient to elicit the effect. In terms of the *late posterior positivity/P600*, it suggests that the presence and successful detection of a semantic anomaly (a selectional restriction violation) is not sufficient to elicit this effect. In the absence of a prior context, participants in this experiment may have failed to establish a situation model. Instead of engaging in deep comprehension, they may have achieved high behavioral accuracy in this task by attending primarily to the surface level of the text, monitoring for a match or mismatch between the selectional constraints of the verb and the semantic features of the critical noun.

While the results of Experiment 1 are suggestive, they leave open the question of exactly *what* features of the prior context are necessary for eliciting these two late ERP responses. One possibility is that the presence of *any* discourse context — even one that is non-constraining — is sufficient to encourage participants to establish a situation model and engage in deep comprehension (Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). Another possibility is that comprehenders require a rich and globally constraining linguistic

context in order to elicit these two ERP effects. For example, in the high constraint contexts used by Kuperberg, et al. (2020), the lexical constraint at the point of the critical noun was derived from the entire preceding context, rather than just the immediately preceding verb. This type of rich, globally constraining context would lead comprehenders to build a *rich* situation model that constrains for an upcoming event, and it is possible that the late frontal and posterior positivities are only elicited within these types of rich, globally constraining linguistic contexts.

Our goal in Experiment 2 was to help adjudicate between these two possibilities.

Experiment 2

Experiment 2: Introduction

In Experiment 2a and 2b, we again measured ERPs to expected, unexpected and anomalous nouns. This time, however, they appeared in two-sentence discourse contexts. We varied these contexts such that the source of lexical constraint came either from only the preceding verb (*locally constraining*: Experiment 2a) or from the entire discourse context (*globally constraining*: Experiment 2b).

In Experiment 2a, we presented the same locally constraining sentences used in Experiment 1. However, these sentences were now presented at the end of two-sentence discourse contexts. The events described in these two introductory sentences were generally vague and they were not strongly associated with the event described in sentence three. Thus, just as in Experiment 1, any lexical constraints at the point of the critical noun were determined primarily by the lexical properties of the preceding verb. We refer to these items as *locally constraining* discourse scenarios.

Locally constraining:

He was thinking about what needed to be done on his way home. He finally arrived. James unlocked the (door/laptop/gardener)...

In Experiment 2b, we used rich, globally constraining discourse scenarios, which were a subset of the high constraint items used by Kuperberg, et al. (2020). Unlike the items in Experiment 2a, these discourse scenarios were semantically rich and interconnected, and the expected event in the final sentence followed naturally from the set of events described in sentences one and two. Notably, the verb in the final sentence was always non-constraining in isolation (e.g. *cautioned*). Thus, these scenarios were also lexically constraining at the point of the critical noun, but this constraint stemmed from the entirety of the preceding discourse context rather than the preceding verb. We refer to these items as *globally constraining* discourse scenarios.

Globally constraining:

Tim really enjoyed baking apple pie for his family. He had just finished mixing the ingredients for the crust. To proceed, he flattened the (dough/foil/onlookers)...

We were careful to match the strength of lexical constraint between the locally constraining contexts (used in Experiment 2a) and the globally constraining contexts (used in Experiment 2b), as quantified by our cloze norms. This enabled us to examine how the nature of the prior context (locally vs. globally constraining) influenced the processing of expected, unexpected and anomalous sentence continuations. Specifically, by contrasting Experiment 1 with Experiment 2a, we were able to isolate the influence of the presence of prior context, and by comparing the results of Experiments 2a and 2b, we were able to isolate the effects of global vs. local constraint on the N400, *late frontal positivity*, and *late posterior positivity/P600*.

As noted in the Introduction, there is a general consensus that extra-sentential context can immediately influence the difficulty of accessing the semantic features of incoming words, as indexed by the N400 (Van Berkum et al., 1999, 2003; Kuperberg, Paczynski & Ditman, 2011). However, there have been no ERP studies that have directly compared the effects of local and global context when the probability of the critical word is matched across conditions. In one behavioral study addressing this question (Fitzsimmon & Drieghe, 2013), readers showed similar effects of local and global contextual constraints on eye-movement behavior. Specifically, reading times and word skipping rates were equivalent on words that were globally predictable (*I looked up after hearing a chirping noise. I saw a bird...*) and locally predictable (*I looked up to the sky. I saw a feathered bird...*), with evidence of facilitation in both conditions relative to an unpredictable baseline. Therefore, assuming that early eye-movement measures and the N400 both reflect the difficulty of lexico-semantic retrieval, we predicted similar N400 facilitation effects on expected critical words in both locally constraining discourse contexts (Exp. 2a) and globally constraining discourse contexts (Exp. 2b).

Of primary interest was how the prior discourse context would influence modulation of the two late post-N400 positivities. As discussed earlier, one possibility was that the short, impoverished contexts presented in Experiment 1 encouraged participants to engage in a shallow processing strategy. On this account, participants relied purely on matching the lexical-thematic properties of the verb with the semantic and syntactic features of the noun in order to provide acceptability judgments. If this is the case, then the addition of a two-sentence contexts in Experiment 2a should encourage readers to establish a basic situation model and re-engage in deep comprehension, which in turn could enhance the amplitude of these two late positive components.

Alternatively, it may be that the critical factor for eliciting these late positivities is the establishment of a *rich* and interconnected situation model that globally constrains for a specific upcoming event. On this account, only a rich and globally constraining context would afford participants the opportunity to build this type of situation model, and these two late positivities should only emerge in Experiment 2b.

Finally, within the hierarchical generative framework outlined in the Introduction, these two manipulations may have dissociable effects on the two late positivities. It may be the case that even an impoverished context is sufficient to establish a full communication model, with constraints over what sorts of events can be successfully incorporated. If this is the case,

then any input that is incompatible/conflicts with these constraints would lead to a failure to incorporate the input into the situation model, evoking a *late posterior positivity/P600*. In contrast, if the *late frontal positivity* indexes a large update or reinterpretation of the prior situation model, then this ERP response may only be observed in Experiment 2b in which the lexical constraint of the critical word stems from a rich, globally constraining discourse.

Experiment 2: Methods

Materials

For this experiment we used two different sets of materials. The first was a set of *locally* constraining three-sentence discourse scenarios that we developed based on the single sentences used in Experiment 1, which we refer to as Experiment 2a. The second was a set of *globally* constraining materials, which we refer to as Experiment 2b.

To develop our locally constraining discourse scenarios (Exp. 2a), a two-sentence introductory context was added to each sentence used in Experiment 1. These introductions were written to be plausible but relatively unassociated with the expected event occurring in the final critical sentence. Thus, just as in Experiment 1, the predictability of the expected critical noun was driven purely by the lexical properties of the preceding verb (*The group always split up their responsibilities evenly. Everyone did their share. Charlotte swept the floor...*). To confirm that these discourse contexts did not alter the lexical constraint of the critical noun, we conducted a second cloze norming study on these discourse materials using a new pool of participants, using the same procedures as those described under Experiment 1. After exclusions, a minimum of 50 cloze responses were collected for each discourse context. This confirmed that the cloze probability of the expected critical noun did not differ between Experiment 1 and Experiment 2a (Exp. 1, single sentences: 63%, SD = 16%; Exp. 2a, locally constraining discourse contexts: 63%, SD = 17%, $t < 1$).

For the globally constraining discourse materials (Exp. 2b), we used a different set of discourse scenarios — a subset of the ‘high constraint contexts’ used by Kuperberg, et al. (2020). In these scenarios, the final sentence was similar in structure to those used in the locally constraining materials described above, but they included verbs that, when presented in isolation, did not constrain strongly for a particular noun (average constraint = 16%, e.g. *worried, designed, greeted*). Unlike the locally constraining contexts used in Experiment 2a, these sentences were preceded by a rich two-sentence discourse context. These contexts often contained links establishing coherence across sentences, including co-reference, temporal relationships, intentionality, and causation (Zwaan & Radvansky, 1998). Thus, in combination with the first few words of the final sentence, these rich discourse contexts made a particular upcoming critical noun highly predictable (e.g. *The aircraft was behaving strangely. Something on the control panel seemed off. Certainly, this worried the pilot...*).

For both Experiment 2a and 2b, each constraining context was again paired with three possible continuations: an expected noun, an unexpected but plausible noun, or a semantically anomalous noun. As before, the unexpected continuations were always lexically unpredictable (<1%) but plausible, and both the unexpected and anomalous nouns were counterbalanced across items. As in Experiment 1, the critical nouns were never

sentence-final and the words immediately following the critical word were always held constant across conditions (see Appendix and Kuperberg, et al., 2020 for additional information).

Offline cloze ratings confirmed that the lexical constraint of the locally constraining discourse contexts in Experiment 2a and the globally constraining contexts used in Experiment 2b were matched (Exp. 2a, locally constraining: 63%, SD = 16%; Exp. 2b, globally constraining: 66%, SD = 10%, $t(163) = 1.44$, $p = .15$). Thus, these two stimulus sets differed in the *source* of lexical constraint (Experiment 2a: preceding verb; Experiment 2b: global context), but not in the degree of lexical constraint just prior to the critical noun.

In addition, we also wished to verify our intuition that the globally constraining contexts were more semantically rich and interconnected than the locally constraining contexts in Experiment 2a. Therefore, we conducted an additional rating study with a new group of fifteen participants recruited from Mechanical Turk. Participants read the expected version of each three-sentence scenario (truncated after the critical word), and they were asked to rate (1–5) how “associated or connected the final sentence is to the previous context” (5 = “Very Connected”, 1: “Very Disconnected”). As expected, there were large differences in “connectedness” ratings across the two stimulus sets (Exp 2a: $M = 2.7$, $SD = 0.8$; Exp 2b: $M = 4.7$, $SD = 0.2$). In addition, although each scenario always contained two introductory sentences, these two sentences had more words on average in the globally constraining scenarios (Exp 2a: mean = 14.5, $SD = 3$; Exp 2b: mean = 21.7, $SD = 4$).

In the ERP experiment, these two sets of experimental stimuli were presented in a randomized order in the same experimental session. In addition to these 174 experimental trials (29 per condition), participants also saw 116 filler scenarios, which balanced the proportion of plausible and implausible trials across the experiment.

Participants

In Experiment 2, a new group of 30 participants was recruited from Tufts University and the surrounding community using the same screening criteria as Experiment 1 (21 females, mean age = 23 years). None of these participants were excluded due to excessive artifact or low comprehension accuracy. All participants provided written informed consent and were paid for their time.

Procedure

Participants read each three-sentence discourse scenario, while EEG was recorded from the scalp. The first two sentences of each scenario were presented in full, and participants pressed a button when they were ready to continue to the next sentence. The final sentence was then presented one word at a time, with the same timing as Experiment 1. At the end of each trial, participants indicated via button-press whether the preceding scenario made sense or not. In addition, on 10% of trials, participants were given a True/False comprehension question to ensure that they were attending to all three sentences of the preceding discourse. The experimental stimuli were split into 10 blocks, and the entire recording session lasted approximately 1.5 hours. EEG recording, filtering, and artifact rejection parameters were identical to Experiment 1. On average, following artifact rejection, 23 trials remained in each

condition (range: 15–26), and there were no significant differences in artifact rejection rates across conditions, $F(5,145) = 1.5, p = .21$.

Analysis Strategy

We analyzed Experiment 2a and 2b using the same strategy as that described for Experiment 1 (three spatiotemporal regions corresponding to the N400, the *late frontal positivity*, and the *late posterior positivity/P600*; see Figure 2). Because the critical nouns in Experiment 2a (locally constraining contexts) and Experiment 2b (globally constraining contexts) differed in sentence position, the main effect of “Experiment” was not particularly informative. Therefore, our analyses focused first on the main effect of Condition within each experiment, followed by cross-experiment comparisons to determine whether these main effects of Condition differed reliably (Exp. 1 vs. Exp. 2a, and Exp. 2a vs. Exp. 2b).

Experiment 2: Results

Behavioral Results

As in Experiment 1, accuracy in the acceptability judgment task was high, both in the locally constraining (91%) and the globally constraining discourse contexts (94%). Comprehension question accuracy was also high (93%), suggesting that participants were carefully attending to all three sentences of the discourse scenarios.

ERP Results: Experiment 2a (locally constraining discourse contexts)

N400 (300–500ms, Central region)—On the N400, we saw a main effect of Condition, $F(2,58) = 21.1, p < .001$, reflecting a smaller N400 to expected critical words than to both unexpected, $t(29) = -5.58, p < .001$, and anomalous, $t(29) = -5.91, p < .001$, critical words. The amplitude of the N400 did not differ between the unexpected and anomalous critical words, $t(29) = -0.15, p = .88$.

Late frontal positivity (600–1000ms, Prefrontal region)—Just as in Experiment 1, there was no main effect of Condition on the *late frontal positivity*, $F(2,58) = 0.22, p = .79$, with planned comparisons confirming that there was no difference between unexpected and anomalous critical words, $t(29) = 0.58, p = .57$.

Late posterior positivity/P600 (600–1000ms, Posterior region)—In contrast, we observed clear differences across conditions in the amplitude of the *late posterior positivity/P600*, $F(1,29) = 19.7, p < .001$. Unlike in Experiment 1, the *late posterior positivity/P600* was larger to semantically anomalous than expected critical words, $t(29) = 4.75, p < .001$. The amplitude of the *late posterior positivity/P600* did not differ between the unexpected and expected critical words, $t(29) = 0.56, p = .58$.

ERP Results: Experiment 2b (rich, globally constraining contexts)

N400 (300–500ms, Central region)—Just as in Experiments 1 and 2a, there was a main effect of Condition, $F(2,58) = 44.4, p < .001$, reflecting a smaller N400 to expected than to unexpected, $t(29) = -7.21, p < .001$, and anomalous, $t(29) = -8.26, p < .001$, critical words. Although smaller in magnitude, there were also significant differences in the N400 between

the unexpected and anomalous critical words, $t(29) = -2.50$, $p = .02$, which is generally consistent with prior findings (see Kuperberg, et al., 2020 for further discussion).

Late frontal positivity (600–1000ms, Prefrontal region)—Unlike in Experiment 2a, in rich, globally constraining contexts, the *late frontal positivity* did show significant differences across the three conditions, $F(2,58) = 13.4$, $p < .001$, due to a larger *late frontal positivity* on unexpected than expected critical words, $t(29) = 5.33$, $p < .001$. There was no difference in this component between the semantically anomalous and expected words, $t(29) = 0.72$, $p = .48$.

Late posterior positivity/P600 (600–1000ms, Posterior region)—Finally, in these rich globally constraining contexts, just as in Experiment 2a, we observed clear differences across the three conditions on the *late posterior positivity/P600*, $F(1,29) = 14.5$, $p < .001$, again due to a significantly larger *late posterior positivity/P600* on anomalous than expected critical words, $t(29) = 4.55$, $p < .001$. There was also a difference in this spatiotemporal region between unexpected nouns and expected critical word, $t(29) = 3.52$, $p = .001$, although this likely reflects the widespread *late frontal positivity* evoked by unexpected words, which was also visible over some posterior electrodes (Figure 3).

Comparisons across Experiments

Comparison between Experiment 1 and Experiment 2a—While the pattern of facilitation on the N400 and the null effects on the *late frontal positivity* were similar across Experiment 1 and Experiment 2a, the effect of semantic anomaly on the *late posterior positivity/P600* differed across these two experiments. Specifically, while semantically anomalous words did not produce a *late posterior positivity/P600* effect in Experiment 1 (short sentence contexts), this effect was present in Experiment 2a, when these same anomalies appeared in discourse contexts (see Figure 3).

To determine whether these between-experiment differences were statistically reliable, we directly compared all three effects — the N400 effect (unexpected *minus* expected), the *late frontal positivity* effect (unexpected *minus* expected) and the *late posterior positivity/P600* effects (anomalous *minus* expected) between the two experiments. This between-subjects analysis confirmed a significant difference between Experiment 1 and Experiment 2a in the magnitude of the anomaly effect on the *late posterior positivity/P600* (anomalous *minus* expected: *Exp1*: $-0.2\mu\text{V}$, *Exp2a*: $2.4\mu\text{V}$, $t(58) = 3.54$, $p < .001$). In contrast there were no differences between the two experiments in the magnitude of the N400 effect (unexpected *minus* expected: *Exp1*: $-2.6\mu\text{V}$, *Exp2a*: $-3.1\mu\text{V}$, $t(58) = 0.59$, $p = .56$), or the *late frontal positivity* effect (unexpected *minus* expected: *Exp1*: $0.6\mu\text{V}$, *Exp2a*: $0.2\mu\text{V}$, $t(58) = 0.53$, $p = .60$).

Comparison between Experiment 2a and Experiment 2b—While modulation of the N400 effect and *late posterior positivity/P600* effect appeared to be similar between Experiment 2a (locally constraining discourse contexts) and Experiment 2b (globally constraining discourse contexts), we saw clear differences between the two experiments in the modulation of the *late frontal positivity*. Specifically, in Experiment 2a, no *late frontal*

positivity effect was observed when comparing unexpected and expected critical words, but this effect was present in Experiment 2b.

To determine whether these differences were significant, we compared all three effects across the two experiments. This analysis confirmed a significant difference between Experiment 2a and Experiment 2b in the *late frontal positivity* effect (unexpected *minus* expected: *Exp2a*: 0.2 μ V, *Exp2b*: 2.6 μ V, $t(29) = 3.78$, $p < .001$). In contrast, there were no differences between the two experiments in the magnitude of the N400 effect (unexpected *minus* expected: *Exp2a*: -3.1 μ V, *Exp2b*: -3.5 μ V, $t(29) = -0.68$, $p = .50$), or the *late posterior positivity/P600* effect (anomalous *minus* expected: *Exp2a*: 2.4 μ V; *Exp2b*: 3.4 μ V, $t(29) = 1.16$, $p = .25$). See Figures 3 and 4 for a summary of these effects across experiments.

Experiment 2: Discussion

The findings of Experiment 2a and 2b were quite clear. As in Experiment 1, words that were expected produced a smaller N400 response than words that were unexpected. This N400 effect did not differ significantly between Experiment 2a when critical words appeared in locally constraining contexts, and Experiment 2b when critical words appeared in globally constraining contexts.

However, unlike in Experiment 1, the addition of a prior discourse context resulted in clear ERP differences in the post-N400 time-window (600–1000ms). Importantly, the pattern of these late effects differed depending on whether the context was impoverished and locally constraining, or rich and globally constraining. While a large *late posterior positivity/P600* effect was produced by anomalous (versus expected) words following *both* locally and globally constraining discourse contexts (i.e. in both Experiment 2a and 2b), a larger *late frontal positivity* effect was produced by unexpected (versus expected) words only in rich, globally constraining discourse contexts in Experiment 2b.

In conjunction with the results of Experiment 1, these findings suggest that readers' engagement in comprehending a prior linguistic context was linked to the production of a *late posterior positivity/P600* effect to semantic anomalies. Second, they suggest that the global constraint of the prior context is an important predictor of whether readers will produce a *late frontal positivity* to unexpected critical words that violate strong lexical predictions.

General Discussion

In the present study, we examined online language comprehension while participants read sentences and judged whether or not they made sense. Specifically, we examined neural responses to words that were either expected, unexpected, or semantically anomalous, given the constraints of the prior context. In two experiments, we asked whether words that confirmed or violated a comprehender's prior predictions about a specific upcoming event, or about what is possible in the real world, are processed differently depending on the extent and nature of the preceding linguistic context.

In Experiment 1, participants read very short sentences in which lexical constraint stemmed entirely from a single preceding verb (*James unlocked the...*). Consistent with the prior literature, highly predictable words (*door*) produced a smaller N400 than the other two conditions. However, relative to these expected words, no *late frontal positivity* effect was produced by unexpected but plausible continuations (*James unlocked the laptop...*), and no *late posterior positivity/P600* was produced by semantically anomalous words that violated the selectional constraints of the verb (*James unlocked the gardener...*). These null results differed from previous ERP studies which have reported both of these post-N400 ERP effects in extended, multi-sentence discourse contexts (DeLong, Quante & Kutas, 2014; Kuperberg, et al., 2020). These findings therefore provided some support for the claim that certain aspects of online language processing proceed differently in impoverished linguistic environments (Kuperberg, 2007; Szewczyk & Schriefers, 2011).

In Experiment 2a, these same locally constraining sentence contexts were presented as part of a longer, three-sentence discourse. Importantly, the first two sentences of these discourse scenarios were relatively semantically impoverished and were not strongly connected to the predicted event in the final critical sentences. Therefore, just as in Experiment 1, the lexical constraint of the context, prior to the critical noun, was determined almost entirely by lexical properties of the preceding verb. We found that, in comparison with Experiment 1, the added discourse context made no difference to the modulation of the N400 or the *late frontal positivity*. However, within these longer discourse contexts, the semantic anomalies now elicited a robust *late posterior positivity/P600* effect. This finding suggests that the elicitation of the *late posterior positivity/P600* to semantically anomalous words is associated with the presence of an extended linguistic context.

Finally, in Experiment 2b, participants read a separate set of rich discourse scenarios, which also contained expected, unexpected, or anomalous continuations. In these scenarios, instead of the contextual constraint stemming from a single verb, it was derived from the entirety of a rich, interconnected discourse passage. Just as in Experiment 1 and 2a, we saw effects of contextual constraint on the N400, and, just as in Experiment 2a, we saw a robust *late posterior positivity/P600* effect to semantic anomalies. Critically though, this was the only experiment in which we also observed a robust *late frontal positivity* to the unexpected but plausible continuations.

Taken together, these findings suggest that the comprehension of an extended context played an important role in producing both late positivity effects in the present study. Obviously, we do not wish to claim that multi-sentence discourse contexts are always required to elicit these effects. Several prior studies using extended, single sentence materials have reported *late posterior positivity/P600* effects (e.g. Kuperberg et al., 2003) and *late frontal positivity* effects (e.g. Federmeier et al., 2007). However, these findings provide strong evidence that both the presence of an extended linguistic context, and its semantic richness can influence the elicitation of these two ERP responses.

In the following sections, we discuss the pattern of dissociations among these three ERP components across the three experiments. We then consider how these findings might shed

light on the functional underpinnings of the ERP components themselves, as well as the broader theoretical implications of our findings.

The N400

One robust finding from the present set of studies was the consistent reduction in N400 amplitude to contextually expected critical words. The magnitude and timing of this N400 reduction did not differ depending on whether the constraint of the context stemmed from a single verb (as in Experiment 1 and 2a) or from a rich discourse scenario (as in Experiment 2b). Although these locally and globally constraining contexts were carefully matched in lexical constraint (as determined by cloze ratings) this did not necessarily ensure that we would observe equivalent N400 reductions across contexts. Indeed, there are many examples in the literature of dissociations between offline measures of predictability and N400 amplitudes (e.g. Chow, Smith, Lau & Phillips, 2016; Urbach & Kutas, 2010; Xiang & Kuperberg, 2015; see Kuperberg, 2016, for discussion), suggesting that, under some circumstances, the full set of information available within a context is not mobilized quickly enough to predict upcoming semantic information or facilitate semantic processing. The fact that no such dissociation occurred in the present dataset suggests that sentence boundaries alone do not present a meaningful barrier to anticipatory semantic processing during language comprehension (Fitzsimmon & Drieghe, 2013; Van Berkum, et al., 1999). If anything, we saw a small but non-significant trend toward greater N400 facilitation effects in the globally constraining discourse contexts (Experiment 2b).

We should note, however, that this pattern may not hold across all communicative environments or in all groups of comprehenders. For example, it has been shown that individuals with reduced working memory capacity show enhanced N400 priming from local, lexical associates (*arms and legs*), even when these associates are incongruent with the global discourse context (Van Petten, Weckerly, McIssac & Kutas, 1997; Boudewyn, Long & Swaab, 2013). Similarly, individuals with schizophrenia show impaired use of global (relative to local) contextual information (e.g. Swaab et al., 2013; see Kuperberg, 2010 for a review). Thus, while the present set of findings suggest that local and global constraints exert similar effects in skilled comprehenders, this equivalency may break down under high levels of processing load and in certain populations.

The late posterior positivity/P600

Our results have important implications for understanding the functional significance of the 'semantic' *late posterior positivity/P600*. As noted in the Introduction, this component is often observed in response to words that are anomalous given the constraints of the prior context (e.g. Kuperberg, 2007; van de Meerendonk, 2009; Kuperberg, et al., 2020). In Experiment 1 however, participants showed no evidence of a *late posterior positivity/P600* to semantic anomalies. This mirrors the findings of prior studies that used short, generally non-constraining sentence contexts (Friederici et al., 1993, 1999; Ainsworth-Darnell et al., 1998; Gunter et al., 1999; Osterhout & Nicol, 1999; Kos, Vosse, van den Brink & Hagoort, 2010). Importantly, in our study, this null effect was found despite participants' high accuracy in classifying the sentences as anomalous at the end of each trial (see also Gunter et al., 1999;

Osterhout & Nicol, 1999). This suggests that merely detecting an anomaly is insufficient to produce the *late posterior positivity/P600*.

The appearance of the *late posterior positivity/P600* in Experiments 2a and 2b suggests that comprehending an extended language context played a key role in triggering this effect. We suggest that this is because prior context motivated comprehenders to construct a situation model of the discourse and to engage in deep comprehension. When new bottom-up information conflicted with constraints of the situation model (i.e. constraints concerning semantic possibility/impossibility), this resulted in a failure to incorporate this new information, triggering a late posterior positivity/P600 (see Shetreet et al., 2019; Kuperberg, et al., 2020). We suggest that this failure, in turn, led the comprehender to engage in second-pass attempts to make sense of the input (see below).

In contrast, when processing the minimal contexts presented in Experiment 1, we suggest that comprehenders failed to establish a situation model at all, and instead processed the surface structure of the text (cf. Kintsch & van Dijk, 1978; Graesser, Singer & Trabasso, 1994), with the primary goal of detecting the match or mismatch between the thematic structure of the verb and its argument. This interpretation is consistent with previous findings that no *late posterior positivity/P600* is produced by syntactic violations in jabberwocky sentences (in which content words were replaced by pseudowords), even though participants were highly accurate in detecting these violations (Münste, Matzke & Johannes, 1997; Ericsson, Olofsson, Nordin, Rudolfsson & Sandstrom, 2008).

It is worth noting that these findings are incompatible with theories that attribute the *late posterior positivity/P600* simply to a binary task-relevant categorization of whether the critical word is plausible or anomalous (e.g. Bornkessel-Schlesewsky, et al., 2011; see also Sassenhagen, Schlewsky & Bornkessel-Schlesewsky, 2014). In Experiment 1, participants carried out the same judgment task as that used in Experiments 2a and 2b, but no hint of this effect was observed at the critical word or subsequent words of the sentence (data not shown). Rather than simply indexing the successful detection of a semantic anomaly, we believe the *late posterior positivity/P600* instead reflects the commitment to engage in additional processing of the bottom-up input, in service of the broader goal of successful comprehension. These additional processing stages may involve a reanalysis of the prior context (van de Meerendonk et al., 2009), attempts to repair the prior context, and/or second-pass attempts to come to new representation of meaning (Kuperberg et al., 2006; Kuperberg, 2007; Paczynski & Kuperberg, 2012; Brouwer, Fitz & Hoeks, 2012). We return to this point below when discussing open questions and future directions.

The late frontal positivity

The present pattern of results also has implications for understanding the functional significance of the *late frontal positivity*. As noted in the Introduction, this effect is often elicited by words that are plausible but still highly unexpected given the constraints of the prior context (Federmeier, et al., 2007; Van Petten & Luka, 2012). One possibility is that the *late frontal positivity* purely indexes the violation of lexico-semantic constraint. However, the absence of a *late frontal positivity* in Experiment 1 and Experiment 2a suggests that this ERP effect cannot solely reflect the detection of a lexical prediction violation or activity

occurring only at the lexico-semantic level (see also Lau, Holcomb & Kuperberg, 2013; Brothers, Swaab & Traxler, 2015; Kuperberg, et al., 2020). In both Experiment 1 and Experiment 2a, comprehenders had enough information to predict a specific upcoming word, and this lexical prediction was ultimately violated by the lexico-semantic features of the bottom-up input. However, despite producing differences in the amplitude of the N400, these unexpected words failed to generate a *late frontal positivity*.

The appearance of a robust *late frontal positivity* to the unexpected continuations in Experiment 2b suggests that the rich, globally constraining context played a critical role in producing this effect. We suggest that this is because, within these contexts, readers had established a rich situation model into which they had already incorporated a likely upcoming event (e.g. inferring that someone was rolling out dough in a pie making scenario). When the bottom-up input (e.g. *foil*) was inconsistent with this situation model, it led to a large shift or re-evaluation of this model for the comprehender to come to a new interpretation of the passage as a whole (see Kuperberg, et al., 2020). Within such constraining sentence contexts, the completion of this high-level updating process is likely to have entailed top-down suppression of an incorrect lexico-semantic prediction (e.g. the predicted word *dough*, being suppressed as the unexpected target is accessed and selected; see also Ness, Meltzer-Asscher, 2018).

On this account, the reason why no *late frontal positivity* was produced in Experiment 1 is that the context was so minimal that comprehenders failed to engage in deep comprehension at all — that is, they failed to set up any sort of situation model at the point of the critical word. The reason why it was not produced in Experiment 2a is because the anticipated event (*unlocking the door*) was not strongly connected to the full set of events that came before. Therefore, at the point of the critical noun, this predicted event had not been fully incorporated (pre-updated) into the prior situation model, and the unexpected continuation (*laptop*) did not trigger a large update at the level of the situation model, relative to the expected continuation.

Theoretical Implications

Several researchers have argued that the N400 primarily reflects the effects of confirmed semantic predictions, while the late positivities are associated with prediction violations. Moreover, it has been noted that the topography of these late components (frontal vs. posterior) is linked to whether the resulting interpretation of the unexpected input is semantically plausible or anomalous (Kuperberg, et al., 2020; Van Petten & Luka, 2012). The present findings build upon these observations by suggesting that offline measures of lexical probability and plausibility, when viewed in isolation, cannot provide a full account of comprehenders' online neural responses to predictable and unpredictable content during comprehension.

Instead, these findings provide strong evidence for the idea that the *late posterior positivity/P600* cannot simply reflect the detection of a semantic anomaly, and that the *late frontal positivity* cannot simply reflect the detection of a lexical prediction violation. Instead, these findings demonstrate that an extended prior context play an important role in eliciting these effects. We argue that this is because, by comprehending these contexts, readers established

a situation model. We suggest that both late positivity effects reflect activity at this high level of representation.

We next argue that the full set of findings can be understood within a hierarchical generative framework of language processing (see also, Kuperberg, 2016; Kuperberg & Jaeger, 2016; Kuperberg, et al., 2020). As noted in the Introduction, within this framework, the particular subset of stored linguistic and non-linguistic knowledge that an agent brings to any language task — her *hierarchical generative model* — will depend on her overall communicative goal (see Figure 5).

We suggest that in Experiment 1, when the prior context was extremely short, participants assembled a generative model that constituted the minimal number of levels required to fulfill the requirements of the task — that is, to decide whether or not each sentence made sense. The model included levels of representation that encoded the semantic and syntactic features of verbs and nouns, which were combined into event structures defining the broad thematic-semantic roles of “who does what to whom”. Critically, this generative network did *not* include a still higher level of representation that encoded a situation model.

We suggest that, in Experiment 1, information was passed up the network primarily in a bottom-up fashion. Thus, constraints at the verb pre-activated fine-grained semantic properties, as well as the part of speech of the upcoming noun phrase. Together, these allowed for the bottom-up anticipation of a specific upcoming event structure (e.g. *Agent unlocked inanimate Patient*). Expected critical nouns (e.g. *unlocked the door*) matched both fine-grained semantic predictions as well as the predicted event structure. Unexpected critical nouns mismatched the predicted semantic features, producing a larger N400, but were consistent with the verb’s predicted event structure. Anomalous critical nouns again mismatched the predicted semantic features producing a larger N400, and they *also* mismatched the predicted event structure. Being able to distinguish between the match or mismatch at the level of the event structure was sufficient for participants to meet the requirements of the task. Crucially, because there was no high-level situation model, no late positivities were produced.

In contrast, we argue that in Experiments 2a and 2b, participants engaged in a deeper mode of comprehension, with the implicit goal of establishing coherence within and across sentences (Graesser, Singer & Trabasso, 1994). To accomplish this goal, they assembled a generative network that included not only semantic, syntactic, and event structure representations, but also a still higher level of representation that encoded the situation model. This complete *communication model* established the constraints of the situation model – the assumption that the communicator would convey events that were *possible* in light of the preceding context, while ruling out potential event structures that conveyed impossible events (e.g. *unlocking a gardener*).

In both Experiment 2a and 2b, in the anomalous condition, the bottom-up input conflicted with the constraints of the situation model, leading to an initial failure to incorporate this input and a temporary arrest in comprehension. However, because, within this generative framework, the overarching goal is to *explain* the bottom-up input, the comprehender likely

attempted to resolve this conflict through a second-pass analysis – i.e. reanalyzing or repairing the input, and/or revising the communication model itself. We suggest that the *late posterior positivity/P600* reflected this decision to engage in second-pass attempts to explain the input.

In Experiment 2b, participants not only established a general situation model, but, prior to encountering the incoming word, they had established a rich and interconnected situation model, into which they had already incorporated a specific upcoming event. When the unexpected, bottom-up input reached this top layer of the hierarchy, it led to a shift/update in the participants' prior interpretation. We suggest that the *late frontal positivity* reflected this large shift in the situation model to reach a new coherent interpretation.

Open questions and future directions

Our findings highlight a number of important questions for future research. With regards to the *late frontal positivity*, they suggest that rich globally constraining contexts play an important role in eliciting this effect following a prediction violation. Critically though, there are several examples in the literature of enhanced *late frontal positivities* to unexpected words, even within lexically non-constraining sentence contexts (Federmeier & Kutas, 2005; Thornhill & Van Petten, 2012; Zirnstein, van Hell & Kroll, 2018). These findings can be explained within a hierarchical generative framework if it is assumed that, in such cases, unexpected incoming words are highly informative, producing a large update in the comprehenders' situation model, even though no specific event had been previously predicted. This hypothesis could be tested more directly in future studies by comparing ERP responses to contextually informative and uninformative words while holding other factors, such as cloze probability and lexical constraint, constant.

A second set of questions concerns the precise functional role of the *late posterior positivity/P600*. Here we have argued that the key trigger for this effect, at the point of the critical word, is not simply a dichotomous decision or classification of sentences as well-formed or semantically anomalous (Bornkessel-Schlesewsky, et al., 2011; Sassenhagen, Schlesewsky & Bornkessel-Schlesewsky, 2014), but, rather the downstream consequences of this decision. This has some interesting parallels with analogous discussions about the functional significance of the well-known P300 ERP component (Donchin & Coles, 1988), which shares common computational principles with the *late posterior positivity/P600* (Coulson, King & Kutas, 1998; Osterhout, Kim & Kuperberg, 2012; Sassenhagen, Schlesewsky & Bornkessel-Schlesewsky, 2014). Although the posterior P300 is classically produced when participants make categorical task-relevant decisions about an eliciting event, it has been argued that simply making an task-relevant categorical decision is not sufficient to produce this effect; rather, what is critical is that this decision must have some utility for future action, described broadly by Donchin as 'context updating' (Donchin & Coles, 1988). This commitment to engage in this future-oriented activity (contextual updating) may be linked to the top-down allocation of attention, functioning to increase the gain on information processing, and possibly mediated by the phasic release of the neuromodulator norepinephrine (Nieuwenhuis, Aston-Jones & Cohen, 2005; Yu & Dayan, 2005). In future studies, it will be important to test whether the *late posterior positivity/P600* elicited to

anomalous inputs during language comprehension is similarly linked to this type of top-down modulation of attentional gain.

Finally, the findings of this study tie into a large literature suggesting that comprehenders can approach a text with different goals and “standards of coherence”, depending on task demands and their own internal motivations (van den Broek, Bohn-Gettler, Kendeou, Carlson & White, 2011; Graesser, Singer & Trabasso, 1994). This has important implications for understanding the neurocognitive basis of poor reading comprehension. Skilled comprehenders are able to monitor and detect breaks in coherence during online processing, engaging in reanalysis and repair strategies to prevent comprehension failures, while poor readers are less accurate at detecting these coherence breaks and are less able to compensate accordingly (Garner, 1980; McInnes, Humphries, Hogg-Johnson & Tannock, 2003; Wagoner, 1983). It will be important for future studies to investigate whether these two late positivities might provide a useful neural marker for probing these coherence monitoring processes during online language comprehension.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Appendix: Example Stimuli

Experiment 1 and Experiment 2a

1a) He was thinking about what needed to be done on his way home. Finally, he arrived.

James unlocked the (door/laptop/gardener) and started his homework.

1b) The couple came home later that evening. The door was slightly ajar.

Judy alerted the (police/gardener/laptop) about the situation.

2a) It was a warm but rainy evening. Something didn't feel right.

Trevor robbed the (bank/commuter/monitor) in the middle of the night.

2b) Given the circumstances, something had to be done.

No one was quite sure what the best solution would be.

Jimmy dimmed the (lights/monitor/commuter) before finally going home.

3a) It was finally time for the annual trip. There were activities for everyone.

Michael milked the (cow/goat/tank) and fed some sheep.

3b) It had been a strenuous day. She was almost finished with the last task.

Kathleen parked the (car/tank/goat) and returned to the barracks.

Experiment 2b

1a) Tim really enjoyed baking apple pie for his family.

He had just finished mixing the ingredients for the crust.

To proceed he flattened the (dough/foil/engineer) on the counter

1b) The crew was navigating through harsh waters.

One of the sailors spotted a gigantic iceberg.

At once they warned the (captain/engineer/foil) about the danger.

2a) The lifeguards received a report of sharks right near the beach.

Their immediate concern was to prevent any incidents in the sea.

Hence, they cautioned the (swimmers/trainees/drawer) to be wary.

2b) Jason left Wells Fargo for a credit union.

After many persistent letters, the bank persuaded him to return by offering lots of benefits.

A week later, he reopened the (account/drawer/trainees) and again reconsidered.

3a) The groom was asked to write his own vows and read them aloud at the wedding.

This wasn't something he was excited about.

In the end he humored his (wife/father/presentation) at the ceremony

3b) At the conference, the UN's Security Council President addressed the recent issues.

Interpreters from all over the world sat in the front row.

Carefully, they translated the (speech/presentation/father) into many languages.

References

- Ainsworth-Darnell K, Shulman HG, & Boland JE (1998). Dissociating brain responses to syntactic and semantic anomalies: Evidence from event-related potentials. *Journal of Memory and Language*, 38, 112–130.
- Bentin S, McCarthy G, & Wood CC (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology*, 60, 343–355. [PubMed: 2579801]
- Bornkessel-Schlesewsky I, Kretzschmar F, Tune S, Wang L, Genç S, Philipp M, Schlewsky M (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language*, 117, 133–152. [PubMed: 20970843]
- Boudewyn Megan Ann, Long DL, & Swaab TY (2013). Effects of working memory span on processing of lexical associations and congruence in spoken discourse. *Frontiers in Psychology*, 4.
- Brothers T, Swaab TY, & Traxler MJ (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, 136, 135–149. [PubMed: 25497522]
- 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]
- Brysbaert M, & New B (2009). Moving beyond Ku era and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41, 977–990. [PubMed: 19897807]
- Camblin CC, Gordon PC, & Swaab TY (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language*, 56, 103–128. [PubMed: 17218992]

- Chow W-Y, Smith C, Lau E, & Phillips C (2016). A “bag-of-arguments” mechanism for initial verb predictions. *Language, Cognition and Neuroscience*, 31, 577–596.
- Coulson S, King JW, & Kutas M (1998). Expect the Unexpected: Event-related Brain Response to Morphosyntactic Violations. *Language and Cognitive Processes*, 13, 21–58.
- DeLong KA, Quante L, & Kutas M (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, 61, 150–162. [PubMed: 24953958]
- DeLong KA, Urbach TP, Groppe DM, & Kutas M (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, 48, 1203–1207. [PubMed: 21457275]
- Donchin E, & Coles MG (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11, 357–374.
- Dowty DR (1989). On the semantic content of the notion of thematic role In Chierchia G, Partee B, & Turner R (Eds.), *Properties, types and meaning* (pp. 69–129). Norwell, MA: Kluwer
- Ericsson E, Olofsson JK, Nordin S, Rudolfsson T, Sandstrom G, (2008). Is the P600/SPS affected by the richness of semantic content? A linguistic ERP study in Swedish. *Scandinavian Journal of Psychology*, 49, 1–9. [PubMed: 18190397]
- Federmeier KD, & Kutas M (2005). Aging in context: age-related changes in context use during language comprehension. *Psychophysiology*, 42, 133–141. [PubMed: 15787850]
- Federmeier KD, Kutas M, & Schul R (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115, 149–161. [PubMed: 20728207]
- Federmeier KD, Wlotko EW, De Ochoa-Dewald E, & Kutas M (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, 1146, 75–84. [PubMed: 16901469]
- Fitzsimmons G, & Drieghe D (2013). How fast can predictability influence word skipping during reading? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 1054–1063.
- Frank MC, Goodman ND (2012). Predicting pragmatic reasoning in language games. *Science*, 336, 998. [PubMed: 22628647]
- Friederici AD, Pfeifer E, & Hahne A (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, 1, 183–192. [PubMed: 8257874]
- Friederici AD, Steinhauer K, & Frisch S (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory & Cognition*, 27(3), 438–453. [PubMed: 10355234]
- Friston KJ (2005) A theory of cortical responses. *Philosophical Transactions of the Royal Society B*. 360, 815–836.
- 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*, 8, 89–106. [PubMed: 23971417]
- Garner R (1980). Monitoring of understanding: An investigation of good and poor readers’ awareness of induced miscomprehension of text. *Journal of Reading Behavior*, 12, 55–63.
- Graesser AC, Singer M, & Trabasso T (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371–395. [PubMed: 7938337]
- Gunter T, Friederici A, & Hahne A (1999). Brain responses during sentence reading. *Neuroreport*, 10, 3175–3178. [PubMed: 10574555]
- Hagoort P, Brown C, & Groothusen J (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8(4), 439–483.
- Holcomb PJ (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language*, 35, 66–85. [PubMed: 3179703]
- Jackendoff R (1987). The status of thematic relations in linguistic theory. *Linguistic Inquiry*, 18, 369–411.
- Kim A, & Osterhout L (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52, 205–225.
- Kintsch W (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163–182. [PubMed: 3375398]

- Kintsch W, & van Dijk TA (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.
- Kos M, Vosse TG, van den Brink D, & Hagoort P (2010). About edible restaurants: Conflicts between syntax and semantics as revealed by ERPs. *Frontiers in Psychology*, 1, 222. [PubMed: 21833277]
- Kim A, & Osterhout L (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52, 205–225.
- Kuperberg GR (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23–49. [PubMed: 17400197]
- Kuperberg GR (2010). Language in schizophrenia part 2: An introduction. *Language and Linguistics Compass*, 4, 576–589. [PubMed: 20936080]
- Kuperberg GR (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, 31, 602–616.
- Kuperberg GR, Brothers T, Wlotko EW (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, 32, 12–35. [PubMed: 31479347]
- Kuperberg GR, Caplan D, Sitnikova T, Eddy M, & Holcomb PJ (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes*, 21, 489–530.
- Kuperberg GR, & Jaeger TF (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31, 32–59.
- Kuperberg GR, Paczynski M, & Ditman T (2011). Establishing causal coherence across Sentences: An ERP study. *Journal of Cognitive Neuroscience*, 23, 1230–1246. [PubMed: 20175676]
- Kuperberg GR, Sitnikova T, Caplan D, & Holcomb PJ (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17, 117–129. [PubMed: 12763198]
- Kutas M (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, 8, 533–572.
- 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, 621–647.
- Kutas M, & Hillyard SA (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205. [PubMed: 7350657]
- Kutas M, & Hillyard SA (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition*, 11, 539–550. [PubMed: 6656613]
- Kutas M, & Hillyard SA (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161. [PubMed: 6690995]
- Lau EF, Holcomb PJ, & Kuperberg GR (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, 2, 484–502.
- Lau E, Namyst A, Fogel A, & Delgado T (2016). A direct comparison of N400 Effects of predictability and incongruity in adjective-noun combination. *Collabra: Psychology*, 2, 13.
- Ledoux K, Camblin CC, Swaab TY, & Gordon PC (2006). Reading words in discourse: The modulation of lexical priming effects by message-level context. *Behavioral and Cognitive Neuroscience Reviews*, 5, 107–127. [PubMed: 16891554]
- Levin B (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press.
- Li X, Shu H, Liu Y, & Li P (2006). Mental representation of verb meaning: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, 18, 1774–1787. [PubMed: 17014380]
- McInnes A, Humphries T, Hogg-Johnson S, & Tannock R (2003). Listening comprehension and working memory are impaired in attention-deficit hyperactivity disorder irrespective of language impairment. *Journal of Abnormal Child Psychology*, 31, 427–443. [PubMed: 12831231]
- Morey RD (2008) Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, 4, 61–64.

- Münte TF, Heinze H-J, Matzke M, Wieringa BM, & Johannes S (1998). Brain potentials and syntactic violations revisited: no evidence for specificity of the syntactic positive shift. *Neuropsychologia*, 36, 217–226. [PubMed: 9622187]
- Münte TF, Matzke M, & Johannes S (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *Journal of Cognitive Neuroscience*, 9, 318–329. [PubMed: 23965010]
- Mumford D (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological Cybernetics*, 66, 241–251. [PubMed: 1540675]
- Ness T, Meltzer-Asscher A (2018). Lexical inhibition due to failed prediction. Behavioral evidence and ERP correlates. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 44, 1269–1285.
- Nieuwenhuis S, Aston-Jones G, & Cohen JD (2005). Decision making, the P3, and the locus coeruleus–norepinephrine system. *Psychological Bulletin*, 131, 510. [PubMed: 16060800]
- Nieuwland MS, & Van Berkum JJA (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24, 691–701. [PubMed: 15894468]
- Osterhout L, & Holcomb PJ (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785–806.
- Osterhout L, & Nicol J (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, 14, 283–317.
- Osterhout L, Kim A, & Kuperberg GR (2012). The neurobiology of sentence comprehension In Spivey M, Joannisse M, McRae K (Eds), *The Cambridge handbook of psycholinguistics* (pp. 365–389). Cambridge: Cambridge University Press.
- Paczynski M, & Kuperberg GR (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb–argument processing. *Language and Cognitive Processes*, 26, 1402–1456. [PubMed: 22199415]
- Paczynski M, & Kuperberg GR (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. [PubMed: 23284226]
- Rao RP, & Ballard DH (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2, 79–87. [PubMed: 10195184]
- Rayner K, & Well AD (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3, 504–509. [PubMed: 24213985]
- Rayner K, Garrod S, & Perfetti CA (1992). Discourse influences during parsing are delayed. *Cognition*, 45, 109–139. [PubMed: 1451412]
- Rugg MD (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology*, 22, 642–647. [PubMed: 4089090]
- Sassenhagen J, Schlesewsky M, & Bornkessel-Schlesewsky I (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, 137, 29–39. [PubMed: 25151545]
- Shetreet E, Alexander EJ, Romoli J, Chierchia G, & Kuperberg G (2019). What we know about knowing: Presuppositions generated by factive verbs influence downstream neural processing. *Cognition*, 184, 96–106. [PubMed: 30584974]
- Swaab TY, Boudewyn MA, Long DL, Luck SJ, Kring AM, Ragland JD, Carter CS (2013). Spared and impaired spoken discourse processing in schizophrenia: Effects of local and global language context. *Journal of Neuroscience*, 33, 15578–15587. [PubMed: 24068824]
- Szewczyk JM, & Schriefers H (2011). Is animacy special?: ERP correlates of semantic violations and animacy violations in sentence processing. *Brain Research*, 1368, 208–221. [PubMed: 21029726]
- Thornhill DE, & Van Petten C (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, 83, 382–392. [PubMed: 22226800]
- Till RE, Mross EF, & Kintsch W (1988). Time course of priming for associate and inference words in a discourse context. *Memory & Cognition*, 16, 283–298. [PubMed: 3210969]

- Urbach TP, & Kutas M (2010). Quantifiers more or less quantify on-line: ERP evidence for partial incremental interpretation. *Journal of Memory and Language*, 63, 158–179. [PubMed: 20640044]
- Van Berkum JJ, Zwitterlood P, Hagoort P, & Brown CM (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, 17, 701–718. [PubMed: 14561457]
- Van Berkum JJ van, Hagoort P, & Brown CM (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11, 657–671. [PubMed: 10601747]
- van de Meerendonk N, Kolk HHJ, Vissers C Th WM, & Chwilla DJ (2008). Monitoring in Language Perception: Mild and Strong Conflicts Elicit Different ERP Patterns. *Journal of Cognitive Neuroscience*, 22, 67–82.
- van de Meerendonk ND, Kolk HHJ, Chwilla DJ, & Vissers CTWM (2009). Monitoring in language perception. *Language and Linguistics Compass*, 3, 1211–1224.
- van den Broek P, Bohn-Gettler CM, Kendeou P, Carlson S, & White MJ (2011). When a reader meets a text: The role of standards of coherence in reading comprehension In *Text relevance and learning from text* (pp. 123–139). Charlotte, NC, US: IAP Information Age Publishing.
- Van Dijk TA, Kintsch W, (1983); *Strategies of Discourse Comprehension*. New York, US: Academic Press.
- Van Petten C, & Luka BJ (2012). Prediction during language comprehension: benefits, costs, and ERP components. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 83, 176–190. [PubMed: 22019481]
- Van Petten C, Weckerly J, McIsaac HK, & Kutas M (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science*, 8, 238–242.
- Wagoner SA (1983). Comprehension monitoring: What it is and what we know about it. *Reading Research Quarterly*, 18, 328–346.
- Xiang M, & Kuperberg G (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, 30, 648–672.
- Yu J. Aa, & Dayan P (2005). Uncertainty, neuromodulation, and attention. *Neuron*, 46, 681–692. [PubMed: 15944135]
- Zirnstein M, van Hell JG, Kroll JF, (2018). Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition*, 176, 87–106. [PubMed: 29549762]
- Zwaan RA, & Radvansky GA (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162–185. [PubMed: 9522683]

Experiment 1, Short Sentences

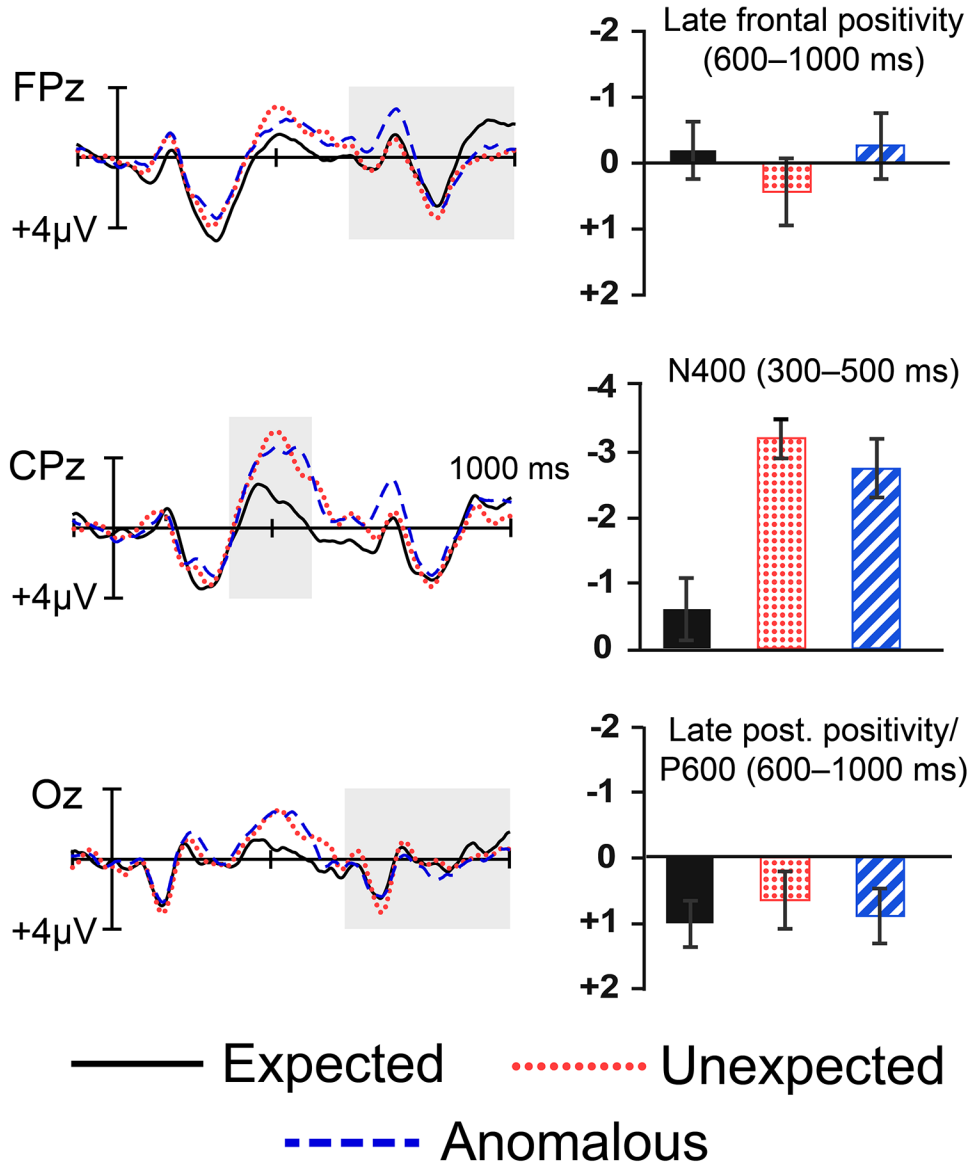


Figure 1. Grand-average ERPs for the three conditions, plotted at three midline electrode sites. In this and subsequent figures, negative is plotted up, and all waveforms were low-pass filtered at 15Hz for presentation purposes. Bar graphs to the right show average voltages within each spatiotemporal region of interest (see text) with ± 1 SEM error bars, calculated within-subjects (Morey, 2008).

Experiment 2a Locally Constraining Discourse

Experiment 2b Globally Constraining Discourse

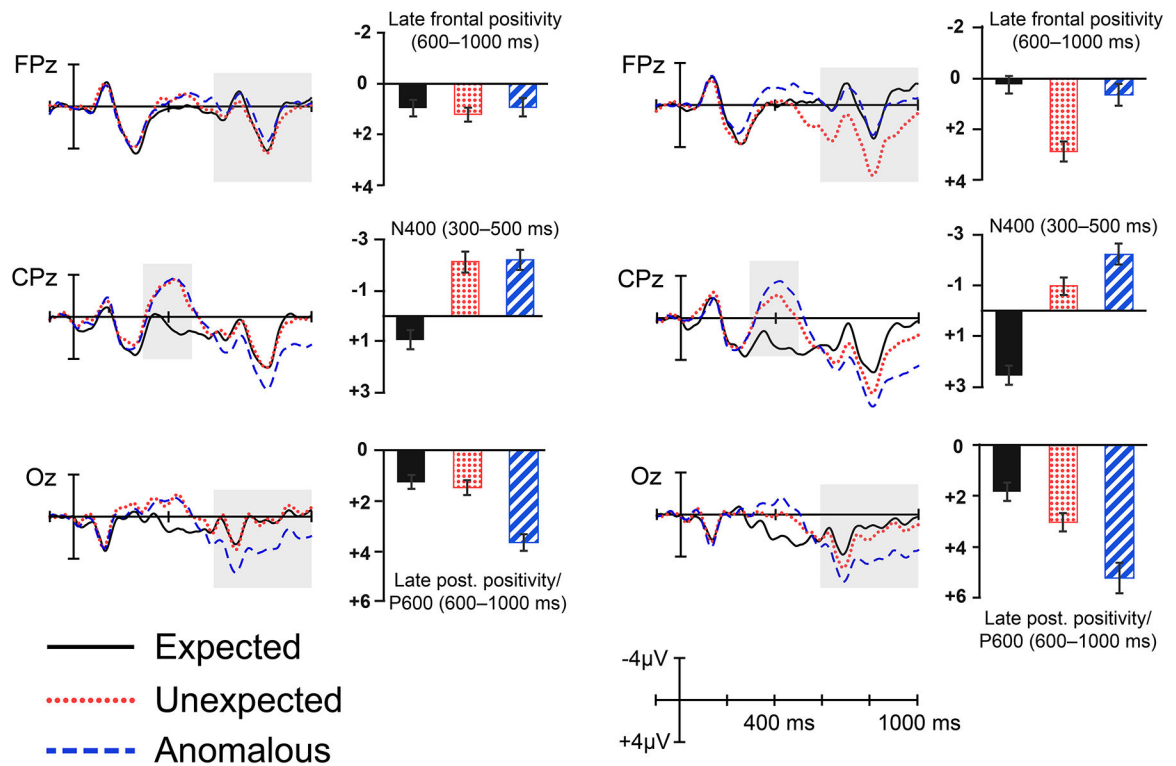


Figure 2. Grand-average ERPs in Experiment 2a and 2b, plotted at three midline electrode sites. The bar graphs show average voltages within each spatiotemporal region of interest (see text) with ± 1 SEM error bars, calculated within-subjects.

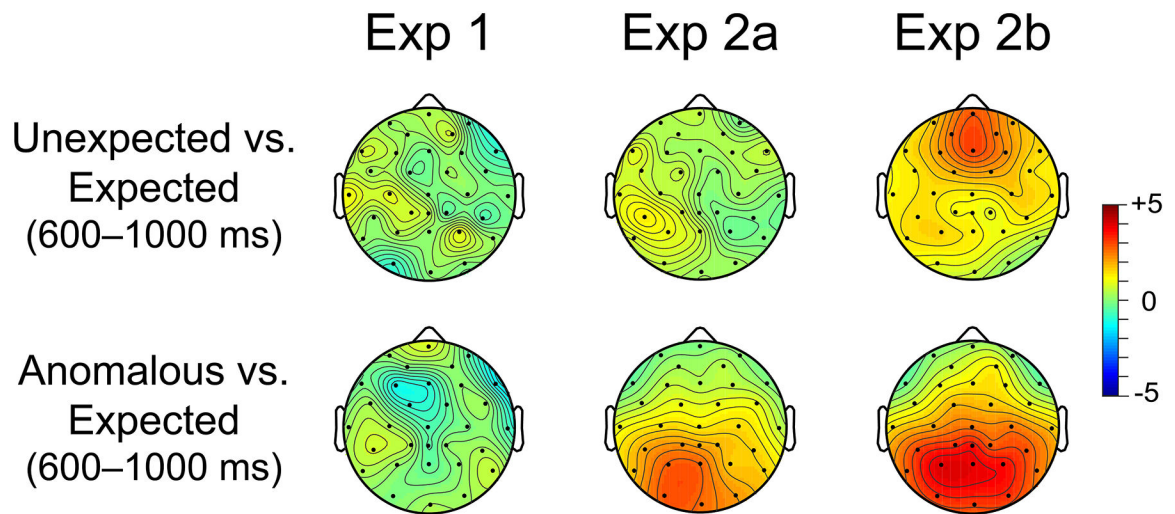


Figure 3. Topographic difference maps demonstrating the magnitude (μV) and scalp distribution of late ERP effects (600–1000ms) across the three experiments.

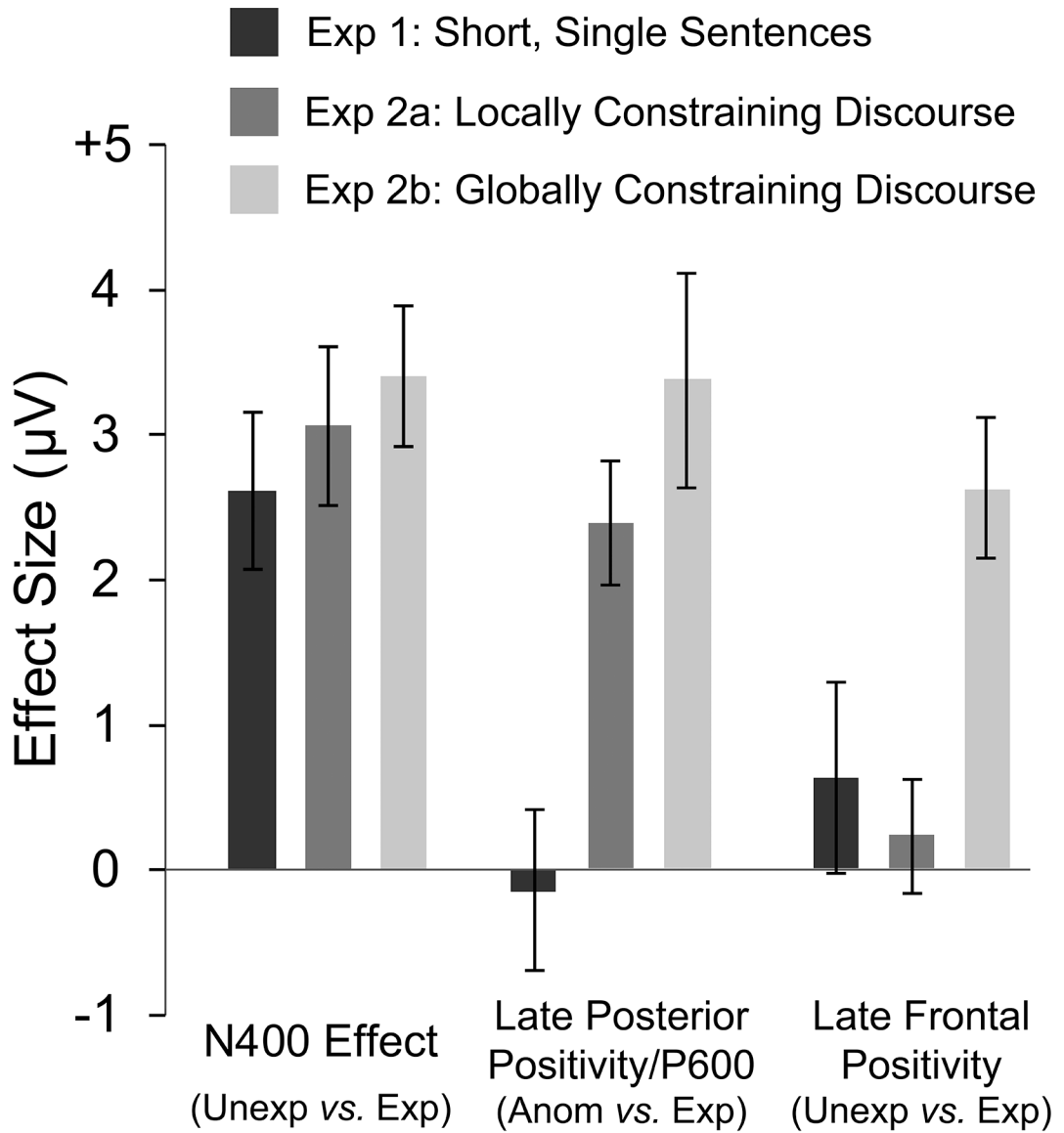
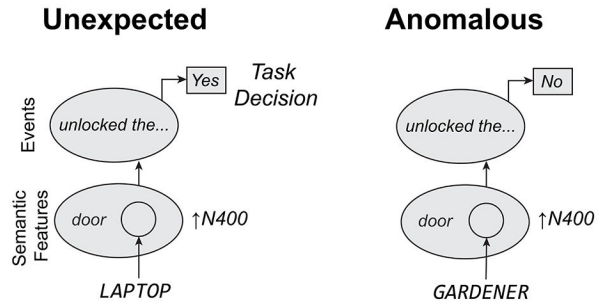


Figure 4. Mean amplitude of the N400 effect (Unexpected *minus* Expected), the *late posterior positivity/P600* effect (Anomalous *minus* Expected), and the *late frontal positivity* effect (Unexpected *minus* Expected) across the three experiments. Error bars represent ± 1 SEM.

**Experiment 1:
Short Sentence
Contexts**

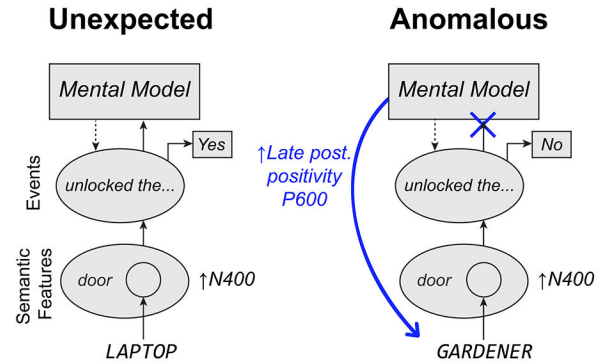
James unlocked the...
door (expected)
laptop (unexpected)
gardener (anomalous)



**Experiment 2a:
Locally Constraining
Discourse**

He was thinking about what
needed to be done on his
way home. He finally arrived.

James unlocked the...
(door/laptop/gardener)



**Experiment 2b:
Globally Constraining
Discourse**

Tim really enjoyed baking
apple pie for his family. He
had just finished mixing the
ingredients for the crust.

To proceed, he flattened the...
(dough/foil/onlookers)

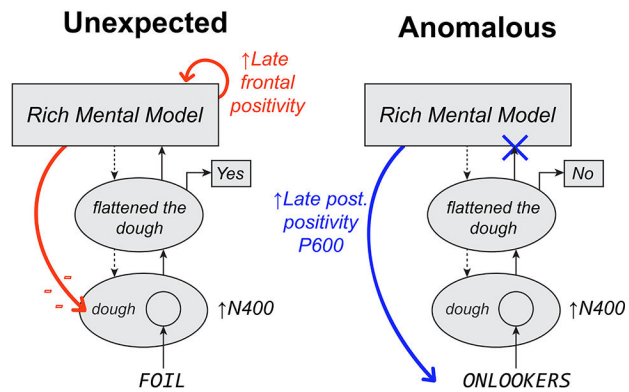


Figure 5. A schematic diagram illustrating differences in the structure and state of reader’s hierarchical generative networks across experiments and conditions. Dotted arrows represent anticipatory information that has been passed down through the hierarchy prior to word onset, while solid, upward arrows represent bottom-up information being passed from the lexical input through higher levels of the hierarchy. Increased N400 responses are thought to arise at the semantic feature level, as new unpredicted semantic information is accessed/ decoded. The two late positivities are thought to arise when new information leads to large updates in the prior situation model (*late frontal positivity*, red), or when information conflicts with the constraints of the communication model, triggering re-analysis/repair (*late posterior positivity/P600*, blue).