A Tale of Two Positivities (and the N400): Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation

Gina R. Kuperberg [1,2] and Edward W. Wlotko[1, 3] ,

[1]Department of Psychology, Tufts University, Medford, MA USA

[2]Department of Psychiatry and the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, MA USA

[3]Moss Rehabilitation Research Institute, Elkins Park, PA USA

Corresponding Author:

Gina R. Kuperberg MD PhD

Department of Psychology

Tufts University

490 Boston Avenue

Medford, MA 02155

Tel: 617-627-4959

e-mail: gkuperberg@mgh.harvard.edu

Word count: 8,871

Keywords: comprehension, ERP, late positivity, predictive coding, P600

Abstract

Hierarchical prediction has been proposed as a fundamental computational principle underlying neurocognitive processing. Here we ask whether the brain engages distinct neurocognitive mechanisms in response to inputs that fulfill versus violate strong predictions at different levels of representation during language comprehension. Participants read three-sentence scenarios in which a verb in the third sentence constrained for a broad event structure and the animacy of an upcoming noun ("they cautioned the…"). High Constraint (HC) contexts additionally constrained for a specific event and lexical item (e.g. following a two-sentence context about a beach, lifeguards and sharks, "they cautioned the…"). Low Constraint (LC) contexts did not constrain for any specific event/lexical item. We measured ERPs on critical nouns that fulfilled and/or violated each of these predictions. We found clear dissociable effects to fulfilled semantic predictions (a reduced N400), to event/lexical prediction violations (an increased anteriorly-distributed late positivity), and to event structure/animacy prediction violations (an increased posteriorly-distributed late positivity/semantic P600). We argue that the anterior late positivity reflected the successful updating of the comprehender's existing mental model with new unpredicted input, which necessarily entailed the suppression of the incorrect lexical prediction. We suggest that the posterior late positivity/P600 reflected a failure to incorporate new unpredicted input into the comprehender's current mental model, and possibly a reanalysis of the previous inputs in attempts to revise or repair this model. Together, these findings provide strong evidence that confirmed and violated predictions at different levels of representation manifest as distinct spatiotemporal neural signatures.

Introduction

The goal of language comprehension is to extract the communicator's intended message. This is challenging. Linguistic inputs unfold quickly, they are often ambiguous, and our communicative environments are noisy. It therefore helps if we can use context to mobilize our stored linguistic (and non-linguistic) knowledge to predict[1] upcoming information. If incoming information matches these predictions, its processing will be facilitated. On the other hand, there will be times when we encounter inputs that violate strong predictions. In the present study, we use event-related potentials (ERPs) to show that the brain engages distinct neurocognitive mechanisms in response to inputs that fulfill and that violate strong predictions at different levels of representation.

ERPs have provided some of the strongest evidence that brain is sensitive to predictive processes during language comprehension. The component that is primarily sensitive to *fulfilled* semantic predictions is the N400 — a centro-parietally distributed negative-going waveform that is largest between 300-500ms after the onset of an incoming word. The N400 is highly sensitive to the probability of encountering a word's semantic features given the preceding context (Kutas & Hillyard, 1984; Kutas & Federmeier, 2011; Kuperberg, 2016).[2] If these features match features that have already been predictively pre-activated by a highly constraining context (e.g. as in "birthday" following context (1)), then this word will evoke a smaller (less negative) N400 than a word appearing in a low constraint context (e.g. "collection" following context (3)). Importantly, however, the N400 is not a direct index of prediction violation: its amplitude is just as large to "collection" following a low constraint context (as in (3)) as to "collection" when it violates a high

[1]The term *prediction* has been used in many different ways in the language processing literature (Kuperberg & Jaeger, 2016). In the present study, we use it in its minimal sense — to imply that a prior context has changed the state of the language processing system before new bottom-up input is observed. We use the term *predictive pre-activation* to refer more specifically to the top-down flow of information from a higher to a lower level of representation that pre-activates information at the lower level before new bottom-up information reaches this lower level (see Kuperberg & Jaeger, 2016, Section 3).

[2]We distinguish the N400 (between 300-500ms) from earlier ERP components that may be more sensitive to the prediction of phonological/orthographic properties of incoming words. For example, the N250 component peaks between 200-300ms after word onset and has a wider (more frontal) scalp distribution than on the N400 (Lau, Holcomb & Kuperberg, 2013; Brothers, Swaab, & Traxler, 2015; see Grainger & Holcomb, 2009 for its characterization using masked priming paradigms). The N250 may be the same component as that previously characterized as the 'phonological mismatch negativity' (Connolly & Phillips, 1994) or the N200 (van den Brink, Brown, & Hagoort, 2001), which are seen to inputs that mismatch expected phonemes during spoken language comprehension. It may also be related to a component as that has previously been characterized as the P2 and linked to the top-down attentionally-mediated extraction of visual features during sentence processing (Federmeier, Mai, & Kutas, 2005, see also Paczynski & Kuperberg, 2012).

constraint context (as in (2)), see Kutas & Hillyard, 1984; Federmeier, Wlotko, De Ochoa-Dewald & Kutas, 2007.

(1) He bought her a pearl necklace for her birthday.

(2) He bought her a pearl necklace for her collection.

(3) He looked worried because he might have broken his collection.

In contrast to the N400, another set of ERP components, which we refer to as the *Late Positive Components* (LPCs) do appear to be differentially sensitive to words that violate strong predictions (when accompanied by an N400 effect, they have also been referred to as Post-N400 Positivities, PNPs, see Van Petten & Luka, 2012). The LPCs have a positive-going polarity and are visible at the scalp surface after the N400 time window between approximately 500-1000ms, although they are likely to reflect neurocognitive processes that begin earlier within the N400 time window. At least two types of LPCs have been distinguished (Van Petten & Luka, 2012; Kuperberg, 2013; DeLong, Quante and Kutas, 2014). First, LPCs with an *anterior* scalp distribution (maximal at frontal sites) can be evoked by words that violate strong predictions about a specific event/lexical item, but that still yield plausible interpretations (such as in sentence (2) above) (e.g. Federmeier et al., 2007; Van Petten & Luka, 2012; see also Kutas, 1993). We refer to these as *ant-LPCs*. Second, LPCs with a *posterior* scalp distribution (maximal at parietal and occipital sites), otherwise known as 'semantic P600s', can be evoked by words that violate strong contextual constraints for alternative interpretations but that are perceived as highly implausible/impossible (e.g. "Every morning at breakfast the eggs would *eat…", e.g. Kuperberg, Sitnikova, Caplan & Holcomb, 2003; for a review, see Kuperberg, 2007). We refer to these as *post-LPCs/P600s*.[3]

[3] Posteriorly distributed LPCs have gone by many names in the language processing ERP literature. As discussed by Van Petten & Luka (2012), in the earlier literature, despite frequently being observed, along with the N400 effect, on implausible words, they were either attributed to non-linguistically-relevant processes or not discussed at all. In the early 1990s, several groups reported that posteriorly distributed late positivities were evoked by syntactically anomalous or dispreferred words (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992; 1993). This syntactically sensitive component was referred to as the P600. When it later became apparent that certain types of semantic violations also evoked a posteriorly distributed late positivity, sometimes in the absence of an N400 effect (e.g. Hoeks, Stowe and Doedens, 2004; Kim & Osterhout, 2005; Kolk, Chwilla, van Herten, & Oor, 2003; Kuperberg, Sitnikova, Caplan & Holcomb, 2003), the component was descriptively termed a 'semantic P600' to distinguish it from the 'syntactic P600' (see Kuperberg, 2007 for a review). In the present study, we use term post-

Similar mechanisms have been proposed to underlie the ant-LPC and post-LPC/P600 effects. These include the detection of conflict between the predicted and bottom-up input (ant-LPC: DeLong, Urbach, Groppe & Kutas, 2011; post-LPC/P600: van de Meerendonk, Kolk, Chwilla & Vissers, 2009; Kuperberg, 2007), the selection of the bottom-up input by enhancing its activation and suppressing/overriding incorrectly predicted information (ant-LPC: Federmeier et al., 2007; Kutas, 1993; post-LPC/P600: van de Meerendonk et al., 2009), and attempts to integrate the violating bottom-up input to reach a new higher-level interpretation (ant-LPC: Federmeier, Kutas & Schul, 2010; DeLong, Quante & Kutas, 2014; post-LPC/P600: "continued combinatorial processing", see Kuperberg, 2007). As discussed by Van Petten & Luka (2012), the main factor that appears to distinguish the ant-LPC and the post-LPC/P600 is the plausibility of the resulting interpretation, with the ant-LPC produced by words that violate expectations but that still result in a plausible interpretation, and the post-LPC/P600 produced by words that violate expectations but that are highly implausible. This raises important questions. For each of these effects, precisely what representations are predicted and violated by the bottom-up input? To trigger each effect, at what level of representation is conflict between the predicted information and bottom-up input detected? What exactly is meant by 'integration' and how is this linked to whether we perceive an incoming word as being plausible or highly implausible/impossible in relation to its context?

One reason why these questions remain unanswered is that the ERP research characterizing the ant-LPC and the post-LPC/P600 has proceeded quite independently from different research groups working within different theoretical frameworks. The studies characterizing each of these effects have generally used different sets of stimuli with different syntactic structures, different positions of critical words and different tasks – all factors that are known to modulate the N400 and late positivity components. This makes it difficult to compare across studies (but see DeLong, Quante and Kutas, 2014 for a dissociation between the ant-LPC and post-LPC on the basis of plausibility in the same study). The primary goal of the

---

LPC/P600 to make it clear that we are referring to the same component. Our discussion is primarily concerned with the post-LPC/P600 that is evoked by semantic violations (the 'semantic P600'). While some of the ideas we propose are relevant to understanding the syntactic P600 (as well as posteriorly distributed late positivities that are evoked by other types of linguistic and non-linguistic violations), a full discussion of these relationships is outside the scope of this paper.

present study was to bring these literatures together under a common theoretical umbrella. We aimed to dissociate the ant-LPC and post-LPC/P600 from one another, and from the N400, by manipulating predictions that are fulfilled and violated at different levels of representation in a single experiment.

In explaining the logic of our design, we appeal to two *hierarchical levels of representation* – a higher *message level*, which encodes the current message that the comprehender believes is being conveyed, and a lower *lexico-semantic level*, which encodes the semantic features and properties associated with an incoming word. We assume that the comprehender draws upon information at both these representational levels to incrementally build a *mental model*, which is based on her knowledge and experiences about the world, her linguistic knowledge, her beliefs about the communicator, her communicative goals, and the linguistic and non-linguistic context that she has thus far encountered. We also assume that, at both the message and lexico-semantic levels, information can be encoded at different *grains*. At the higher message level, comprehenders can represent coarser-grained *event structures* and finer-grained specific *events*. By *event structures*, we refer to *sets* of events that the comprehender believes are possible/compatible with her current mental model (e.g. <Animate Agent *acts on* Inanimate Patient>). By *events*, we refer to finer-grained representations of individual specific events, states and facts that are compatible with a given event structure (e.g. the event, <lifeguards warn swimmers>). At the lexico-semantic level, comprehenders can encode information at coarser grains (e.g. sets of semantic properties that together define particular semantic categories such as <animate> and <inanimate>), and at finer grains (e.g. the semantic features that are linked to a single lexical item (e.g. <swimmers>).[4]

We propose that a primary *trigger* of both the ant-LPC and the post-LPC/P600 effects is the arrival of new bottom-up information at the message level that strongly violates/conflicts with information that has already been strongly predicted at this level and incorporated into the comprehender's mental model. We suggest that the *post-LPC/P600* is triggered by an input that violates/conflicts with a strongly predicted

[4] Throughout this manuscript, we will use <> around words when we refer to their conceptual features. For example, <animate> is used to denote a set of conceptual features that are typically associated with animate entities (e.g. <sentient>, <able to move> etc), while <swimmers> is used to denote the set of conceptual features that are typically associated with the specific lexical item, *swimmers* (e.g. <in water>, <staying afloat> etc).

event structure. Because this event structure constrains the boundaries of the comprehender's current mental model, this new input cannot be incorporated into this existing model, leading comprehenders to perceive it as being highly implausible/impossible (see Xiang & Kuperberg, 2015; Shetreet, Alexander, Romoli, Chierchia & Kuperberg, under review, for recent discussions). In contrast, we suggest that the *ant-LPC* is triggered by an input that violates a specific predicted event, but not a predicted event structure. This input *can* therefore be used to update the existing mental model, and so comprehenders perceive it as plausible.

Design of the current study

As a step towards testing these hypotheses, we created a set of well-controlled three-sentence discourse scenarios, with either Low Constraint (LC) or High Constraint (HC) contexts. Both LC and HC contexts included a verb in the third sentence whose thematic-semantic structure constrained strongly for a coarse event structure at the higher message level (either <Animate Agent *acts on* Animate Patient> or <Animate Agent *acts on* Inanimate Theme>), and whose animacy-based selection restrictions constrained for coarse-grained animacy features (either <animate> or <inanimate>) at the lexico-semantic level. LC contexts, like (1), had impoverished discourse contexts and so they did not additionally constrain for any specific upcoming event/lexical item. In contrast, HC contexts, like (2), had rich discourse contexts that constrained strongly for a specific event (e.g. <cautioned the swimmers>), as well as for fine-grained semantic features associated with a specific lexico-semantic item (e.g. <swimmers>). These constraints are depicted schematically in Figure 1.

(1) *Low Constraint (LC) context:* Eric and Grant received the news late in the day. They mulled over the information, and decided it was better to act sooner rather than later. Hence, they cautioned the…(trainees/drawer).

(2) *High Constraint (HC) context:* 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).

Critical words in the third sentence either confirmed or violated each of these constraints. This yielded five types of discourse scenarios, see Table 1 for examples. In the High Constraint Predicted scenarios (*HC_Pred*: Condition 1 in Table 1), the critical word ("swimmers") was highly predictable because it satisfied both coarse-grained and fine-grained constraints at both the message-level and the lexico-semantic level. In all four other conditions, critical words were unpredictable, but each for a different reason. In the Low Constraint Unpredicted scenarios (*LC_Unpred*: Condition 2 in Table 1), the critical word was unexpected, but it did not violate event structure/animacy constraints (because it confirmed the verb's thematic structure/animacy selection restrictions), and it did not violate constraints for a specific event/lexical item (because the discourse context did not constrain for any single event/lexical item). In the Low Constraint Selection Restriction violation scenarios (*LC_SRviol*: Condition 3 in Table 1), the critical word violated event structure/animacy constraints (because it violated the verb's thematic structure/animacy selection restrictions), but it did not violate constraints for a specific event/lexical item (because the discourse context did not constrain strongly for any single event/lexical item). In the High Constraint Lexical violation scenarios (*HC_Lexviol*: Condition 4 in Table 1), the critical word ("trainees") violated constraints for a specific event (<cautioned the swimmers>) and a specific lexico-semantic item (<swimmers>), but it did not did not violate event structure/animacy constraints (because it confirmed the verb's thematic structure/animacy selection restrictions). Finally, in the High Constraint Selection Restriction violation scenarios (*HC_SRviol*: Condition 5, Table 1), the critical word violated *both* event structure/animacy constraints and specific event/lexical constraints. Critical words in the four unpredictable were matched on their general semantic relatedness with preceding content words in their contexts (as assessed using Latent Semantic Analysis, LSA: Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998). We measured ERPs on critical words in all five conditions.

Based on numerous previous studies, we expected that the amplitude of the N400 would be smallest on critical words in the HC_Pred scenarios in which all predictions were confirmed, and larger in the four unpredictable conditions. We hypothesized that the N400 would not be modulated by contextual constraint,

over and above lexico-semantic predictability (as shown by Federmeier et al., 2007 with other stimuli). Finally, we were interested in whether the N400 would be larger on critical words that mismatched the verb's animacy selection restrictions than on critical words that matched these restrictions. This would provide evidence that the N400 can be sensitive to whether an incoming word matches or mismatches coarse-grained semantic (animacy) features, even when its cloze probability and semantic relatedness with previous content words are fully matched across conditions (see Paczynski & Kuperberg, 2011, 2012).

Our main set of questions concerned the late positivities. We hypothesized that we would be able to dissociate the scalp distribution of the LPC evoked by words that violated strong event structure/animacy predictions from the LPC evoked by words that violated strong event/lexical predictions in same set of participants carrying out the same task. Specifically, we hypothesized that the HC_Lexviol critical words, which violated strong predictions for a specific event/lexical item, would selectively evoke an ant-LPC (larger than the ant-LPC evoked by critical words in the HC_Pred, LC_Unpred, or LC_SRviol conditions), while the LC_SRviol critical words, which violated strong predictions for an event structure/animacy features, would selectively evoke a post-LPC/P600 (larger than the post-LPC/P600 evoked by critical words in the HC_Pred, LC_Unpred or HC_Lexviol conditions). This would provide strong evidence that distinct neurocognitive mechanisms can be triggered by violations at these different grains of representation.

In addition, our design allowed us to ask not only whether the neurocognitive mechanisms underlying the ant-LPC and the post-LCP are distinct, but also whether they are independent of one another. The key condition to address this question was the doubly violated scenarios (HC_SRviol) in which the critical words violated *both* constraints for a specific event/lexical item as well as for a broader event structure/animacy features. If event/lexico-semantic and event structure/animacy violations elicit completely independent effects, then the doubly violated HC_SRviol critical words should evoke *both* an ant-LPC and a post-LPC/P600 (assuming additivity of independent ERP effects, see Osterhout & Nicol, 1999). If, on the other hand, the neurocognitive processes underlying the ant-LPC and the post-LPC/P600

are interdependent, then we should see non-additive effects. For example, if the detection of an event structure/animacy violation overrides processes reflected by the ant-LPC, then we should see a post-LPC/P600 but no ant-LPC on the double violations. Such a finding would, in turn, constrain our interpretation of the precise mechanisms underlying each of these effects (see Discussion)

Methods

Development and norming of materials

Participants read five types of scenarios (see Table 1). In all scenarios, the first two sentences introduced a context. The third sentence began with an adjunct phrase of 1-4 words, followed by a pronominal subject that referred back to the first two sentences, followed by a verb, a determiner, a *critical word* (always a direct object noun), and then three additional words. The scenarios varied by the constraint of the context (i.e., the combination of the first two sentences and the third sentence until just before the critical word) and by whether the critical word matched or violated the event/lexico-semantic constraints of the HC contexts and/or the event structure/animacy selection restrictions of the previous verb, see Table 1 and Introduction for examples and a description of each condition. To construct these scenarios, we began with a set of 100 preferentially transitive verbs (50% selecting for animate direct objects; 50% for inanimate direct objects), which, without any discourse context, were not highly predictive of any particular upcoming event or direct object. These verbs were selected and cloze normed in isolation of any context, as described below. We then wrote high and low constraint two-sentence discourse contexts each verb (matching the number of words across these contexts) and, based on a second cloze norming study of these discourse contexts, we created the five conditions as described in the legend to Table 1.

In both cloze norming studies, participants were recruited through Amazon Mechanical Turk (AMT). They were asked to complete each context with the first word that came to mind (Taylor, 1953), and in an extension of the standard cloze procedure, to then provide two additional words that could complete the sentence (see Federmeier et al, 2007; Schwanenflugel & LaCount, 1988). Responses were excluded if the first language learned was anything other than English or if there were any self-reported psychiatric or

neurological disorders. In addition, "catch" questions were used to identify and omit bots, and participants were manually excluded if their responses indicated a failure to comply with instructions.

Cloze norming study 1: Selection of non-constraining verbs. We started with a set of 600 transitively-biased verbs, compiled from various sources including Levin (1993) and materials from previous studies conducted in our laboratory (Paczynski & Kuperberg, 2011, 2012). Verbs with log Hyperspace Analogue to Language (HAL) frequency (Lund & Burgess, 1996) of two standard deviations below the mean (based on English Lexicon Project database, Balota et al. 2007) were excluded. For each verb, we constructed a simple active, past tense sentence stem that consisted of only a proper name, the verb, and a determiner (e.g., "Harry explored the…"). These sentences were divided into six lists in order to decrease the time demands on any individual participant during cloze norming. Between 89 and 106 participants (depending on list) who met inclusionary criteria provided completions for each verb. All 100 verbs included in the final stimuli had cloze values of less than 22% and the average constraint was 16.4% (SD=0.73%).

Cloze norming study 2: Selection and characterization of the final set of discourse stimuli. We began with an initial set of 198 pairs of HC and LC contexts (the combination of the first two sentences and the third sentence, including the verb and the determiner). These were pseudorandomly divided into two lists, such that each list contained only one of the two contexts associated with each verb. The two lists were then divided into thirds to decrease time demands on any individual participant during cloze norming. Between 51 and 69 participants who met inclusionary criteria provided completions for each scenario. The resulting values were used to create the five types of discourse scenarios as described in the legend to Table 1.

The 500 scenarios (100 sets of five) were rotated across the five lists, with 20 items per condition in each list. Counterbalancing worked such that (a) the same adjunct phrase plus verbs in the sentences appeared in all conditions across lists, (b) the same critical words were counterbalanced across all of the unpredictable scenarios (HC_Lexviol, HC_SRviol, LC_Unpred, and LC_SRviol scenarios), (c) the same HC contexts were counterbalanced across the HC_Pred, HC_Lexviol and HC_SRviol scenarios, and (d) the same LC contexts were counterbalanced across the LC_Unpred and LC_SRviol scenarios. To each list,

we then added an additional 20 HC_SRviol, 20 LC_Unpred, and 20 LC_SRviol scenarios so that each participant saw 160 scenarios in total. This ensured that each participant saw an equal number of high and low constraint contexts, and that, at each level of contextual constraint (HC versus LC), half of the scenarios were plausible, and the other half were highly implausible (violating selection restrictions).

Participants

We report data from thirty-nine native English speakers (age 18-32 years, mean=21.6, SD=3.6; 21 male). Forty participants were originally recruited but one failed to complete the session. All participants were recruited from Tufts University and the surrounding communities. Participants were screened on the basis of the following exclusionary criteria: significant exposure to any language other than English before age 5; history of psychiatric or neurological diagnoses or injury; use of psychoactive medication within the preceding 6 months. Participants were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). 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.

Stimulus presentation and Task

Participants sat in a dimly lit room while stimuli were presented approximately 150cm from the computer screen (see Figure 2 and legend for details). Their task was to press either one of two buttons after seeing a ‘‘?’’ cue to indicate whether they judged each scenario be “make sense” or not. Following approximately 32 trials (all fillers), at random, they also answered a yes/no comprehension question about the preceding scenario (to ensure that they comprehended the scenarios as a whole rather than just the third sentence). The experiment was divided into 4 blocks, with block order randomized in each participant. Participants were given 10 practice trials.

EEG recording and processing

EEG was recorded from 32 active electrodes in a modified 10/20 system montage. Signals were digitized with the Biosemi Active-Two acquisition system at 512 Hz and a passband of DC - 104 Hz. Data were processed using the EEGLab (Delorme & Makeig, 2004) and ERPLab (Lopez-Calderon & Luck,

2014) analysis packages in the MATLAB environment. EEG channels were referenced offline to the average of the left and right mastoid channels. A 0.05 Hz highpass filter was applied offline. EEG was segmented into epochs spanning 300ms before to 1400ms after the presentation of critical words. Independent Components Analysis (using the extended infomax algorithm implemented in EEGLab) was then applied to the epoched data. Components reflecting EOG artifacts were removed from the dataset. Further artifactual contamination was detected using a semi-automated procedure employing ERPLab software algorithms, with detection thresholds set individually for each participant. Trials were removed from further analysis (6.9% on average) when artifacts were identified in a window of -100 to 1200ms around the onset of the critical word. Remaining trials were averaged using a -100-0ms baseline.

ERP statistical analysis

We first carried out two omnibus repeated measures ANOVA: one within the N400 time window (300-500ms) and one within the late positivity time window (600-1000ms) to determine whether there were overall differences among the five scenarios types across the anterior posterior distribution of the scalp. This omnibus ANOVA included five levels of Scenario Type and four levels of Region (Regions are shown in Figure 3). All follow-up analyses focused on testing our *a priori* hypotheses through planned pair-wise comparisons between ERPs evoked by critical words in the different scenario types within spatiotemporal regions of interest. The N400 was operationalized as the average voltage between 300-500ms within the central region; the ant-LPC was operationalized as the average voltage between 600-1000ms within the prefrontal region, and the post-LPC/P600 was operationalized as the average voltage between 600-1000ms within the posterior region. For the N400 component, we also carried out a 2 x 2 ANOVA on the unpredicted critical words that crossed two levels of Contextual constraint (HC, LC) and two levels of Selection Restriction confirmation (SR confirmed, SR violated).

Results

Behavioral findings

Overall, participants' acceptability judgements corresponded to the experimental conditions on 89.6%

of trials on average (SD=0.05%, range 76.2% -96.9%). Participants were most accurate in responding "YES" to HC_Pred items (97.6%), followed by responding "NO" to HC_SRviol (91.0%) and LC_SRviol (89.8%) items. Participants were least accurate at responding "YES" to HC_Lexviol and LC_Unpred items (both 86.3%). Participants correctly answered 86.4% of the comprehension questions on average (27.6 out of 32 questions; SD=2.0; range 23 - 32 or 71.9%-100%), indicating that they attended to the context information in the scenarios, rather than just to the final sentences.

ERP findings

Omnibus ANOVAs revealed main effects of Scenario Type and Scenario Type by Region interactions in both the 300-500ms and the 600-1000ms windows (see Table 2). We therefore proceeded straight to planned analyses within our spatiotemporal regions of interest.

N400: 300-500ms. Within the central region of interest, the amplitude of the N400 evoked by highly predictable critical words in HC contexts (HC_Pred) was smaller (less negative) than that evoked by unexpected critical words in each of the four other scenario types (see Table 3 and Figure 4). A 2 (Contextual constraint: HC, LC) x 2 (Selection Restriction confirmation: SR confirmed, SR violated) ANOVA on the N400 evoked by the critical words in the four unpredictable scenario types revealed a main effect of Selection Restriction confirmation [$F(1,38)=18.52$, $p=0.0001$] due to a larger N400 on critical words that violated the animacy-based selection restrictions of their preceding verb compared to those that matched these restrictions. However, there was no main effect of Contextual constraint [$F(1,38)=0.32$, $p=0.5738$] or interaction between the two factors [$F(1,38)=0.27$, $p=0.608$].

Anterior Late Positivity (ant-LPC): 600-1000ms. Within the prefrontal region of interest, the ant-LPC was larger (more positive) to critical words that violated strong event/lexical predictions in high constraint contexts (HC_Lexviol) than to critical words in all other scenario types (see Table 4, Figure 4).

Posterior Late Positivity (Post-LPC/P600): 600-1000ms. Within the posterior region of interest, the post-LPC/P600 was larger (more positive) to critical words that violated the animacy selection restrictions of their preceding verbs (in the HC_SRviol and the LC_SRviol scenarios) than critical words in the three

other sentence types (see Table 4, Figure 4). None of these contrasts produced an ant-LPC effect. In addition, Contextual constraint impacted the amplitude of the post-LPC/P600 evoked by the animacy SR violations: its amplitude was larger in HC than the LC contexts (HC_SRviol > LC_SRviol).

Discussion

We used a single set of well-controlled stimuli to vary predictions at different levels of representation within a single experiment. We found that each manipulation was associated with the modulation of a distinct ERP component. Confirmed predictions of lexico-semantic features were reflected by an attenuation of the N400; violated event/lexical predictions were associated with an enhanced anteriorly distributed late positivity (the ant-LPC), while violated event structure/animacy predictions elicited an enhanced posteriorly distributed late positivity (the post-LPC), also known as the semantic P600.

The N400: Confirmed predictions at the level of semantic features

As expected, the N400 was significantly smaller on predictable critical words in the HC_Pred scenarios than on unpredictable words in each of the four other conditions. Importantly, however, the N400 evoked by unpredicted words was not sensitive to the lexical constraint of the preceding context over and above cloze probability: N400 amplitudes on critical words in the HC_Lexviol and LC_Unpred scenarios were indistinguishable. This replicates Federmeier et al. (2007) in a new stimulus set (three-sentence discourse scenarios as opposed to single sentences). We further show that the insensitivity of the N400 to contextual constraint extended to selection restriction violations: N400 amplitudes on critical words in the HC_SRviol and LC_SRviol were also indistinguishable. Taken together, these findings support the hypothesis that the N400 primarily reflects the degree of match between semantic features that were pre-activated by the context and semantic features associated with an incoming word, rather than the detection or recovery from violations of strong lexico-semantic predictions.

In addition, critical words that violated the animacy selection restrictions of their preceding verbs elicited a slightly larger N400 amplitude than critical words that were consistent with these animacy constraints (i.e., the N400 pattern was HC_SRviol = LC_SRviol > HC_Lexviol = LC_Unpred), even

though these four conditions were matched on lexical predictability (cloze) and on overall semantic relatedness with their preceding contexts (as operationalized using LSA). We suggest that this larger N400 amplitude reflected the mismatch between animacy features that were pre-activated by the verb and the animacy features associated with the critical nouns.[5] This highlights the fact that the N400 can be modulated even on low cloze lexically unpredictable words if such words differ in the predictability of their semantic features (see also Federmeier & Kutas, 1999; Paczynski & Kuperberg, 2011, Wlotko & Federmeier, 2012).

The Late Positivity Components (LPCs): Violations of strong predictions

In contrast to the N400, the late positivity components were selectively modulated by *violations* of strong predictions. While many previous studies have reported "post-N400" positivity effects on contextually unexpected or highly implausible/anomalous words, the eliciting conditions and functional significance of these effects have remained unclear. One proposal is that the anterior-posterior scalp distribution of the LPC dissociates on the basis of plausibility (DeLong, Quante & Kutas, 2014; Kuperberg, 2013; Van Petten & Luka, 2012). Our results are consistent with this proposal. We extend it by specifying the representations that must be violated for the comprehender to perceive a word as being highly implausible/impossible. Specifically, we argue that if cues within the context lead the comprehender to constrain/restrict her current mental model to a particular set of possible/compatible events (an event structure), and an incoming word violates this predicted event structure, then it cannot be incorporated into the existing mental model. We suggest that the conflict between the input and the predicted event structure at the message level is what triggers a post-LPC/P600 (as in the LC_SRviol and HC_SRviol scenarios), and what leads the comprehender to classify the resulting representation as impossible/highly implausible[6]. In

[5] Similar to previous findings (Paczynski & Kuperberg, 2011), the degree of N400 modulation depended on whether the verb constrained for an animate or inanimate direct object, and on the animacy of the argument. A detailed discussion of these interactions is beyond the scope of the current paper.

[6]The post-LPC/P600 is most commonly evoked by words that are judged offline to result in an *impossible*, rather than an implausible-but-possible interpretation (see Kuperberg, 2007, section 3.4 page 32 for discussion). For example, the word "plant" in the sentence, "Every morning at breakfast the eggs would plant…" violates animacy-based selection restrictions, produces a post-LPC/P600, and is judged to be highly implausible/impossible. In contrast, the word "plant" in the sentence, "Every morning at breakfast the boys would plant…" does not violate selection restrictions, does not evoke a post-LPC/P600 and is judged to be

contrast, if a new input to the message level violates a strong prediction of a specific event, but does not violate an event structure, then the new event can be incorporated into the existing mental model and it will be perceived as plausible, triggering an ant-LPC effect, as in the plausible HC_Lexviol scenarios.

To the degree that the different scalp distributions of the ant-LPC and the post-LCP/P600 reflect distinct underlying neural sources, these findings suggest that the neurocognitive mechanisms underlying the two effects are at least partially distinct. The brain recognized and distinguished the different grains of representation that were predicted and violated as a consequence of our experimental manipulations. This of course raises the question of precisely what neurocognitive processes are indexed by each of the late positive components. One possibility is that they each reflect distinct and independent mechanisms that map directly on to the grains of representation that were violated. This account, however, would predict that the double violations (HC_SRviol) should produce *both* an ant-LPC and a post-LPC/P600 effect because they violated *both* event/lexical and event structure/animacy constraints. In fact, the double HC_SRviol critical words failed to produce an ant-LPC effect, producing only a post-LPC/P600 effect. This finding constrains the functional interpretation of the ant-LPC. It suggests that it cannot simply reflect processes that are limited to the lexico-semantic level – either the detection of lexico-semantic conflict (conflict between the predicted word and the incoming critical word) or a process of gaining unique access to the critical word's meaning through 'automatic' competitive lexico-semantic inhibition of the incorrectly predicted word. This is because, in the double violated HC_SRviol scenarios, the predicted word also conflicted with the critical word, but this lexico-semantic conflict did not evoke an ant-LPC effect.

---

implausible but not impossible (Kuperberg et al., 2007; see also Kuperberg et al., 2003; van de Meerendonk, Kolk, Vissers & Chwilla, 2010). However, there are some situations in which a post-LPC/P600 can be evoked by words that, in offline tasks, are judged to be implausible but not impossible. For example, a post-LPC/P600 is evoked by "served" in the semantically reversible sentence, "The restaurant owner forgot which waitress the customer had served ... " (Chow, Smith, Lau & Phillips, 2015; see also Kolk et al., 2003). We have argued that this is because, during fast, online processing, comprehenders do not necessarily use all contextual cues to predict an upcoming event structure. If certain cue combinations very reliably predict and lead the comprehender to constrain her interpretation to a particular event structure, this will override the effects of weaker contextual cues. However, when the full set of contextual cues do become available to the message level, the resulting interpretation will conflict with the predicted event structure, triggering the post-LPC/P600 effect during online comprehension. With more time to make offline judgments, however, comprehenders are able to use the full set of contextual cues, and they will judge such sentences to be implausible but not impossible (see Kuperberg, 2016 for discussion).

Our findings also suggest that the ant-LPC cannot simply reflect *attempts* to integrate the critical word into the comprehender's existing mental model. This is because such attempts are likely to have occurred in both the HC_SRviol and the LC_SRviol scenarios, and neither of these violations evoked an ant-LPC effect. Rather, as we discuss further below, our data suggest that the ant-LPC produced by the HC_Lexviol critical words primarily reflected the successful updating of the comprehender's existing mental model with the newly inferred event, and that this successful update necessarily entailed top-down suppression of the incorrect lexico-semantic prediction. On this account, the reason why no ant-LPC effect was evoked by the HC_SRviol critical words is that, having detected the event structure/animacy violation, comprehenders were unable to successfully update their existing mental model and there was no need to suppress the predicted word's lexico-semantic features.

Finally, it is worth noting that not only did the double violations (HC_SRviol) fail to produce an ant-LPC effect, but they produced a post-LPC/P600 effect that was larger than that produced by selection restriction violations in low constraint contexts. We suggest that this is because the comprehender's original high certainty prediction for a specific event also increased her certainty of the upcoming event structure, thereby making it easier to detect the event structure/animacy violation. This result is consistent with existing evidence that the post-LPC/P600 evoked by syntactic violations is also larger in high constraint than low constraint contexts (Gunter, Friederici, Schriefers, 2000).

As we discuss next, the full pattern of ERP findings can be understood within a *hierarchical generative framework* of language comprehension (Kuperberg & Jaeger, 2016; Kuperberg, 2016; see also Xiang & Kuperberg, 2015). Below we outline this framework and offer functional interpretations of the N400, ant-LPC and post-LPC/P600 components within this framework. We then discuss these interpretations in relation to previous proposals.

A hierarchical generative framework of language comprehension

Within a hierarchical generative framework of language comprehension (see Kuperberg & Jaeger, 2016, section 5, page 16), the comprehender draws upon a network of stored representations (and the

statistical contingencies between them) that she believes are relevant to her current goals – an internal generative model[7]. Comprehension is conceptualized as an incremental process of testing and updating 'hypotheses' at each of these levels of representation, with the overarching goal being to build a mental model by inferring the hypothesis at the highest level of the hierarchy – the message-level representation -- that can best explain statistical structure of the full set of linguistic and non-linguistic inputs she has thus far encountered. The comprehender cannot achieve this goal all at once because the input unfolds linearly over time and it takes time to pass up new information from lower to higher levels of representation. However, if she has strong hypotheses at a higher level of representation, she can test them by generating top-down predictions and propagating this information down to pre-activate information at lower levels of the hierarchy. Then, as new bottom-up information (evidence) is passed up the model and reaches these lower levels, any information that confirms the comprehender's top-down predictions is facilitated ('explained'), and any information that is unpredicted ('unexplained') is propagated back up the generative model and used to update hypotheses at higher levels of the hierarchy. In this way, through cycles of prediction and hypothesis updating, the comprehender is able to 'home in on' the communicator's underlying message with increasing certainty as the linguistic (and non-linguistic) inputs progressively unfold in real time. We now illustrate these principles with the examples of stimuli used in the present study. The example LC and HC contexts given in the Introduction are repeated here again for convenience.

(1) *Low Constraint (LC) context:* Eric and Grant received the news late in the day. They mulled over the information, and decided it was better to act sooner rather than later. Hence, they cautioned the…(swimmers/trainees/drawer).

[7] Generative models are classically described within Bayesian probabilistic frameworks of cognition (Griffiths, Kemp & Tenenbaum, 2008) at Marr's first computational level of analysis. Here, the term 'hypothesis' refers to a possible latent (hidden) cause of an observation. At any given time, the agent has multiple hypotheses, with each hypothesis being held with a particular degree of belief, expressed as a probability value. The beliefs of all possible hypotheses add up to 100%. Generative models, however, can also be described Marr's second algorithmic level. Here, one way of conceptualizing a hypothesis is as a particular pattern (or path) of connections within a neural network, with belief corresponding to its (settled) state of activation (see McClelland 1998 & 2013).

(2) *High Constraint (HC) context:* 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...(trainees/drawer).

After reading both these LC and the HC contexts, but before encountering the critical words, we assume that the comprehender has predicted the event structure, <Animate Agent cautioned the Animate Patient>, at the message level, constraining her current mental model with this event structure, and that she has also pre-activated coarse-grained <animate> features at the lexico-semantic level. After reading the HC context, we assume that the comprehender has additionally predicted and incorporated the specific event being communicated (<*cautioned the swimmers*>) into her mental model (see Hare, Jones, Thomson, Kelly, & McRae, 2009; McRae & Matsuki, 2009), and that she has pre-activated finer-grained semantic features that are associated with the word, <swimmers> (see Figure 1).[8]

Within this framework, the N400 is naturally understood as reflecting the consequences of hypothesis testing at the lexico-semantic level, as illustrated in Figure 5. Specifically, it reflects the degree of match between semantic features that have been pre-activated and those that have become available when the bottom-up input is decoded at the lexico-semantic level (see Kuperberg, 2016 for further discussion). In the HC_Pred scenarios both the animacy features and finer-grained semantic features of the incoming word match feature that have been pre-activated, and so the amplitude of the N400 is small. In all four other conditions, however, critical words encode semantic features that have not already been pre-activated, and they therefore produce a larger amplitude N400 (the N400 is largest to the selection restriction animacy violations because these critical words mismatch animacy features that have been pre-activated by the verb (see Paczynski & Kuperberg, 2011, 2012).

[8] A generative model also assumes that the comprehender has predicted/pre-activated information at other levels of representation. In the present study, this would include other lexical properties associated with the upcoming critical word. Specifically, in both HC and LC contexts, she is likely to have predicted its syntactic representation (a noun-phrase). And, in the HC contexts, she is also likely to have predicted its phonological/orthographic representation (*swimmers*). However, because this experiment did not explicitly manipulate information at either of these levels of representation, we do not discuss this further.

As illustrated in Figure 5 (bottom left), the ant-LPC can be understood as reflecting the process of successfully updating the existing mental model with new unpredicted input. In the case of the HC_Lexviol scenarios, this would have involved shifting from the original high certainty belief that the specific event being communicated was <lifeguards cautioned the swimmers> to a new high certainty belief, <lifeguards cautioned the trainees>, and incorporating this new belief into the mental model. We suggest that this process of hypothesis updating was *initiated* as soon as unpredicted fine-grained semantic features associated with <trainees> reached the message level of representation, even though, at this stage, at the lexico-semantic level, the incorrectly predicted features of <swimmers> were still active. However, in order to *complete* this message-level inference and update the mental model with the newly inferred event, activation of the semantic features associated with <swimmers> had to be suppressed. In other words, we argue that the ant-LPC evoked by the HC_Lexviol critical words reflected activity at *both* the higher message level (updating the mental model in the light of new unpredicted information) as well as at the lower lexico-semantic level (the top-down suppression of the incorrectly predicted semantic features associated with <swimmers> in order to uniquely select the features associated with <trainees>).

Finally, as illustrated in Figure 5 (right), we suggest that the post-LCP/P600 evoked by the LC_SRviol and the HC_SRviol critical words, at the very least, reflects the detection of conflict between the predicted event structure (<lifeguards caution animate Patient>) and the new input (<drawer>) at the message level. Because this input violates the predicted event structure – the boundary of the current mental model -- the comprehender cannot incorporate it into this existing model. Consequently, in the HC_SRviol scenarios, comprehenders make no attempts to suppress the semantic features associated with the incorrectly predicted word, <swimmers>, explaining why no ant-LPC was evoked by these double violations (see Figure 5, bottom right). As we discuss further below, the post-LPC/P600 may *also* reflect the re-evaluation (reanalysis), revision and/or attempts to repair the existing mental model in prolonged attempts to incorporate the critical words and come to a plausible representation of meaning. In the present study,

however, these attempts failed, which is why the event structure/animacy violated scenarios were judged to be highly implausible.

Relationship with other proposals of these ERP components

A classic debate about the N400 is whether it reflects 'lexical access' (e.g. Lau, Phillips, Poeppel, 2008) or (combinatorial) 'integration' (e.g. Hagoort, Baggio, & Willems, 2009)[9]. This debate has tended to assume that 'lexical access' and 'integration' are distinct and independent processes. The generative framework described above makes no such assumptions. Within this framework, unpredicted semantic features can be passed up to the message-level, triggering shifts in belief at this level (that is, initiating 'integration') *before* the comprehender has *uniquely* accessed a single lexical item (involving the suppression of any competitors), as in the HC_Lexviol, LC_SRviol and HC_SRviol scenarios. Thus, whether or not the N400 reflects full lexical access and whether it reflects integration depends on what has already been predicted. For example, in the LC_Unpred scenarios, the large N400 can be taken to reflect *both* full lexical access (there are no competitors to suppress) *and* message-level integration (the event structure has already been predicted at the message level and so, to integrate the new input, the comprehender only needs to access the unpredicted fine-grained semantic features associated with <trainees>). In the HC_Lexviol scenarios, however, the N400 reflects access to the semantic features of the input (although not *unique* access) but not message-level integration (which is associated with the ant-LPC effect), see Figure 5 (top left). This interpretation of the N400 is closest to proposals that the N400 reflects the retrieval of, or access to, the *semantic features* associated with incoming words (e.g. Kutas & Federmeier, 2011), with the assumption that semantic retrieval/access is dissociable from the process of accessing a specific stored lexical entity (e.g. Federmeier & Kutas, 1999; Laszlo & Federmeier, 2009).

These data are also consistent with some previous proposals that the ant-LPC reflects the suppression of the incorrectly predicted item (Kutas, 1993; Federmeier, et al., 2007) and/or higher message level

[9] Others have used the term 'integration' in relation to the N400 simply to refer to a delayed ('post-lexical') use of context to facilitate semantic processing of incoming words, without assumptions about whether this includes combinatorial or unification processes to build new propositional meaning (e.g. Chwilla, Hagoort & Brown, 1998;Van Petten & Luka, 2012).

integration (Federmeier et al., 2010; DeLong, Quante & Kutas, 2014). The generative framework outlined above emphasizes that, at least for the HC_Lexviol scenarios that triggered the ant-LPC effect in the present study, top-down suppression and message-level integration are tightly intertwined: suppression of the incorrectly predicted competitor is an inherent component of being able to successfully incorporate a lexically violated item into the comprehender's existing mental model. As noted above, our findings on the double violations (HC_SRviol) also rule out the idea that the ant-LPC *only* reflects the detection of lexico-semantic conflict between the predicted and bottom-up semantic features (or the 'automatic' inhibition of incorrectly predicted lexico-semantic competitors in the absence of higher-level activity).

Finally, our interpretation of the post-LPC/P600 is broadly consistent with proposals that it reflects the detection of conflict between alternative interpretations (e.g. Kuperberg, 2007; Kim & Osterhout, 2005; van de Meerendonk, et al., 2009), triggering prolonged combinatorial processing in attempts to come up with a new message-level interpretation (Kuperberg, 2007; Paczynski & Kuperberg, 2012). Here, we conceptualize conflict to be at the message level, between the predicted event structure and new bottom-up input that is incompatible with this event structure. We also distinguish the processes reflected by the post-LPC/P600 (a failure to incorporate the bottom-up input into the *existing* mental model) from those reflected by the ant-LPC (successfully updating the existing mental model with new bottom-up input).

Open Questions

Our findings leave open several questions. The first set concerns the ant-LPC. We have argued that this component primarily reflects the process of updating the comprehender's mental model by shifting to a new high certainty event representation based on the new input. In the present study, this process was triggered by the violation of a prior strong event/lexico-semantic prediction -- a classic trigger of the ant-LPC effect (Federmeier et al., 2007). However, a lexical prediction violation may not be the *only* trigger for the effect. For example, encountering an incoming critical word that is highly informative about the nature of a specific event may lead a comprehender to infer this new event with high certainty and to update her mental model accordingly, even if she hasn't previously strongly predicted a different event. And, indeed,

anteriorly distributed late positivities have sometimes been reported in less constraining contexts (e.g. Chow, Lau, Wang & Phillips, 2018; Thornhill & Van Petten, 2012; Zirnstein, van Hell & Kroll, 2018). It will be important for future studies to investigate this systematically.

A second set of questions concerns the post-LPC/P600. We have argued that, at the very least, the post-LPC/P600 reflects a failure to incorporate unpredicted input into the *existing* mental model. It is also likely to reflect additional processes that can, in some situations, potentially allow the comprehender to come to a coherent interpretation – specifically, a reanalysis of preceding words within the context that could potentially allow her to revise or repair her existing mental model (although such attempts may not necessarily be successful, e.g. Kuperberg et al., 2006). For example, in the event structure/animacy violated scenarios, the post-LPC/P600 might reflect a re-analysis of the verb to check whether it did indeed select for the animacy features that were violated by the new input. The present study cannot disentangle these possibilities. This is because participants were able to carry out the coherence judgment task without being required to engage in any of these additional processes.[10] Future experiments, however, may be able to further probe the neurocognitive function of the post-LPC/P600 by using tasks that explicitly ask participants to repair the input or attempt to create representations of impossible events.

Another set of open questions relates to the role of context in producing the late positivities during language comprehension. We have argued that both the ant- and post-LPC/P600 effects are driven by shifts in belief at the message level of representation, with differing consequences for whether or not any unpredicted input can be incorporated into the existing mental model. This implies that an initial mental model needs to be in place to produce these effects. In the present experiment, the critical words appeared in mini-discourse scenarios that contained two complete sentences and an introductory phrase that referred back to these context sentences. If the ant-LPC and the post-LPC/P600 effects are indeed related to

[10] The post-LPC/P600 is largest when participants are explicitly asked to make coherence judgments (e.g, Xiang & Kuperberg, 2015; Kolk et al., 2003; discussed by Kuperberg, 2007, section 3.5, page 33). However, this is by no means the only factor that influences its amplitude: plausibility judgment is neither necessary nor sufficient to produce a P600 effect. We conceive of coherence monitoring as an important component of successful reading comprehension (Garner, 1980; Wagoner, 1983; Cain, 2016 ; van de Meerendonk et al., 2009; see also Sanford et al. 2011 for evidence that the detection of incoherence can play an important role in triggering a P600 effect during passive reading comprehension).

consequences of violated predictions for discourse-level interpretation, they may not necessarily be seen on words that violate strong constraints for an event structure and/or event, if the prior context is shorter such that no mental model is in place prior to the onset of the critical word (see Kuperberg, 2007, section 3.6 page 34 for discussion).

Finally, a central question is *why* the brain engages different neurocognitive mechanisms in response to predictions that are confirmed versus predictions that are strongly violated? It is sometimes assumed that the neurocognitive mechanisms reflected by the late positivities reflect unwanted *costs* of violating strong predictions (e.g. Federmeier, 2007, and Kutas, DeLong, & Smith, 2011 for discussion). However, another way of understanding their functional role is that they provide a strong signal that the statistical structure of the broader communicative environment has changed, and that, to continue predicting efficiently, comprehenders need to quickly *adapt* to their new communicative environment (Kuperberg, 2013). Adaptation may involve modifying the comprehender's existing generative model, or it may involve switching to (or inferring) a new, previously stored generative model (for a detailed discussion of language adaptation within this type of generative framework, see Kleinschmidt & Jaeger, 2015; Qian, Jaeger & Aslin, 2012, 2016). Some existing evidence links the post-LPC/P600 to adaptive processes (Coulson, King & Kutas, 1998; Hanulíková, Alphen, van Goch & Weber, 2012), and it will be important for future studies to systematically investigate the role of the late positivities in adaptation to the broader communicative environment.

Conclusion

In conclusion, we have shown that the brain evokes responses that distinguish between confirmed semantic predictions, violations of predictions of specific events and of whole event structures. These findings can be interpreted within a hierarchical generative architecture of language comprehension, in which confirmed and violated predictions at different levels and grains of representation manifest as distinct spatiotemporal neural signatures.

References

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English lexicon project. *Behav Res Methods, 39*(3), 445-459.

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

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behav Res Methods, 46*(3), 904-911.

Cain, K. (2016). Reading comprehension development and difficulties: an overview. *Perspectives on Language and Literacy, 42*(2), 9.

Chow, W.-Y., Lau, E., Wang, S., & Phillips, C. (2018). Wait a second! delayed impact of argument roles on on-line verb prediction. *Lang Cogn Neurosci, 33*(7), 803-828.

Chow, W. Y., Smith, C., Lau, E., & Phillips, C. (2015). A “bag-of-arguments” mechanism for initial verb predictions. *Lang Cogn Neurosci, 31*(5), 577-596.

Chwilla, D. J., Hagoort, P., & Brown, C. M. (1998). The mechanism underlying backward priming in a lexical decision task: Spreading activation versus semantic matching. *Q J Exp Psychol-A, 51*(3), 531-560.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *J Cogn Neurosci, 6*(3), 256-266.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain responses to morphosyntactic violations. *Lang Cogn Process, 13*(1), 21-58.

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

DeLong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology, 48*(9), 1203-1207.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods, 134*(1), 9-21.

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491-505.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *J Mem Lang, 41*(4), 469-495.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain Lang, 115*(3), 149-161.

Federmeier, K. D., Mai, H., & Kutas, M. (2005). Both sides get the point: hemispheric sensitivities to sentential constraint. *Mem Cognit, 33*(5), 871-886.

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

Garner, R. (1980). Monitoring of understanding: An investigation of good and poor readers' awareness of induced miscomprehension of text. *Journal of Reading Behavior, 12*(1), 55-63.

Grainger, J., & Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Lang Linguist Compass, 3*, 128-156.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge Handbook of Computational Psychology* (pp. 59-100). New York: Cambridge University Press.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *J Cogn Neurosci, 12*(4), 556-568.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences, 4th Ed.* (pp. 819-836). Cambridge, MA: MIT Press.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Lang Cogn Process, 8*(4), 439-483.

Hanulikova, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *J Cogn Neurosci, 24*(4), 878-887.

Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition, 111*(2), 151-167.

Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Brain Res Cogn Brain Res, 19*, 59-73.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *J Mem Lang, 52*(2), 205-225.

Kleinschmidt, D. F., & Jaeger, F. T. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychol Rev, 122*(2), 148-203.

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain Lang, 85*(1), 1-36.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Res, 1146*, 23-49.

Kuperberg, G. R. (2013). The proactive comprehender: What event-related potentials tell us about the dynamics of reading comprehension. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling Reading Comprehension: Behavioral, Neurobiological, and Genetic Components* (pp. 176-192). Baltimore, MD: Paul Brookes Publishing.

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Lang Cogn Neurosci, 31*(5), 602-616.

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes, 21*(5), 489-530.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Lang Cogn Neurosci, 31*(1), 32-59.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: evidence from event-related potentials. *Brain Lang, 100*(3), 223-237.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Brain Res Cogn Brain Res, 17*(1), 117-129.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Lang Cogn Process, 8*, 533-572.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190-207). New York: Oxford University Press.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol, 62*, 621-647.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychol Rev, 104*(2), 211-240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Process, 25*(2-3), 259-284.

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *J Mem Lang, 61*(3), 326-338.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *J Cogn Neurosci, 25*(3), 484-502.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Rev Neurosci, 9*(12), 920-933.

Levin, B. (1993). *English Verb Classes And Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front Hum Neurosci, 8*.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav Res Meth Instr, 28*(2), 203-208.

McClelland, J. L. (1998). Connectionist models and Bayesian inference. In M. Oaksford & N. Chater (Eds.), *Rational Models of Cognition* (pp. 21-52). New York: Oxford University Press.

McClelland, J. L. (2013). Integrating probabilistic models of perception and interactive neural networks: a historical and tutorial review. *Front Psychol, 4*, 503.

McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Lang Linguist Compass, 3*(6), 1417-1429.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Osterhout, L., & Holcomb, P. J. (1992). Event-related potentials elicited by syntactic anomaly. *J Mem Lang, 31*(6), 785-806.

Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. In S. M. Garnsey (Ed.), *Language and Cognitive Processes. Special Issue: Event-related brain potentials in the study of language* (Vol. 8 (4), pp. 413-437). Hove: Lawrence Erlbaum Associates.

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence and time course of the brain responses to syntactic and semantic anomalies. *Lang Cogn Process, 14*, 283-317.

Paczynski, M., & Kuperberg, G. R. (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb argument processing. *Lang Cogn Process, 26*(9), 1402-1456.

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *J Mem Lang, 67*(4), 426-448.

Peirce, J. W. (2007). PsychoPy--Psychophysics software in Python. *J Neurosci Methods, 162*(1-2), 8-13.

Qian, T., Jaeger, T. F., & Aslin, R. N. (2012). Learning to represent a multi-context environment: more than detecting changes. *Front Psychol, 3*, 228.

Qian, T., Jaeger, T. F., & Aslin, R. N. (2016). Incremental implicit learning of bundles of statistical patterns. *Cognition, 157*, 156-173.

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. . (2011). Anomalies at the borderline of awareness: an ERP study. *J Cogn Neurosci, 23*, 514-523.

Schwanenflugel, P. J., & Lacount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *J Exp Psychol Learn Mem Cognit, 14*(2), 344-354.

Shetreet, E., Alexander, E. J., Romoli, J., Chierchia, G., & Kuperberg, G. R. (Under review). What we know about knowing: Presuppositions generated by factive verbs influence downstream neural activity processing.

Taylor, W. (1953). 'Cloze' procedure: A new tool for measuring readability. *Journal Q, 30*, 415-433.

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: frontal positivity and N400 ERP components. *Int J Psychophysiol, 83*(3), 382-392.

van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. T. W. M. (2009). Monitoring in language perception. *Lang Linguist Compass, 3*(5), 1211-1224.

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring language perception: mild and strong conflicts elicit different ERP patterns. *J Cogn Neurosci, 22*(1), 67-82.

van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *J Cogn Neurosci, 13*(7), 967-985.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *Int J Psychophysiol, 83*(2), 176-190.

Wagoner, S. A. (1983). Comprehension monitoring: What it is and what we know about it. *Read Res Q*, 328-346.

Wlotko, E. W., & Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology, 49*(6), 770-785.

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Lang Cogn Neurosci, 30*(6), 648-672.

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

Acknowledgments

This work was funded by the National Institute of Child Health and Human Development (R01 HD08252 to G.R.K.). E.W.W. was supported by an Institutional Research and Academic Career Development Award from the National Institute of General Medical Sciences (K12GM074869 to Tufts University, PI C. Moore). We thank Maria Luiza Cunha Lima, Margarita Zeitlin, and Connie Choi for their contributions to constructing the experimental materials, and Margarita Zeitlin and Simone Riley for their assistance with data collection. We also thank Trevor Brothers for his help with data analysis, Lena Warnke and Arim Choi Perrachione for their help in making figures, and Trevor Brothers, Lotte Schoot and Lin Wang for their insightful comments on the manuscript.

Scenario Type	Example	Lexical constraint of context*	Cloze of critical word*	SSV of critical word‡
1. High Constraint, predictable (HC_Pred)	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...	71% (13%)	71% (13%)	0.31 (0.12)
2. Low Constraint, unpredictable (LC_Unpred)	Eric and Grant received the news late in the day. They mulled over the information, and decided it was better to act sooner rather than later. Hence, they cautioned the trainees...	20% (9%)	0.2% (0.8%)	0.18 (0.09)
3. Low Constraint, Selection Restriction violation (LC_SRviol)	Eric and Grant received the news late in the day. They mulled over the information, and decided it was better to act sooner rather than later. Hence, they cautioned the drawer...	20% (9%)	0% (0%)	0.17 (0.09)
4. High Constraint, lexical prediction violation (HC_Lexviol)	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 trainees...	71% (13%)	0.1% (0.5%)	0.18 (0.10)
5. High Constraint, Selection Restriction violation (HC_SRviol)	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 drawer...	71% (13%)	0% (0%)	0.16 (0.09)

Table 1. Examples of five experimental conditions created around the same verb (here, "cautioned"). To create the *High Constraint predictable* (HC_Pred) scenarios, the HC context was paired with the noun with the highest cloze probability. To create the *High Constraint lexical violation* (HC_Lexviol) scenarios, the HC context was paired with a direct object noun of low cloze, but that was still plausible in relation to this context. To create the *Low Constraint unpredictable* (LC_Unpred) scenarios, same noun was paired with the LC context. To create the *Selection Restriction violations*, the HC and LC contexts were paired with nouns that violated the animacy restrictions of the verbs. The critical word in each of the example sentences is underlined (although this was not the case in the experiment itself). The sentences continued with additional words, as indicated by the three dots. The same critical words appeared in all conditions 2-5 across lists.
Means are shown with range in square brackets. Stimulus characteristics of these unpredictable critical words were as follows (M=mean, SD=standard deviation), retrieved from the English Lexicon Project(Balota et al., 2007), except Concreteness retrieved from Brysbaert et al. (2014). Number of letters: M=7.3, SD=2.21; Log Frequency from the SUBTLEX database: M=2.44, SD=0.78; Orthographic Levenshtein Distance: M=2.6, SD=0.86; Concreteness: M=4.1, SD=0.7. Stimulus characteristics of the predictable critical words in condition 1 are as follows. Number of letters: M=5.7, SD=1.56; Log Frequency from the SUBTLEX database: M=3.2, SD=0.66; Orthographic Levenshtein Distance: M=1.9, SD=0.56; Concreteness: M=4.3, SD=0.69.
*Lexical constraint of each context was calculated by identifying the most common completion across participants who saw that context in the cloze norming studies and tallying the proportion of subjects who provided this completion. Cloze probabilities of critical words were calculated based on the percentage of respondents providing the critical noun used in the experiment. Alternate word forms (e.g., singular/plural or masculine/feminine gendered nouns) were collapsed, but synonyms or lexical alternatives were not collapsed (e.g. couch/sofa).

‡ Latent Semantic Analysis (LSA, Landauer & Dumais 1997; Landauer et al. 1998) was used to calculate Semantic Similarity Values [SSVs] between the critical word and all the preceding words in the stimulus using term-to-term comparisons derived from the TASA "General Reading up to 1st year college" corpus with the maximum number of dimensions available. In five cases where the LSA calculator did not recognize the critical word, close synonyms were used instead.

Effect	*df*	**300-500ms**			**600-1000ms**		
		F	*p*	*ε*	*F*	*p*	*ε*
Scenario Type	4,152	19.38	<0.0001	0.9066	5.85	0.0002	0.8835
Scenario Type x Region	12,456	9.66	<0.0001	0.4468	14.55	<0.0001	0.4730
Region	3,114	15.02	<0.0001	0.5191	16.43	<0.0001	0.5909

Table 2. Results of the omnibus ANOVA showing main effects of Scenario Type, interactions between Sentence Type and Region, and main effects of Region in time windows of interest.
All p-values are reported with Greenhouse-Geisser correction for sphericity applied, along with the associated *ε* (epsilon) value.

Contrast of Interest	**N400** (300-500ms; Central region)	
	F	*p*
HC_Pred < LC_Unpred	41.54	<0.0001
HC_Pred < HC_Lexviol	27.33	<0.0001
HC_Pred < LC_SRviol	61.66	<0.0001
HC_Pred < HC_SRviol	59.56	<0.0001

Table 3. Results of planned pair-wise contrasts (paired t-tests) between the amplitude of the N400 evoked by predicted critical words in highly lexically constraining contexts (HC_Pred) and by unexpected critical words in each of the four other scenario types: (1) unpredictable critical words in low constraint contexts (LC_Unpred); (2) critical words that violated constraints for a specific event/lexical item based on high constraint contexts (HC_Lexviol); (3) critical words that violated the verb's constraints for a particular event structure and its animacy-based selection restrictions in low constraint contexts (LC_SRviol); (4) critical words that violated both the verb's constraints for a particular event structure/animacy-based selection restrictions as well constraints for a specific event/lexical item based on high constraint contexts (HC_SRviol).
Degrees of Freedom: 1,38.
< indicates that the amplitude of the negativity is smaller (less negative) to the HC_Pred critical words than to the critical words in the contrasting scenario.
Significant effects are indicated using the following symbols: ^ p < .10. *p < .05. **p < .01.
The Central region was defined as described in Figure 3.

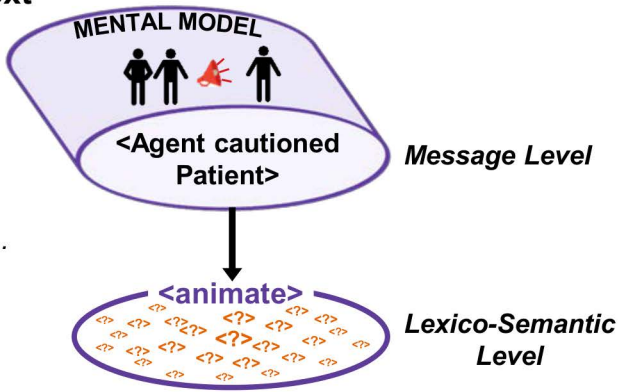
Contrast of Interest	Late Anterior Positivity (600-1000ms; Prefrontal region)		Late Posterior Positivity/P600 (600-1000ms; Posterior region)	
	F	*p*	*F*	*p*
(1) HC_Lexviol > HC_Pred	**27.15**	**<0.0001**	**5.09**	**0.0299**
(2) HC_Lexviol > LC_Unpred	**14.38**	**0.0005**	0.82	0.3722
(3) HC_Lexviol > HC_SRviol	**15.63**	**0.0003**	#9	
(4) HC_Lexviol > LC_SRviol	**17.98**	**0.0001**	#10	
(5) LC_SRviol > HC_Pred	0.31	0.5791	**26.43**	**<0.0001**
(6) HC_SRviol > HC_Pred	0.06	0.8102	**41.36**	**<0.0001**
(7) LC_SRviol > LC_Unpred	0.44	0.5097	**13.23**	**0.0008**
(8) HC_SRviol > LC_Unpred	0.6	0.4427	**31.69**	**<0.0001**
(9) HC_SRviol > HC_Lexviol	#3		**16.26**	**0.0003**
(10) LC_SRviol > HC_Lexviol	#4		**6.53**	**0.0147**
(11) HC_SRviol > LC_SRviol	0.06	0.8098	**6.17**	**0.0175**

Table 4. Results of planned pair-wise contrasts (paired t-tests) between the amplitude of the Late Anterior Positivity and the Late Posterior Positivity/P600 evoked by each type of unpredicted critical word. Prefrontal and Posterior regions were defined as shown in Figure 3.
(1) – (4) contrast critical words that violate predictions of a specific event/lexical item (HC_Lexviol) with critical words in each of the other four scenario types. Each contrast shows a significant Late Anterior Positivity effect.
(5) – (10) contrast critical words that violate predictions of an event structure/animacy features, based on the verb's animacy-based selection restrictions (either in high constraint contexts, HC_SRviol or low constraint contexts, LC_SRviol) with critical words in each of the two other scenario types. Each contrast shows a significant Late Posterior Positivity/P600 effect.
(11) contrasts critical words in the two types of scenarios that violated predictions of an event structure/animacy features, and shows that the Late Posterior Positivity/P600 is larger in high constraint than low constraint contexts.
Degrees of Freedom: 1,38.
> indicates that the amplitude of the positivity is larger (more positive) than in the contrasting scenario.
Significant effects are indicated using the following symbols: ^ p < .10. *p < .05. **p < .01.

Figure 1. Schematic illustration of the state of the language comprehension system at two hierarchical levels of representation after reading a Low Constraint and a High Constraint context, but before encountering the upcoming critical noun. We depict information at a higher *message* level and a lower *lexico-semantic* level. We assume that the comprehender draws upon information encoded at both these levels to incrementally build a *mental model*. The *message* level of representation encodes the full set of upcoming events, states or facts that the comprehender believes are possible/compatible with her current mental model. In these example scenarios, this set of events can be described as the broad *event structure* that is constrained by the thematic-semantic structure of the verb in the third sentence ("cautioned"), i.e. <Agent cautions Patient>. This event structure defines the boundaries of the current mental model. The *lexico-semantic* level of representation encodes the full set of semantic features and properties that the comprehender believes are possible/compatible with the upcoming input. In these example scenarios, this is the set of conceptual features that are associated with the broad semantic category, <animate> (e.g. <sentient>, <able to move> etc) of the upcoming noun. In the Low Constraint context, the first two sentences introduce two novel referents ("Eric" and "Grant", depicted by the two human figures). Together with the verb in the third sentence, this constrains for an upcoming event structure, <they [Eric and Grant] cautioned Patient> (depicted with the warning icon and a third human figure). However, this Low Constraint context does not otherwise constrain for any specific event or state at the message level, or for any specific set of semantic features at the lexico-semantic level (although different sets of semantic features may be activated with slightly different strengths based on their frequencies, indicated by <?s>). In the High Constraint context, however, the first two sentences convey a rich semantic schema (a beach, sharks, lifeguards and the worry about future incidents in the sea), and, together with the verb in the third sentence, this strongly predicts a specific event (<lifeguards cautioned swimmers>) at the message level, which can therefore be incorporated into the current mental model ahead of encountering the upcoming word (as depicted by the icons for lifeguards, warning sign, swimmers and sharks). This High Constraint context also constrains strongly for the specific semantic features and properties associated with <swimmers> at the lexico-semantic level (e.g. <in water>, <staying afloat> etc), which are strongly pre-activated. In Figure 5 (see Discussion), we show schematic illustrations of activity within the language system *after* encountering critical words in the four unpredictable conditions.
Icons used in this figure were taken from the www.thenounproject.com

Low Constraint context

Eric and Grant received the news late in the day. They mulled over the information, and decided it was better to act sooner rather than later. Hence they cautioned the…



High Constraint context

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…

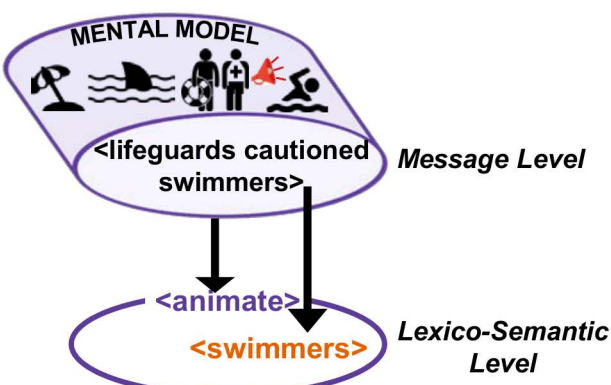


Figure 2. Presentation of stimuli. Stimuli were presented using PsychoPy 1.83 software (Peirce, 2007), displayed on a LCD monitor using white Arial font set to 0.1 of the screen height, on a black background. Each trial began with the prompt "READY?" (centered in green font). When participants initiated the trial with a button press, the first two sentences were presented successively in full on the screen. Participants advanced through these sentences at their own pace using a button press. Then a centered fixation cross appeared (750ms; 100ms inter-stimulus interval, ISI), followed by the third sentence, which was presented word-by-word (each word 450ms; 100 msISI). After the sentence-final word (SFW), a "?" appeared (centered, magenta font), prompting participants to make an acceptability judgement about whether or not the preceding scenario "made sense" or not. The "?" remained on the screen for an unlimited amount of time, until the response was registered. Following 32 of the 160 trials (all fillers) a yes/no comprehension question appeared about the preceding scenario. For example, the scenario "Damien was not at all like himself that day. Something in the atmosphere made him act differently, and everyone could tell. Relentlessly, he ridiculed the patients until they cried." was followed by the question, "Is this behavior typical of Damien?", with the intended answer being "NO".

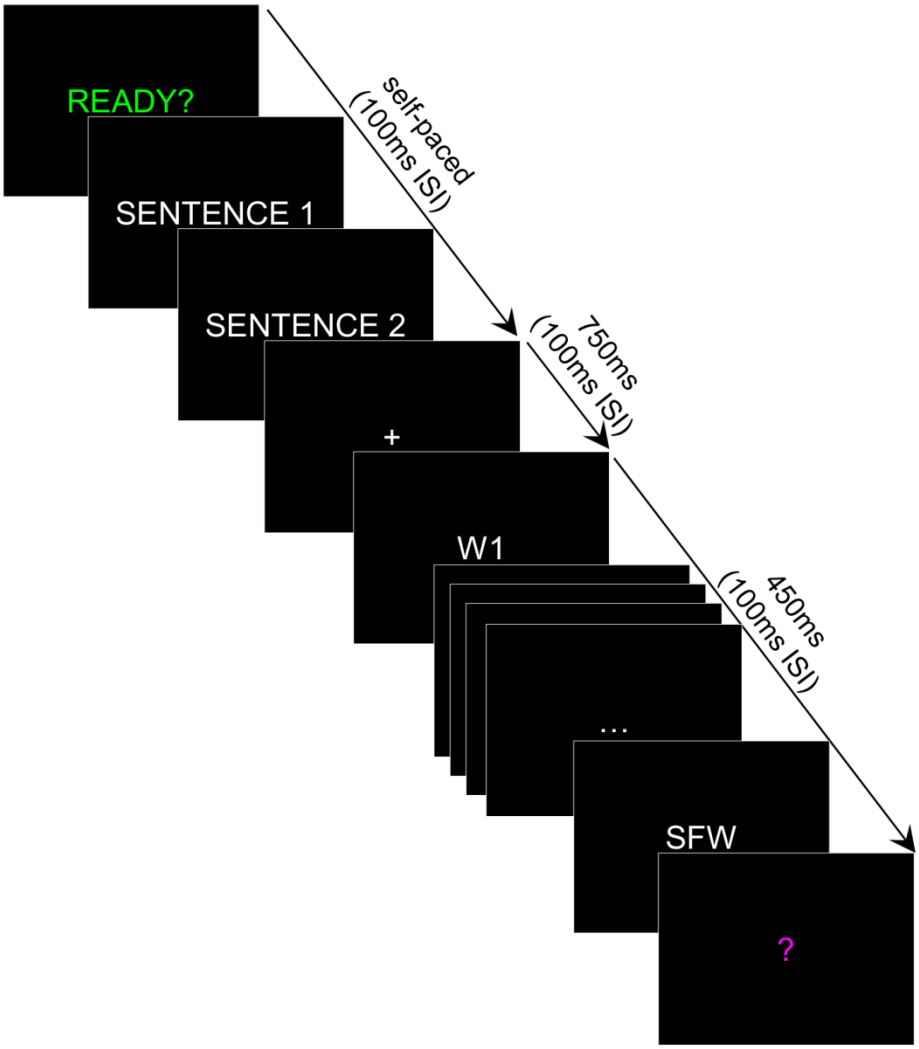


Figure 3. Spatial regions of interest used for analysis. The N400 was operationalized as the average voltage between 300-500ms within the Central region. The ant-LPC was operationalized as the average voltage between 600-1000ms within the Prefrontal region. The post-LPC/P600 was operationalized as the average voltage between 600-1000ms within the Posterior region.

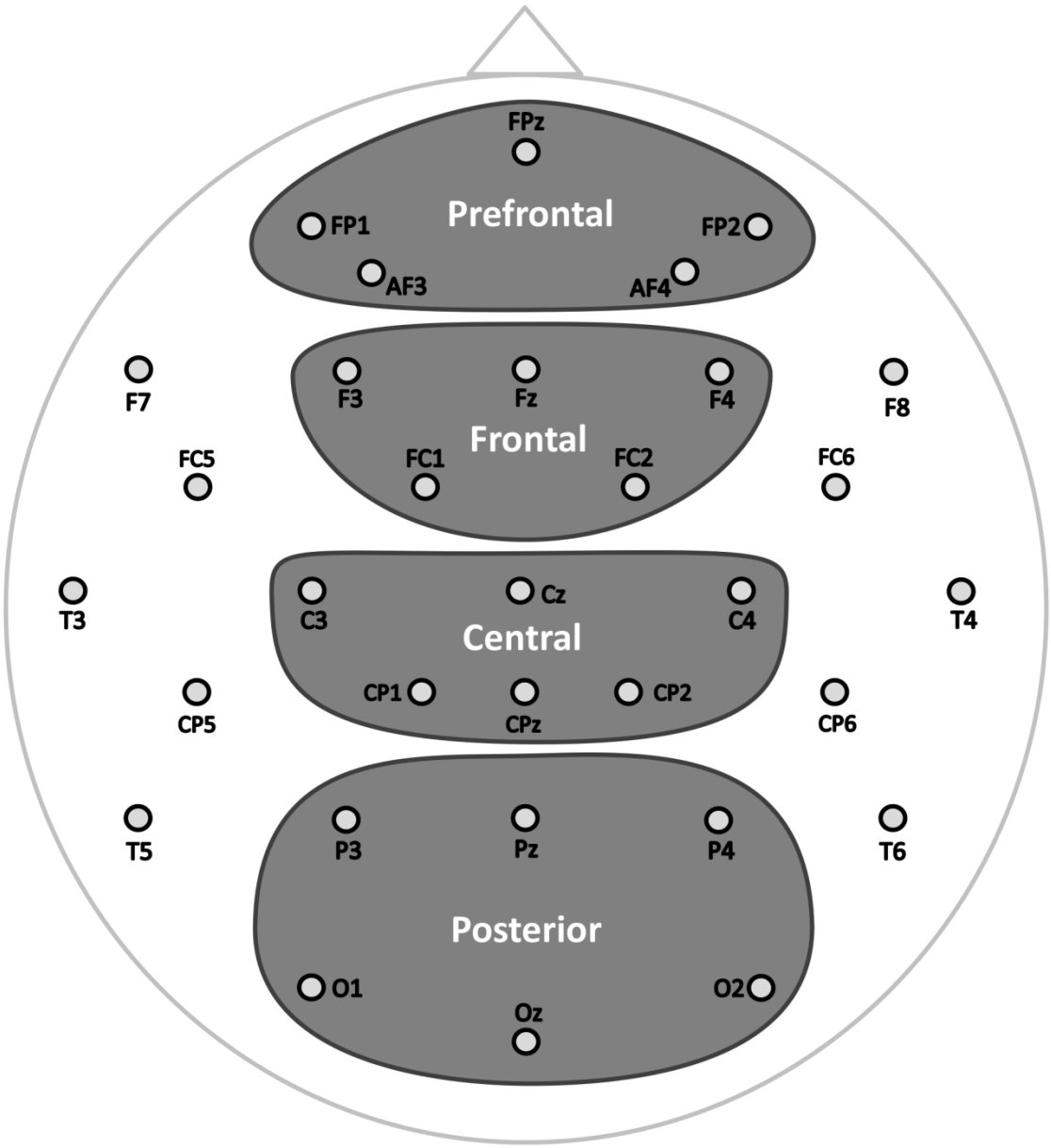
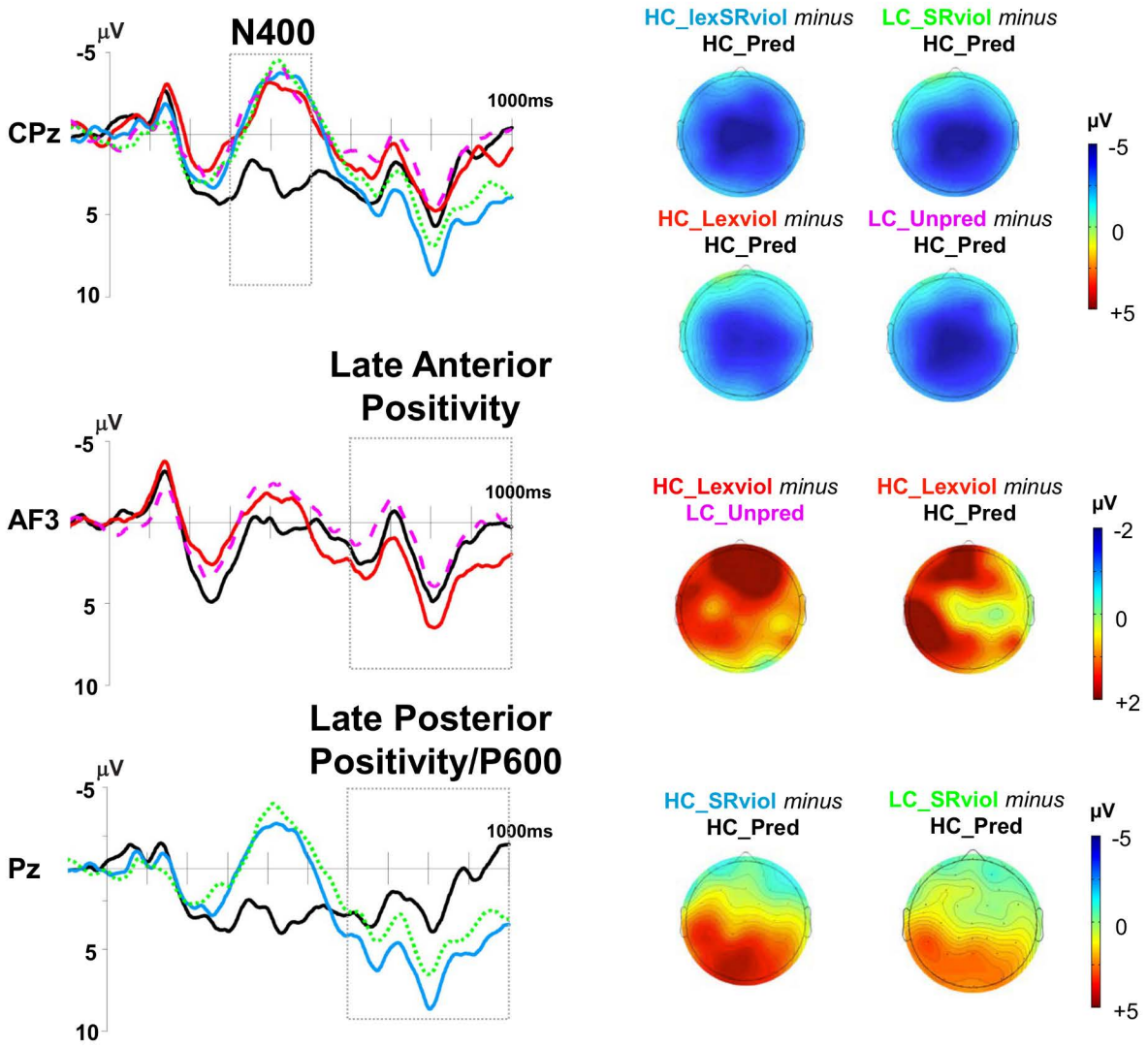


Figure 4. Grand-averaged waveforms and voltage maps. Top: Waveforms evoked by critical words in all five conditions are shown at CPz. The N400 was smaller (less negative) to critical words in the HC_Pred scenarios than to unpredicted critical words in all four other scenarios. Voltage maps show that this N400 effect was maximal at central and centroparietal sites. *Middle:* Waveforms evoked by critical words in the HC_Lexviol, the HC_Pred and the LC_Unpred scenarios are shown at AF3. The anterior late positivity effect (ant-LPC) was larger (more positive) to critical words in the HC_Lexviol scenarios than to critical words in the HC_Pred and LC_Unpred scenarios. Voltage maps show that this effect was evident at prefrontal sites in both these contrasts. *Bottom:* Waveforms evoked by critical words in the HC_SRviol, the LC_SRviol and the HC_Pred scenarios are shown at Pz. The posterior late positivity/P600 effect (post-LPC/P600) was larger (more positive) to critical words in the HC_SRviol and the LC_SRviol scenarios than to critical words in the HC_Pred scenarios. The post-LPC/P600 was also larger to critical words in the HC_SRviol than in the LC_SRviol scenarios. Voltage maps show that this effect was maximal at posterior sites in both contrasts.



HC_Pred	High Constraint Predicted	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**…
LC_Unpred	Low constraint Unpredicted	Eric and Grant received the news late in the day. They decided it was better to act sooner than later. Hence, they cautioned the **trainees**…
LC_SRviol	Low constraint Selection Restriction violated	Eric and Grant received the news late in the day. They decided it was better to act sooner than later. Hence, they cautioned the **drawer**…
HC_Lexviol	High Constraint Lexically violated	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 **trainees**…
HC_SRviol	High Constraint Selection Restriction violated	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 **drawer**…

Figure 5. Schematic illustration of the state of the language comprehension system at two hierarchical levels of representation after encountering critical nouns in the four types of unpredictable scenarios (conditions 2-5 in Table 1). Each quadrant of the figure illustrates activity within the N400 time window on the left, and within the late positivity time window, on the right. Descriptions of the higher message level and the lower lexico-semantic level of representation are given in the legend of Figure 1 in the Introduction, which illustrates the state of the language system after reading High Constraint and Low Constraint contexts, but *before* encountering the critical nouns.

Top left: Low Constraint Unpredicted (LC_Unpred): *N400 time window:* The semantic features associated with <trainees> do not match any pre-activated semantic features in this Low Constraint context (except for those associated with its animacy), and so a large N400 is evoked. The unpredicted semantic features associated with <trainees> are passed up to the message level, as depicted by the blue upward curved arrow. *Late positivity time window:* The event structure, <Agent cautioned Patient>, has already been predicted (based on the strong thematic constraints of the verb), and so once the semantic features of <trainees> reach the message level, the comprehender has effectively inferred the event, <they cautioned trainees>, and so no late positivity effect is evoked.

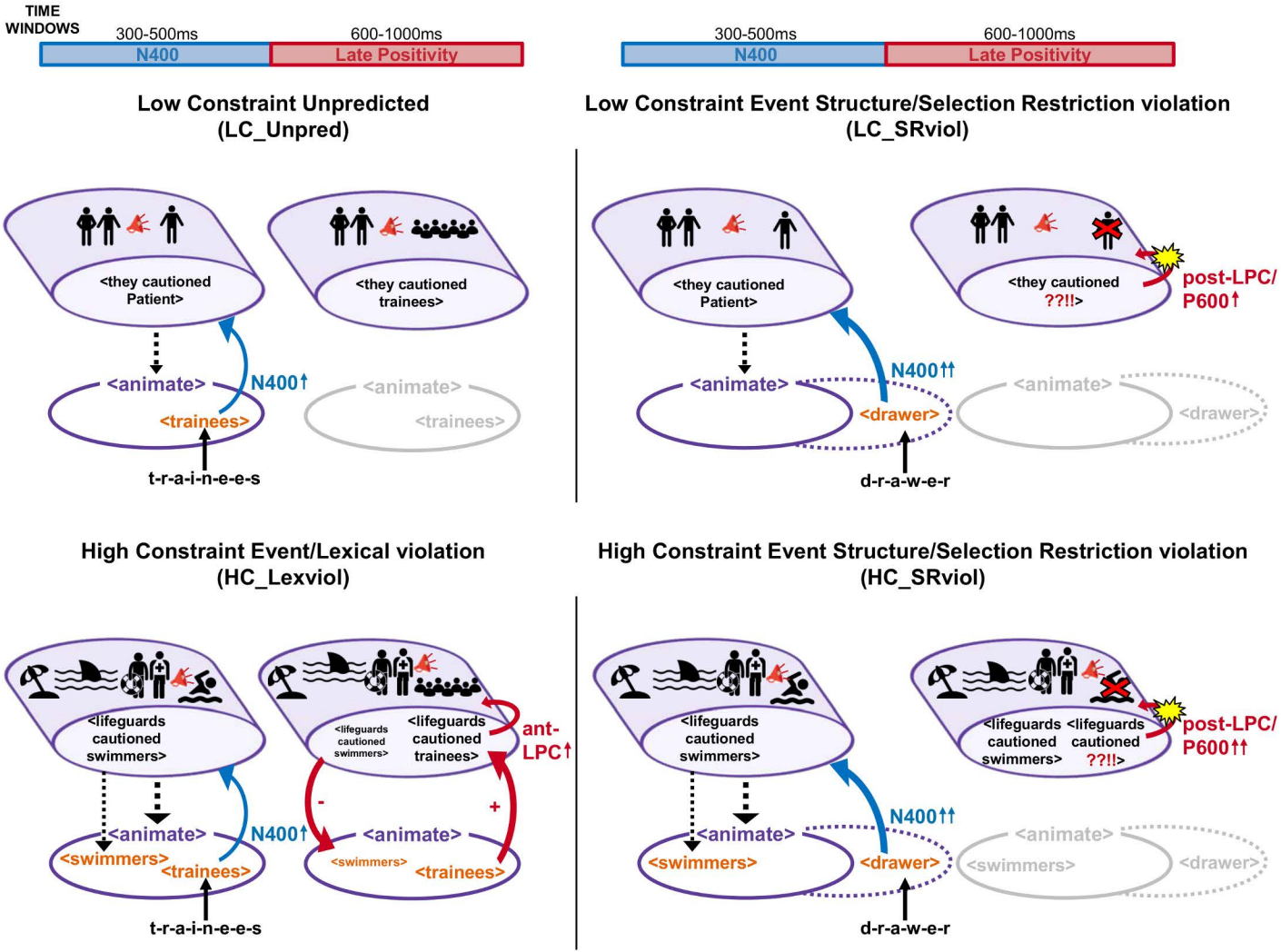
Bottom left: High Constraint Event/Lexical violation (HC_Lexviol): *N400 time window:* Prior to the onset of the critical noun, the specific event, <lifeguards cautioned swimmers>, and the semantic features associated with <swimmers> have already been predicted. However, specific semantic features associated with <trainees> do not match these pre-activated semantic features, and so a large N400 is evoked. The unpredicted semantic features associated with <trainees> are passed up to the message level, depicted by the blue upward curved arrow. *Late positivity time window:* As soon as the unpredicted semantic features associated with <trainees> reach the message level, the comprehender begins to shift from the incorrectly predicted event, <lifeguards cautioned swimmers> to the new event, <lifeguards cautioned trainees>. In order to *complete* this message-level inference and update the current mental model with the newly inferred event (depicted by the small upward curved red arrow into the mental model), the comprehender must suppress activation over <swimmers> and enhance activation over <trainees> at the lexico-semantic level (depicted by the downward and upward red arrows between the message and lexico-semantic levels). We suggest that the *anteriorly distributed LPC (ant-LPC)* evoked by these critical words reflects process of successfully updating the discourse model, which, in this case, involves top-down suppression of the incorrectly predicted semantic features.

Top right: Low Constraint Event structure/Selection Restriction violation (LC_SRviol): *N400 time window:* The semantic features associated with <drawer> do not match any pre-activated semantic features, and they also mismatch the animate features that have been pre-activated, and so an even larger N400 is evoked by "drawer" than by the plausible unpredicted critical words described above. The unpredicted semantic features associated with <drawer> are passed up to the message level, depicted by the blue upward curved arrow. *Late positivity time window:* When the unpredicted semantic features associated with <drawer> reach the message level, a conflict is detected between the predicted event structure, <Agent cautions Patient>, and this new input, indicated by the red cross. Because the predicted event structure formed the boundaries the comprehender's current mental model, she cannot incorporate the new input into her existing mental model, depicted by the yellow starburst. This evokes the posteriorly distributed late positivity/P600 (post-LPC/P600). The post-LPC/P600 may also reflect a

reanalysis of the previous input in attempts to revise or repair the comprehender's existing mental model (not shown here).

Bottom right: High Constraint Event structure/Selection Restriction violation (HC_SRviol): *N400 time window:* The semantic features associated with <drawer> do not match the pre-activated semantic features associated with <swimmers>, and they also mismatch the animate features that have been pre-activated, and so an even larger N400 is evoked by "drawer" than by plausible unpredicted critical words. The unpredicted semantic features associated with <drawer> are passed up to the message level, depicted by the blue upward curved arrow. *Late positivity time window:* When the unpredicted semantic features associated with <drawer> reach the message level, a conflict is detected between the predicted event, <lifeguards caution swimmers>, and this new input, indicated by the red cross. The comprehender abandons her prediction but she cannot incorporate the new input into her existing mental model, depicted by the yellow starburst. This evokes the posteriorly distributed late positivity/P600 (post-LPC/P600). The post-LPC/P600 may also reflect a reanalysis of the previous input in attempts to revise or repair the comprehender's existing mental model (not shown here).

Icons used in this figure were taken from the www.thenounproject.com


TIME WINDOWS
300-500ms
600-1000ms
N400
Late Positivity
Low Constraint Unpredicted (LC_Unpred)
<they cautioned Patient>
<they cautioned trainees>
<animate>
<trainees>
N400↑
t-r-a-i-n-e-e-s
Low Constraint Event Structure/Selection Restriction violation (LC_SRviol)
<they cautioned ??!!>
post-LPC/ P600↑
N400↑↑
<drawer>
d-r-a-w-e-r
High Constraint Event/Lexical violation (HC_Lexviol)
<lifeguards cautioned swimmers>
<lifeguards cautioned trainees>
ant-LPC↑
<swimmers>
High Constraint Event Structure/Selection Restriction violation (HC_SRviol)
<lifeguards cautioned ??!!>
post-LPC/ P600↑↑