Multiple predictions during language comprehension: Friends, foes, or indifferent companions?

Trevor Brothers[1, 2], Emily Morgan[3], Anthony Yacovone[2,4], Gina Kuperberg[2,4]

[1] Department of Psychology, North Carolina A&T

[2] Department of Psychology, Tufts University

[3] Department of Linguistics, University of California, Davis

[4] Department of Psychiatry and the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital

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

Abstract

To comprehend language, we continually use prior context to pre-activate expected upcoming information, resulting in facilitated processing of incoming words that confirm these predictions. But what are the consequences of disconfirming prior predictions? To address this question, most previous studies have examined unpredictable words appearing in contexts that constrain strongly for a single continuation. However, during natural language processing, it is far more common to encounter contexts that constrain for multiple potential continuations, each with some probability. Here, we ask whether and how pre-activating both higher and lower probability alternatives influences the processing of the lower probability incoming word. One possibility is that, similar to language production, there is continuous pressure to select the higher-probability pre-activated alternative through competitive inhibition. During comprehension, this would result in relative costs in processing the lower probability target. A second possibility is that if the two pre-activated alternatives share semantic features, they mutually enhance each other's pre-activation. This would result in greater facilitation in processing the lower probability target. To distinguish between these accounts, we recorded ERPs as participants read three-sentence scenarios that constrained either for a single word or for two potential continuations – a higher probability expected candidate and a lower probability *second-best* candidate. We found no evidence that competitive pre-activation between the *expected* and *second-best* candidates resulted in costs in processing the *second-best* target, either during lexico-semantic processing (indexed by the N400) or at later stages of processing (indexed by a later frontal positivity). Instead, we found only benefits of pre-activating multiple alternatives, with evidence of enhanced graded facilitation on lower-probability targets that were semantically related to a higher-probability pre-activated alternative. These findings are consistent with a previous eye-tracking study by Luke and Christianson (2016, *Cogn Psychol*) using corpus-based materials. They have significant theoretical implications for models of predictive language processing, indicating that routine graded prediction in language comprehension does not operate through the same competitive mechanisms that are engaged in language production. Instead, our results align more closely with hierarchical probabilistic accounts of language comprehension, such as predictive coding.

Keywords: explaining away, prediction, predictive coding, competition, N400, late frontal positivity

Introduction

One of the most robust findings in the study of language comprehension is that the more predictable the input, the easier it is to process (Kuperberg & Jaeger, 2016; Ehrlich & Rayner, 1981; Rayner & Well, 1996). Numerous studies have shown that, relative to less predictable words, more predictable words are processed faster (Staub, 2015) and produce smaller evoked neural responses (Kutas & Hillyard, 1984; Delong, Urbach, & Kutas, 2005).

The most common explanation for this graded effect of predictability is that the prior context predictively *pre-activates* upcoming lexico-semantic information before new bottom-up input becomes available.[1] When a new word is encountered, its processing is facilitated to the degree that its semantic features have already been pre-activated. So long as these predictions are generated probabilistically, based on the statistics of the communicative environment, the ease of processing each incoming word should be inversely related to its prior probability, given the preceding context (DeLong, Urbach, & Kutas, 2005; Federmeier, 2007; Kuperberg & Jaeger, 2016). Indeed, we know from numerous studies that the magnitude of the N400 — an ERP component that is thought to reflect the ease of accessing (or retrieving) the semantic features associated with an incoming word (Kutas & Federmeier, 2011; Van Berkum, 2009; Kuperberg 2016) — is inversely related to that word's contextual predictability, regardless of whether this is estimated using standard cloze procedures (cf. Taylor, 1953; e.g. Kutas & Hillyard, 1984; DeLong, Urbach & Kutas, 2005; Wlotko & Federmeier, 2012), or using large language models

[1] We use the term "lexico-semantic" to refer to the semantic features associated with a particular word. We provide a more precise discussion about the relationship between these features and a word's conceptual and lexical representations towards the end of the Discussion. By "predictive pre-activation", we mean "the pre-activation of information at lower representational level(s) on the basis of information at higher levels within our internal representations of context, ahead of the bottom-up input reaching these lower levels" (Kuperberg & Jaeger, 2016, section 3, page 39). We do not make any assumptions about whether comprehenders pre-activate an upcoming word's orthographic or phonological features (see DeLong, Urbach, & Kutas, 2005, and Nieuwland et al., 2018 for debate).

(cf. Brown et al., 2020; e.g. Michaelov, Coulson, & Bergen, 2021; Szewczyk & Federmeier, 2022; Heilbron, Armeni, Schoffelen, Hagoort & De Lange, 2022).

This predictive pre-activation account can also explain why, in plausible sentences, the amplitude of the N400 produced by unpredictable words appearing in contexts that strongly constrain for an alternative continuation (e.g. *"He bought her a pearl necklace for her…collection"*) is no larger than the N400 produced by unpredictable words appearing in low constraint[2] contexts that do not strongly predict any single continuation (e.g. *"He looked worried because he might have broken his...collection"),* e.g. Kutas & Hillyard (1984); Federmeier, Wlotko, De Ochoa-Dewald & Kutas (2007); Kuperberg, Brothers, & Wlotko (2020). In both these situations, the incoming word's lexico-semantic representation has received no pre-activation, and so it will be relatively harder to access/retrieve, resulting in a relatively large N400 response.

In most accounts of predictive pre-activation, it is assumed that, rather than predicting one word at a time, comprehenders pre-activate *multiple* potential continuations in parallel, outside conscious awareness. For example, when reading "*Johnathan brewed the…*", readers might pre-activate the lexico-semantic representations of "*beer*", "*coffee*" and "*tea*" simultaneously, each with a different strength that is related to the probability of each lexico-semantic representation. As a consequence, encountering any of these words would produce a smaller N400 than a lower probability continuation (e.g. "*Johnathan brewed the <u>poison</u>*"). This, however, raises a question that has not yet been addressed in the prior literature: During this pre-activation phase, what influence do these multiple, pre-activated alternatives exert on one another (excitatory and/or inhibitory), and what impact does this have on processing the

[2] Contextual constraint is usually operationalized as the probability of the context's best (most probable) completion.

incoming word when it subsequently becomes available? In principle, there are three possibilities.

The first is that there are minimal interactions between the multiple pre-activated candidates until the new bottom-up input arrives. On this account, when the incoming word is encountered, the degree of facilitation it receives should depend solely on its own probability, *regardless* of the probability of any pre-activated alternatives. We refer to this as the *independent pre-activation* account.

The second possibility is that, during the pre-activation phase, these multiple pre-activated alternatives begin to compete, *mutually inhibiting* one another through a winner-take-all mechanism. A consequence of this mutual inhibition is that when an incoming word subsequently becomes available, it should receive *less* facilitation than one would expect given its estimated probability. We refer to this as the *competitive pre-activation* account.

Competitive interactions of this kind are implemented in classic Interactive Activation and Competition (IAC) models, which have been proposed as accounts of written word recognition (McClelland & Rumelhart, 1981; McClelland & Elman, 1986), spoken word recognition (Dahan, Magnuson, Tanenhaus & Hogan, 2001) and syntactic parsing (competitive ranked parallel models: Spivey & Tanenhaus, 1998; MacDonald, Pearlmutter & Seidenberg, 1994). Importantly, IAC architectures have also been proposed to implement language *production*, with mutual inhibition between lexical candidates playing a central role in selecting a single candidate for later articulation (e.g. Chen & Mirman, 2012). Recent work suggests that an IAC model (Chen & Mirman, 2012) can simulate times to produce predicted upcoming words in a speeded cloze completion task in which participants first comprehend a sentence context and then produce the most likely upcoming word; see Ness and Meltzer-Asscher (2021a) and Nakamura and Phillips

(2022). Thus, if this type of competitive mutual inhibition operates during the pre-activation phase of language *comprehension*, this would provide evidence that top-down prediction during language comprehension is *routinely* implemented through one of the same processing mechanisms that is employed in language production (Pickering & Garrod, 2013; Fitz & Chang, 2019; see also Van Petten & Luka, 2012; Thornhill & Van Petten, 2012)..[3]

The third possibility is that instead of acting as competitors, multiple pre-activated lexico-semantic candidates (e.g. *beer, coffee, tea*) serve to *reinforce* one another if they share semantic features. On this account, the presence of a semantically related pre-activated alternative would actually lead to *more* facilitation and a *smaller* N400 response than one would expect based on that word's lexical probability. We will refer to this as the *friendly pre-activation* account.

Evidence consistent with friendly pre-activation comes from the ERP studies showing that the pre-activation of semantic features can facilitate the processing of incoming words during language comprehension, even if these words are lexically unpredictable. For example, the N400 is reduced in response to *implausible* words that are semantically related to a predictable alternative (Kutas and Hillyard 1984; Federmeier & Kutas, 1999; for recent replications, see DeLong, Chan & Kutas, 2019; Ito, Corley, Pickering, Martin & Nieuwland, 2016). This anticipatory semantic facilitation effect on the N400 has also been described on unexpected (zero cloze) *plausible* continuations (Thornhill & Van Petten, 2012; DeLong & Kutas, 2020). However, no previous study ERP has asked whether, in contexts that constrain for *multiple* continuations, less expected but non-zero probability words can receive facilitation from a higher probability pre-activated alternative as a function of semantic overlap.

Finally, we note that the *competitive and friendly pre-activation* accounts are not mutually

[3] Note, this wouldn't imply that production-like mechanisms are *never* engaged in implementing top-down prediction during language comprehension, particularly in very high constraint contexts (see Federmeier, 2022).

exclusive. For example, some IAC architectures include both intra-lexical competition *and* mutual reinforcement from shared semantic features (Chen & Mirman, 2012). These architectures might predict that a lower probability word would receive *less* facilitation than expected if it is semantically *unrelated* to a higher probability alternative, but *more* facilitation than expected if it shares semantic features with the higher probability alternative. Indeed, Ness and Meltzer-Asscher (2021a) recently showed behavioral evidence consistent with this type of hybrid account in a speeded cloze production task.

Understanding whether and how pre-activated lexico-semantic alternatives interact with one another is important not only for understanding the nature of routine predictive processing during language comprehension (including its relationship with language production), but also because of its ecological validity. In natural language, it is relative rare to encounter low probability words that violate a very high-constraint context. In contrast, we frequently encounter moderately constraining contexts that are predictive of multiple alternatives. Moreover, in these contexts, readers often encounter inputs that disconfirm the most probable continuation but confirm a lower probability continuation.

This pattern was demonstrated in a key study by Luke and Christianson (2016), who measured cloze probability in a corpus of mixed-genre texts by asking readers to predict each word in turn. The authors found that most words in these naturalistic texts were only somewhat predictable, with content words having an average cloze probability of 13%. Nonetheless, readers were more consistent, on average, in their *expectations* about upcoming content words, resulting in an average lexical constraint of 36%. Strikingly, although most incoming words disconfirmed an individual reader's *most common* prediction, 79% of words matched a cloze continuation that was produced by at least some participants.

Luke and Christianson also examined eye-movement data as participants read these texts for comprehension. Reading times were analyzed as a function of (a) the cloze probability of each word, and (b) whether or not each word was the most probable continuation produced by all participants in the cloze task. They found only an effect of cloze probability, but no interaction with whether or not the word was the most probable completion. In addition, the authors calculated the semantic relationship between each content word and the full set of offline cloze responses produced in response to the prior context. They found enhanced behavioral facilitation on words that were more semantically related to these alternative predictions. Thus, taken together, Luke and Christianson's findings suggest that, during natural reading, parallel lexico-semantic pre-activation provides benefits (*friendly pre-activation*), but no costs (no *competitive pre-activation*), on reading times.

Goals of the present study

Given the theoretical and ecological importance of Luke and Christianson's behavioral findings, we wanted to carry out a conceptual replication of their work. We had three main goals.

First, we wanted to determine whether Luke and Christianson's results held up using controlled experimental materials. We see the use of naturalistic and controlled experimental stimuli as complementary approaches. Luke and Christianson (2016) provided key data about the distributions of cloze and constraint values within naturalistic text by virtue of using texts gathered from a variety of real-world sources such as news articles and fiction. Moreover, their use of extended multi-sentence texts provides the most ecologically valid reading experience for experimental participants. However, as the authors discuss, with these types of stimuli, it is difficult to dissociate effects of lexical properties such as word length and frequency from effects of context-specific predictability, since these effects are inherently confounded in naturally occurring texts. Naturalistic corpus studies also introduce the possibility of uncontrolled spill-over effects, or other effects arising from uncontrolled properties of the text (Rayner, Pollatsek, Drieghe, Slattery & Reichle, 2007; Brothers, Hoversten & Traxler, 2017; Angele et al., 2015; see Brothers & Kuperberg, 2021 for recent discussion). Thus, the use of controlled experimental stimuli offers an opportunity to specifically test the hypotheses motivated by the theories described above, while controlling for lexical factors such as word length and frequency.

Second, we were interested in replicating Luke and Christianson's behavioral findings using a different technique — ERPs instead of reading times. ERPs provide a time-sensitive measure of online comprehension. The N400 component, in particular, is known to be sensitive to many of the same factors that influence reading times, including frequency, predictability, and semantic overlap, and this component has played a central role in debates on the role of prediction and misprediction in language comprehension (Van Petten & Luka, 2012; see also Federmeier, 2007; DeLong, Urbach & Kutas, 2005; Nieuwland et al., 2018). Moreover, although many classic connectionist and neural network models of language processing were originally developed to simulate behavioral findings, ERPs provide an important test case of the computational principles implemented by these models (see Nour Eddine, Brothers, & Kuperberg, 2022 for a comprehensive review). ERPs therefore provide an important and complementary perspective to behavioral findings.

Third, ERP methods allow us to examine not just the initial stages of lexico-semantic processing, indexed by the N400, but also later ERP components that might be particularly sensitive to the disconfirmation of prior predictions. Previous behavioral studies have found little evidence of late processing costs on lower probability continuations that are inconsistent with a prior higher probability prediction (Luke and Christianson, 2016; Frisson, Harvey, & Staub, 2017; Steen-Baker, Ng, Payne, Anderson, Federmeier & Stine-Morrow, 2017; Fischler & Bloom, 1979, 1985; Schwanenflugel and LaCount,1988; but see Ness & Meltzer-Asscher, 2021b, who showed that participants took longer to make speeded congruency decisions in two-word phrases in which the second word violated a prediction, e.g. "*rearview camera*" where mirror was predicted vs. "*desert storm*" where there was no strong prediction). However, several ERP studies have reported that in plausible sentences, unexpected (zero-cloze) words appearing in contexts that constrain strongly for a single alternative (e.g. *"He bought her a pearl necklace for her…collection"*) can sometimes produce a larger late frontally-distributed positive component between 500–1000ms, in comparison with unexpected words appearing in low constraint contexts (e.g. *"He looked worried because he might have broken his...collection"*, see Federmeier et al., 2007; Kuperberg, Brothers, & Wlotko, 2020; Lai, Rommers, & Federmeier, 2021).

One possible interpretation of this late frontal positivity effect is that it indexes late processing "costs" associated with suppressing an incorrect lexical prediction (Kutas, 1993; Ness & Meltzer-Asscher, 2018). This would follow from an account in which the prediction of a higher-probability pre-activated alternative remains active in the late time window and, in order to successfully integrate the lower-probability incoming word into its prior context, it is necessary to suppress/inhibit this incorrect prediction within this later time window. This *late suppression* account would therefore predict a larger late frontal positivity on *less* versus *more* probable words in contexts that constrain for multiple continuations, where there would also be additional demands to inhibit an incorrectly pre-activated competitor.

Design and questions addressed in the present study

To address the questions outlined above, we developed a set of materials with two types of contexts, each continuing with either a more or a less expected critical word.

First, *WithCompetitor* contexts, such as (1), always constrained for *two* upcoming words — a more probable expected alternative (e.g. *hearts*) and a less probable alternative (e.g. *flowers*). Following these contexts, participants saw either the *Expected* critical word ("hearts") or the *SecondBest* critical word ("flowers").

Second, *NoCompetitor* contexts, such as (2), always constrained for just *one* upcoming continuation (e.g. *roses*). Participants saw either this *Expected* critical word or a *ZeroCloze* but plausible critical word (e.g. *rocks*).

(1) *WithCompetitor context:*

> Stephen wanted to do something special for his girlfriend. He decided to make her a hand-made card. On it, he drew some… *Expected:* **hearts** / *SecondBest:* **flowers**

(2) *NoCompetitor context*:

> Alexis was thrilled with her new garden. All of the flowers had bloomed overnight. In particular, she loved the… *Expected:* **roses / ***ZeroCloze:* **rocks**

In the *WithCompetitor* contexts, the cloze probabilities of the *SecondBest* critical words were, on average, lower than those of the *Expected* critical words in both the *WithCompetitor* and *NoCompetitor* contexts. However, as discussed in the Methods, the range of cloze probability values within each condition was wide, allowing us to address our hypotheses at varying degrees of predictability. Specifically, we addressed three sets of questions.

First, is there any evidence that *competitive pre-activation* influences the magnitude of the N400? The *competitive pre-activation* account predicts that in *WithCompetitor* contexts, prior to the presentation of the incoming word, there should be some degree of *mutual inhibition* between the two pre-activated alternatives (e.g. between <*hearts*> and <*flowers*> in (1) above). This should result in *less* facilitation and a *larger* N400 in response to both the *SecondBest* and the *Expected* critical word than one would expect based only on their cloze probabilities. In contrast, if the N400 response to each of these words remains proportional to its cloze probability, with no penalty for having pre-activated a competitor, this would provide evidence for the *independent pre-activation* account. For specific details about how we statistically tested these hypotheses, see Models 1 and 2 in the Results section.

Second, is there any evidence that *friendly pre-activation* influences the N400? The *friendly pre-activation* account predicts that there should be a facilitatory effect on the N400 produced by a lower probability word whenever it shares semantic features with a higher probability pre-activated alternative. To address this question, we began by considering all scenarios in which a lower probability critical word appeared in place of a potentially pre-activated alternative—namely, the *NoCompetitor ZeroCloze* and the *WithCompetitor SecondBest* scenarios, asking if there was any additional facilitation on the N400 as a function of how semantically related the observed word was to the more expected continuation (Model 3, Results). In addition, because previous ERP work has only demonstrated this type of facilitatory effect on zero-cloze words (Kutas and Hillyard 1984; Federmeier & Kutas, 1999; Thornhill & Van Petten, 2012; DeLong & Kutas, 2020), we separately tested for evidence of friendly pre-activation on the subset of *WithCompetitor SecondBest* continuations, which always had non-trivial cloze probabilities (Model 4, Results).

Our investigation of the N400 across these scenarios provided a strong test of whether pre-activating multiple alternatives results in *competition* or *facilitation*. However, as noted above, *competitive* and *friendly pre-activation* are not mutually exclusive, and, in principle, both can influence the degree of facilitation on an incoming word, jointly impacting the amplitude of the N400. Therefore, we also wanted to test for an effect of competitive pre-activation in isolation of any facilitatory effects. To do this, we conducted an analysis on the subset of *SecondBest* words that were *unrelated* to the expected continuation (Model 5, Results).

Third, and finally, we turned to the question of whether there was any evidence of late suppression on the late frontal positivity produced by *SecondBest* critical words in the *WithCompetitor* contexts. According to the *late suppression* account, late costs should be incurred when suppressing an unobserved strongly pre-activated alternative in order to integrate an observed lower-probability incoming word into the prior context. In the *WithCompetitor* contexts, this would predict a larger late frontal positivity on *SecondBest* than on *Expected* completions. We therefore compared the amplitude of the late frontal positivity to these two types of completions (Model 6, Results).

Methods

Materials

Overall Design

Our stimuli consisted of plausible, three-sentence discourse scenarios. This three-sentence stimulus design was based on prior studies (Kuperberg, Brothers, & Wlotko, 2020; Brothers, Wlotko, Warnke, & Kuperberg, 2020) and provided a slightly more natural reading experience than one-sentence stimuli. In each scenario, the first two sentences introduced the scenario using a variety of sentence structures. The third sentence was more controlled and consisted of an adverbial phrase, the subject, a transitive verb, an optional determiner, a direct object critical noun, and 3–4 words to conclude the sentence.

As described in the Introduction, the prior context either constrained for either *one* or *two* upcoming words, i.e., *WithCompetitor* and *NoCompetitor* contexts, respectively (see Table 1). In the *WithCompetitor* contexts, participants either saw the *Expected* (A1) or *SecondBest* (A2) continuation. In the *NoCompetitor* contexts, participants either saw the *Expected* continuation (A3) or a *ZeroCloze* but plausible (A4) continuation. We refer to these four conditions as the *TargetScenarios*. Because the specific critical words varied across the four types of *TargetScenarios*, to control for low-level lexical differences across items, we also constructed a set of *ControlScenarios*, which used the same four critical words as those used in the *TargetScenarios* (B1–B4). These were generated by writing two new introductory sentences and pairing these with the same final sentences used in the *TargetScenarios.* In these *ControlScenarios*, there were no clear preferences for a particular continuation, making them relatively non-constraining and non-competitive in nature.

In Table 1, we present all of the eight conditions described above—namely, the four conditions in the *TargetScenarios* group (A1: *WithCompetitor Expected*, A2: *WithCompetitor SecondBest*, A3: *NoCompetitor Expected*, A4: *NoCompetitor ZeroCloze*), and the four conditions with lexically-matched critical words in the *ControlScenarios* group (B1–B4). All scenarios were written to be semantically plausible (see below for a rating study that verified that this was the case).

Table 1. Stimuli

Stimulus Group	Context Type	Continuation Type	Average Cloze (SD)	Example Contexts with both Continuation Types
Target Scenarios	*WithCompetitor*	*Expected* (A1)	57.4% (14.7%)	Stephen wanted to do something special for his girlfriend. He decided to make her a hand-made card. On it, he drew some… **hearts** (A1) / **flowers** (A2)
		SecondBest (A2)	16.3% (8.5%)	
	NoCompetitor	*Expected* (A3)	60.9% (15.0%)	Alexis was thrilled with her new garden. All of the flowers had bloomed overnight. In particular, she loved the… **roses** (A3) / **rocks** (A4)
		ZeroCloze (A4)	0.1% (0.4%)	
Control Scenarios	Controls for *WithCompetitor*	Control for A1 (B1)	4.5% (7.2%)	Stephen always doodled in class. He took out a fresh sheet of paper. On it, he drew some… **hearts** (B1) / **flowers** (B2)
		Control for A2 (B2)	3.0% (5.0%)	
	Controls for *NoCompetitor*	Control for A3 (B3)	5.5% (7.9%)	Alexis had just moved to a new city. She enjoyed exploring new sites. In particular, she loved the… **roses** (B3) / **rocks** (B4)
		Control for A4 (B4)	0.3% (1.3%)	

Cloze norming and item selection

To create and classify our stimuli into the eight conditions described in Table 1, we carried out cloze norming studies. For this, we recruited participants from within the United States through Amazon Mechanical Turk. Based on self-report, all participants were between the ages of 18–35 and their native language was English. All participants provided informed consent, and they were compensated for their time.

On each trial in this task, participants saw one stimulus item up to (but not including) the critical word. They were then asked to respond with “the most likely next word” (Taylor, 1953). After providing their response, participants were then asked to give two further possible responses for each context, each with the prompt, “Please enter another likely next word.” In this way, we obtained the first, second, and third best continuations from each subject for each item (see also Federmeier et al, 2007; Schwanenflugel, Harnishfeger & Stowe, 1988). All participants completed a guided practice before viewing the experimental stimuli. Stimuli were normed in batches of different sizes (from 4–74 stories), which took approximately 3–60 minutes. Participants were paid up to $6 per hour for their time. At least 50 participants provided continuations for each context. In total, over 800 vignettes were normed for consideration.

For each item, we determined its contextual constraint by identifying the most common completion for each context, and then calculating the percentage of participants who provided that particular word as their first response. The cloze probability of each critical word was calculated as the percentage of respondents providing that word as their first response. We will refer to this as the *top-1 cloze probability*. In addition, for each item, we calculated a measure of cloze probability that was based on the percentage of respondents for whom the critical word was *one of their three responses.* As discussed below, this was particularly important in allowing us to identify *WithCompetitor* contexts, even in contexts where most of the *top-1 probability* mass was taken up by the most expected target word. We will refer to this second measure as the *top-3 cloze probability*.

Using these norming data, we then constructed a set of 60 *WithCompetitor* and 60

NoCompetitor contexts with a wide range of contextual constraints (33–92%). As described above, the *NoCompetitor* contexts (*Mean constraint*: 61%) did *not* have a second continuation with non-trivial cloze probability (i.e. all alternative continuations provided in the cloze task had a top-1 cloze probability below 10%). The *NoCompetitor Expected* (A3) words had a top-1 cloze probability of at least 33% and were always the modal completion for their context. The *NoCompetitor ZeroCloze* (A4) words were chosen to have a cloze probability of near 0%.[4]

As also described above, the *WithCompetitor* contexts (*Mean constraint*: 57%) could be followed by at least two continuations with non-trivial cloze probabilities. The *WithCompetitor Expected* (A1) critical words had a top-1 cloze probability of at least 36% and were always the modal completion for their context. The *WithCompetitor SecondBest* (A2) words either had a top-1 cloze probability greater than 10% *or* a top-3 cloze probability greater than 25%.[5] These *SecondBest* completions were typically listed as being "second most likely" by participants. However, in a few cases, we instead took the third best completions in order to avoid repeating a critical word within the experiment. To the extent possible, all critical words were matched across conditions in length, log word frequency, orthographic neighborhood size, and semantic concreteness.

Across the *ControlScenarios* (60 for the *WithCompetitor* conditions and 60 for the *NoCompetitor* conditions: B1–B4), the average constraint was 19% with a range of 8–52%. The critical words, when presented in these control discourse scenarios, had an average top-1 cloze probability of below ~5%.

[4] 3 out of 60 words in this condition were chosen by a single participant (out of >50 participants) in cloze norming, giving them a non-zero cloze probability of <2%. The remainder had 0% cloze probability.

[5] The choice of the exact values of 10% and 25% was arbitrary, reflecting our intuition of approximately what constitutes a non-trivial competitor. However, we do not expect there to be a true categorical cutoff for what constitutes a non-trivial competitor.

Across the full stimulus set, the contexts had an average constraint of 39% and an average cloze probability of 18.5% for the target and control critical words. Therefore, unlike some prior studies, these levels of constraint and predictability were highly similar to those encountered in naturalistic texts (Luke & Christianson, 2016).

Plausibility norming

Late frontally distributed positivities are only produced by unexpected words when they can be plausibly integrated into their prior contexts (e.g. Van Petten & Luka, 2012; Kuperberg, Brothers, & Wlotko, 2020). In contrast, highly implausible/anomalous critical words often produce a late posteriorly distributed positivity effect, known as the P600 (Kuperberg, 2007, section 3.4 page 32; Kuperberg et al., 2003; van de Meerendonk, Kolk, Vissers & Chwilla, 2010; Paczynski & Kuperberg, 2012; Kuperberg, Brothers, & Wlotko, 2020), while mildly implausible words tend not to produce robust late positivities at all (e.g. Kuperberg, Sitnikova, Caplan & Holcomb, 2003; van de Meerendonk, Kolk, Vissers & Chwilla, 2010). We therefore wanted to verify that our scenarios were indeed plausible. To do this, we conducted a plausibility norming study using the online platform, Prolific (www.prolific.co). We recruited a set of participants in the US and UK who listed their first language as English, and then asked them to rate various scenarios on a scale of 1–7 (1 = "makes no sense at all"; 7 = "makes perfect sense").

In this norming study, we included not only all the scenarios from the current study, but also sets of three-sentence scenarios from prior studies run in our lab that had previously been normed to be *plausible,* highly *implausible/anomalous*, and *semi-implausible.* Specifically, we included (a) the *high constraint expected* and *high constraint unexpected* scenarios from a study by Kuperberg, Brothers and Wlotko (2020), which had been normed to be *plausible,*[6] (b) the *high constraint anomalous* scenarios, also from Kuperberg, Brothers and Wlotko (2020), which contained selection restriction violations and were therefore *highly implausible*, and (c) a set of scenarios from Greene, Brothers, Weber, Noriega, and Kuperberg (2020), without selection restriction violations, that had previously been normed to be *semi-implausible*.

To keep the task length reasonable, each participant saw a subset of items: 60 from the current experiment, 38 or 39 from the previous studies, as well as 10 sanity check items that were designed to be either very plausible or highly implausible. Items were counterbalanced such that no participant saw more than one version of each scenario. Each version was rated by 10 participants.

The results are given in Table 2. They confirm that the *TargetScenarios* (A1–A4) and *ControlScenarios* (B1–B4) in the present study all received plausibility scores that were higher than those of the *highly implausible/anomalous* and *semi-implausible* scenarios used in our previous work.

Table 2. Mean plausibility ratings for all conditions in the present study and those from Kuperberg, Brothers, and Wlotko, 2020 (KBW20) and Greene, Brothers, Weber, Noriega, and Kuperberg, 2020 (GBWNK20).

Experiment	Condition	Plausibility
Present study	*WithCompetitor Expected* (A1)	6.66
Present study	*WithCompetitor SecondBest* (A2)	6.45
Present study	*NoCompetitor Expected* (A3)	6.55
Present study	*NoCompetitor ZeroCloze* (A4)	5.10
Present study	*ControlScenario* for A1 (B1)	5.55
Present study	*ControlScenario* for A2 (B2)	5.35
Present study	*ControlScenario* for A3 (B3)	5.20
Present study	*ControlScenario* for A4 (B4)	4.84

[6] 33 items from Kuperberg, Brothers, & Wlotko (2020) were not included because they were extremely similar or identical to the scenarios used in the present study.

KBW20	*Expected*	6.57
KBW20	*Unexpected* (plausible)	5.29
GBWNK20	*Semi-implausible* (no selection restriction violations)	2.89
KBW20	*Anomalous* (selection restriction violations)	1.89

Counterbalancing of stimuli and construction of final stimulus lists

In the ERP experiment, we created four counterbalanced lists. Each participant saw each context in the *TargetScenarios* and *ControlScenarios* just once. However, our counterbalancing scheme worked to ensure that they never saw the same critical word twice; that is, if a participant saw a critical word in a *WithCompetitor* or *NoCompetitor* context, they saw a different critical word in the *ControlScenario* for that item. In addition, participants did not see a *WithCompetitor* or *NoCompetitor* trial and its corresponding *ControlScenario* in the same half of the experiment. Presentation order was counterbalanced across participants, and participants were randomly assigned to one of these four lists.

Participants

We report data from 32 native English speakers who were recruited from Tufts University and the surrounding community. Fifteen further participants were tested but were subsequently excluded because of excessive noise in the ERP recording (see Data Preprocessing for cutoff criteria)..[7] The final set of participants were between the ages of 18 and 35 (*Mean age* = 24.0; *SD*: 4.1). All were right-handed and had normal or corrected-to-normal vision. All participants reported having no significant exposure to any language other than English before the age of 5, no history of neurological disorder(s), and no current use of psychoactive medication. All participants

[7] This relatively high exclusion rate was because data were collected on a new high impedance system, and in these participants, we used minimal scalp abrasion prior to data collection.

provided written informed consent and were paid for their time at a rate of $15 per hour. All protocols were approved by Tufts University Social, Behavioral, and Educational Research Institutional Review Board.

Stimulus Presentation

Participants sat in a comfortable chair in a dimly lit room approximately 150cm from the LCD computer monitor. They were asked to minimize muscle activity, eye movements, and blinking, particularly while reading the sentences. Stimuli were presented word-by-word using PsychoPy 1.83 software (Peirce, 2007). Each word was presented in a white Arial font on a black background, with 3 characters covering approximately 1 degree of visual angle. Each trial began with the prompt “READY?” in green font presented at the center of the screen, and the participant pressed a button to advance. The first two sentences of each context appeared in full, separated by a button press. After participants read the second sentence, a fixation marker (“++++”) appeared for 750ms before the third sentence appeared, one word at a time, in the center of the screen (450ms word duration, 100ms inter-stimulus interval).

In 128 trials, a *yes/no* comprehension question appeared immediately after the third sentence. These sentences often required readers to draw inferences based on the entire scenario, and they never referred specifically to the critical words. Their purpose was to encourage participants to attend to and deeply comprehend the scenarios.

Within each half of the experiment, the order of item presentation was randomized individually for each participant. Trials were presented in blocks of 30 items, which generally took 8–10 minutes, with a break between each block. Prior to seeing the experimental trials, participants saw 12 practice items with a similar structure to the experimental items.

ERP Acquisition and Preprocessing

We recorded ERPs using the BioSemi ActiveTwo EEG system with ActiView v7.05 EEG acquisition software (http://www.biosemi.com/). We recorded from 32 active Ag/AgCl electrodes in an elastic cap placed according to a modified international 10-20 system. Additional electrodes were placed below the left eye and beside the right eye to monitor for blinks and eye movements, as well on each mastoid to serve as reference. The EEG signal was amplified, digitally filtered online with the Biosemi Active-Two acquisition system using a low pass 5th order sinc response filter with a half-power cutoff at 104 Hz, and continuously sampled at 512Hz.

Data was processed using the EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) toolboxes in MATLAB. EEG channels were referenced offline to the average of the left and right mastoid channels. A 2nd order Butterworth IIR filter with a half-amplitude high pass cutoff of 0.1 Hz was applied offline. The ERP was then segmented into epochs spanning from –300ms to 900ms, time-locked to the critical word. Only trials free from ocular, muscular, and electrical artifacts were included in analysis, as determined by preprocessing routines from the EEGLAB and ERPLAB toolboxes using participant-specific artifact detection thresholds, combined with manual inspection. To be included in the analysis, participants had to have at least 15 artifact-free trials per condition, and at least 160 artifact-free trials overall (across the 8 conditions). On average, 18% of trials were rejected for artifacts for the included participants, and artifact rejection rates did not different significantly across the eight conditions, $F(7,255) = 0.58$, $p = 0.77$.

ERP statistical analysis

We extracted single trial artifact-free ERP data, using a baseline of –300 to 0ms, by averaging across electrode sites and time windows in two spatiotemporal regions of interest, which were selected *a priori*, based on previous studies using a similar design (Kuperberg, Brothers, & Wlotko, 2020; Brothers, Wlotko, Warnke, & Kuperberg, 2020). The N400 was operationalized as the average voltage between 300–500ms across five central electrode sites (Cz, CPz, C3/4, CP1/2). The late frontal positivity was operationalized as the average voltage between 600-900ms across five prefrontal electrode sites (FPz, FP1/2, AF3/4).

We analyzed these trial-level data using a series of linear mixed-effects regression models, which allowed us to look for effects of categorical predictors as well as continuous item-level predictors. This random effect for items refers to contexts (not individual target words). Thus, all trials that use the same introductory scenarios (and their control contexts) have the same item label.

All regression analyses were conducted in R (R Core Team, 2022), using the *lme4* package version 1.1–31 (Bates et al., 2015) and *lmerTest* version 3.1–3 (Kuznetsova et al., 2017). Following Barr et al. (2013), all regression analyses included the maximal random effects structures justified by the design both by subjects and by items. Random effect correlations were included by default. However, if we encountered issues with model convergence or singular fits, we removed these correlations. If this step did not resolve the issues, we continued to simply the random effects structure until reaching convergence without singular fits.

For the analyses designed to test the *friendly pre-activation* account (Models 3 and 4*),* we calculated the semantic relatedness between the lower and higher probability words in all of our *TargetScenarios*. Specifically, we calculated the semantic relatedness (1 – cosine distance) on an item-by-item basis between the *WithCompetitor SecondBest* (A2) and the *WithCompetitor Expected* (A1) words, and between the *NoCompetitor ZeroCloze* (A4) and *NoCompetitor Expected*

(A3) words. The semantic vectors used for these computations were obtained using a predictive Continuous Bag of Words model (see Mandera, Keuleers, & Brysbaert, 2017; http://meshugga.ugent.be/snaut-english, 300 dimensions, window size = 6).

We also conducted two additional analyses that are only reported in the Supplementary Materials/OSF. First, we conducted a series of Bayesian analyses. This is because, to foreshadow our findings, we report a series of null main effects and interactions that are critical to distinguishing the theoretical accounts under discussion. For these supplementary analyses, we calculated the Bayes Factor (BF_{01}) to quantify the evidence in favor or against these null findings. Second, to explore the possibility of other late ERP effects related to misprediction (e.g. a left frontally distributed negativity described by Wlotko and Federmeier's, 2012), which might occur outside our spatiotemporal regions of interest, we implemented a series of Mass Univariate analyses across all time points from 600–900ms and all electrode sites (except for temporal sites), correcting for multiple comparisons using a cluster-based approach.

Results

Behavioral results

Comprehension question accuracy across all conditions was 91% (on average), suggesting that readers were attending carefully to the discourse contexts.

ERP Results

In Figure 1, we show grand-average ERPs produced by the *WithCompetitor Expected* (A1) and *SecondBest* (A2) critical words in the *TargetScenarios*, along with the collapsed grand-averages produced by the same critical words in the *ControlScenarios* (B1–B2). Between 300–500ms, the

WithCompetitor Expected critical words elicited the smallest N400 responses; the *WithCompetitor SecondBest* critical words, which, on average had moderate cloze probabilities, produced a larger N400 responses, and the *ControlScenario* continuations, which, on average, had low cloze probabilities, produced the largest N400 response. Beyond the N400 time-window, between 600–900ms, the *WithCompetitor SecondBest* words appeared to produce a slightly larger positivity at frontal sites than the critical words in the *ControlScenarios.*

In Figure 2, we show grand-average ERPs from the *NoCompetitor Expected* (A3) and *ZeroCloze* (A4) conditions, as well as ERPs produced by the same critical words appearing in the *ControlScenarios.* Here, the critical words in the two unexpected conditions (*ZeroCloze*, *ControlScenarios*) had similarly low cloze probabilities, and produced a larger N400 than the *NoCompetitor Expected* critical words. Beyond the N400 time-window, there again appeared to be a slightly larger positivity at frontal sites between 600–900ms to the *NoCompetitor ZeroCloze* words relative to the critical words in the *ControlScenarios*.

ERP plots for all conditions individually at all electrodes sites and all time points are included in Supplementary Materials.

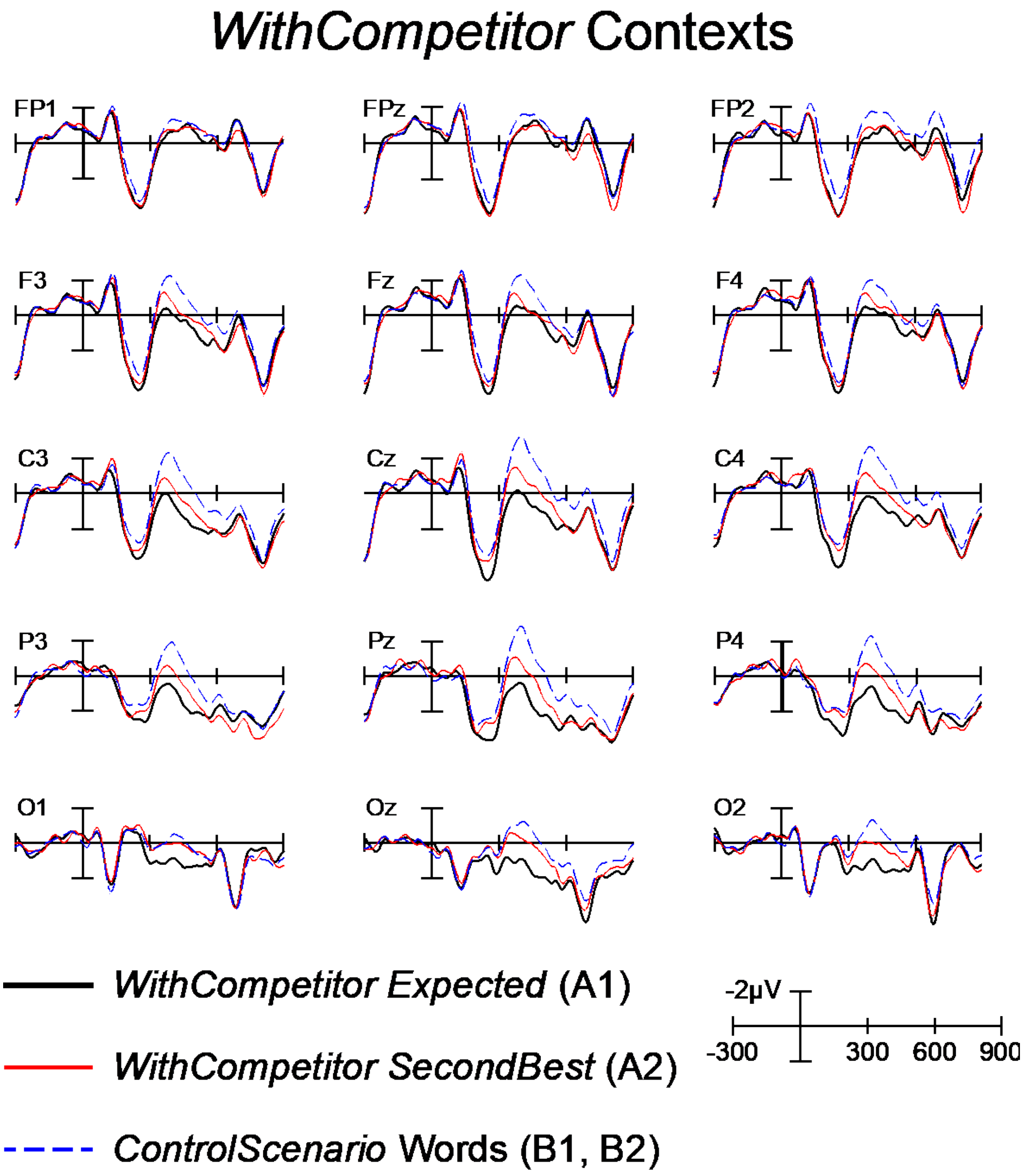

Figure 1. Grand-average event-related potentials, time-locked to the onset of *Expected* and *SecondBest* critical words in the *WithCompetitor* contexts, and to the same critical words appearing the *ControlScenarios.*

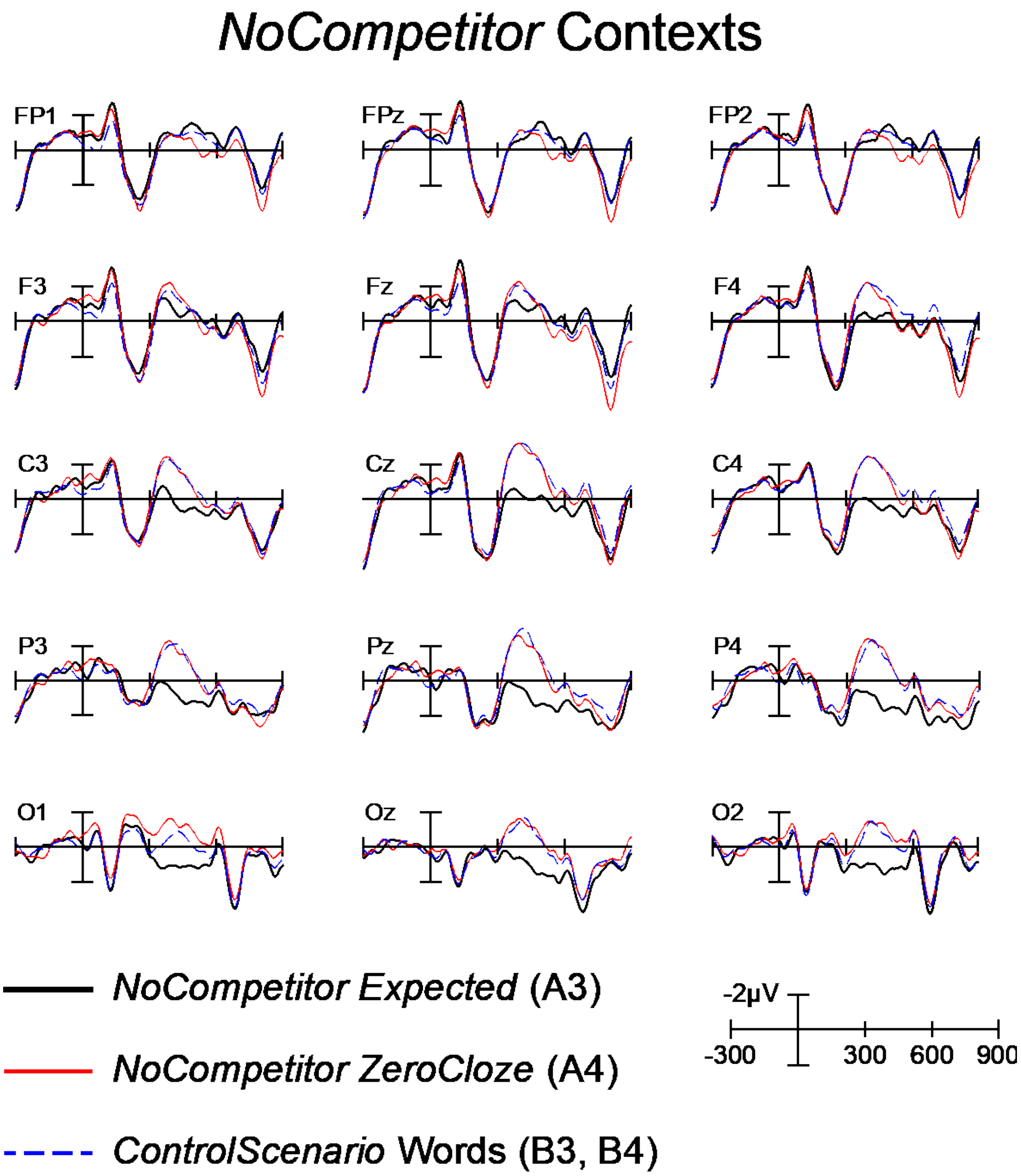

Figure 2. Grand-average event-related potentials, time-locked to the onset of *Expected* and *ZeroCloze* critical words in the *NoCompetitor* contexts, and to the same critical words appearing the *ControlScenarios.*

It is important to note that although there were clear differences between conditions in the *mean* cloze probability of the critical words, which was reflected by the differences in N400 shown in Figures 1 and 2, there was considerable variability in cloze values *within* each of these conditions. Indeed, several conditions actually overlapped in their ranges of cloze probabilities (see Methods). For example, the range of *top-1* cloze probabilities for critical words in the *WithCompetitor SecondBest* condition was 0–36%, which overlapped with the ranges in the *WithCompetitor Expected* (36–92%), *NoCompetitor Expected* (33.3–88%), *NoCompetitor ZeroCloze* (0–1.9%) conditions, as well as with the range of critical words in the *ControlScenarios* (range across B1–B4, 0–36%). Therefore, to test the divergent predictions made by the *competitive pre-activation*, *independent pre-activation*, and *friendly pre-activation* accounts, we used condition labels as categorical predictors (where relevant), while controlling for item-specific cloze probability as a continuous predictor variable.[8] This enabled us to address three sets of questions.

[8] For all analyses that included the *WithCompetitor* contexts, we also carried out the same set of analyses with an independent measure of lexical probability, taken from the large language model, GPT-3 (Brown et al., 2020). When we estimate lexical probability via cloze responses from multiple individuals, we assume that each individual response represents a noisy sample from an internal probability distribution of representation within an individual brain, and so multiple averaged guesses should provide a more precise estimation of these probabilistic representations than responses from a single individual – the so-called "wisdom of the crowd" effect (Galton, 1907; Stroop, 1932). Thus, if two words in two different contexts are matched in cloze probability, we tend to assume that their levels of pre-activation are also matched. However, it is possible that this assumption may not hold for the *WithCompetitor* contexts. This is because competition amongst pre-activated representations could, in principle, directly influence the probability with which these items are *produced* during the cloze task. For example, inhibition from the more probable alternative may reduce pre-activation of the *SecondBest* representation, making it less likely that this *SecondBest* continuation is produced than if it appeared in a *NoCompetitor* context. If this was the case, then it could lead to a systematic underestimation of the lexical probability of these *SecondBest* items, which would, in turn, reduce the likelihood of detecting evidence of competitive pre-activation during language *comprehension*. Our use of an independent corpus-based measure of lexical probability helped rule out this potential confound: For all tests of theoretical interest, we observed the same pattern of results using these GPT-3 estimates (see Supplementary Materials/OSF for results of these analyses in the annotated R script).

Question 1: Competitive or independent pre-activation? Evidence from the N400

We began by asking whether there was any evidence of *competitive pre-activation*. This account predicts that in contexts with more than one likely continuation, there should be some degree of mutual inhibition between pre-activated alternatives prior to the incoming word becoming available. For example, in the *WithCompetitor* context from Table 1, the pre-activated lexico-semantic representations, <hearts> and <flowers>, should inhibit one another during the prediction phase, resulting in *less* facilitation and a *larger* N400 than one would expect (based only on cloze probability) to the incoming *SecondBest* and *Expected* critical words. In contrast, the *independent pre-activation* account predicts that the amplitude of the N400 evoked by these words should depend solely on their cloze probabilities, regardless of the presence of a co-activated alternative. We tested these divergent predictions with Models 1 and 2.

Model 1: Does competitive pre-activation result in an increased N400 to SecondBest words?

In Model 1, we focused on the N400 in response to the *WithCompetitor Expected* (A1), and *WithCompetitor SecondBest* (A2) critical words, while controlling for Item-specific cloze probability. In order to control for low-level lexical variables, we also included the N400 responses from the identical set of critical words appearing in their associated *ControlScenarios* (B1 and B2, respectively). Thus, this model included fixed effects of Item-specific Cloze, Continuation Type (*Expected* = –0.5, *SecondBest* = 0.5), Stimulus Group (*TargetScenarios* = 0.5, *ControlScenarios* = –0.5), and an interaction between Continuation Type and Stimulus Group.

If there is competitive pre-activation in the *WithCompetitor* contexts, then we should see larger N400 responses to the *SecondBest* words, relative to *Expected* words, after controlling for item-specific cloze probability. In contrast, we should see no such difference between the N400 produced by the same critical words appearing in their *ControlScenarios*. In other words, we should see an interaction between Continuation Type and Stimulus Group.

Our results showed a clear main effect of Item-specific Cloze on the N400 (b = 3.58, t = 2.47, p = .014). Critically, however, we saw no significant interaction between Continuation Type and Stimulus Group (see Table 3). These findings suggest that all critical words in our study evoked N400 amplitudes in proportion to their cloze probabilities, with no penalties for being in the presence of a pre-activated alternative.

Table 3. *WithCompetitor Expected, SecondBest* and *ControlScenario* conditions

Model 1 (N400)	Estimates			
Predictor	*b*	*SE*	*t-value*	*p-value*
Item-specific Cloze	3.58	1.45	2.47	.01 *
Continuation Type	–0.18	0.41	–0.44	.66
Stimulus Group	0.71	0.60	1.18	.24
Continuation Type × Stimulus Group	–0.50	0.77	–0.65	.52

Random Effects Structure: (1 + continuation*stimulus_group || subject) + (1 + stimulus_group + continuation:stimulus_group || item)

Model 2: Does competitive pre-activation result in an increased N400 to Expected words?

In Model 2, we compared the N400 responses to the *WithCompetitor Expected* (A1) and the *NoCompetitor Expected* (A3) critical words. Again, to control for lexical variables, we included the N400 responses from the identical critical words appearing in their associated *ControlScenarios* (B1, B3). In this model, we included fixed effects of Item-specific Cloze, Contextual Competition (*WithCompetitor* = 0.5, *NoCompetitor* = –0.5), Stimulus Group (*TargetScenario* = 0.5, *ControlScenario* = –0.5), and an interaction between Contextual Competition and Stimulus Group.

If *Expected* words in *WithCompetitor* contexts were slightly inhibited by their *SecondBest* competitors during the pre-activation phase, then they should produce slightly larger N400s than the cloze-matched *Expected* words in the *NoCompetitor* contexts. As in Model 1, we would expect to see no such differences between the same critical words in their associated *ControlScenarios*. In other words, there should be an interaction between Contextual Competition and Stimulus Group (see Table 4).

In this model, however, we found only a main effect of Item-specific Cloze on the N400 (*b* = 3.51, *t* = 2.87, *p* = .005). There were no other significant main effects, nor interactions. Taken together with Model 1, these results fail to provide evidence for the *competitive pre-activation* account; that is, the degree of pre-activation for a given word did not appear to be sensitive to the presence or absence of a pre-activated alternative.

Table 4. *WithCompetitor Expected, NoCompetitor Expected, and ControlScenario conditions*

Model 2 (N400)	Estimates			
Predictor	*b*	*SE*	*t-value*	*p-value*
Item-specific Cloze	3.51	1.22	2.87	.005 *
Contextual Competition	0.41	0.28	1.45	.15
Stimulus Group	0.10	0.71	0.14	.89
Contextual Competition × Stimulus Group	0.75	0.58	1.30	.20

Random Effects Structure: (1 + cloze + contextual_competition:stimulus_group || subject) + (1 + cloze + stimulus_group || item)

Question 2: Friendly pre-activation due to shared semantic features: Evidence from the N400

Our second aim was to determine whether there is any evidence for *friendly pre-activation* as a result of overlap between the semantic features associated with the observed lower probability critical word and an unobserved higher probability pre-activated alternative. In contrast to the *competitive pre-activation* account, which predicted *less* facilitation and a *larger* N400 than that one would expect based on cloze probability alone, the *friendly pre-activation* account predicts *more* facilitation and a *smaller* N400 than one would expect based only on cloze. Moreover, this account also predicts that the reduction in the N400 should be *graded* with increasing levels of relatedness between the critical word and its unobserved, higher probability alternative. To address this question, we first assessed the effects of friendly pre-activation in all lower probability continuations (Model 3). We then restricted our analyses to the subset of *SecondBest* continuations in the *WithCompetitor* contexts (Models 4 and 5).

Model 3: For all lower probability words, is there a graded effect of friendly pre-activation?

In Model 3, we included the N400 responses to *all* lower probability continuations in contexts with a potentially pre-activated higher probability alternative—namely, the *NoCompetitor ZeroCloze* and the *WithCompetitor SecondBest* conditions. According to the *friendly pre-activation* account, the mere pre-activation of a higher probability alternative should provide facilitation on the N400 evoked by a lower probability continuation, but only if the two words are semantically related to one another. Thus, if there is friendly pre-activation, the N400 produced by lower probability continuations should become *smaller* (than predicted by cloze) as semantic relatedness increases.

To test this hypothesis, Model 3 included a continuous fixed effect of Semantic Relatedness, which was calculated, item-by-item, between each critical word and its higher probability competitor (see Methods). The model also included a fixed effect of Item-specific Cloze to control for differences between the two lower probability conditions.

Consistent with a *friendly pre-activation* account, we observed a significant effect of Semantic Relatedness such that the N400 produced by lower probability continuations (*ZeroCloze*, *SecondBest*) decreased as semantic relatedness increased (see Table 5; $b = 6.01$, $t = 3.55$, $p <$

.001).[9] In Figure 3 (left panel), we show the averaged N400 responses for each lower probability word as a function of its semantic relatedness with its higher probability alternative. Consistent with the model results, the amplitude of the N400 increases with increasing similarity between the critical word and its higher probability alternative.

Table 5. *ZeroCloze and SecondBest continuations*

Model 3 (N400)	Estimates			
Predictor	*b*	*SE*	*t-value*	*p-value*
Item-specific Cloze	2.78	2.52	1.10	.28
Semantic Relatedness	6.01	1.69	3.55	< .001 *

Random Effects Structure: (1 + cloze + semantic_relatedness | subject) + (1 | item)

To further visualize this effect, we used a median split to subdivide the trials into those that were semantically related to a higher probability continuation (*Mean semantic relatedness*: 0.44; *SD*: 0.10) and those that were semantically unrelated to a higher probability continuation (*Mean semantic relatedness:* 0.19; *SD*: 0.07). We then computed grand-average waveforms for each of these conditions across central electrode sites. Figure 3 (right panel) shows these waveforms, together with the waveforms produced by *Expected* critical words (averaged across conditions A1 and A3) and critical words in the *ControlScenarios* (averaged across conditions B2 and B4). As expected, we found graded N400s such that *Expected* words evoked the smallest responses, followed by a larger N400 to critical words that were *Related* to a higher probability alternative, and the largest N400s to critical words that were *Unrelated* to a higher probability alternative, and

[9] In Model 3, as well as in Model 4, which included only a subset of the non-modal continuations, there was no significant effect of Cloze, perhaps because of the relatively restricted range of cloze probability in these analyses. However, the fact that Semantic Relatedness predicted N400 amplitude, while controlling for cloze probability, supports the friendly pre-activation account.

that appeared in the *ControlScenarios.* Taken together, these findings suggest that, instead of reducing the degree of lexico-semantic facilitation, the pre-activation of a higher probability alternative can further facilitate the processing of a lower probability critical word.

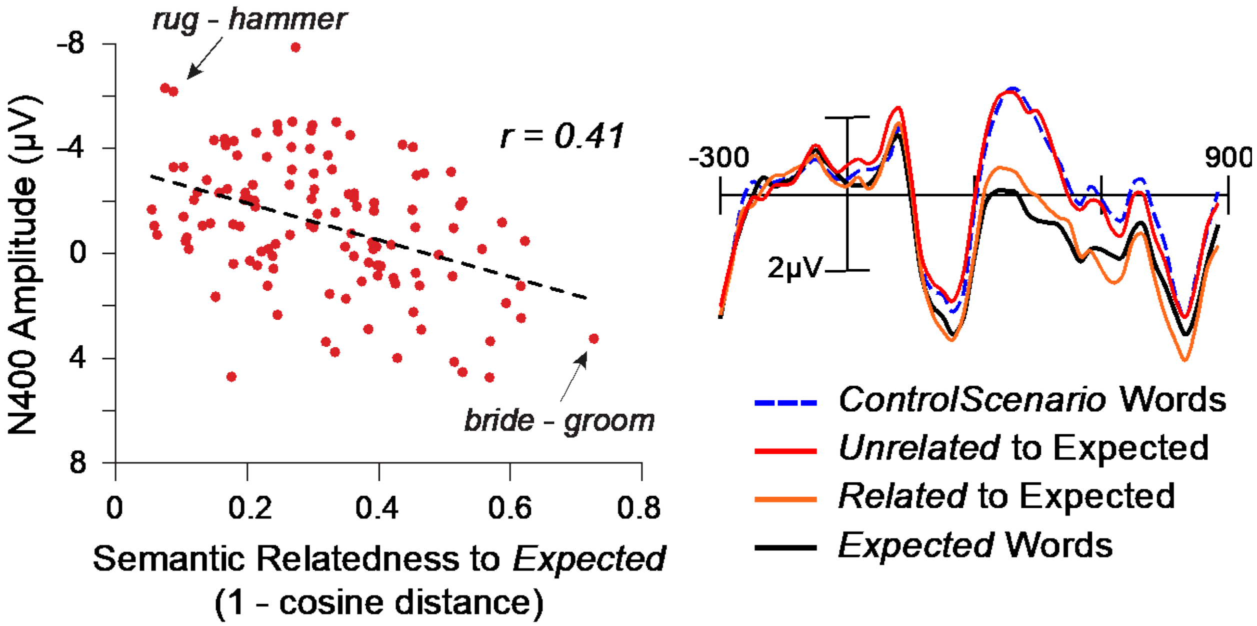

Figure 3. *Left*: The averaged N400 response (300–500ms) for each lower probability critical word in conditions A2 (*WithCompetitor SecondBest*) and A4 (*NoCompetitor ZeroCloze*) plotted as a function of its semantic relatedness to its more probable alternative (taken from the *WithCompetitor Expected* condition, A1, and the *NoCompetitor Expected* condition, A3).
Right: Grand-averaged ERPs within the N400 spatiotemporal region in response to (a) critical words in the *ControlScenarios* (averaged across conditions B2 and B4, blue dotted), (b) critical words in conditions A2 and A4 that were semantically unrelated to a higher probability alternative (*Unrelated* to Expected, red), (c) critical words in conditions A2 and A4 that were semantically related to a higher probability predicted alternative (*Related* to Expected, orange), and (d) *Expected* critical words (averaged across conditions A1 and A3, black)

Model 4: Does friendly pre-activation also enhance facilitation on SecondBest critical words in WithCompetitor contexts?

In Model 3, we grouped together responses to *all* lower probability continuations in contexts with a potentially pre-activated alternative, including zero-cloze continuations where previous work has already demonstrated *friendly pre-activation* (Kutas and Hillyard 1984;

Federmeier & Kutas, 1999; Thornhill & Van Petten, 2012; DeLong & Kutas, 2020). Here, in Model 4, we focused solely on the N400 produced by the *WithCompetitor SecondBest* continuations, which have not been examined in previous studies. This provided a more direct test of whether, in these *WithCompetitor* contexts, *SecondBest* critical words receive friendly pre-activation from their higher probability alternatives, which, as discussed earlier, could have potentially acted as competitors. Similar to Model 3, this model included fixed effects of Semantic Relatedness and Item-specific Cloze (see Table 6). We again found a significant effect of Semantic Relatedness on the N400 on the *WithCompetitor SecondBest* continuations (*b* = 7.37, *t* = 2.87, *p* = .006), when controlling for cloze probability.

Table 6. *WithCompetitor SecondBest* only

Model 4 (N400)	Estimates			
Predictor	*b*	*SE*	*t-value*	*p-value*
Item-specific Cloze	0.03	4.04	0.01	.99
Semantic Relatedness	7.37	2.57	2.87	.006 *

Random Effects Structure: (0 + cloze + semantic_relatedness || subject) + (1 | item)

Model 5: When minimizing friendly pre-activation, is there any evidence of a competition effect on SecondBest completions?

As noted in the Introduction, the *competitive pre-activation* and *friendly pre-activation* accounts are not mutually exclusive. Thus, even though, as shown in Model 4, *SecondBest* critical words that were semantically related to a more expected continuation received *more* facilitation than one would expect based on their cloze probability, it remained possible that *SecondBest* critical words that were semantically *unrelated* to a more expected continuation might receive *less* facilitation than expected, as a result of *competitive pre-activation* (see Ness and Meltzer-Asscher, 2021a, for evidence during a speeded cloze task). On this account, one explanation for why we found no evidence of *competitive pre-activation* in Model 1 could be that any effects of competitive pre-activation were outweighed by friendly pre-activation.

To explore this possibility, we further restricted our analysis to the subset of *WithCompetitor* items in which the *Expected* and *SecondBest* continuations were unrelated to one another. All trials were split into *Related* and *Unrelated* groups using a median split on Semantic Relatedness (*Median* = 0.39). We then selected the *SecondBest* words from these semantically *Unrelated* trials and paired them with their associated *ControlScenarios*. We then implemented another model with fixed effects of Stimulus Group (*TargetScenarios* = 0.5, *ControlScenarios* = –0.5) and Item-specific Cloze as a control variable. If there was any evidence of *competitive pre-activation* on the N400, we should see an effect of Stimulus Group in this model. However, similar to the models above, we observed no such categorical effect of Stimulus Group; if anything the beta-weight for this model went in the opposite direction, with greater facilitation on unrelated *SecondBest* words in the *WithCompetitor* contexts than when the same words appeared in the *ControlScenarios* (see Table 7).

Table 7. *WithCompetitor SecondBest* continuations from semantically *Unrelated* trials

Model 5 (N400)	Estimates			
Predictor	*b*	*SE*	*t-value*	*p-value*
Item-specific Cloze	–4.29	4.39	–0.98	.33
Stimulus Group	1.44	0.82	1.75	.09

Random Effects Structure: (1 + cloze + stimulus_group || subject) + (1 + stimulus_group || item)

Question 3. Late costs of suppressing higher probability alternatives: Evidence from the late frontal positivity.

Our final aim was to determine whether, in the *WithCompetitor* contexts, there was any evidence for costs associated with late inhibition when processing *SecondBest* versus *Expected* continuations. To address this question, we focused on the late frontal positivity, which has sometimes been interpreted as an index of inhibiting or suppressing an incorrect lexical prediction in order to successfully integrate an observed input into the unfolding context (Kutas, 1993; Ness & Meltzer-Asscher, 2018).

Model 6: Are there inhibitory costs on the late frontal positivity produced by SecondBest relative to Expected continuations in WithCompetitor contexts?

Similar to Model 1 for the N400, Model 6 compared the late frontal positivities produced by *WithCompetitor Expected* (A1) and *WithCompetitor SecondBest* (A2) critical words, as well as the same critical words appearing in the *ControlScenarios* (B1 and B2, respectively), i.e., it included fixed effects of Item-specific Cloze, Continuation Type (*Expected* = –0.5, *SecondBest* = 0.5), Stimulus Group (*TargetScenarios* = 0.5, *ControlScenarios* = –0.5), and an interaction between Continuation Type and Stimulus Group.

If late costs are incurred when comprehenders suppress/inhibit an incorrectly pre-activated alternative, then the *SecondBest* completions should produce a larger frontal positivity than the *Expected* continuations, and there should be no such differences between the same critical words appearing in the *ControlScenarios*, i.e., a there should be an interaction between Continuation Type and Stimulus Group. However, our results did not indicate any such interaction (see Table 8). Thus, in *WithCompetitor* contexts, there was no evidence of any late penalty in processing lower probability words in the presence of a higher probability alternative.

Table 8. *WithCompetitor Expected, WithCompetitor SecondBest, and ControlScenario conditions*

Model 6 (Late Frontal Positivity)	Estimates			
Predictor	*b*	*SE*	*t-value*	*p-value*
Item-specific Cloze	–1.49	1.47	–1.02	.31
Continuation Type	0.45	0.40	1.12	.26
Stimulus Group	0.79	0.57	1.37	.17
Continuation Type × Stimulus Group	–0.53	0.83	–0.63	.53

Random Effects Structure: (1 + cloze + stimulus_group || subject) + (0 + cloze + continuation_type:stimulus_group || item)

Models 7 and 8: Are there continuous effects of Constraint on the late frontal positivities to lower probability critical words?

One reason why some researchers have suggested that the late frontal positivity reflects late costs associated with suppressing an incorrectly predicted alternative is that this component is sometimes larger in response to zero-cloze words appearing in high constraint contexts, which constrain for a single alternative, relative to zero-cloze words appearing in low constraint contexts that do not constrain for any specific word (e.g. Federmeier et al., 2007; Kuperberg, Brothers, & Wlotko, 2020).

In light of these findings, we attempted to replicate (and extend) these previous findings by investigating how the late frontal positivity to lower probability continuations varies as a function of Contextual Constraint in the present study (Models 7 and 8). In comparison with previous studies, which have contrasted ERPs to zero-cloze words appearing in *high constraint* versus *low constraint* contexts, the present study included discourse contexts with a fairly wide range of constraint across conditions. Thus, even though, the *average* constraint of the contexts in the *ControlScenarios* was lower (*Mean constraint* = 19.5%, *SD* = 7.8%) than that of the *TargetScenarios (Mean constraint* = 59.1%, *SD* = 14.8%), there was still a considerable amount of variability in Contextual Constraint within both these stimulus groups (*ControlScenarios Range* = 8–52%, *TargetScenarios Range* = 33.3–92%). We therefore decided to combine these two conditions and use a continuous item-level measure of constraint, rather than categorical measure of constraint, for these analyses.

In Model 7, we examined the effect of Item-specific Constraint on the late frontal positivity produced by zero-cloze items (analogous to what has been examined in previous studies). In the present study, this included all items in the *NoCompetitor ZeroCloze* (A4) and their associated *ControlScenarios* (B4). This model included fixed effects of Item-specific Cloze and Item-specific Constraint (see Table 9). In this analysis, we found that the effect of Item-specific Constraint trended toward significance in the predicted direction (*b* = 1.46, *t* = 1.90, *p* = .06).

Table 9. *ZeroCloze and their associated ControlScenarios*

Model 7 (Late Frontal Positivity)	Estimates			
Predictor	*b*	*SE*	*t-value*	*p-value*
Item-specific Cloze	9.96	19.09	0.52	.60
Item-specific Constraint	1.46	0.77	1.90	.06

Random Effects Structure: (1 + cloze + constraint || subject) + (1 | item)

Because of this near replication, and given the interest of an anonymous reviewer, we then ran another model (Model 8) with the same predictors to explore whether there was an influence of Item-specific Constraint on the late frontal positivities produced by the *WithCompetitor SecondBest* critical words (A2) and their corresponding controls (B2). Results indicated a significant effect of Item-specific Constraint (see Table 10) such that late frontal positivities became larger as the constraint of the context increased (*b* = 2.46, *t* = 2.51, *p* = .014).

The results from these two analyses (Models 7 and 8), taken together with the prior literature, suggest that lower probability continuations show larger late frontal positivities in higher relative to lower constraint contexts.[10] We discuss this finding further in the Discussion.

Table 10. *SecondBest and their associated ControlScenarios*

Model 8 (Late Frontal Positivity)	Estimates			
Predictor	*b*	*SE*	*t-value*	*p-value*
Item-specific Cloze	–1.16	2.14	–0.54	.59
Item-specific Constraint	2.46	0.98	2.51	.014 *

Random Effects Structure: (1 + constraint || subject) + (0 + constraint || item)

Discussion

It is well-established that linguistic predictions are probabilistic, and that the processing of incoming words is facilitated in a graded fashion to the degree that they have been pre-activated by the prior context. This raises an obvious question: Are there "costs" associated with generating lexico-semantic predictions during language comprehension?

To address this question, most researchers have focused on the specific situation in which a plausible but lexically unpredictable target word violates a strong lexical prediction that is generated in a highly constraining context (e.g. *"He bought her a pearl necklace for her...collection"*). It has been hypothesized that such prediction violations might incur processing costs as a result of competition between the bottom-up input and the incorrect

[10] For completeness, we also ran identical models to Models 7–8 on the N400 response. Consistent with the prior literature, we did not see any effects of constraint on the N400 to the *NoCompetitor ZeroCloze* continuations (*b* = 0.40, *t* = 0.49, *p* = .62) nor the *WithCompetitor SecondBest* continuations (*b* = 1.93, *t* = 1.71, *p* = .09).

prediction. To test this hypothesis, researchers have contrasted these prediction-violating-target words with target words that are equally unpredictable, but that do not violate a strong prediction (e.g. *"He looked worried because he might have broken his...collection").*

ERP and behavioral studies examining this contrast have found no evidence of costs associated with violating strong predictions during lexico-semantic processing (*ERP studies examining the N400:* Kutas & Hillyard, 1984; Federmeier, Wlotko, De Ochoa-Dewald & Kutas, 2007; Kuperberg, Brothers, & Wlotko, 2020; *behavioral studies:* Frisson et al., 2017; Steen-Baker, Ng, Payne, Anderson, Federmeier & Stine-Morrow, 2017; Fischler & Bloom, 1979, 1985; Schwanenflugel and LaCount, 1988; see also Wong, Veldre & Andrews, 2022). However, some ERP studies have shown that, in plausible sentences, this contrast sometimes reveals a frontally-distributed positivity effect at a later stage of processing (e.g. Federmeier et al., 2007; Kuperberg, Brothers, & Wlotko, 2020). This effect has sometimes been interpreted as indexing later costs as a result of competition between the predicted word and the bottom-up input (Kutas, 1993; Ness & Meltzer-Asscher, 2018).

In the present study, we ask a different question that has received far less attention in the literature: Are there any consequences of disconfirming a prior prediction in contexts that constrain for *multiple* possible candidates?

In contrast to prediction violations on zero-cloze words, where any competition would begin only once the bottom-up input is encountered, in these *WithCompetitor* contexts, the pre-activated alternative candidates could begin to compete with one another *before* the onset of the target word. This would lead to *relative* costs in processing the bottom-up input; that is, even if the target is partially predictable because it confirms one of the pre-activated alternatives, it may still be more difficult to process than if there had been no pre-activated competitor. These relative costs could be incurred during the initial stages of lexico-semantic processing, predicting a larger N400 than would be expected based only on cloze probability. Alternatively, they might manifest at a later stage of processing on the late frontal positivity.

In a naturalistic eye-tracking study, Luke and Christianson (2016) showed (a) that such “high competition” contexts followed by “second best” continuations are encountered frequently in natural texts, but (b) when these texts are read for comprehension, the second-best continuations incurred no behavioral processing costs on either early or late reading time measures. Instead, these authors found evidence of increased facilitation when less predictable words were semantically related to a higher probability alternative.

In the present study, we conceptually replicated Luke and Christianson’s findings using a different method — ERPs — and using a controlled experimental design, which allowed us to control for potential lexical confounds. Contrary to the predictions of a *competitive pre-activation* account, we found that in *WithCompetitor* contexts, the N400s produced by *SecondBest* and *Expected* continuations were no larger than would be predicted given their cloze probabilities alone. Instead, we found evidence for *friendly pre-activation* on the N400: Extending previous ERP findings showing that zero-cloze words that are semantically related to a predicted continuation produce facilitation on the N400 (e.g. Federmeier & Kutas, 1999), we found that when a *SecondBest* critical word was semantically related to the higher probability alternative, it produced a *smaller* N400 than expected given its cloze probability (i.e. enhanced facilitation). Finally, we found no evidence for differences in the responses produced by the *SecondBest* and the *Expected* continuations on a later frontal positivity, even though this component was influenced by the constraint of the prior context.

These findings have important theoretical implications for understanding the mechanisms underlying probabilistic prediction during language comprehension. In the remainder of this discussion, we first consider the lack of evidence for competitive effects on the N400 in the *WithCompetitor* contexts. We then consider the lack of evidence for competitive effects on the late frontal positivity and consider reasons why this late effect was nonetheless sensitive to contextual constraint. Third, we discuss how our findings extend previous work that has demonstrated the role of *friendly pre-activation* on the processing of zero-cloze words. Finally, we discuss the computational principles of a parallel, interactive framework of predictive processing that can accommodate these findings, and how these principles might be implemented by a biologically plausible architecture and algorithm known as predictive coding.

No evidence for effects of *competitive pre-activation* on the N400

During language comprehension, prediction is often viewed as a top-down process in which comprehenders use their current high-level interpretation to pre-activate lower-level lexico-semantic representations of upcoming words. At face value, this top-down process is similar in many respects to language production, in which producers use an intended high-level message to activate lower-level lexical representations for later articulation. This has led some researchers to propose that top-down prediction during language comprehension *routinely* employs the same mechanisms that are employed during language production (e.g. Pickering & Garrod, 2013; Fitz & Chang, 2019; Van Petten & Luka, 2012).

In testing this theory, a critical factor to consider is that, to produce language, there is continuous top-down pressure to select a *single* word (e.g. Levelt, 2001). One well-known mechanism of top-down lexical selection during language production is *mutual competitive inhibition* in which co-activated lexical representations each exert lateral inhibition on one another until a single candidate is selected. This type of “winner-takes-all” selection mechanism is a fundamental characteristic of classic Interactive Activation and Competition (IAC) models that have been used to model language production (e.g. Chen & Mirman, 2012). For example, in a recent study, Ness and Meltzer-Asscher (2021a) used an IAC model (Chen & Mirman, 2012) to simulate production times in a speeded cloze tasks. These authors showed that competitive inhibition between unrelated pre-activated lexical representations could explain the longer-than-expected times to produce upcoming words (see also Nakamura & Phillips, 2022).

To the extent that the same IAC principles have been proposed to underlie aspects of language comprehension (e.g. McClelland & Rumelhart, 1981; McClelland & Elman, 1986; see also Spivey & Tanenhaus, 1998; MacDonald, Pearlmutter & Seidenberg, 1994), then this would predict that the presence of a competing alternative during the predictive phase of language comprehension should reduce lexico-semantic facilitation on incoming words when they become available. This would result in larger N400s on critical words in *WithCompetitor* contexts than one might expect based on cloze probability alone. This reduced facilitation should be particularly apparent on *SecondBest* continuations (e.g. “flowers”) because, in IAC models, lateral inhibition scales non-linearly, such that more strongly activated lexical units exert the greatest inhibitory pressure. Specifically, during the pre-activation phase, the more weakly pre-activated second-best alternative (e.g. <flowers>) would receive strong inhibition from the more strongly pre-activated alternative (e.g. <hearts>), and so when this *SecondBest* input (“flowers”) actually appears, it should be relatively more difficult to access its lexico-semantic representation. In addition, one might *also* expect to see some effect of top-down competitive inhibition on the *Expected* words in the *WithCompetitor* contexts (“hearts") compared to *Expected* words in *NoCompetitor* contexts with the same cloze probability, although this inhibition effect should be smaller in magnitude.

However, we found no evidence for these types of competitive effects in the *WithCompetitor* contexts: First, the *SecondBest* words produced N400s in proportion to their cloze probabilities, with no additional effect of Continuation Type (*Expected* vs. *SecondBest*). Second, the *Expected* words in *WithCompetitor* context had comparable N400s to the *Expected* words in *NoCompetitor* contexts. These findings therefore do not support the idea that mutual inhibition between multiple pre-activated candidates influences subsequent lexico-semantic processing of incoming words between 300–500ms.

No evidence for *late suppression* costs to *SecondBest* versus *Expected* critical words on the late frontal positivity

We also found no evidence for neural costs in processing *SecondBest* relative to *Expected* continuations at a *later* stage of processing, either on the late frontal positivity or on any other late ERP component (see Supplementary Material).[11] This provides evidence against a *late suppression* account, which claims that in order to integrate a lower probability word into its prior context, it is necessary to suppress an alternative predicted (but unobserved) representation that remain active past the N400 time window (Kutas, 1993; Ness & Meltzer-Asscher, 2018).

The original motivation for the *late suppression* account of the late frontal positivity was that this ERP component is sometimes larger to plausible zero-cloze words appearing in very high constraint *versus* low constraint contexts (e.g. Federmeier, Wlotko, De Ochoa-Dewald &

[11] Wlotko and Federmeier (2012) proposed that another ERP component might reflect late costs associated with competition: In an exploratory post-hoc analysis, these authors observed a left-lateralized frontal *negativity* in response to medium-high (75–90%) cloze probability words, particularly those with an alternative competing continuation. These authors speculated that this effect indexed working memory resources necessary to deal with multiple competing possibilities during lexical selection. We did not find any evidence of this effect, either in the analyses presented in this manuscript or in further analyses reported in Supplementary Materials, despite the fact that our *NoCompetitor Expected* versus *WithCompetitor Expected* contrast closely resembled the contrast where Wlotko and Federmeier found their late left-lateralized frontal negativity.

Kutas, 2007; Kuperberg, Brothers, & Wlotko, 2020; although this is not always the case, e.g. see Thornhill and Van Petten, 2012; Zirnstein, van Hell & Kroll, 2018). In the present study, we replicate and extend this original finding by showing significant (and near significant) *graded* effects of Item-specific Constraint on the late frontal positivity (see Models 7 and 8). Therefore, our findings raise the question of what neurocognitive processes the late frontal positivity *does* index, and why this component is sometimes sensitive to contextual constraint, but not to lexical-level inhibition.

In recent work, we have argued that, rather than indexing processes that operate over individual lexical items (such as lexical suppression), the late frontal positivity indexes processes related to the successful updating of the comprehender's higher-level *situation model* upon encountering new unpredicted input (Brothers, Wlotko, Warnke, & Kuperberg, 2020; Kuperberg, Brothers, & Wlotko, 2020; Brothers, Greene, & Kuperberg, 2020). On this account, the reason why the late frontal positivity is often enhanced on *unexpected plausible* words that violate a higher probability prediction is that these types of unexpected words tend to trigger larger updates/shifts of the situation model (by retrieving new schema-relevant information from long-term memory). For example, when reading "*He bought her a pearl necklace for her collection*", the final word (*collection*) may produce a large late frontal positivity because the comprehender updates the situation model by inferring new schema-relevant events that are related to the collection of jewelry.

Critically, this account of the late frontal positivity implies that the presence of a strong lexical-level competitor is neither necessary nor sufficient to induce updates of the situation model and produce this effect. In the present study, this would explain why the *WithCompetitor SecondBest* completions (e.g. "flowers") did not produce a larger late frontal positivity than the

WithCompetitor Expected completions. We suggest that *both* these critical words produced some degree of update in the comprehender's situation model, which was, on average, greater than that produced by critical words in the *ControlScenarios*.

This *situation model updating* account can also explain why the late frontal positivity effect is *not* produced by unexpected words that violate strong predictions in very short sentences where comprehenders are unlikely to engage in building a situation model (e.g. "*James unlocked the…[door]/laptop"*, see Brothers, Wlotko, Warnke, & Kuperberg, 2020, Experiment 1). In addition, it can explain why, relative to *expected* words, a robust late positivity *is* sometimes produced by unexpected words appearing in *low constraint* contexts (e.g. Chow, Lau, Wang & Phillips, 2018; Freunberger & Roehm, 2016; Davenport & Coulson, 2011), sometimes with an amplitude that is, in fact, just as large as to high constraint unexpected continuations (e.g. Thornhill & Van Petten, 2012; Ng, Payne, Steen, Stine-Morrow, & Federmeier, 2017; Hubbard, Rommers, Jacobs, & Federmeier, 2019; Zirnstein, van Hell & Kroll, 2018). In these cases, *both* types of unexpected words may be informative enough to induce fairly large updates of the situation model (see Brothers, Greene & Kuperberg, 2020).

We emphasize, however, that the present study was not designed to directly test this situation model updating account of the late frontal positivity, and so it will be important for future studies to further explore the function role of this late frontal effect.

Friendly pre-activation

In contrast to the lack of evidence for *competitive pre-activation*, we did find clear evidence for *friendly pre-activation* on lexico-semantic processing. The N400 response produced by lower probability words was reduced when these words shared semantic features with a more strongly pre-activated alternative (*bride – groom*). This finding is consistent with previous ERP studies that have reported this type of "anticipatory semantic overlap effect" both on unexpected *implausible* words (Federmeier & Kutas, 1999; DeLong, Chan, & Kutas, 2019; Ito, Corley, Pickering, Martin, & Nieuwland, 2016) as well as on unexpected zero-cloze *plausible* words (Thornhill & Van Petten, 2012; DeLong & Kutas, 2020; for consistent behavioral results, see Frisson et al., 2017, Experiment 2; Roland, Yun, Koenig & Mauner, 2012; Wong, Veldre & Andrews, 2022).[12]

Importantly, by conceptually replicating Luke and Christianson's (2016) behavioral findings in their natural corpus, our findings extend this previous ERP work in two ways: First, we show that the anticipatory effect of semantic relatedness on the N400 is graded, with a linear relationship between degree of semantic relatedness and degree of facilitation. Second, we demonstrate that this additional facilitation occurs on *SecondBest* continuations (with non-trivial cloze probabilities) that are encountered in *WithCompetitor* contexts. Unlike a zero-cloze word, which will receive facilitation from a higher-probability predicted alternative only *after* it is encountered, *SecondBest* completions are likely to have already received some pre-activation from the semantically related alternative *before* the onset of the bottom-up input (see below for further discussion). As discussed earlier, during this pre-activation phase, the higher probability alternative could have, in principle, acted as a competitor. Therefore, by showing that these higher probability alternatives can facilitate, rather than inhibit, subsequent lexico-semantic access, we provide important evidence that semantic overlap from alternative pre-activated items

[12] In a previous eye-tracking study, Frisson, Harvey and Staub, 2017 (Experiment 2) observed semantic overlap effects on plausible prediction violations, but only on late eye-tracking measures. This led the authors to conclude that lexical predictability and semantic overlap influenced distinct processing stages (word recognition and integration). The current findings, however, suggest that lexical predictability and semantic relatedness both modulated the same underlying ERP response (the N400) with a similar time course. Again, this finding supports the claim that semantic overlap effects are anticipatory in nature and can influence the initial stages of lexico-semantic retrieval (for additional supporting evidence, see Wong, Veldre & Andrews, 2022).

can support everyday language processing. Indeed, as noted earlier, Luke and Christianson (2016) showed that such *SecondBest* continuations in *WithCompetitor* contexts occur frequently in natural language.

We should note that finding an anticipatory semantic overlap effect on critical words in *WithCompetitor* contexts is compatible with some IAC architectures. For example, in Chen & Mirman's IAC model (2012), in addition to receiving inhibitory lateral connections from other localist lexical items, each lexical item also receives cross-layer excitatory connections from distributed sets of semantic features. If two pre-activated candidates share semantic features, then this shared excitation can sometimes outweigh any mutual lexical-level inhibition. Evidence that this can impact language production comes from a recent study by Ness & Meltzer-Asscher (2021a), who carried out simulations using Chen and Mirman's IAC model, and showed that the additional pre-activation received by an expected lexical unit that shared semantic features with its second best "competitor" was able to explain peoples' faster response latencies to produce this word in a speeded cloze task.

However, as discussed earlier, Ness & Meltzer-Asscher (2021a) also found that the mutual lateral inhibition between lexical units in the IAC model was able to explain why producers took longer to produce expected words the presence of an unrelated competitor. In the present study, however, we found no evidence that this type of mutual inhibition between pre-activated competitors influenced *comprehension*: Even when we considered only the subset of *WithCompetitor* contexts in which the *SecondBest* continuation was semantically unrelated to the *Expected* continuation (based on a median split), the N400 produced by *SecondBest* continuation was no larger than that produced by the same critical words in the *ControlScenarios*.

Taken together, these findings provide strong evidence for *friendly pre-activation*, but no evidence for *competitive pre-activation* during language comprehension.

Explaining parallel, graded and friendly pre-activation within a hierarchical probabilistic generative framework

In the sections above, we discussed how researchers have appealed to architectures such as IAC in which lateral inhibitory connections between lexical representations play a key role in inhibiting pre-activated competitors. These frameworks, however, are incompatible with our current findings, as we found no evidence that inhibition between pre-activated lexical competitors influences lexico-semantic processing during comprehension. In this section, we will argue that our findings can be better understood within a probabilistic generative framework of language processing (see Kuperberg & Jaeger, 2016 for an overview). We first discuss the computational principles of this framework at Marr's first level of analysis (Marr, 1971), and consider how these principles can explain the present set of findings. We then consider how this framework could be implemented at the algorithmic level, highlighting *predictive coding* as a particularly promising architecture and algorithm for achieving this goal.

Marr level 1: Probabilistic Inference: Explaining the bottom-up input and explaining away alternatives

At the heart of all probabilistic generative frameworks is the *generative model* — an internal network of hierarchically organized knowledge that encodes the agent's probabilistic assumptions about how latent causes (also called *hypotheses*) cause or "generate" observations from the environment (see Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010). At each level of representation, each hypothesis is held with a particular probability, referred to as a *belief*, and at any given time, an agent can hold *multiple* beliefs in parallel, which, together can be described as a probability distribution. When new input (evidence) becomes available from the environment, prior beliefs at each level of representation are *updated* through Bayes's rule, with belief flowing dynamically up and down the hierarchical generative network until it settles on the latent causes that best "explain" the statistical structure of the input (Pearl, 1982).

Within this probabilistic framework, the process of deep language comprehension can be understood as the process of inferring the high-level message that the producer intended to communicate from a sequence of linguistic inputs that unfold in real time. We will refer to this high-level interpretation as a situation model — a representation of the set of events being communicated, including the referential, spatial, temporal, motivational and causal coherence relationships that link them (Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). We assume that this situation model lies at the top of the comprehender's internal generative model, and that the network below it comprises all relevant information that is needed to infer this interpretation. This network encodes information at multiple levels of linguistic and non-linguistic representation (e.g. event structures, syntax, semantics, phonology, orthography). However, for the purpose of explaining our current findings, we will primarily focus on just four of these levels of representation: concepts, semantic features, lexical items, and orthographic features. Within this part of the generative network, each individual concept (e.g. {lime}) serves as a latent cause of a unique combination of distributed semantic features (e.g. the combination of <sour> and <edible> and <squeezable> and <green>). Each unique combination of semantic features, in turn, serves as a latent cause of a specific lexical representation (e.g. /lime/), which similarly serves as the latent cause for a particular set of distributed form features (e.g. "L-I-M-E"). Note that, with these assumptions, each lexical representation describes a mapping function that uniquely links a particular unique combination of semantic features with a particular set of form features.[13]

During word-by-word language comprehension, the situation model continually propagates belief down to lower levels of the generative hierarchy. Because this high-level situation model represents information over a long time span, these beliefs will reach the conceptual, semantic, and lexical levels *before* new bottom-up input becomes available to these levels. Thus, within this framework, a *lexico-semantic prediction* corresponds to the comprehender's prior beliefs about the particular concepts, the particular sets of semantic features, and the particular lexical items that are most likely to have caused/generated the orthographic features of the upcoming word. For example, when reading the scenario, "At the restaurant, Anthony got his food. He squeezed the fresh....", a comprehender may have a 70% prior belief that Anthony squeezed a {lemon}/lemon/ and its unique set of semantic features, and a 30% belief that Anthony squeezed a {lime}/lime/ and its unique set of semantic features. (Note that, as discussed later, at the level of semantic features, a comprehender's prior belief about a particular *combination* of features does *not* necessarily equate to the average of their prior belief of encountering each of these features individually).

Then, when new orthographic/phonological input is encountered (e.g. the orthographic features, L-I-M-E), this provides strong new evidence that is compatible with only one candidate hypothesis (only one latent cause) at each level of representation. As a result, over multiple cycles of belief updating, the agent's belief over the conceptual representation, {lime}, its unique combination of semantic features (<sour> and <edible> and <squeezable> and <green>), and its

[13] Within a *Bayesian belief network,* each individual lexical item would correspond to a particular value of a variable that functions to "d-separate" these semantic and form features such that they are conditionally independent (Pearl, 1988; see Narayanan & Jurafsky, 2001 for discussion in relation to a different aspect of language comprehension).

lexical representation, /lime/, will each rise to nearly 100%, while belief will fall over all other mutually exclusive hypotheses at each of these levels of representation. This process of belief updating can be conceptualized as a type of "competitive selection" in that it involves selecting one latent cause from multiple mutually exclusive hypotheses. However, as we discuss next, and as illustrated in Figure 4, this type of "selection-by-inference" is quite different from the competitive selection that is implemented by IAC networks.

As explained earlier, in an IAC architecture, selection is implemented through mutual lateral inhibition between active units at a single lexical level of representation. In contrast, in probabilistic inference, latent causes compete to *explain* observations at the level below. To win this competition, each possible hypothesis at the conceptual and lexical levels must be evaluated in relation to each possible *combination* of observed semantic and orthographic features in order to determine which hypothesis/latent cause provides the best possible explanation of the particular combination of features that are observed. For example, at the lexical level, /lime/ and /dime/ both match the observed orthographic features, "I-M-E", but only /lime/ can additionally explain the specific combination "L-I-M-E", with the presence of "L" and the absence of "D" in the input's first position, and so it will win the competition. Analogously, at the conceptual level, both {lemon} and {lime} can explain the observed semantic features, <sour>, <edible>, and <squeezable>, but only the conceptual representation {lime} can account for the specific combination of observed features, including the presence of <green> and the absence of <yellow>.

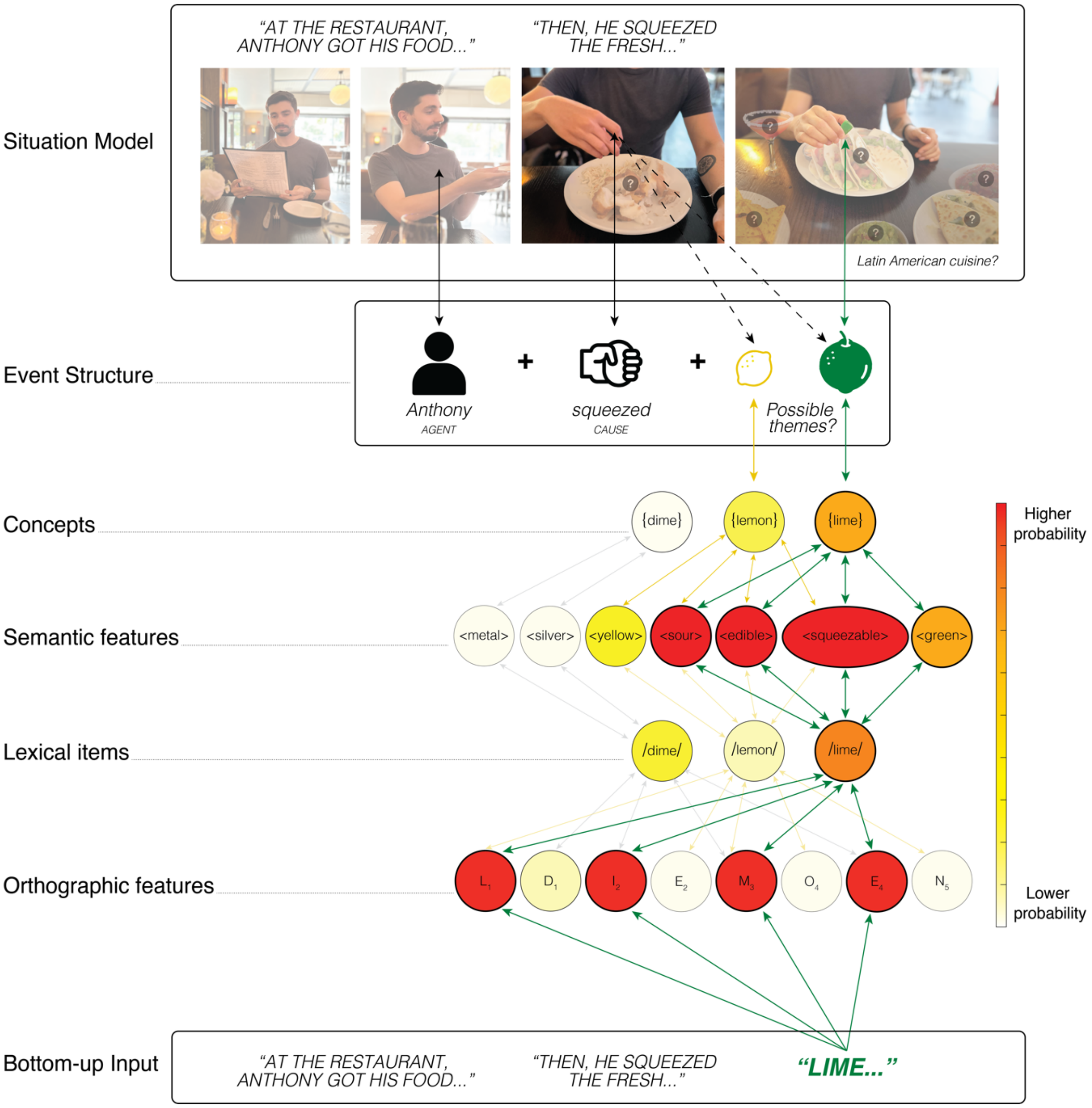

Figure 4. Schematic illustration of the state of one part of a comprehender's internal generative model at approximately 400ms after observing the new bottom-up orthographic input, L-I-M-E, following the context, "At the restaurant, Anthony got his food. He squeezed the fresh…".

At the level of lexical items: Most belief is centered over the lexical item, /lime/, which (a) provides the best explanation for the full set of observed orthographic features, L-I-M-E, and (b) is best explained by the unique set of semantic features that is being inferred (or retrieved) at the level above. At this point, there is also some belief over /lemon/, which was the more likely prior lexical candidate *before* the bottom-up input was encountered, but belief in /lemon/ will continue to fall because it cannot explain all the observed orthographic features, L-I-M-E. Finally, at this point, there is some belief over /dime/, which is able to explain several of the observed orthographic features ("I-M-E"). However, as belief continues to rise over /lime/, it will fall over this competing orthographic lexical neighbor, which is said to be "explained away".

At the level of semantic features: Belief is rising over the unique combination of semantic features (<sour> and <edible> and <squeezable> and <green>), which provides the best explanation of the most probable lexical candidate, /lime/, that is being inferred at the lexical level below. At this point in time, there is also some belief over the unique combination of semantic features associated with the concept, {lemon} (<yellow> and <sour> and <edible> and <squeezable>), which was the most probable concept *before* the bottom-up input was encountered. However, belief over this particular combination will continue to fall because it does not provide the best explanation for the more probable lexical candidate, /lime/. Note that the probability of the *individual* semantic features that are shared by the conceptual representations of {lime} and {lemon}, i.e., <sour>, <edible>, and <squeezable>, will each remain high (at 100%); it is primarily the probability of the feature, <yellow>, that will fall. Therefore, at the level of semantic features, any *change* in belief induced by the bottom-up input (or, equivalently, the amount of "work" of retrieving or accessing these features from semantic memory) will be less than the change in belief that is induced either at the lexical level below or at the conceptual level above. Finally, at this point in time, there is also some belief over the particular set of semantic features that can explain /dime/, which as noted above, is being inferred with lower probability at the level below. However, the comprehender's belief in this unique combination of semantic features will continue to fall (a) because it cannot explain the more probable lexical candidate, /lime/, that is being inferred at the level below, and (b) because it cannot be explained by the more probable unique concept, {lime}, that is being inferred at the level above.

At the level of concepts: Most belief is centered over the concept, {lime} — the latent cause that (a) provides the best explanation of the most probable unique combination of semantic features (<sour> and <edible> and <squeezable> and <green>) that is being inferred at the level below, and (b) is also explained by the event structure that is being inferred at the level above. At this point, any belief over the concept {dime} is minimal because of the lack of both bottom-up and top-down support. Finally, at this point in time, there is some remaining belief over {lemon}, which was the more likely conceptual candidate *before* the bottom-up input was observed. However, as belief continues to rise over {lime}, it will fall over this competing conceptual candidate, which is said to be "explained away".

At the level of event structures: As belief rises over the concept, {lime}, it will increase over the specific event {Anthony squeezed the fresh lime}. Note, given the preceding context, the correct syntactic structure has already been inferred. Therefore, within this framework, the successful "access" of a word's semantic features and its underlying conceptual representation by 500ms (at the end of the N400 time-window) will often equate to successful "lexical integration" — the integration of a word into its local event/proposition.

At the level of the situation model: Note that integrating a word into its local event/proposition does not necessary equate to integrating the newly inferred event into the entire situation model. Updating the entire situation model may involve the additional inference (or retrieval) of new schema-relevant information. For example, in the present case, the reader is likely to have already inferred that Anthony is eating fish. In addition, as the reader becomes increasingly certain of the specific event, {Anthony squeezed the fresh lime}, she may additionally infer that Anthony is eating tacos, which may, in turn, lead her to retrieve more details about other items that he is eating (e.g. Mexican or Latin American cuisine). We suggest that this process of successfully updating belief at the level of the situation model by retrieving additional schema-relevant features may be linked to another ERP component that peaks at a later stage of processing — the late frontal positivity.

Moreover, as discussed by Lee and Mumford in their foundational paper describing hierarchical inference in the visual system, because belief flows dynamically up and down the hierarchical generative model, competitive inference over latent causes at higher levels of the hierarchy will continually influence competitive inference at lower-levels, and vice versa (Lee & Mumford, 2003, p. 1437). For example, as illustrated schematically in Figure 3, as belief rises over the lexical representation, /lime/, this will lead the comprehender to infer/retrieve its unique set of semantic features from long-term memory. The rise in belief over this unique set of semantic features will, in turn, provide new bottom-up evidence that leads the comprehender's prior conceptual beliefs to shift from {lemon} to {lime}. This, in turn, will provide new top-down evidence that induces a further rise in belief over /lime/ (*versus* /dime/) as the most likely lexical hypothesis.[14]

Critically, as belief rises over the correct latent conceptual and lexical representations ({lime} and /lime/), it falls over their competing neighbors ({lemon} and /dime/). This phenomenon exemplifies a form of Bayesian reasoning known as "explaining away" in which a rise in belief over one latent cause results in a reduction in belief over competing latent causes that share overlapping outcomes/observations. "Explaining away" is classically illustrated by the sprinkler problem: if we see wet grass and then find out that the sprinkler was on, our belief in the more likely hidden cause — that it was raining — decreases (Pearl, 1988). As we discuss later, explaining away can be implemented in neural networks that approximate Bayesian inference, where it provides a more "natural" mechanism for competitive selection than lateral inhibition (see Smolensky, 1986; Gaskell & Marslen-Wilson, 1997).

These basic principles of probabilistic inference can explain two key aspects of the current findings. First, they can explain why, in the *WithCompetitor* contexts that constrained for upcoming words that were semantically related to one another, we found evidence of friendly but not competitive pre-activation. As explained above, in a Bayesian framework, multiple prior beliefs can be maintained in parallel. Thus, after reading the context, "At the restaurant, Anthony got his food. He squeezed the fresh....", the conceptual/lexical representations of *both*

[14] This has important implications for theories of language comprehension: It implies that "lexical access" (inferring the lexical item that best explains a set of form features) is inherently intertwined with, and inseparable from inferring its conceptual representation. Moreover, to the degree that a word's conceptual representation (e.g. {lime} is part of an event, e.g. {Anthony squeezed the lime}, then this also implies that "lexical access" is inherently linked to "lexical integration". However, we emphasize that integrating a single word into its local context to infer an event within a single proposition is not the same as updating a still-higher-level situation model, based on this newly inferred event. The latter process may additionally involve activating/retrieving new schema-relevant information from long-term memory. For example, inferring, {Anthony squeezed the lime} may lead the comprehender to update her situation model by increasing belief over (or, equivalently, retrieving) information related to Mexican or Latin American cuisine, see Figure 3. As discussed earlier, the successful update of the comprehender's higher-level situation model may be linked to the late frontal positivity ERP component, rather than the N400.

{lemon}/lemon/ and {lime}/lime/ are pre-activated in parallel, without any pressure to select between them. Moreover, because these pre-activated representations share common semantic features (i.e. <sour>, <edible>, and <squeezable>), and because the prior probability of each of these shared features is 100%, the average prior probability of encountering the set of *individual* semantic features that correspond to "lime" will be *greater* than 30% (i.e. greater than the prior probability of the *unique combination* of these features). As a result, when this lower probability target, "lime" is encountered, and its lexical and conceptual representation is actually inferred, the total *change* in probability *at the level of semantic features* will be less than the change in belief at either the lexical or conceptual levels.

Crucially, behavioral measures of processing and the N400 are primarily sensitive to changes at the level of semantic features, rather than at the lexical or conceptual levels. This has important implications for using estimates of *lexical* probability, based either on cloze or large language models, to predict behavioral and ERP measures of processing: Given that so many of the words that we encounter in natural text have semantically related alternatives (Luke and Christianson, 2016), these measures are likely to systematically *underestimate* the probability of encountering their *semantic* features and therefore their processing difficulty.

Second, these probabilistic principles can explain why, in the *WithCompetitor* contexts that constrained for upcoming words that were semantically *unrelated* to each other, we also saw no evidence of costs due to competitive pre-activation; that is, why the process of inferring the correct lexical and conceptual representation of the target was no more difficult than if the same target had been encountered with the same probability in a *NoCompetitor* context. To illustrate why this was the case, imagine reading a context like "Gina was about to eat her fish and fries. She squeezed the…". At this point, we may have a 70% prior belief in /ketchup/ and a 30% prior belief in /lemon/. As discussed earlier, in an IAC architecture, these two pre-activated unrelated lexical representations would begin to compete *before* the bottom-up input is encountered. Moreover, once “lemon” is encountered, its pre-activated lexical representation, /lemon/, would continue to receive lateral inhibition from the more strongly pre-activated item, /ketchup/. For both these reasons, it will be harder to select the correct target, /lemon/, in this *WithCompetitor* context than if there had been no competing alternative.

In contrast, in a Bayesian framework, there is no prior pressure to select between the pre-activated representations, /ketchup/ and /lemon/.[15] until new bottom-up evidence, becomes available. Moreover, once the new target input is encountered, the total change in belief that it induces is determined primarily by its own prior probability. For example, in the above example, encountering the word “lemon” should induce an increase in belief of 70% over its lexical representation, i.e. a change from 30% to almost 100%. This will be accompanied by a fall in belief of 70% to nearly 0% over /ketchup/; that is, the total change of belief is 70% in both directions. Similarly, in a *NoCompetitor* context, observing “lemon” will induce a 70% increase in belief over /lemon/. Because within this framework, all lexical probabilities must add up to 1, this will be accompanied by a 70% decrease in belief over the full set of alternative lexical representations in the lexicon.

Marr level 2: A neural network that approximates Bayesian inference: Predictive coding

Of course, principles that are specified at Marr's first level of analysis are not always applicable at Marr's second algorithmic level. However, there are some connectionist networks

[15] Within this framework, the only time when a comprehender would “pre-select” upcoming candidates during the pre-activation phase is if the context constrains for representations that are mutually incompatible with one another. For example, the selection restrictions of a verb can constrain either for semantic features associated with animate or inanimate entities (see Wang, Wlotko, Alexander, Schoot, Kim, Warnke, & Kuperberg, 2020 for recent evidence for this type of distributed pre-activation).

and algorithms that *can* approximate Bayesian inference, and, in these cases, the fundamental probabilistic principles outlined above should also apply. Specifically, within these types of neural networks, one can think of each lexical representation as corresponding to a specific pattern or "blend" of activity (cf. Smolensky, 1986) over the particular set of connectionist units that encode its unique set of semantic features. On the assumptions that (a) comprehenders pre-activate upcoming lexico-semantic information based on the full situation model they have constructed prior to encountering an incoming word, and (b) they allocate a fixed amount of resources for this pre-activation, each unique blend would be pre-activated in parallel, with a strength that mirrors its estimated prior probability.[16] Then, upon encountering new bottom-up input, activity would increase over the unique pattern that encodes the semantic features of the incoming word, while, at the same time decreasing over all other blends. Once again, however, if the incoming word's semantic features are compatible with semantic features that have been pre-activated as part of another word's unique blend, then that word should still receive additional facilitation, evoking a smaller N400 than one would expect based only on its lexical probability (the probability of encountering its *unique set* of semantic features).

There are several types of neural networks and algorithms that can approximate Bayesian inference. For example, a recent modification of the original IAC model – the Multinomial Interactive Activation model – has been shown to implement optimum Bayesian inference through Gibbs' sampling (McClelland, Mirman, Bolger & Khaitan, 2014). However, we believe

[16] This proposal assumes fully incremental language comprehension in which the comprehender (a) continually updates her situation model based the prior context, and (b) uses this updated situation model to generate top-down predictions that reach lower levels of representation before new lexico-semantic information becomes available from the bottom-up input. This, however, will not always the case, and will depend on multiple factors, including the presence of discourse coherence markers that can influence updates to the situation model (e.g. Xiang & Kuperberg, 2015), the comprehender's goals (e.g. Brothers, Wlotko, Warnke, & Kuperberg, 2020), the broader communicative environment (e.g. Delaney-Busch, Morgan, Lau & Kuperberg, 2019), the presentation rate of the linguistic stimuli (e.g. Camblin, Ledoux, Boudewyn, Gordon, & Swaab, 2007, Wlotko & Federmeier, 2015), as well as the speed of information flow across the cortex, see Kuperberg & Jaeger, 2016, Section 3.4, pp. 42-45 for discussion.

that a particularly promising approach for understanding both language comprehension, and the functional role of the N400, is *predictive coding* – a biologically plausible neural architecture and algorithm that has been proposed to approximate Bayesian inference in the brain (Mumford, 1992; Rao & Ballard, 1999; Rao & Ballard, 1997; Friston, 2005; Spratling, 2016a, 2016b).

In predictive coding, probabilistic inference is approximated by a particular dual-unit connectionist architecture that implements a particular optimization algorithm. Specifically, at each level of the representational hierarchy, "state units" actively generate top-down predictions that attempt to explain (or reconstruct) information that is observed at the level below. Any observed information at the lower level that fails to match these top-down predictions (residual information) produces activity within lower-level "error units", which is termed, "prediction error". This prediction error is then passed back up to the higher level where it is used to update the representations encoded within the state units. These updated state units will therefore produce more accurate predictions/reconstructions on the next iteration of the algorithm. This process repeats over multiple iterations and proceeds in parallel at multiple levels of the hierarchy until prediction error is minimized. At this point, the brain will have converged on the representations that best explain the bottom-up input.

In recent work, we have developed and implemented a predictive coding model of lexico-semantic processing in which we directly link the N400 component to the summed activity produced by lexical and semantic error units (i.e. the magnitude of lexico-semantic prediction error) as the model infers the conceptual and lexical representation from bottom-up orthographic inputs (Nour Eddine, Brothers, Wang, Spratling & Kuperberg, 2023; Nour Eddine, Brothers, & Kuperberg, 2022).

As in Chen and Mirman's (2012) IAC model, in our predictive coding model, each lexical unit is linked to a unique set of distributed semantic features. However, in contrast to this IAC architecture, there are no lateral inhibitory connections between state units within the lexical layer. Therefore, there are no competitive interactions between pre-activated conceptual or lexical representations. Instead, the selection of the correct lexical and conceptual representation begins only *after* the bottom-up input is observed. And, at this point, instead of competing through mutual lateral inhibition within any single layer of the network, the correct representation is selected through the type of global competition described above, in which all possible combinations of features at multiple levels of the hierarchy are considered and competing lexical and conceptual neighbors are explained away.

In predictive coding, explaining away occurs because state units at each level of representation suppress prediction error at the level below, thereby depriving their competing neighbors of their own inputs (see Spratling, De Meyer, & Kompass, 2009; Spratling, 2016a for discussion). For example, at the lexical level, /lime/ generates top-down predictions that suppress orthographic prediction error, thereby depriving potential lexical competitors (orthographic neighbors, e.g. /dime/) of their initial source of bottom-up activation, while at the conceptual representation (e.g. {lime}) generates semantic predictions that suppress semantic prediction error, thereby depriving potential conceptual neighbors (semantic competitors, e.g. {lemon}) of activity.

Our predictive coding model is able to simulate the time course of the N400, as well its sensitivity to multiple lexical and contextual variables. Notably, consistent with the empirical data, the magnitude of lexico-semantic prediction error is highly sensitive to an incoming word's contextual probability, but not the constraint of the prior context (the probability of the most likely lexical candidate). Also consistent with the present findings, prediction error is smaller to unexpected words that share semantic features with a predicted alternative (Nour Eddine, Brothers, Wang, Spratling & Kuperberg, under review; Nour Eddine, Brothers, & Kuperberg, 2022). This correspondence suggests that predictive coding may provide a promising theoretical account of the neural computations that support lexico-semantic processing and give rise to the N400 response.

Conclusion

To sum up, we find no evidence that, in contexts that constrain for more than one continuation, competitive interactions between pre-activated parallel graded predictions reduces lexico-semantic processing of incoming words, as indexed by the N400. We also find no evidence that competition from a higher probability candidate induces costs in processing a lower probability candidate at a later stage of processing, as indexed by the late frontal positivity. Instead, readers show processing benefits when they encounter lower-probability incoming words that are semantically related to a higher-probability alternative. These findings have important theoretical implications for informing models of predictive language processing, suggesting that routine top-down prediction does not rely on precisely the same mechanisms as those employed in language production. Finally, our results are consistent with hierarchical accounts of language comprehension based on probabilistic inference, such as predictive coding.

Acknowledgments

This work was funded by the National Institute of Child Health and Human Development (R01 HD082527 to G.R.K.) and the National Science Foundation (SPRF-FR 1715072 to E.M.). We would like to thank several people who contributed to constructing the experimental materials and to data collection, including Minjae Kim, Sofia Vizitiu, and Lena Warnke. We also thank

Nathaniel Delaney-Busch and Edward Wlotko for their technical guidance and insightful comments on experimental design.

Declarations of interest: none

Bibliography

Angele, B., Schotter, E. R., Slattery, T. J., Tenenbaum, T. L., Bicknell, K., & Rayner, K. (2015). Do successor effects in reading reflect lexical parafoveal processing? Evidence from corpus-based and experimental eye movement data. *Journal of Memory and Language, 79*, 76-96.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. doi: 10.1016/j.jml.2012.11.001

Bates, D. M., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi: 10.18637/jss.v067.i01

Brothers, T., Greene, S., & Kuperberg, G. R. (2020). *Distinct neural signatures of semantic retrieval and event updating during discourse comprehension.* Paper presented at the 27th Annual Meeting of the Cognitive Neuroscience Society, Boston, MA.

Brothers, T., Hoversten, L. J., & Traxler, M. J. (2017). Looking back on reading ahead: No evidence for lexical parafoveal-on-foveal effects. *Journal of Memory and Language, 96*, 9-22. doi: 10.1016/j.jml.2017.04.001

Brothers, T., & Kuperberg, G. R. (2021). Word predictability effects are linear, not logarithmic: Implications for probabilistic models of sentence comprehension. *Journal of Memory and Language, 116*. doi: 10.1016/j.jml.2020.104174

Brothers, T., Wlotko, E. W., Warnke, L., & Kuperberg, G. R. (2020). Going the extra mile: Effects of discourse context on two late positivities during language comprehension. *Neurobiology of Language, 1*(1), 135-160. doi: 10.1162/nol_a_00006

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., . . . Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 1877-1901. doi: 10.5555/3495724.3495883

Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., & Swaab, T. Y. (2007). Processing new and repeated names: Effects of coreference on repetition priming with speech and fast RSVP. *Brain Research, 1146*, 172-184. doi: 10.1016/j.brainres.2006.07.033

Chen, Q., & Mirman, D. (2012). "Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors": Correction to Chen and Mirman (2012). *Psychological Review, 119*(4), 898-898. doi: 10.1037/a0030049

Chow, W. Y., Lau, E. F., Wang, S., & Phillips, C. (2018). Wait a second! delayed impact of argument roles on on-line verb prediction. *Language, Cognition and Neuroscience, 33*(7), 803-828. doi: 10.1080/23273798.2018.1427878

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*(5-6), 507-534. doi: 10.1080/01690960143000074

Davenport, T., & Coulson, S. (2011). Predictability and novelty in literal language comprehension: An ERP study. *Brain Research, 1418*, 70-82. doi: 10.1016/J.Brainres.2011.07.039

Delaney-Busch, N., Morgan, E., Lau, E., & Kuperberg, G. R. (2019). Neural evidence for Bayesian trial-by-trial adaptation on the N400 during semantic priming. *Cognition, 187*(June 2019), 10-20. doi: 10.1016/j.cognition.2019.01.001

DeLong, K. A., Chan, W. H., & Kutas, M. (2019). Similar time courses for word form and meaning preactivation during sentence comprehension. *Psychophysiology, 56*(4), e13312. doi: 10.1111/psyp.13312

DeLong, K. A., & Kutas, M. (2020). Comprehending surprising sentences: Sensitivity of post-N400 positivities to contextual congruity and semantic relatedness. *Language, Cognition and Neuroscience, 35*(8), 1044-1063. doi: 10.1080/23273798.2019.1708960

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

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. doi: 10.1016/J.Jneumeth.2003.10.009

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*(6), 641-655. doi: 10.1016/s0022-5371(81)90220-6

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491-505. doi: 10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D. (2022). Connecting and considering: Electrophysiology provides insights into comprehension. *Psychophysiology, 59*(1), e13940. doi: 10.1111/psyp.13940

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469-495. doi: 10.1006/Jmla.1999.2660

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

10.1016/j.brainres.2006.06.101

Fischler, I. S., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior, 5*, 1-20. doi: 10.1016/S0022-5371(79)90534-6

Fischler, I. S., & Bloom, P. A. (1985). Effects of constraint and validity of sentence contexts on lexical decisions. *Memory and Cognition, 13*, 128-139.

Fitz, H., & Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology, 111*, 15-52. doi: 10.1016/j.cogpsych.2019.03.002

Freunberger, D., & Roehm, D. (2016). Semantic prediction in language comprehension: evidence from brain potentials. *Language, Cognition and Neuroscience, 31*(9), 1193-1205. doi: 10.1080/23273798.2016.1205202

Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language, 95*, 200-214. doi: 10.1016/j.jml.2017.04.007

Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360*(1456), 815-836. doi: 10.1098/Rstb.2005.1622

Galton, F. (1907). Vox Populi. *Nature, 75*, 450-451. doi: 10.1038/075450a0

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12*(5-6), 613-656. doi: 10.1080/016909697386646

Greene, S., Brothers, T., Weber, E., Noriega, S., & Kuperberg, G. R. (2020). *The time course of predictability and plausibility effects during discourse comprehension.* Paper presented at the 27th Annual Meeting of the Cognitive Neuroscience Society, Boston, MA.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences, 14*(8), 357-364. doi: 10.1016/j.tics.2010.05.004

Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & de Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences of the United States of America, 119*(32), e2201968119. doi: 10.1073/pnas.2201968119

Hubbard, R. J., Rommers, J., Jacobs, C. L., & Federmeier, K. D. (2019). Downstream behavioral and electrophysiological consequences of word prediction on recognition memory. *Front Hum Neurosci, 13*, 291. doi: 10.3389/fnhum.2019.00291

Ito, A., Corley, M., Pickering, M., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language, 86*, 157-171. doi: 10.1016/j.jml.2015.10.007

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23-49. doi: 10.1016/j.brainres.2006.12.063

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

Kuperberg, G. R., Brothers, T., & Wlotko, E. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience, 32*(1), 12-35. doi: 10.1162/jocn_a_01465

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

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 17*(1), 117-129. doi: 10.1016/S0926-6410(03)00086-7

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

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163. doi: 10.1038/307161a0

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software, 82*(13).

Lai, M. K., Rommers, J., & Federmeier, K. D. (2021). The fate of the unexpected: Consequences of misprediction assessed using ERP repetition effects. *Brain Res, 1757*, 147290. doi: 10.1016/j.brainres.2021.147290

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A, 20*(7), 1434. doi: 10.1364/josaa.20.001434

Levelt, W. J. (2001). Spoken word production: a theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America, 98*(23), 13464-13471. doi: 10.1073/pnas.231459498

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, 213. doi: 10.3389/fnhum.2014.00213

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology, 88*, 22-60. doi: 10.1016/j.cogpsych.2016.06.002

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676-703. doi: 10.1037/0033-295X.101.4.676

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language, 92*, 57-78.

Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 262*(841), 23-81. doi: 10.1098/rstb.1971.0078

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1-86. doi: 10.1016/0010-0285(86)90015-0

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition. *Cognitive Science, 38*(6), 1139-1189. doi: 10.1111/cogs.12146

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*(5), 375-407. doi: 10.1037//0033-295x.88.5.375

Michaelov, J. A., Coulson, S., & Bergen, B. K. (2021). So Cloze yet so Far: N400 amplitude is better predicted by distributional information than human predictability judgements. *arXiv* *preprint arXiv:2109.01226*.

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological Cybernetics, 66*(3), 241-251. doi: 10.1007/BF00198477

Nakamura, M., & Phillips, C. (2022). *Pre-activated lexical items inhibit each other: a quantitative analysis of speeded cloze data.* Paper presented at the 35th Annual Conference on Human Sentence Processing.

Narayanan, S., & Jurafsky, D. (2001). *A Bayesian model predicts human parse preference and reading times in sentence processing.* Paper presented at the Advances in Neural Information Processing Systems 14 (NIPS 2001), Vancouver, BC.

Ness, T., & Meltzer-Asscher, A. (2018). Lexical inhibition due to failed prediction: Behavioral evidence and ERP correlates. *J Exp Psychol Learn Mem Cogn, 44*(8), 1269-1285. doi: 10.1037/xlm0000525

Ness, T., & Meltzer-Asscher, A. (2021). Love thy neighbor: Facilitation and inhibition in the competition between parallel predictions. *Cognition, 207*, 104509. doi: 10.1016/j.cognition.2020.104509

Ness, T., & Meltzer-Asscher, A. (2021). Rational adaptation in lexical prediction: The influence of prediction strength. *Frontiers in Psychology, 12*, 622873. doi: 10.3389/fpsyg.2021.622873

Ng, S., Payne, B. R., Steen, A. A., Stine-Morrow, E. A. L., & Federmeier, K. D. (2017). Use of contextual information and prediction by struggling adult readers: Evidence from reading times and event-related potentials. *Scientific Studies of Reading, 21*(5), 359-375. doi: 10.1080/10888438.2017.1310213

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., . . . Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *Elife, 7*, e33468. doi: 10.7554/eLife.33468

Nour Eddine, S., Brothers, T., & Kuperberg, G. R. (2022). The N400 in silico: A review of computational models. In K. Federmeier (Ed.), *Psychology of Learning and Motivation* (Vol. 76, pp. 123-206): Academic Press.

Nour Eddine, S., Brothers, T., Wang, L., Spratling, M., & Kuperberg, G. R. (under review). A predictive coding model of the N400. *bioRxiv*. doi: 10.1101/2023.04.10.536279

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language, 67*(4), 426-448. doi: 10.1016/j.jml.2012.07.003

Pearl, J. (1982). Reverend Bayes on inference engines: A distributed hierarchical approach. In D. Waltz (Ed.), *Proceedings of the American Association for Artificial Intelligence National Conference on AI* (pp. 133-136). Pittsburgh, PA: AAAI Press.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, California: Morgan Kaufmann.

Peirce, J. W. (2007). PsychoPy--Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13. doi: 10.1016/j.jneumeth.2006.11.017

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(04), 329-347. doi: 10.1017/S0140525X12001495

R Core Team. (2022). R: A language and environment for statistical computing (Version 4.2.2). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Rao, R. P. N., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation, 9*(4), 721-763. doi: 10.1162/neco.1997.9.4.721

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79-87. doi: 10.1038/4580

Rayner, K., Pollatsek, A., Drieghe, D., Slattery, T. J., & Reichle, E. D. (2007). Tracking the mind during reading via eye movements: comments on Kliegl, Nuthmann, and Engbert (2006). *Journal of Experimental Psychology: General, 136*(3), 520-529; discussion 530-527. doi: 10.1037/0096-3445.136.3.520

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin and Review, 3*(4), 504-509. doi: 10.3758/BF03214555

Roland, D., Yun, H., Koenig, J. P., & Mauner, G. (2012). Semantic similarity, predictability, and models of sentence processing. *Cognition, 122*(3), 267-279. doi: 10.1016/j.cognition.2011.11.011

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. doi: 10.1037//0278-7393.14.2.344

Smolensky, P. (1986). Neural and conceptual interpretation of PDP models. In D. E. Rumelhart, J. L. McClelland & PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol 2: Psychological and Biological Models* (pp. 390-431). Cambridge, MA: MIT Press.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1521-1543. doi: 10.1037/0278-7393.24.6.1521

Spratling, M. W. (2016). A neural implementation of Bayesian inference based on predictive coding. *Connection Science, 28*(4), 346-383. doi: 10.1080/09540091.2016.1243655

Spratling, M. W. (2016). Predictive coding as a model of cognition. *Cognitive Processing, 17*(3), 279-305. doi: 10.1007/s10339-016-0765-6

Spratling, M. W., De Meyer, K., & Kompass, R. (2009). Unsupervised learning of overlapping image components using divisive input modulation. *Computational Intelligence and Neuroscience, 2009*, 381457. doi: 10.1155/2009/381457

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: critical review and theoretical interpretation. *Language and Linguistics Compass, 9*(8), 311-327. doi: 10.1111/lnc3.12151

Steen-Baker, A. A., Ng, S., Payne, B. R., Anderson, C. J., Federmeier, K. D., & Stine-Morrow, E. A. L. (2017). The effects of context on processing words during sentence reading among adults varying in age and literacy skill. *Psychol Aging, 32*(5), 460-472. doi: 10.1037/pag0000184

Stroop, J. R. (1932). Is the judgment of the group better than that of the average member of the group? *Journal of Experimental Psychology, 15*(5), 550-562. doi: 10.1037/h0070482

Szewczyk, J. M., & Federmeier, K. D. (2022). Context-based facilitation of semantic access follows both logarithmic and linear functions of stimulus probability. *Journal of Memory and Language, 123*. doi: 10.1016/j.jml.2021.104311

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

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: frontal positivity and N400 ERP components. *International Journal of Psychophysiology, 83*(3), 382-392. doi: 10.1016/j.ijpsycho.2011.12.007

Van Berkum, J. J. A. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and Pragmatics: From Experiment to Theory* (pp. 276-316). Basingstoke: Palgrave Macmillan.

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. *Journal of Cognitive Neuroscience, 22*(1), 67-82. doi: 10.1162/jocn.2008.21170

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176-190. doi: 10.1016/j.ijpsycho.2011.09.015

Wang, L., Wlotko, E., Alexander, E. J., Schoot, L., Kim, M., Warnke, L., & Kuperberg, G. R. (2020). Neural evidence for the prediction of animacy features during language comprehension: Evidence from MEG and EEG Representational Similarity Analysis. *Journal of Neuroscience, 40*(16), 3278-3291. doi: 10.1101/709394

Wlotko, E. W., & Federmeier, K. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex, 68*, 20-32. doi: 10.1016/j.cortex.2015.03.014

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *Neuroimage, 62*(1), 356-366. doi: 10.1016/j.neuroimage.2012.04.054

Wong, R., Veldre, A., & Andrews, S. (2022). Are there independent effects of constraint and predictability on eye movements during reading? *J Exp Psychol Learn Mem Cogn*. doi: 10.1037/xlm0001206

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience, 30*(6), 648-672. doi: 10.1080/23273798.2014.995679

Zirnstein, M., van Hell, J. G., & Kroll, J. F. (2018). Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition, 176*, 87-106. doi: 10.1016/j.cognition.2018.03.001

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162-185. doi: 10.1037/0033-2909.123.2.162