Unexpected guests: When disconfirmed predictions linger

Stephanie Rich (skrich@ucsc.edu)
Department of Linguistics, 1156 High Street
Santa Cruz, CA 95064 USA

Jesse A. Harris (jharris@humnet.ucla.edu)
Department of Linguistics, 3125 Campbell Hall
Los Angeles, CA 90095 USA

Abstract

A vast amount of literature suggests that the language processor generates expectations about upcoming material. Several studies have found evidence for a prediction error cost in cases where the comprehender encountered not the predicted word but a plausible unexpected continuation instead. This cost is argued to be a result of an inhibitory process that suppresses activation of the originally predicted word. Other studies have found no such evidence for a prediction cost. In a probe recognition memory task, we find evidence for interference from an incorrectly predicted word, and in a self-paced reading study, we find evidence for facilitation when the originally predicted word is encountered later on in the sentence. Taken together, our results provide evidence against a strong version of the suppression account, in which all incorrectly predicted words are fully inhibited. Instead we argue in favor of a passive lingering activation account, in which activation for the disconfirmed prediction gradually decays over time.

Keywords: lexical prediction; suppression and inhibition; language processing

Introduction

A growing stream of recent studies has converged on the idea that the language processing system predicts upcoming input. Although evidence in favor of prediction has mounted in many areas of language processing (Kuperberg & Jaeger, 2016), lexical prediction has received by far the most attention. While the mechanisms of prediction are still up for debate, the central findings are that given a sufficiently constraining context, comprehenders may predict a likely upcoming word or words, thus facilitating lexical processing. However, comprehenders are not always presented with the expected word form, in which case, the prediction is disconfirmed. Fewer studies have explicitly investigated cases in which an initial prediction was found to be incorrect (but see Kutas, 1993; Delong et al., 2011; Frisson et al., 2017; Ness & Meltzer-Asscher, 2018, and others, discussed below).

Lexical prediction

The predictability of a word in context is often operationalized in terms of its cloze probability, i.e., the proportion that the specific word form is provided as the continuation to a given fragment (Taylor, 1953). Highly predictable words have been shown to facilitate sentence processing across methods and experimental designs. In eye-tracking while reading studies, readers spend less time fixating highly predictable words, and are more likley to skip them entirely (Ehrlich & Rayner, 1981; Morris, 1994; Rayner & Well, 1996). In visual world studies, listeners anticipatorily look to images that correspond to likely theme objects (*a cake*) following highly constraining verbs (*eat*) compared to controls (Altmann & Kamide, 1999). Finally, electroencephalography (EEG) studies have found an inverse relationship between the cloze value of a word and the amplitude of the corresponding N400 waveform, an event-related potential (ERP) component reflecting a spike in negativity roughly 400ms post-stimulus (Kutas & Hillyard, 1984; Kutas & Federmeier, 2011, for review). In these studies, the amplitude of the N400 is significantly decreased following a high cloze word. In addition, the N400 amplitude appears to show a graded sensitivity to predictability: low cloze words that are semantically related to a high cloze words elicit lower N400 peaks than unrelated controls (Federmeier & Kutas, 1999).

There are already several different conceptualizations of the mechanisms underlying lexical prediction in the literature, though it is currently unclear whether these accounts are mutually exclusive. Prediction may be conceptualized as a *passive* process, involving spreading activation across a number of semantically related words (e.g., Staub, 2015), or as an *active* one, in which specific representations are actively predicted, compatible with findings that words of a similar phonological form to the predicted word are also activated (e.g., DeLong et al., 2005). In the case of active prediction, accounts differ on whether a specific form of the word or a collection of lexical features is predicted. Additional debate concerns whether the language processing system *pre-updates* the sentence representation with the predicted word, and under what circumstances, or if the lexical representation is merely *activated* in advance of the encounter. In theory, both are possible. For example, a sufficiently deep commitment to a prediction could lead to pre-updating the sentence with the predicted word (Lau et al., 2013). Otherwise, the predicted word or feature set would merely receive additional activation. It may also be possible that prediction is sensitive to higher-level information from the broader discourse, such as event or situation models, which would then provide top-down information about likely upcoming events (Kuperberg, 2021).

The present studies were not designed to test the possible mechanisms of prediction. We simply rely on an assumption that there is *some* pre-activation for likely upcoming words, and then explore the mechanisms implicated in modulating the activation of disconfirmed predictions.

Disconfirmed predictions

Studies of lexical prediction often exploit highly constraining contexts resulting in rather high cloze values for a specific word. However, it is unclear how often comprehenders encounter situations with these extremes, which are relatively rare in naturally-occuring text (Luke & Christianson, 2016). In any event, there are often many unlikely, but still plausible, words that may felicitously follow a highly constraining context. An open question in the literature is whether there is a *prediction error cost* for disconfirmed lexical predictions.

In one account of prediction error, the predicted word form is actively inhibited or suppressed when another word is found in its stead (Kutas, 1993), an idea which has has been employed in other areas of sentence processing research. For example, Gernsbacher & Faust (1991) proposed that the irrelevant meanings of ambiguous homophones are suppressed once a likely meaning has been selected for integration.

Differences in methods may contribute to different accounts, as distinct findings have been observed in eye-tracking compared to EEG research. While encountering an implausible or anomalous continuation may disrupt reading (Rayner et al., 2004), there is currently no evidence in reading that a disconfirmed prediction results in a processing penalty (Luke & Christianson, 2016; Frisson et al., 2017). In contrast, EEG studies reveal not only a posterior spike in positivity ('post-N400-positivity' or PNP) in response to anomalous continuations, but also a frontal PNP following unexpected but plausible continuations (Van Petten & Luka, 2012; DeLong et al., 2014; Ness & Meltzer-Asscher, 2018). It has been proposed that this component reflects an inhibitory process in which the previously predicted word(s) are suppressed (Kutas, 1993; Van Petten & Luka, 2012; Federmeier, 2007; Delong et al., 2011), supported further by evidence that this component is correlated with inhibitory effects found in cross modal lexical priming (Ness & Meltzer-Asscher, 2018).

Present study

Here we explore three hypotheses regarding disconfirmed predictions. Each hypothesis assumes that a representation of the predicted word is pre-activated in some way by prior context. The hypotheses differ in the status of the representation once the prediction is disconfirmed. The hypotheses make distinct experimental predictions that are tested in a probe recognition task and a self-paced reading study.

We first consider a strong version of the *suppression hypothesis*, in which the activation for a highly likely word is actively inhibited by the processor in all cases. In this case, a disconfirmed prediction would result in a prediction error cost. Furthermore, as a result of suppressed activation of the predicted word, this hypothesis predicts that there should be no difficulty in rejecting the predicted word in a probe recognition task, and no facilitation for encountering the predicted word later on in a sentence during reading.

We also consider a hypothesis in which activation for the predicted word lingers, and may therefore influence later processing. We entertain two versions of the *lingering activation hypothesis*: an *active* version in which heightened activation for the predicted word is actively maintained in memory, and a *passive* version in which activation for the predicted word passively and gradually decays following the disconfirmation. Both versions of the lingering activation hypothesis predict difficulty in rejecting the predicted word in a probe recognition task. They further predict some facilitation for encountering the predicted word later on in the sentence during reading. The facilitation would decrease as distance from the prediction site to the predicted word increases under the passive version only.

Memory probe recognition task

Materials

In a 3 × 2 design, a memory probe task crossed constraint and probe word. Thirty triplets (Table 1) contrasted neutral context baselines (Neutral) with contexts that provided constraint towards a target word (*spider*) either through an immediately preceding adjective (*hairy*; Local-Constraint), or through broader context in the preceding clause (*he was afraid of creepy crawlers*; Global-Constraint), adapted in part from materials in Fitzsimmons & Drieghe (2013). Items were normed in a cloze task with a set of 30 subjects from the same population, but who did not participate in the probe recognition task. A final set of 18 items were selected from the original set, such that the predicted word had a high cloze probability across constraining contexts ($M_{Global} = 0.74$, $M_{Local} = 0.69$, $M_{Neutral} = 0.25$). The predicted target word was replaced by an unpredicted but semantically compatible word (*mouse*), matched with the predicted word on lexical characteristics known to affect reading and lexical decision times such as length, frequency, number of syllables and morphemes, and orthographic neighborhood size. The probe word following the sentence was either the encountered control word or the predicted target word.

The materials were interspersed with 50 filler items, 18 of which were similar in structure to the experimental items. The remaining 32 fillers constituted the experimental items from an unrelated manipulation and unrelated fillers. A total of 68 items were presented in a counterbalanced and individually randomized order.

Participants

Forty-eight self-reported native speakers of English participated. Subjects were recruited from a Psychology SONA Subject Pool, and received one course credit as compensation for a single session lasting no longer than thirty minutes. Participants who reported learning English after the age of eight or who scored below 80% on comprehension questions were excluded from the final dataset and were replaced.

Constraint	Sentence			Probe word	
	Preceding context	*Target*	*Post-target*	*Predicted*	*Encountered*
Neutral	John was afraid of many things and screamed when he saw the	mouse	in the corner.	SPIDER RT (M =1082, SE = 22) Acc (M = 98%, SE =1)	MOUSE RT (M = 965, SE = 17) Acc (M = 97%, SE =2)
Local	John was afraid of many things and screamed when he saw the hairy	mouse	in the corner.	SPIDER RT (M = 1236, SE =26) Acc (M = 96%, SE =2)	MOUSE RT (M =1002, SE = 17) Acc (M = 99%, SE =1)
Global	John was afraid of creepy crawlers and screamed when he saw the	mouse	in the corner.	SPIDER RT (M = 1196, SE =29) Acc (M = 94%, SE =2)	MOUSE RT (M =953, SE = 16) Acc (M = 97%, SE =2)

Table 1: Experiment 1. Sample sextet of items with reaction times and accuracy for probe words. Sentences were presented word by word.

Procedure

Participants read sentences presented one word at a time in rapid serial visual presentation (RSVP) with 200ms between each word. The experiment was presented with Linger software. Words appeared in the center of the screen in black monospaced text on a white background on a CRT monitor. Responses were recorded using a PS/2 keyboard. Before the sentence was displayed, a fixation cross appeared on screen. Following each sentence, a probe word in CAPS was presented. Participants were instructed to respond as quickly and as accurately as possible as to whether or not they had seen the probe word in the preceding sentence using the 'f' and 'j' keys, respectively; see Figure 1. Half of the sentences were followed by an additional two-alternative forced-choice comprehension question.

The experiment was conducted in an isolated room dedicated to testing. Subjects wore headphones playing pink noise at a low volume to mask unexpected sounds. Untimed breaks were provided every 10 trials. Experimental sessions lasted approximately 30 minutes on average.

To perform this task correctly, participants needed to reject the predicted word as a word they had seen in the sentence. A suppression account predicts that there should be no interference from a disconfirmed predicted, as there would be no lingering activation for the word. Therefore, it would be equally easy to reject the probe in a condition with highly constraining context and in condition with a neutral context.

In contrast, both the active and passive lingering activation accounts predict interference from the still-active predicted word. Under either version, rejecting the predicted word probe should be more difficult in highly constraining contexts than in unconstrianing, neutral contexts.

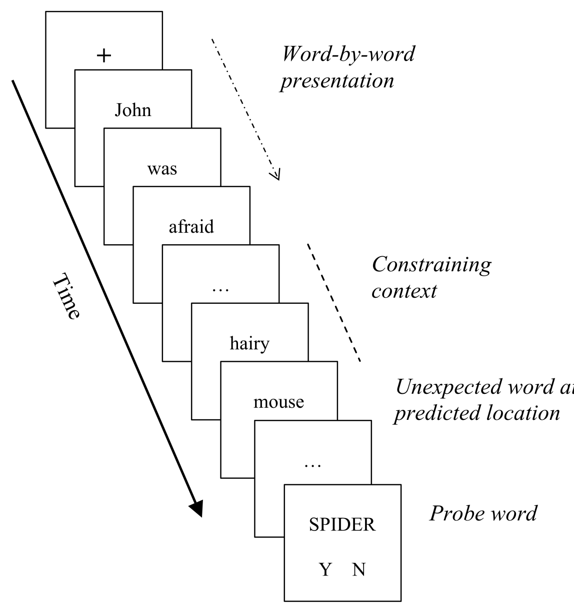

Figure 1: Rapid serial visual presentation with post-sentence memory probe.

Results

Prior to analysis, response latencies below the 5th or above the 95th percentile were censored, i.e., recoded as values at the 5th or 95th percentile, in each condition. Mean response times, as well as accuracy data, can be found in Table 1. The data were fit as linear mixed regression effects models with Constraint and Probe word as fixed effect predictors. Random effect structures were specified with by-subjects and by-items random intercepts, after random slope models failed to converge.

Response times to the predicted word probe were significantly slower than response times to the encountered word probes ($\hat{\beta} = 99.24, SE = 7.50, t = 13.23, p < .001$). It should be noted that these response types may be qualitatively different; a correct response to the predicted word is a rejection ('no' response), whereas a correct response to the encountered word requires acceptance ('yes' response). One might hypothesize that rejections yielded longer response times for reasons independent of the manipulated factors.

However, response times to reject the predicted word also varied by condition. The Constraint variable was Helmert coded to first compare any form of constraint (Global and Local combined) against the Neutral context, and then to compare between Local-Constraint and Global-Constraint. Two main effects were found. First, responses following any kind of constraint were slower than those following unconstraining contexts ($\hat{\beta} = 48.81, SE = 10.59, t = 4.61, p < .001$). Second, responses to Local-Constraint were slower than Global-Constraint ($\hat{\beta} = 23.55, SE = 9.21, t = 2.56, p < .05$). Most importantly, there was an interaction in which any form of constraining context yielded slower responses to the expected word, but did not elicit different response latencies to the encountered word ($\hat{\beta} = 40.89, SE = 10.59, t = 3.86, p < .001$).

Discussion

Unsurprisingly, subjects were faster to accept words that had appeared previously in the sentence than reject those that had not. This result could reflect a general advantage of overt exposure to the encountered probe word, as well as a penalty for rejecting a word not present in the sentence, even though the decision is correct. However, the penalty for correctly rejecting the predicted word was greater in constraining contexts, with no difference between Global and Local conditions, than in the neutral constraint condition.

Assuming that longer latencies reflect some degree of interference from the predicted word, the results are compatible with either an active or passive lingering activation account, but are unexpected under a suppression account. According to lingering accounts, residual activation from the predicted word should interfere with rejections of the word. In contrast, the suppression account predicts no interference for the predicted word in any of the constraining contexts.

It is not yet possible to tease apart the lingering activation accounts from these results alone, as both make similar predictions for the probe recognition task. However, the accounts can be distinguished by how long activation from the predicted would be expected to linger in memory. To test this prediction, we conducted a second study examining potential facilitation from lingering activation of a disconfirmed prediction in self-paced reading. The following study manipulated the distance between the target word and the highly constraining context that would have provided bias towards it. The active lingering activation account predicts facilitation regardless of distance, while the passive account would expect facilitation to diminish as distance increased.

As both types of constraint contributed to interference for a predicted word in the probe recognition study, only one form of constraint was used, simplifying the design and increasing power. Furthermore, the design allowed for identical critical regions and response patterns, addressing a potential concern about the different types of responses being made in Experiment 1.

Self-paced reading

Materials

The self-paced reading study was designed to examine the effect of reading a previously predicted (target) word after that prediction had been disconfirmed. A 2×2 Latin square design was employed crossing Constraint (Global, Neutral) and Distance to predicted word (Near, Far). Sixteen items from Experiment 1 were modified to separate the context and target word into two sentences. Contextually constraining information appeared in the first sentence. The second sentence contained a verb frame that, following a biasing Global-Constraint context, would strongly bias towards the target word (*spider*). Instead, an unexpected, but plausible continuation was presented (*He screamed when he saw the **mouse***), and the presentation of the target word was delayed until the next clause (*but he didn't notice the **spider***). To probe the time course of lingering activation, the distance between the prediction site and the target word was also manipulated (Near, Far). The Far condition was distinguished by an additional adverbial in its own region, which was unlikely to affect the overall meaning of the sentence. A sample item can be found in Table 2.

Materials were interspersed with 78 other items, 52 from unrelated experimental manipulations, and presented in counterbalanced and individually randomized order.

Participants

A distinct set of 48 self-reported native speakers of English participated in Experiment 2. Recruitment and replacement procedures were identical to Experiment 1.

Procedure

Participants read sentences in black 11 pt monospaced font on a white background on a CRT monitor in the same room as Experiment 1. Sentences were presented in a non-cumulative self-paced moving window reading task presented with Linger, using the presentation regions illustrated in Table 2. Participants progressed through the sentences using a space bar and were asked to minimize the movements of their fingers. Half of the sentences were followed by comprehension questions as in Experiment 1.

Results

Outliers were identified and censored in a procedure identical to Experiment 1. Mean reading times after outlier transformation for the critical region can be found in Table 2. Linear mixed effect regression models were constructed with Constraint, Distance, and their interaction as fixed effects, and by-subjects and by-items random intercepts, after random slope models failed to converge. Fixed effects were given deviation contrast coding with the Neutral context and Near position as reference levels for their respective conditions.

Analysis focused on the critical region, the target word (e.g. *the spider*), which was constant across all conditions. Overall, the Near condition was read more slowly than the

Constraint	**Preceding sentence context**			
Neutral	It was obvious / that John was afraid / of many things.			
Global	It was obvious / that John was afraid / of creepy crawlers.			
Distance	**Target sentence by region**			
	Prediction site	*Intervening*	*Target*	*Post-target*
Near	He screamed when he saw / the **mouse** / in the corner	but he didn't notice	the **spider** *Neutral:* $M = 558, SE = 21$ *Global:* $M = 507, SE = 13$	crawling on the wall / of the bedroom.
Far	He screamed when he saw / the **mouse** / in the corner	but somewhat surprisingly / he didn't notice	the **spider** *Neutral:* $M = 519, SE = 19$ *Global:* $M = 533, SE = 19$	crawling on the wall / if the bedroom.

Table 2: Experiment 2. Sample quartet of items with reading times for the target word. The unexpected but semantically congruous word (***mouse***) and the word predicted in its location (***spider***) are provided in bold for illustration, but were not bolded in the experiment. Region breaks for presentation and analysis are indicated by column breaks or '/' within a column.

Far condition ($\hat{\beta} = -9.38, SE = 3.91, t = -2.40, p < .05$) independent of the context manipulation. This effect was qualified by an interaction in which the difference between Neutral and Global-constraint was larger in the Near condition than the Far condition ($\hat{\beta} = 11.86, SE = 3.91, t = 3.04, p < .01$). In other words, the penalty for the Near condition was reduced following Global Constraint. To further probe the interaction, the `emmeans` package (Lenth, 2020) was used to estimate the marginal means of the model to compare the difference between Neutral and Global-constraint contexts at Near and Far positions. Compared to the Neutral context condition, Global constraining context elicited faster reading times in the Near ($\hat{\beta} = -34.80, SE = 11.0, 95\% CI[-56.51, -13.10], p < .01$) but not the Far ($95\% CI[-9.07, 34.30]$) condition; see Panel B in Figure 2.

Discussion

Reading time data indicated that disconfirmed lexical predictions facilitate processing when encountered later in the sentence. A strong version of the suppression account would not predict the presence of any lingering activation to later facilitate lexical processing or integration, again providing evidence in support of either an active maintenance or passive decay account. The difference between Neutral and Global Constraint contexts emerged in the Near condition, where Globally constraining context elicited faster reading times in the context of an overall slowdown on the critical region.

Although visually the interaction appears to be driven by the Neutral condition in Figure 2, it is important to note that an additional region intervened between the prediction site and target region in the Far condition. As a result, the critical region occurred in a later presentation window in the Far condition compared to the Near condition. Thus, baseline reading times for each presentation window may differ, and, indeed, there was a main effect of Distance in the critical region in which the Near condition was read more slowly overall.

It appears the interaction in this region reflects an advantage for Globally constraining context over the Neutral baseline in the Near condition, but not the Far condition, when more material intervened between the target word and the prediction site, allowing more time for activation of the predicted word to decay. We interpret this result as preliminary evidence against an active lingering activation account, which would predict a consistent benefit for Global constraint throughout the sentence.

General Discussion

In a probe recognition memory task and self-paced reading study, we set out to understand whether activation for a predicted word lingers after the prediction was disconfirmed. We entertained two versions of the lingering-activation hypothesis: one in which activation for the predicted word is *actively maintained*, and one in which the activation may *passively decay* over the course of the following sentence.

The predictions of these accounts were tested in two studies. In a probe recognition memory task, we found that a predicted word was harder to reject when it was highly probable in preceding context compared to neutral contexts. In a self-paced reading study, the predicted word elicited faster reading times, even after the prediction had already been disconfirmed. However, this advantage quickly faded. Both effects are predicted by an account in which activation for an incorrectly predicted word lingers following disconfirmation. We take these findings as evidence against the strongest version of the suppression account, in favor of one of the lingering-activation hypotheses. The time course patterns of the self-paced reading study further provide preliminary evidence in support of a *passive* lingering-activation account.

There are several possible interpretations of lingering activation. If prediction is in part a phenomenon in which activation may be passed down from higher, more abstract discourse representations, such as schemas (e.g. Abelson & Schank, 1977) or situation models (e.g. Zwaan & Radvansky,

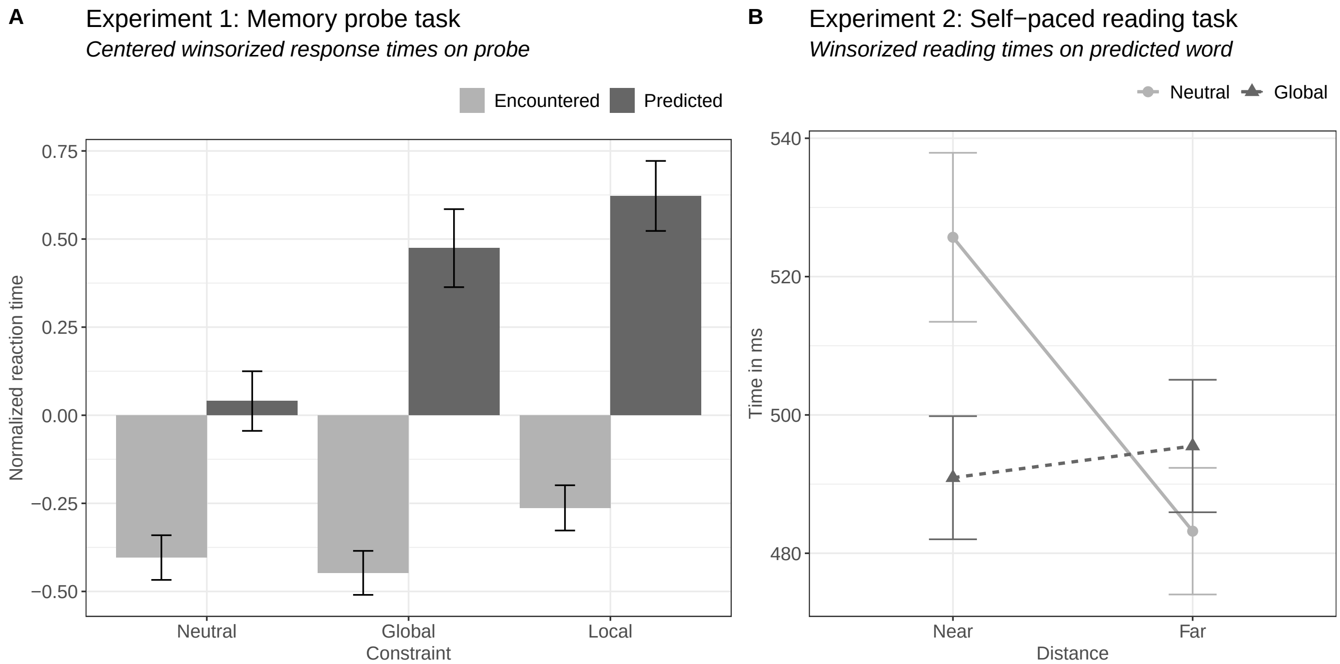

Figure 2: Experiment 1. Response times normalized and centered to the grand mean on the probe task (Panel A). Experiment 2. Reading times on the critical region (Panel B).

1998), a disconfirmed lexical prediction might only relate to the expectation of a particular form and not the full event representation.

As an example, one of our items presented *resort* as an unexpected word following a context about building sandcastles. Even though the target word *beach* is likely to be more expected, encountering *resort* is not incompatible with a situation model containing a resort. In contrast, if *sandbox* had been encountered instead, one might have to update the situation model to represent contexts in which sandboxes are found, e.g., playgrounds or parks instead of beaches. One possibility, then, is that activation from the relevant event is not suppressed if the encountered word does not drastically change the representation of the overall situation.

Our account might also be reinterpreted in terms of the noisy-channel model of sentence processing (Levy, 2008; Gibson et al., 2013), in which comprehenders would maintain some degree of uncertainty for the encountered word if there had been a highly likely alternative. It should be noted, though, that the predicted and encountered words in this study were not close orthographic or phonological neighbors and a comprehender would need to maintain uncertainty despite little perceptual overlap between forms.

As discussed earlier, a number of studies have observed effects that support a suppression mechanism for disconfirmed predictions, whereas others have not. Resolving these apparently conflicting findings remains an important open issue. We explore several possible explanations here. One explanation is that the methodologies used across studies are simply too different and prevent a direct comparison. Most evidence for inhibition is found in ERP studies, which often display stimuli at a slower rate than behavioral tasks. Frisson et al. (2017) speculate that ERP studies may enhance predictive behavior beyond what would be found in normal reading, supported by findings that manipulating the SOA in priming studies influences prediction effects (Lau et al., 2013). Our results therefore can't necessarily speak to the nature of any one ERP component, and in particular the interpretation of the frontal PNP with respect to inhibitory processes. It is entirely possible that an inhibitory process detectable via EEG would not be evident in the behavioral record.

In principle, our results are compatible with any account in which activation for disconfirmed predictions linger. We have argued here for one version of this hypothesis: namely that activation for a falsely predicted word gradually decays over time following the original prediction site. However, our results may also be compatible with a suppression account that results in partial or imperfect suppression, possibly arising from a noisy interpretation process, though it remains to be seen how these accounts would be distinguished on empirical grounds.

Alternatively, suppression may not be applied uniformly, resulting in variable results across studies. To this end, Ness & Meltzer-Asscher (2018) propose that inhibition might only occur if the processor had committed to a specific prediction, and had not only pre-activated the word but pre-updated the context representation as well, in line with Lau et al. (2013). Ness & Meltzer-Asscher (2018) raise the question of whether top-down processes modulate the commitment process, positing that the proportion of disconfirmed predictions in a study may affect whether inhibition takes place. In our self-paced reading study, participants never saw the predicted word in its predicted position, and in the probe recognition task the predicted word never occurred in the sentence. As a result, participants may have been less likely to fully commit to a predicted word in our study as compared to others.

To conclude, our results provide evidence against the strongest interpretation of a suppression hypothesis, which predicts uniform inhibition for disconfirmed lexical predictions. If inhibition is not a process applied uniformly, as these and other results suggest, the central question about activation levels for disconfirmed predictions remains. We present suggestive evidence in favor of a *passive* lingering activation account, in which activation for incorrectly predicted words gradually decays after disconfirmation.

References

Abelson, R., & Schank, R. C. (1977). *Scripts, plans, goals and understanding*. Erlbaum.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

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

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

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.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Jrnl of Vrbl Lrng and Vrbl Behvr*, *20*(6), 641–655.

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

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

Fitzsimmons, G., & Drieghe, D. (2013). How fast can predictability influence word skipping during reading? *JEP:LMC*, *39*(4), 1054–1063.

Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *JML*, *95*, 200–214.

Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: a component of general comprehension skill. *JEP:LMC*, *17*(2), 245–262.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *PNAS*, *110*(20), 8051–8056.

Kuperberg, G. R. (2021). Tea with milk? A hierarchical generative framework of sequential event comprehension. *Topics in Cognitive Science*, *13*(1), 256–298.

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

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *LCP*, *8*(4), 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). *Ann Rev Psych*, *62*, 621–647.

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

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

Lenth, R. V. (2020). emmeans: estimated marginal means, aka least-squares means [Computer software manual]. (R package version 1.5.3)

Levy, R. (2008). A noisy-channel model of human sentence comprehension under uncertain input. In *Emp Meths in NLP* (pp. 234–243).

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cog Psych*, *88*, 22–60.

Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *JEP:LMC*, *20*(1), 92–103.

Ness, T., & Meltzer-Asscher, A. (2018). Lexical inhibition due to failed prediction: Behavioral evidence and erp correlates. *JEP:LMC*, *44*(8), 1269–1285.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *JEP:LMC*, *30*(6), 1290–1301.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psych Bull & Rev*, *3*(4), 504–509.

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *LLC*, *9*(8), 311–327.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Qrtly*, *30*(4), 415–433.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and erp components. *Intl Jrnl of Psychophysiology*, *83*(2), 176–190.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psych Bull*, *123*(2), 162–185.