AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393

https://doi.org/10.1037/xlm0001091

The Power of "Good": Can Adjectives Rapidly Decrease as Well as Increase the Availability of the Upcoming Noun?

Jakub M. Szewczyk[1], Emily N. Mech[1], and Kara D. Federmeier[1, 2, 3]
[1] Department of Psychology, University of Illinois at Urbana-Champaign
[2] Program in Neuroscience, University of Illinois at Urbana-Champaign
[3] Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana–Champaign

OPEN DATA OPEN MATERIALS

Can a single adjective immediately influence message-building during sentence processing? We presented participants with 168 sentence contexts, such as "His skin was red from spending the day at the . . ." Sentences ended with either the most expected word ("beach") or a low cloze probability completion ("pool"). Nouns were preceded by adjectives that changed their relative likelihood (e.g., "neighborhood" increases the cloze probability of pool whereas "sandy" promotes beach). We asked if participants' online processing can be rapidly updated by the adjective, changing the resulting pattern of facilitation at the noun, and, if so, whether updates unfold symmetrically—not only increasing, but also decreasing, the fit of particular nouns. We measured event-related potentials (ERPs) to the adjective and the noun and modeled these with respect to (a) the overall amount of updating promoted by the adjective, (b) the preadjectival cloze probability of the noun and, (c) the amount of cloze probability change for the obtained noun after the adjective. Bayesian mixed-effects analysis of N400 amplitude at the noun revealed that adjectives rapidly influenced semantic processing of the noun, but did so asymmetrically, with positive updating (reducing N400 amplitudes) having a greater effect than negative updating (increasing N400s). At the adjective, the amount of (possible) updating was not associated with any discernible ERP modulation. Overall, these results suggest the information provided by adjectives is buffered until a head noun is encountered, at which point the access of the noun's semantics is shaped in parallel by both the adjective and the sentence-level representation.

Keywords: sentence processing, N400, updating, incremental processing, noun phrases

Supplemental materials: https://doi.org/10.1037/xlm0001091.supp

Sentence comprehension relies on quick access to words' meanings. Whether listening or reading naturally, people take in words at a rate of 2–3 words per second (Levelt et al., 1999). This speed is possible, in part, because the language comprehension system incrementally builds a representation of the meaning conveyed by the sentence and uses this information to make potentially relevant aspects of meaning become available or more easily accessible in advance. The strength of the contextual support for a particular word at a given point in a sentence or longer discourse is often operationalized using *cloze probability* (Taylor, 1953), which is the percentage of individuals that continue a sentence fragment with that word in an offline sentence completion task. The fact that more predictable words are easier to process online is well-attested by, for example, the robust relationship between cloze probability and the N400 component of the event-related potential (ERP) response to words (Federmeier et al., 2007; Kutas & Hillyard, 1984). The N400, which peaks just before 400 ms and is part of the normal response to any word or other complex perceptual stimulus, has been linked to the access of meaning information from long-term semantic memory (Kutas & Hillyard, 1980; for review, see Kutas & Federmeier, 2011). N400 amplitudes are reduced (become less negative) to the extent that the meaning of stimulus has already become active, and there is a strong inverse correlation between cloze probability and the amplitude of N400.

Jakub M. Szewczyk https://orcid.org/0000-0003-4464-082X
Emily N. Mech https://orcid.org/0000-0002-2003-4908
Kara D. Federmeier https://orcid.org/0000-0002-7815-1808

This project received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant agreement 800046. This work was also supported by National Institute on Aging Grant R01-AG026308 to Kara D. Federmeier. Authors wish to thank Gary Dell for valuable suggestions and Gabriel Manfroi for help with materials preparation.

The data are available at https://osf.io/5rtn4

The experiment materials are available at https://osf.io/5rtn4/

Correspondence concerning this article should be addressed to Jakub M. Szewczyk, who is now at Donders Institute for Brain, Cognition and Behaviour, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands. Email: jakub.szewczyk@gmail.com

The robust relationship between offline cloze probabilities and online brain measures indexing facilitation of word processing shows that an incrementally built conceptual representation of the context must make the features of some words more available. However, although it is clear that the availability of word features can be augmented in a supportive context, it is not yet evident whether context can also actively *decrease* the availability of features of less likely words. Answering this question is a central goal of the present study, which examines whether the language comprehension system can rapidly adapt to incoming information that reduces the likelihood of particular words. As we discuss next, there is ample evidence for increases in upcoming words' availability but mixed evidence for decreases in words' availability.

Incremental Accumulation of Positive Evidence for Upcoming Words

The psycholinguistic literature is rife with examples of the accumulation of positive evidence for upcoming words. One line of evidence comes from the measurement of eye movements in response to language while viewing a visual scene (i.e., the visual world paradigm; Tanenhaus et al., 1995). It has been regularly shown that listeners exploit various cues from language input, which rapidly affects their eye-gaze patterns. For example, listeners can immediately integrate information from the subject and a verb to determine the likely object of the sentence, which is evident given that they shift their gaze to the correct object before it occurs in the presented sentence (Borovsky et al., 2012; Kamide, Altmann, et al., 2003; for other examples of anticipatory eye movements see, e.g., Altmann & Kamide, 1999, 2007; Kamide, Scheepers, et al., 2003; see also review in Huettig et al., 2011).

Electrophysiological measures also attest to the rapid, incremental use of accruing context information to facilitate word processing, in the form of N400 amplitude reductions across word position in congruent sentences (Dambacher et al., 2006; Halgren et al., 2002; Payne et al., 2015; Van Petten & Kutas, 1990; Van Petten et al., 1991). When comprehenders start reading a new sentence, little is known about what it describes. Thus, the first few words of a sentence have low cloze probabilities and elicit large (negative) N400s. In coherent sentences, each upcoming word further clarifies what the sentence is describing and the average predictability of words will gradually increase, as evidenced by a consequent reduction in N400 amplitude. In contrast, when comprehenders read syntactic prose—sentences that are grammatical but have no coherent message-level meaning—N400s do not become smaller as a function of word position but remain at levels similar to those seen for initial words (Payne et al., 2015; Van Petten et al., 1991). The contrast between coherent and meaningless sentences clearly shows that information useful for facilitating semantic access accrues in supportive contexts and that the language comprehension system can and does routinely exploit it.

Other ERP studies have examined the ability of an incoming word to update contextual representations and thereby change processing of a subsequent word by examining responses to two linked words in the same sentence, such as a verb and its argument or an adjective and the following head noun (a head word is a word that determines the nature of a phrase, e.g., in the noun phrase "big white elephant," "elephant" is the head noun, preceded by two modifiers). These paradigms show two types of effects. The first type of effect occurs when the first word introduces semantic features that are not yet present in the context but that make the upcoming word more predictable. For example, Maess et al. (2016) compared ERPs to two variants of the verb in sentences such as "He will lead/conduct the orchestra." When compared with the verb "lead," "conducted" is semantically more specific, introduces more new information, and makes "orchestra" more predictable. The resulting amplitudes of the N400 across the verb and the noun show a trade-off: The more specific verb ("conducted") elicits a larger N400 amplitude (because it requires activating more new semantic information) but, consequently, the following noun has a reduced N400 as some of the relevant semantic features have already been accessed at the verb (for similar findings see also Boudewyn et al., 2015; Fleur et al., 2020; Freunberger & Roehm, 2017; Szewczyk & Wodniecka, 2020, incongruent items). These results mechanistically show how information gained at one word eases semantic access for another word.

The second pattern of effects occurs when the first word—for example, an adjective—does not provide new semantic feature information, but nevertheless could be useful for selecting among contextually likely upcoming head nouns (Szewczyk & Wodniecka, 2020, congruent items; for examples and more in-depth description, see section Current Study below). For example, morphological information on an adjective can constrain which contextually–primed nouns are likely to (grammatically) follow. In such a setup, it is possible to quantify how much information useful for selecting the upcoming noun is provided by the adjective. The results show that the N400 to the noun is parametrically explained by the sum of its preadjectival predictability and the change in its predictability driven by the adjective, again supporting the rapid use of the information provided by the adjective. However, notably, the ERPs to the adjective itself do not vary as a function of its informativeness, perhaps because, as we already noted, all semantic features of the adjective were already introduced by the context. We will return to the distinction between prenominal words that do and do not elicit N400 effects when discussing the results of the present study.

Incremental Accumulation of Negative Evidence for Upcoming Words

In all of the cases discussed thus far, new information strengthens support for one or more upcoming words (or, in the case of the visual world paradigm, items on the display), and brain or eye-gaze measures reveal concomitant facilitation for processing or selecting those items. What happens, instead, when a comprehender encounters a word that modifies the message such that words that may previously have been anticipated no longer fit—creating the need for what might be described as "negative updating"? Can the comprehension system move away from information as readily as it can incorporate it during online processing? Surprisingly, there is very little work addressing this question.

Note that all of the demonstrations of incremental language understanding briefly reviewed above do not provide clear evidence for or against negative updating. Most of the ERP experiments testing updating at verbs or adjectives (Freunberger & Roehm, 2017; Maess et al., 2016; Szewczyk & Wodniecka, 2020) used stimuli wherein the first word provided positive information about the second tested word. An exception is the study by Boudewyn and colleagues (Boudewyn et al., 2015), who presented participants with two-sentence items such as: "Frank was throwing a birthday party, and he had made the dessert from scratch. After everyone sang, he sliced up some sweet/healthy and tasty cake/veggies that looked delicious." The target noun was either highly predictable given the context ("cake"; mean CP = .78) or not predictable at all ("veggies"; CP = 0), and the preceding modifier was locally congruent with either the predictable ("sweet") or the unpredictable ("healthy") noun. The goal of the study was to determine if responses at the noun reflect facilitation from both the local context (the modifier) and the global context (the sentence), and the results showed that, indeed, both matter: There was a graded pattern of facilitation on the N400, with the smallest responses to nouns supported by both the local and the global context and the largest responses to nouns supported by neither context type. This pattern could reflect an influence from negative updating when the adjective mismatched the noun that had been rendered most likely by the global context. However, the results could also be explained entirely by positive updating (i.e., "sweet cake" might be more facilitated than "healthy cake" simply because of extra priming from "sweet" and not due to any negative updating induced by "healthy"). To be able to tease apart positive and negative influences, it is necessary to be able to estimate the size of the response in a neutral case (without influence from the adjective).

Similarly, in studies using the visual world paradigm, it is not generally possible to disentangle the contributions of positive updating from possible contributions of negative updating. For example, Sedivy et al. (1999) showed that providing early information that helps to identify an object (e.g., an instruction to "touch the plain red square" when viewing a display in which only one square is plain) speeds up the eye movements to the target object. However, it is impossible to know if this speed up came about solely because the adjective "plain" increased the activation of the target object or because there were also (or instead) decreases in the activation of other objects. Finally, although the overall correlation between word predictability and online measures of word processing is in line with the idea of negative updating, it is inconclusive without targeted measurements focused on words whose predictability has been decreased by the context.

There is some evidence that, at least in certain situations, negative updating does *not* occur, even though it should, as evidenced by offline cloze probability tests. For instance, under many circumstances, N400 responses are insensitive to negation, quantification, and counterfactuals, such that amplitudes to "bird" are similar in the context of "A robin is not a bird" as to "A robin is a bird," despite the very different off-line plausibilities for these two sentences (Ferguson et al., 2008; Fischler et al., 1983; Urbach & Kutas, 2010). Another example of a breakdown of the relationship between cloze probability and N400 amplitude are so-called thematic reversal anomalies, where N400 responses may be similar to the verb "eat" in "For breakfast the boys would eat..." and "For breakfast the eggs would eat..." because the language comprehension system fails to appreciate in time the structural role of "eggs" as a subject versus an object of the verb "eat" (Kuperberg et al., 2003; see, e.g., Bornkessel-Schlesewsky & Schlesewsky, 2008; Brouwer et al., 2012; Kolk & Chwilla, 2007; Kuperberg, 2007; Van De Meerendonk et al., 2009, for reviews and discussions). Comprehenders may also fail to deactivate meanings belonging to an event described in the past after transitioning to a new event (Delogu et al., 2019). Finally, at least in some cases, activations of predictable words can linger, even when those predictions were disconfirmed. For example, the sentence context "Be careful, because the top of the stove is very . . .," leads comprehenders to predict the word "hot." When they are instead presented with the (much less expected) word "dirty," the activation of "hot" remains high, as can be measured several items later in the form of reduced N400 responses when the word "hot" is presented as part of a completely different sentence (Rommers & Federmeier, 2018). Such expected-but-never-encountered words can have such a pervasive influence that they even lead to false memories (Hubbard et al., 2019). The effects described above demonstrate that, at least sometimes, words do not get deactivated in a timely manner.

In contrast to these findings, a recent eye-tracking study by Chow and Chen (2020) showed that comprehenders can quickly revise their predictions to a less predictable sentence continuation, at least when an appropriate visual referent was displayed on the screen. Furthermore, the examples of "failed updating" are each unusual in some way and may not generalize to more typical cases in which comprehenders encounter something unexpected. Negations, counterfactuals and quantifiers may be difficult to process out of context, but when context supports them—for example, when negation is licensed—comprehenders can make use of this information online (Nieuwland & Kuperberg, 2008). Sentences containing thematic role reversals are also a special case, and, although these may not affect the N400, they are appreciated shortly thereafter, showing effects on the semantic P600. Finally, the circumstances in which unrealized predictions have been shown to linger come from highly constrained words that were likely actively predicted—something that we know does not occur for all comprehenders in all circumstances (DeLong et al., 2012; Federmeier et al., 2002, 2010; Wlotko et al., 2012; Wlotko & Federmeier, 2012).

Thus, the question of how the comprehension system accommodates not only additional information but information that entails down-weighting predictions or revising the message remains underexplored. It is a fundamental question because one core function of language is to provide information, which, by definition, means that words that are unexpected and/or violate likely predictions are not uncommon.

The Current Study

In the present study we set out to examine how the system accommodates unexpected information across a variety of conditions, rendering previously predictable words both more and less predictable to varying degrees and over a range of initial constraint. To do this, we recorded electroencephalogram (EEG) as participants read simple sentences in English for comprehension. Sentences varied in how strongly they converged on a likely sentence-final noun (i.e., in their constraint). For each sentence, we selected two target nouns, one of which had the highest (or near-highest) cloze probability and the other of which had a relatively low cloze probability. We will refer to these nouns as HiCP and LoCP nouns, respectively. Prior work using these same materials has demonstrated that N400s to the nouns are graded by cloze probability (Federmeier et al., 2007). Here, critically, each target noun was preceded by an adjective designed to create an update in comprehenders' expectations. For example, consider the following set of sentences:

(1a) He liked lemon and sugar in his herbal tea
(1b) He liked lemon and sugar in his sparkling tea
(1c) He liked lemon and sugar in his sparkling water
(1d) He liked lemon and sugar in his herbal water

Given this sentence context, people tend to predict "tea" (HiCP noun) more than "water" (LoCP). We asked if preceding the noun with an adjective—"herbal" or "sparkling"—could change processing rapidly enough to affect responses to the noun when it arrives. The adjectives were chosen such that they promoted either the HiCP noun ("herbal") or the LoCP noun ("sparkling"). We will refer to these adjectives as pro-HiCP and pro-LoCP adjectives, respectively.

Importantly, when an adjective increased the cloze probability of one noun, it simultaneously decreased the cloze probability of the other noun. For example, "water" following the adjective "sparkling" was much more likely than "water" not following any adjective, but "tea" following "sparkling" was much less likely than "tea" not following any adjective. Of interest is how this additional information from the adjective is—or is not—used during online processing to affect semantic processing, as measured on the N400.

The most basic way in which the adjective could affect the noun would be through simple semantic priming, unfolding without any interaction between the adjective and the context (e.g., "sparkling" could semantically or associatively prime "water"). Note that priming effects of this type would only be expected to create additional facilitation (i.e., via spreading activation; Collins & Loftus, 1975; Quillian, 1967). In other words, the simplest effect of the adjective would look like an additive effect of facilitation from the context and priming from the adjective. In the example above, then, priming from the adjective would augment context-based facilitation for (1a) and (1c), resulting in reduced N400 amplitudes, which would align with the increased cloze probabilities observed for these nouns in the context of that adjective. However, in sentences (1b) and (1d), the adjectives do not prime the nouns; thus, N400 amplitudes would only be affected by the nouns' fit to the larger sentence context. Therefore, for half of the sentence types, N400 amplitudes will be misaligned with (specifically, smaller than) the pattern expected based on the measured postadjective cloze probabilities.

At the other end of the spectrum, it is possible that the adjective is rapidly and incrementally integrated with the unfolding sentence context and used to update the situation model before the noun arrives. In this case, in addition to possible effects of updating elicited at the adjectives themselves, we would expect to see bidirectional effects at the noun—namely, nouns being both more facilitated than would have been the case without the adjective (positive updating) and, crucially, also less facilitated than would have been the case without the adjective (negative updating). Thus, on this account, the N400 should pattern closely with the measured postadjective cloze probabilities. Of course, other patterns are possible as well. What is critical is that by assessing both the strength and direction of the impact that the adjectives have on semantic access of the noun (as measured on the N400), we gain insight into the mechanisms by which updating takes place as sentence processing unfolds.

In using adjectives that match or mismatch critical nouns with different global predictabilities, our approach is thus similar to that adopted by Boudewyn and colleagues (Boudewyn et al., 2015). We expect to replicate their finding that the N400 at the critical noun is influenced by both the global sentence context and the local adjective. However, because the Boudewyn et al. (2015) study used a factorial design that did not contain a neutral condition (i.e., a condition with no adjective-driven updating), it was not possible to quantitatively estimate the relative strength of the influence from the local and global context (although the result pattern suggested that the global context effect was more pervasive), nor to measure to what extent the baseline (premodifier) activation of the noun had increased or decreased as a result of processing the matching and mismatching modifiers. In other words, it did not address our central question of whether representations can be rapidly updated in the negative direction. In the current study we continuously manipulate the influence of the local and global context at the item level. This gives us the ability to estimate the N400 amplitude to the noun when it is not updated by the adjective, and to precisely quantify deflections from the baseline amplitude in both the positive and negative direction.

Because we expect adjective-driven modulations of the N400 at the noun, we will also look for correlates of updating at the adjectives themselves. There are a few alternative patterns of results that we could observe. One possibility is that the adjectives will get immediately integrated with the context and lead to updating of the situation model (the high-level representation of the meaning conveyed by the sentence), leading to the downstream effects at the noun. On surprisal and event prediction error theories (Levy, 2008; Rabovsky et al., 2018), such updating would be expected to elicit effects at the adjective, likely on the N400, such that N400 amplitudes at the adjective would vary with the amount of situation model updating induced by that adjective, with greater model updating resulting in larger N400 responses. To quantify the (potential) updating of the situation model we will use Bayesian surprise (Kullback-Leibler divergence; see Statistical Analysis below), indexing the amount of shift between the distribution of nouns' cloze probability before and after the adjective. The idea that N400 amplitudes elicited by the adjectives might vary with the amount of updating to the situation model is consistent with the pattern observed in studies showing that words that are more semantically informative about an upcoming word elicit larger N400s. This has been seen for verbs that are more informative about upcoming objects (Maess et al., 2016), adverbs that are more informative about upcoming verbs (Freunberger & Roehm, 2017), gender-marked determiners (Fleur et al., 2020) and, in both Boudewyn et al. (2015) and Szewczyk and Wodniecka (2020; incongruent condition), for adjectives that are more informative about upcoming nouns. However, these studies have in common one property that may be relevant for predictions for the current study: The more informative word was likely to bring online completely new semantic information, which would not have been made available directly by the preceding context. For example, in the study by Boudewyn et al. (2015), the noun "veggies" was not predictable given the global context (it had a cloze probability of zero, as "veggies" are not food items associated with birthday parties). Thus, the larger N400 is elicited at a point wherein additional semantic information is likely becoming active—for example,

Table 1
Examples of Experimental Items

Sentence Onset	Adjective	Noun
At night the old woman locked the	pro-HiCP: *front* pro-LoCP: *frosty*	HiCP: *door* LoCP: *window*
It was a dream come true because Jane had always wanted to visit Europe with her	pro-HiCP: *younger* pro-LoCP: *new*	HiCP: *sister* LoCP: *boyfriend*
They carelessly dumped the toxic waste into the	pro-HiCP: *deep* pro-LoCP: *winding*	HiCP: *ocean* LoCP: *river*
I had lost the ticket I needed to pick up my	pro-HiCP: *rental* pro-LoCP: *washed*	HiCP: *car* LoCP: *clothes*

Note. Italics indicate that this is the verbatim text that was displayed on participants' screen.

foods that can fit with the unexpected adjective "healthy". However, there are cases in which adjectives did not lead to any ERP effect of updating, even though they clearly modulated the N400 at the following noun (Szewczyk & Wodniecka, 2020, congruent items). In these cases, the adjective did not introduce completely new semantic information because it fit at least some of the nouns that were predictable in a given sentence context. For example, when Polish participants read the story fragment: "My mother decided that we should have a 'spring clean' in our house. She cleaned the living room and the kitchen, and my father's job was to clean the first floor. My job was to clean the entire$_{\text{MASC-IA}}$ . . .," the gender marking on the adjective "entire" is compatible with at least some noun completions that were made predictable before the adjective. In this case, there is no need for new semantic activation at the adjective. In the present study, the adjectives were always chosen such that they semantically fit with at least one word that was predictable in the context (in that it was generated in the cloze norming). Thus, these adjectives seem more likely to lead to no updating effect at the adjective itself.

We note, however, that prior findings of no N400 to adjectives that match some of the predictable nouns (Szewczyk & Wodniecka, 2020; congruent items) used morphosyntactic cues (gender and number agreement), whereas updating in the present study is based on the semantics of the adjective. It is possible that semantic updating always results in modulations of the N400 at the word introducing the update. In that case, our results should align with other studies that found N400 modulations associated with semantic updating (Boudewyn et al., 2015; Freunberger & Roehm, 2017; Maess et al., 2016).

Finally, the updating at the adjective could instead lead to a P600, as found in the study by Ness and Meltzer-Asscher (2018), wherein verbs that were more informative about the upcoming noun led to a more pronounced P600.[1]

Method

Participants

Thirty-four native speakers of English took part in the study. Participants were remunerated with course credit or received $25. Two participants were excluded because of very low scores on the sentence recognition test (i.e., below .45), leaving 32 participants entering the final analyses (nine male; $M_{\text{age}} = 21$ years, range = 18–34). All participants were right-handed as assessed by the Edinburgh Inventory (Oldfield, 1971) and 14 participants reported nonright-handed relatives in their direct family. All had normal or corrected-to-normal vision and none reported a history of neurological or psychiatric disorders. None of the participants was exposed to a second language before age 5. Procedures were approved by the Institutional Review Board (IRB) of the University of Illinois, and all participants gave written consent to participate.

Materials

Experimental items included 168 sentence frames that have been previously used in other work (e.g., Federmeier et al., 2007), modified for the purpose of the present study. Each of the sentences could be completed by one of two target nouns, both of which occurred in the original cloze norming: the highest or near-highest cloze probability (CP) noun (HiCP noun) and another noun with a CP lower than that of the HiCP noun (see Table 1 and Figure 1 for their mean CP values and ranges).[2] The noun was preceded by either of two adjectives: an adjective increasing the CP of the HiCP noun (pro-HiCP adjective) or an adjective increasing the CP of the LoCP noun (pro-LoCP adjective).[3] Examples of experimental items are given in Table 1 (the complete list of stimuli and all their parameters can be found online at Open Science Framework; https://osf.io/5rtn4/). To determine the CP of each of the two target nouns after each of the two adjectives, as well as to obtain newer CP norming for the target nouns not preceded by any of the adjectives, we conducted cloze probability norming with native English speakers residing in the United States using Amazon Mechanical Turk and Qualtrics. The norming was done in three testing waves

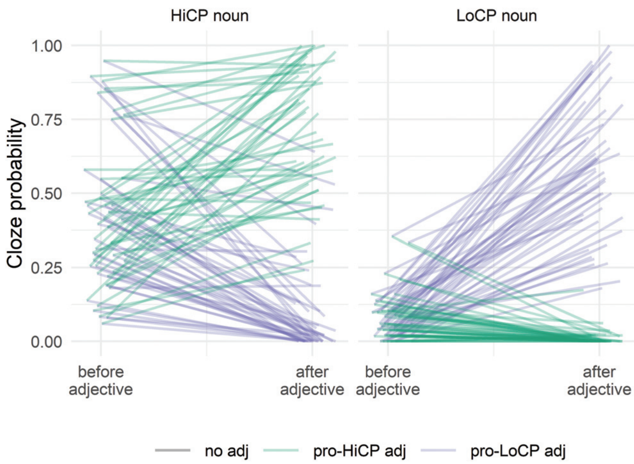
[1] In the above review of ERP studies looking at how adjectives may change the processing at the noun we omitted studies looking at adjective-noun pairs (Fruchter et al., 2015; Lau et al., 2016). Because these studies did not present any context before the adjective, they could not address the question of how adjectives change the on-line predictability of nouns that has been established before the adjectives.

[2] In items with constraint strength >.5, the HiCP noun always had the highest CP of all completions. For 19 items with lower constraint strength, we chose nouns with cloze probability slightly lower (on average, .075 lower) than that of the most likely completion because these alternative nouns were easier to accommodate in the experimental design.

[3] We use the term adjective broadly here, to include all modifiers that have the potential to change comprehenders' expectations about the noun, including noun adjuncts (e.g. "cargo pants") and participial adjectives (e.g. "flying bird").

Figure 1

By-Item Preadjectival Cloze Probability and Its Adjective-Driven Update, Shown for a Random Sample of 30% of the Items



Note. In each panel, values on the left correspond to the cloze probability of the noun when not preceded by the adjective. Each point on the left has two lines extending from it (purple and green) whose other ends (on the right side of each panel) indicate the updated cloze probability of the noun. The right end of the green lines indicates the noun's cloze probability following a pro-HiCP adjective (i.e., adjective increasing the cloze probability of the highest cloze probability noun), while the right end of the purple lines indicates the noun's cloze probability following a pro-LoCP adjective. The left panel shows cloze probability values for HiCP nouns, and the panel on the right shows values for LoCP nouns. Note that pro-HiCP adjectives increase the cloze probability of HiCP nouns while decreasing cloze probability of LoCP nouns (and vice-versa). See the online article for the color version of this figure.

during which we iteratively improved the materials to find adjectives that provide a good distribution of adjective-driven change in CP for each of the nouns (see below). In total, we recruited 1,165 participants. We made sure that none of the cloze norming participants completed the same sentence frame twice (e.g., the same sentence with a different adjective) within the same trimester. We excluded participants whose completions were highly uncorrelated with those of other participants (mean log-odds of completion across items lower than a threshold established individually for each testing wave, in the range between −1.75 and −2.4). In summary, we obtained CP norms for the sentence final noun for three versions of each item: sentences without an adjective, sentences with a pro-HiCP adjective, and sentences with a pro-LoCP adjective. Each variant of each item was completed by at least 33 participants (M = 52 participants). Even though in the experiment proper the target nouns were always preceded by one of the adjectives, the CP norms for sentences without the target adjective were necessary to compute how much each adjective affected the predictability of the target nouns (see Figure 1 and Table 1). The mean constraint strength for the target noun measured before the adjective was .46, but it varied considerably across items (SD = .25; range = .07 – 1.00) and the sentences varied from very weakly constraining to highly constraining.

As shown in Table 2 and Figure 1, pro-HiCP adjectives successfully increased the CP of HiCP nouns (except those nouns that already had a CP close to 1.0, in which case the adjective

Table 2

Descriptive Statistics for HiCP and LoCP Nouns Before an Adjective and After Pro-HiCP and Pro-LoCP Adjectives

ADJ cond	N cond	CP before ADJ	CP after ADJ	D_{KL}
pro-HiCP *front*	HiCP *door*	.45 (.26)	.75 (.20)	0.36 (0.25)
pro-HiCP *front*	LoCP *window*	.07 (.06)	.01 (.03)	
pro-LoCP *frosty*	HiCP *door*	.45 (.26)	.09 (.14)	1.00 (0.45)
pro-LoCP *frosty*	LoCP *window*	.07 (.06)	.55 (.21)	

Note. CP = cloze probability; ADJ = adjective; N = noun; D_{KL} = Kullback-Leibler divergence between pre- and postadjectival distribution over nouns' probabilities (see Statistical Analyses for explanation). Descriptive statistics correspond to all items and not to the provided examples. Values in parentheses indicate standard deviations.

maintained that CP), while reducing the CP of LoCP nouns. Conversely, pro-LoCP adjectives increased the CP of LoCP nouns and reduced the CP of HiCP nouns. Also, as shown by the mean D_{KL} values, pro-LoCP adjectives overall induced a larger change in the probability distribution of the upcoming noun, compared with HiCP adjectives.

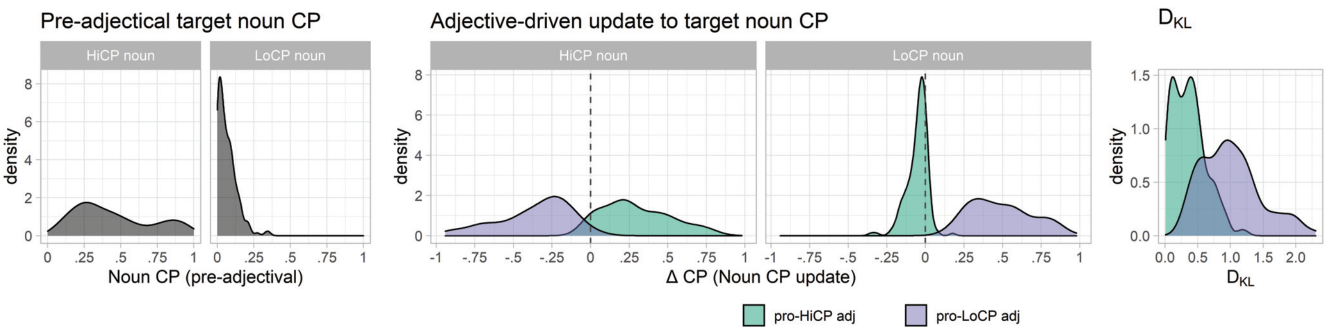
We generated four versions of each item by combining pro-HiCP and pro-LoCP adjectives with HiCP and LoCP nouns. The resulting stimuli were split into four experimental lists, each containing 168 items, with one version of an item per list, and with all combinations of adjectives and nouns for each item occurring across the four lists. In half of the items the noun had a very low cloze probability (i.e., HiCP nouns following pro-LoCP adjectives and LoCP nouns following pro-HiCP adjectives), which sometimes made the sentences implausible when combined with the adjective. As such, to increase the proportion of fully congruent sentences, we added to each of the lists 72 filler sentences taken from norms published by Peelle et al. (2020). The filler items were selected to be similar in structure to experimental items and to end with an adjective followed by a noun. In all filler items the noun had a CP of .5 or higher (M = .77).

Note that the categories of HiCP and LoCP nouns and pro-HiCP and pro-LoCP adjectives were primarily adopted to aid with stimulus creation, ensuring that across the items we obtained a good distribution of the values of CP and the adjective-driven change in CP (see Figure 2). However, from conception, the study was designed for item-level regression analyses. Thus, statistics were always done using the continuous indices of CP and of its change, as an index of updating (although the categories are also sometimes used for data visualization, as when presenting ERP waveforms).

Procedure

Participants were seated 85 cm from a 21″ CRT screen in a dim, quiet testing room. Stimuli were presented in white letters on a dark-grey background. At the beginning of each item's presentation, a fixation marker ("+++") was presented in the center of the screen for 500 ms, followed by a 400–800 ms blank screen, and then by the word-by-word presentation of the sentence. Each word was displayed for 300 ms, centered in the middle of the screen, followed by a 200 ms blank screen. After the offset of the last word, the procedure was paused until the participant pressed the spacebar key, and the screen remained blank for 1 s after the participant made the keypress, after which the fixation marker was

Figure 2
Distributions of the Three Critical Predictors: Preadjectival CP, CP Update, and D_{KL}



Note. Left: Distribution of cloze probabilities (CPs) for the target noun with the highest or near highest CP in a given sentence (HiCP noun) and a lower probability completion (LoCP noun), measured without the experimental adjective. Middle: Distribution of adjective-induced changes in HiCP and LoCP nouns' CP. Pro-HiCP/Pro-LoCP corresponds to adjectives promoting the HiCP/LoCP noun. Note that the x-axis has the same scale as the left panel. Right: Distribution of D_{KL} (Kullback-Leibler divergence) an index measuring the adjective-induced CP updating across all predictable nouns. See the online article for the color version of this figure.

presented marking the onset of the next item. Participants were instructed to read all items carefully, to remain still, and to avoid eye-movements and blinks in the period between the fixation marker and the end of the sentence. They were also told that, after the experiment, they would have to perform a sentence recognition test. After each block of 80 items, participants were offered a break. The sentences were presented in a pseudorandomized order with the restriction that the training and each break is followed by two filler items.

The recognition test was introduced to ensure that the participants maintained their attention while reading the experimental sentences. The test consisted of 36 sentences; 12 sentences were identical to those the subject had read during the experiment, 12 were similar (the target noun was replaced by the other noun associated with a given item), and 12 were completely new. Participants were asked to assign the sentences to one of these three categories (identical, similar, or new). The experiment started with a short training session of four sentences. The complete experiment (including electrode placement and removal and the off-line recognition test) lasted on average 2 hr.

EEG Recording and Preprocessing

EEG was recorded from 26 silver-chloride electrodes mounted in an Electro-Cap using an equidistant montage (see online supplemental materials Figure S1 for electrode layout), amplified through a Brain Products BrainAmpDC amplifier. Recordings were referenced online to the left mastoid and rereferenced offline to the mean of the left and the right mastoids. Additional electrodes were on the outer canthus of each eye to monitor horizontal eye movements, and on the left infraorbital ridge to monitor for vertical eye movements and blinks. Electrode impedances were kept below 5kΩ. The continuous EEG was sampled at 1000 Hz and amplified through a bandpass filter of 0.02–250 Hz. Off-line, the continuous recordings were filtered with a high-pass filter (two-pass Butterworth at 0.1 Hz, 12dB/octave).

The timing of stimuli display was confirmed using a stimtracker and a photodiode. Because our CRT monitor refreshed the screen every 16.67 ms (60 Hz) and because the stimuli were displayed in the middle of the screen, their onset lagged 8ms after the stimulus marker sent to EEG amplifier. For that reason, we shifted all markers 8 ms forward (2 data-points) before conducting the analyses.

Epochs from the continuous EEG in the interval between −100 and 900 ms with respect to the onset of the adjective and the noun were analyzed. Systematic artifacts resulting from eye movements, blinks, and artifacts resulting from poor electrode contact were filtered out using AMICA (Palmer et al., 2011) run on 1-Hz–filtered data restricted to periods when the sentences were displayed. We first removed all trials containing horizontal eye-movements detected using an individualized threshold on a step function convoluted with the ICA channels corresponding to horizontal eye-movements. Vertical eye-movements and blinks were removed using ICA, but only in those participants who had ocular artifacts in more than 30% of trials (15 participants) and only in segments containing the ocular artifacts (detected similarly to horizontal EOG artifacts). In the remaining participants, trials containing ocular artifacts were rejected. In four participants, we removed ICA channels corresponding to occasional single-electrode artifacts. Finally, we low-pass filtered ICA channels that showed consistent muscle noise activity (20 Hz threshold, two-pass Butterworth filter, 24dB/octave). Segments and electrodes that had remaining artifacts (skin potentials, occasional poor electrode contact, etc.) were rejected using a logistic regression-based algorithm trained on manually marked artifacts (the algorithm and weights were the same as used in Szewczyk & Wodniecka, 2020). When an artifact occurred on one target word only, the segment on that electrode was nevertheless rejected on both the adjective and the noun. On average, 5% of trials were rejected (range = 0–13%). All details of artifact correction procedure are available in the scripts online.

Instead of doing classical subtraction-based baseline correction, we regressed out the mean prestimulus baseline amplitude in the main statistical analysis (Alday, 2019). As the baseline, we used the mean amplitude in the 100 ms time-window preceding the target word. For the analysis at the adjectives, we used the preadjectival baseline, while for the analysis at the noun, we used predictors corresponding to both the preadjectival and prenoun baseline. This method did not qualitatively alter the overall results but yielded much more reliable estimates.

Statistical Analyses

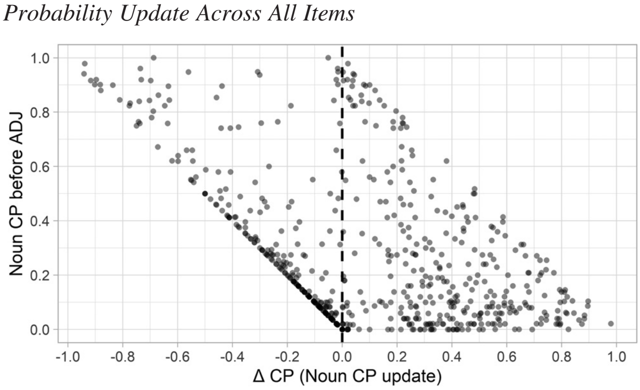
In this study we adopted a regression-based approach. In the analysis of the noun, we focused on two predictors derived from the cloze tests: preadjectival CP and CP update. Preadjectival CP was derived from cloze tests that were comprised of sentences truncated before the critical adjective. It reflects the nouns' predictability before it became affected by the processing of the adjective. The second predictor used in the analysis at the noun, which we will call CP update, corresponds to the change in the noun's CP induced by the adjective. It was computed as the difference between the postadjectival CP, obtained from cloze tests containing a given adjective, and the preadjectival CP. The sum of those two predictors corresponds to the *offline* predictability of the noun following the adjective. We separated this predictability into two subcomponents to test both the influence of the preadjectival context and the specific update from the adjective. As shown in Figure 3, the two predictors were distributed across the full possible range. In addition to the core question of whether updating occurs and is bidirectional, we were also interested in whether updating strength is modulated by the strength of the original prediction for the HiCP noun—that is, by sentential constraint. Finally, we wanted to test if the sensitivity to updating strength changes across the experiment (i.e., trial number) and may be under participants' strategic control.

To test if the amount of updating introduced by the adjective modulates ERPs at the adjective, we needed an index that would reflect the overall change in the distribution of nouns' predictability (note that we could not simply use CP update because it reflects a change in just one noun's predictability, whereas at the adjective the participant does not yet know which noun he or she will get). To this end, we used Kullback-Leibler divergence (D_{KL}; also called cross-entropy), computed using the following formula:

$$H(p,q) = \sum_{x \in X} p(x) \log q\,(x)$$

where X is the distribution of all nouns occurring in the cloze test collected before and after an adjective, q() and p() correspond to pre- and postadjectival cloze probability, respectively. This index

Figure 3
The Distribution of Preadjectival Cloze Probability and Cloze Probability Update Across All Items



Note. The dashed line divides the plot into negative (left) and positive (right) changes in cloze probability (CP) induced by the target adjective.

reflects the amount of Bayesian updating of the noun CP distribution necessary to shift from an initial belief about the probability of obtaining particular nouns (i.e., before the adjective is encountered) to beliefs in nouns' probability updated by seeing the adjective (both distributions estimated using cloze probability tests, one collected on sentences truncated before the adjective, and one collected with sentences truncated after either of the two adjectives). If the likelihood of the nouns gets rationally updated at the adjective, this index reflects the cost of the update (for other examples of uses of D_{KL} in the sentence processing literature and beyond, see Doya et al., 2007; Itti & Baldi, 2009; Levy, 2008; Rabovsky et al., 2018; Szewczyk & Wodniecka, 2020; Yan et al., 2017). Adjectives that do not lead to any change in the distribution of offline predictability yield D_{KL} equal to 0. Because in the norming some nouns had CP = 0 and because D_{KL} is defined only for nonzero entries, we first smoothed the CP distributions within-item using a Bayesian Dirichlet prior:

$$p_i' = \frac{p_i + 1}{N + 1}$$

where N = number of unique nouns in the combined pre- and postadjectival distributions.

The analysis of the N400 to the noun was done using Bayesian linear mixed-effects models with participants and items as random effects, using the "brms" package (Bürkner, 2017) in R (version 4.0.4; R Core Team, 2020).[4] As the dependent variable, we used the mean amplitude in the 300–500 ms time-window averaged over medial, centro-parietal electrodes (midline central and parietal, channels and left and right medial central channels). Following Rouder and colleagues (Rouder et al., 2018), we used two different approaches to Bayesian inference. For all effects we will report posterior intervals, which enables estimation of the effect's size. In addition, for those effects whose "existence" we want to test, we will use Bayes factors. To calculate the Bayes factor, we used bridge sampling (Bennett, 1976; Gronau et al., 2018; Meng & Wong, 1996). Reliable estimation of Bayes Factors requires many sampling iterations and for that reason, we used 14 chains and 28,500 iterations each, 1,500 of which were the warm-up phase (378,000 sampling iterations in total).

The mixed-effects model for the nouns had the following fixed effects: preadjectival baseline, prenoun baseline, preadjectival CP, and CP update. They also included four control variables: orthographic neighborhood (OLD20; Yarkoni et al., 2008), lexical log-frequency (Brysbaert & New, 2009), concreteness (Brysbaert et al., 2014), and the number of preceding words (position in the sentence). Even though CP updating was a within-item manipulation (so these variables were unlikely to be confounded with it), we included them to increase the precision of estimation. As reviewed in the beginning of the article, there is strong evidence for both CP effects, so for those we based our analysis on posterior estimation (we were interested in how the updating effect compares with the preadjectival CP effect). We used amplitude of the baseline as a predictor instead of subtracting it from the N400 amplitude

[4] A previous version of this manuscript contains analyses based on frequentist statistics. It yielded the same pattern of results. The only notable difference is that based on those analyses we could not reach any conclusions concerning null effects. These analyses can be found in version 2 of our preprint: https://psyarxiv.com/ytaz3/.

because this approach improves the model's fit and its ability to precisely estimate the effects of interest (Alday, 2019). Apart from the more standard preword baseline (i.e., prenominal) we also used a preadjectival baseline to accommodate the fact that only before the adjective was the EEG activity truly neutral with respect to the experimental condition. In all models, we used subjects and items as random variables and a maximal random effect structure (Barr et al., 2013) but without correlations between random effects.[5] All predictors were centered before entering the analyses. In addition, control predictors (OLD20, lexical log-frequency, concreteness, and position in the sentence) were normalized. Priors used in the baseline model of the N400 at the noun are shown in Table 3.

For control variables, we used weakly regularizing priors because we did not know what values the coefficients would take in a multiple regression including all the other predictors. The priors assumed that a change of predictor value by 2 *SD* should maximally lead to a change in N400 amplitude by ± 2 μV. For the remaining predictors, we used principled priors. Priors for the baselines were based on our experience with baseline predictors (see, e.g., Szewczyk & Wodniecka, 2020; Szewczyk & Federmeier, 2021, where typically, the preword baseline for the N400 effect for data after a 0.1 Hz high-pass filter had 0.3–0.7 weight. Because the baseline weight decreases with increasing time distance from the baseline, in the analysis of the noun we set the preadjectival prior at half of the prenominal baseline. The priors for CP effects were based on the rich literature assessing N400 effects of CP. Typically, a change in CP = 0 to CP = 1 translates to a 2–4 μV reduction of the N400 amplitude (see, e.g., Kutas & Hillyard, 1984; DeLong et al., 2005; Nieuwland et al., 2018; Szewczyk & Wodniecka, 2020; Szewczyk & Federmeier, 2021). Priors for all predictors were set such that they were also consistent with a null effect. For example, the priors for CP effects were consistent with effects in the -1 to 7 μV range.

At the adjective, we also ran a Bayesian mixed-effects model with N400 amplitude (defined in the same way as at the nouns) as a dependent variable. The baseline model included the preadjectival baseline and three nuisance predictors: lexical log-frequency, orthographic neighborhood, and word position in the sentence. We omitted concreteness as a predictor because the norms did not contain estimates for adjectives in 19 items. The model had the maximal random effects' structure but did not contain random correlations. Priors used in the baseline model at the adjective are shown in Table 3.

Table 3

Priors Used in the Baseline Bayesian Mixed Effects Models Analysis of the N400 at the Noun and at the Adjective

Prior name	Noun	Adjective
Intercept	$\mathcal{N}$ (0, 3)	$\mathcal{N}$ (0, 3)
Prenominal baseline	$\mathcal{N}$ (0.5, 0.3)	—
Preadjectival baseline	$\mathcal{N}$ (0.25, 0.15)	$\mathcal{N}$ (0.5, 0.3)
OLD20 (normalized)	$\mathcal{N}$ (0, 1)	$\mathcal{N}$ (0, 1)
Log-frequency (normalized)	$\mathcal{N}$ (0, 1)	$\mathcal{N}$ (0, 1)
Word position (normalized)	$\mathcal{N}$ (0, 1)	$\mathcal{N}$ (0, 1)
Concreteness (normalized)	$\mathcal{N}$ (0, 1)	—
Preadjectival CP	$\mathcal{N}$ (3, 2)	—
CP updating	$\mathcal{N}$ (3, 2)	—
Residual *SD*	$\mathcal{N}$ (10, 3)	$\mathcal{N}$ (10, 3)
SD in random effects	$\mathcal{N}$ (0, 2)	$\mathcal{N}$ (0, 2)

Note. OLD20 = orthographic neighborhood; CP = cloze probability.

We tested three extensions of this baseline model of the N400 at the adjective. The first extension tested whether Kullback-Leibler divergence (D_{KL}), our index of the amount of updating occurring at the adjective, predicts N400 amplitude to that adjective; in other words, do N400 amplitudes at the adjective directly index the amount of change in the probability distribution of the upcoming nouns? The second extension tested if N400 amplitude is predicted by an index of the adjective's predictability derived from a language model (see below). Finally, we checked if the N400 to the adjective gets reduced when there is a strong similarity between the adjective and the most predictable noun. As in the noun analysis, all three extensions were tested using Bayes factors under a range of realistic priors (sensitivity analysis, see below).

To anticipate the results, no effect of D_{KL} was found in this adjective analysis so we also ran cluster-based permutation tests to determine if there were effects of D_{KL} outside the N400 scalp distribution and/or time window (adapting to R the algorithm described in Maris & Oostenveld, 2007). We looked for clusters using all scalp electrodes in the 100–700 ms time-window. The first level tests were simple linear regressions done on ERPs averaged by-Item (this was a nonhierarchical analysis, so items were defined as each unique combination of context and adjective), with centered D_{KL} and preadjectival baseline as predictors. The cluster mass was based on the sum of *z*-scores of the D_{KL} effect. The permutations were conducted by inverting the centered D_{KL} values for a randomly sampled half of items.

Data and Materials' Availability

All materials, experimental scripts, and the code for running all analyses and generating all figures is available online at https://osf.io/5rtn4. The repository also contains all aggregated data necessary to duplicate our mixed-effects models. Preprocessed segmented EEG data are available via Harvard Dataverse repository at https://doi.org/10.7910/DVN/ZGLTB1, contingent on signing a "Terms of use" agreement.

Software

Data were analyzed and visualized in R (R Core Team, 2020). We used the following R packages: brms (Bürkner, 2017), data.table (Dowle & Srinivasan, 2021), doParallel (Weston & Microsoft Corporation, 2020), eegUtils (Craddock, 2019); ggplot2 (Wickham, 2016), ggridges (Wilke, 2021), lme4 (Bates et al., 2015), patchwork (Pedersen, 2020), readxl (Wickham & Bryan, 2019), scales (Wickham & Seidel, 2020), signal (signal developers, 2014), stringr (Wickham, 2019), TSA (Chan & Ripley, 2020). See online Supplemental Materials Table S6 for versions of the used software.

[5] Random correlations do not affect the estimation of the fixed effects but they quadratically increase the number of parameters. Having too large a number of parameters would make it infeasible to reliably estimate marginal likelihood using bridge sampling, which is necessary to compute Bayes Factors. This would, in turn, make the Bayes Factors estimates imprecise.

Results

Behavioral Sentence Recognition Performance

The average accuracy in the sentence recognition test was .65 (chance level = .33, *SD* = .12, range = .47–.86). We take it to indicate that all (included) participants were attentive and read the sentences for comprehension as directed.

EEG Effects

Noun

While all our manipulations were continuous, Figure 4 shows ERPs averaged into the four pseudo conditions used for the generation of our materials (LoCP and HiCP nouns following pro-LoCP and pro-HiCP adjectives). As can be seen, the ERPs averaged into the pseudo conditions were associated with different N400 amplitudes to the noun (see online Supplemental Materials Figure S2 for the ERP plot across all electrodes). However, based on this visualization it is difficult to precisely ascertain the effect of our experimental manipulations and to separate the effect of preadjectival context from the effect of adjective-driven CP updating. For that, we turn to the Bayesian mixed-effects regression model.

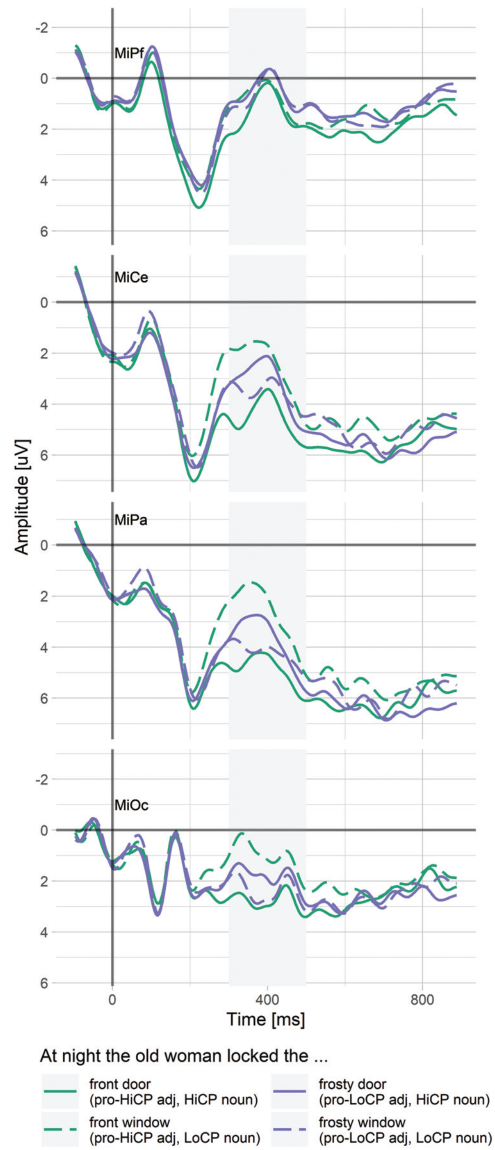
The model confirms that the amplitude of the N400 reflects the sum of the influence of the global, context-based predictability before the adjective (the positive effect of preadjectival CP) and the amount of updating introduced by the adjective (the positive effect of CP update). The estimate for the effect of preadjectival CP effect was $\hat{\beta}$ = 3.18 μV, 95% CrI = [2.25, 4.11], whereas the estimate for the effect of CP updating was $\hat{\beta}$ = 2.14 μV, 95% CrI = [1.49, 2.79] (see online Supplemental Materials Table S1 for a full model output). Positive changes in CP were associated with increased positivity in N400 amplitude to the noun. Thus, adjectives were able to further facilitate the N400 to nouns that they made more predictable. Negative changes in CP were also associated with increased negativity in N400 amplitude —that is, nouns also became less facilitated after the adjective than they would have been based on the global context alone.

However, the model presented above enforces a linear solution for the updating effect, and it is possible that once this constraint is relaxed, the updating effect may become steeper in one direction and flatter in the other. To test whether this is the case, we fitted a series of alternative models. Each of the models assumed that negative updating has a different slope relative to positive updating, including cases in which the relative slope was 0 (null model, no negative updating) and 1 (the same slope as for positive updating). To obtain these models, we created a new predictor of updating strength in which the positive values remained intact but the negative updating values were multiplied by the coefficient of the negative-to-positive relative slope. The merit of this approach is that it enforces continuity between positive and negative updating and allows us to describe the updating effect for both the negative and positive parts using one model parameter that is estimated using all items (no loss of power). Next, by means of Bayes Factors, we compared each of the models assuming some negative updating to the null model assuming no negative updating. The results can be seen in the middle-right column of Figure 5. Model comparisons show that models in which the negative updating is 0.2 to 1.00 of positive updating are better than the null model (BFs > 3), with the strongest evidence for the model assuming that negative updating was 2/3 of that of positive updating (coefficient = .66; see online Supplemental Materials Table S2 for a full model output). This model is visualized in the two leftmost columns in Figure 5. As can be seen, the preadjectival CP effect has a similar slope to the positive part of the CP updating effect.

Finally, we tested if separating CP of the noun following the adjective into preadjectival CP and CP updating improves the model's fit to the data. Using Bayes Factors, we directly compared the

Figure 4

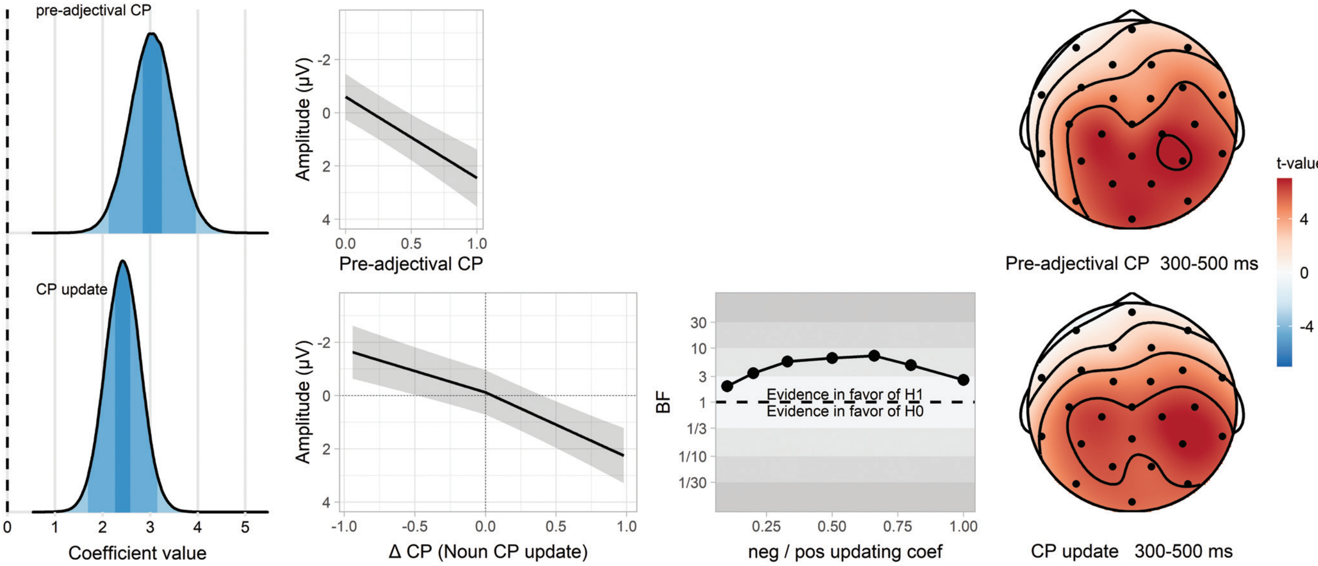
Grand Average Event-Related Potential Waveforms to the Noun Over Four Representative Midline Electrodes for Four Pseudo-Conditions Used for the Generation of the Stimuli



Note. MiPf = midline prefrontal; MiCe = midline central; MiPa = midline parietal, and MiOc = midline occipital), broken down by noun condition (HiCP, LoCP) and adjective condition (pro-HiCP, pro-LoCP). Negative is plotted up. The shaded area highlights the N400 time-window (300–500 ms). See the online article for the color version of this figure.

Figure 5

The Results of Bayesian Mixed-Effects Model of the N400 Amplitude at the Noun



Note. Left: Posterior distributions for the effects of preadjectival CP (cloze probability; top) and CP update (bottom) from the model of the N400 amplitude at the noun assuming a negative-to-positive updating coefficient of 0.66 (i.e., negative updating is one-third weaker than positive updating). The CP update coefficient reflects the slope of the positive part. Fill color corresponds to 66% and 95% credible intervals. Middle-left: Regression slopes for the effects of preadjectival CP (top) and CP update (bottom) from the same model as in the left column. Both panels in the middle-left column have identical scales. Shaded area indicates 95% credible interval. Middle-right: analysis showing Bayes Factors comparing a baseline model assuming no negative updating (H0) with a series of models assuming that negative updating is a fraction of positive updating (H1), for different values of the coefficient of negative-to-positive updating. Right: Scalp maps of the preadjectival CP effect (top) and CP update effect (bottom) in the 300–500 ms time-window at the noun estimated by linear mixed-effects models with the same structure as the main Bayesian model. See the online article for the color version of this figure.

above model in which CP updating is broken into two connected slopes, corresponding to negative and positive updating (with the negative part being 0.66 of the positive part) and compared it to a model that only included postadjectival CP as a predictor. The resulting Bayes Factor was 4.5 in favor of the more complex model, indicating its considerably better fit.

Adjective

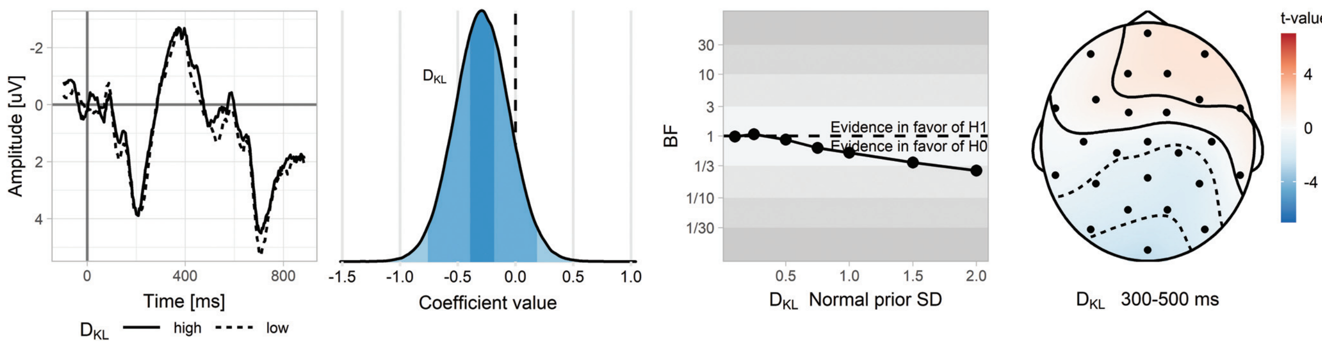
Figure 6, left panel shows that a comparison between ERPs to adjectives with high and low D_{KL} values yielded little modulation of the N400. Accordingly, the posterior distribution of D_{KL} from a Bayesian linear mixed-effects model is close to and overlaps with 0 (see Figure 6, middle-left panel). To estimate if a model assuming a D_{KL} effect should be preferred over the null model that does not assume a D_{KL} effect, we used Bayes Factors and sensitivity analysis. Sensitivity analysis addresses a potential limitation of Bayes factors, namely that, unlike posterior estimates, they critically depend on the priors and that there is little information on how these priors should be set given that we do not even know if the tested effect is different from zero (see Schad et al., 2021 for discussion; for an example of use, see Nicenboim et al., 2020). Sensitivity analysis consists in testing different priors and checking how their choice affects the Bayes factor value. We compared different models against a baseline or null model that did not contain the tested predictor. All tested (i.e., not null) models had the same likelihood, but they varied with respect to the prior on the tested predictor. We were agnostic about the direction of the effect (we always assumed a prior with a normal distribution, with the mean set at 0), and we varied the width of the prior (i.e., its standard deviation). A wide prior is more compatible with larger effects, while a narrow prior is more compatible with smaller effects. Thus, the sensitivity analysis will help us understand the range of conclusions we can draw under different assumptions concerning the size of the effect.

The sensitivity analysis shows that independent of the choice of the prior for D_{KL}, there is evidence in favor of the null model. However, assuming priors consistent with a small effect, the BFs are in the range 0.33–1, which corresponds to "anecdotal" evidence in the classification of Jeffreys (1939). The estimate of the effect had the mean .32, CrI [-0.81, 0.18]. To put these values into context, with a change in D_{KL} equal to the interquartile range of D_{KL} in our dataset (0.64), the resulting N400 modulation would be in the range −0.5 μV to −0.1 μV (see online Supplemental Materials Table S3 for a full model report). To summarize, there is evidence against D_{KL} as a substantial contributor to the N400 to the adjectives, but we cannot exclude the possibility that it exerts a very small effect. Note that even if we assume that the weak effect is not accidental, its scalp distribution does not resemble the typical N400 distribution (compare the scalp map in Figure 6 with the scalp maps in Figures 5, 7, and 8). In summary, we conclude that the N400 to the adjective was not modulated by the amount of updating to the nouns' probability distribution.

While we did not find evidence of updating in the N400 time window and scalp distribution, it is possible that an updating effect could manifest in other time windows or scalp locations. To test

Figure 6

The Results of a Bayesian Model Testing to What Extent the Index of Updating the Noun CP Distribution Affects the N400 to the Adjective



Note. Left: Grand average event-related potentials waveforms to the adjective at a midline parietal electrode (MiPa; data-based). ERPs are split at the median Kullback-Leibler divergence (D_{KL}) over the noun CP distribution (see online Supplemental Material Figure S3 for the ERP plot across all electrodes; see also online Supplemental Materials Figure S4 for the comparison between pro-HiCP and pro-LoCP pseudo-conditions). Middle-left: Posterior distribution of the effect of updating at the adjective indexed by D_{KL}. Fill color corresponds to 66% and 95% credible intervals. Middle-right: Sensitivity analysis comparing the baseline model of the N400 amplitude at the adjective (H0) and models including D_{KL} (H1), under different widths of the normally distributed prior for D_{KL} (all priors centered at zero). Right: Scalp map of the D_{KL} effect in the 300-500 ms time-window at the adjective, as estimated by linear mixed-effects models with the same structure as the main Bayesian model. See the online article for the color version of this figure.

this, we ran a cluster-based permutation procedure. It first identifies all clusters over time and across scalp space in which an effect of interest surpasses a predetermined threshold (in our case, $|t| > 2$) and then tests via a permutation analysis that of these clusters could occur with a probability significantly higher than chance (Maris & Oostenveld, 2007). This exploratory analysis allows us to test whether there are any effects of a predictor at any scalp location, while controlling for Type-II error rate. However, this analysis found no significant clusters in any scalp region or time-window (all *p*s $> .29$). In other words, adjectives induced no significant ERP differences related to the amount of updating they provoked.

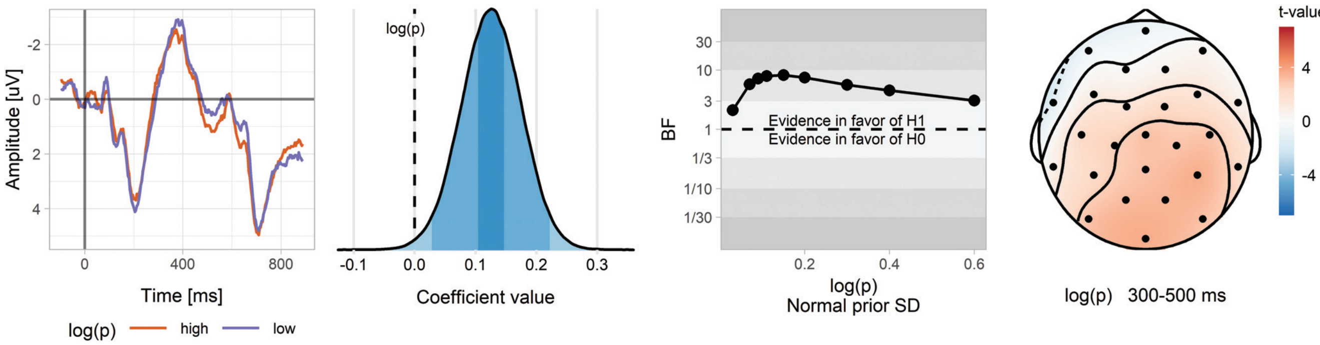
Both the analysis focused on the N400 and the cluster-based analysis failed to find any effect of updating at the adjective. To probe for other sources of variance explaining N400 amplitudes at the adjective position, we turned to exploratory analyses looking for facilitation of the adjective coming from two possible sources: the preceding context and from a highly predictable head noun. As discussed in the beginning of the article, fit to context, as measured by cloze probability, has a well-established, pervasive relationship with N400 amplitudes. However, the CP of adjectives is notoriously difficult to estimate using behavioral measures, as participants do not generally provide an adjective as a likely continuation of a sentence; see, for example, Boudewyn et al. (2015). Here, we instead estimated the predictability of the adjectives using GPT-2 (1558M parameter version), a state-of-the-art Transformer-based neural network model of language (Radford et al., 2018; see Szewczyk & Federmeier, 2021 for validation of this technique as a proxy for CPs). We estimated the log-probability of our adjectives as continuations of the sentences and then tested whether these log-probabilities predict the amplitude of the N400 (see Figure 7). As in the analysis of updating at the adjective (Kullback-Leibler divergence), we conducted a sensitivity analysis. A reanalysis of four experiments that used similar sentences (Szewczyk & Federmeier, 2021) showed that log(p) had a mean effect = 0.22 for target words that were mostly nouns. For adjectives, we expected a similar or a weaker effect. The model shows evidence for a small effect of log (p) on the N400 at the adjectives: $\hat{\beta} = 0.13$ µV, 95% CrI = [0.03, 0.22] (see online Supplemental Materials Table S4 for a full model). The scalp distribution of this effect is consistent with its characterization as an N400 effect.

The second source of facilitation on the adjective could come from the noun itself. Federmeier and Kutas (1999) showed that highly predictable nouns are preactivated and thereby prime other categorically related words, regardless of their congruity. Thus, when an incongruent (but related) word is presented instead of the highly predictable word, it elicits a smaller N400 compared with an unrelated incongruent word. Within the present study, it is possible that when the sentence context makes a noun strongly predictable, it primes associated adjectives.[6] To test this possibility, we constructed an index to estimate the level of support the adjective could receive from the upcoming high-CP noun. First, we estimated the semantic distance between each pair of adjective and HiCP noun, using the cosine of the angle between vector representations of the two words, trained on the Google News corpus using word2vec CBOW algorithm (Mikolov et al., 2013). Next, we multiplied word similarity (1 – distance) with the CP of the HiCP noun (the resulting index had $M = 0.09$, $SD = 0.08$, range = -0.02 to 0.44). Thus, the index of support from the high-CP noun had high values only for items in which there was a strongly predictable noun that, in addition, was strongly associated with the presented adjective. The Bayesian analysis shows that there is strong evidence that the support from the high-CP noun has a small facilitatory effect on the adjective (assuming effect size = 5, a 2 *SD* change in noun support would result in a modulation by 0.8 µV; see Figure 8). Again, the topography of this effect was consistent with its characterization as being on the N400. See online Supplemental Materials Table S5 for a full model.

[6] We thank an anonymous reviewer for the encouragement to look at the semantic distance between the adjective and the noun.

Figure 7

Results of a Bayesian Model Testing the Effect of Adjective's Contextual Predictability on the N400



Note. Left: Event related-potentials (ERPs) to the adjective split at the median log-probability of the adjective estimated using the GPT2-xl neural network language model (data-based). Middle-left: Posterior distribution of the effect of log-probability of the adjective. Fill color corresponds to 66% and 95% credible intervals. Middle-right: Sensitivity analysis showing Bayes factors in favor of the model including the log(p) predictor for different values of standard deviation of the normally distributed prior centered at zero. Right: Scalp map of the log(p) effect in the 300–500 ms time-window at the adjective estimated by linear mixed-effects models with the same structure as the main Bayesian model. See the online article for the color version of this figure.

Finally, we examined the correlation between the two predictors to see if they could explain the same variance in N400 amplitude to the adjectives. The analysis showed that they share only .008 of variance (r = .087), and thus their contribution to explaining N400 amplitude must be independent.

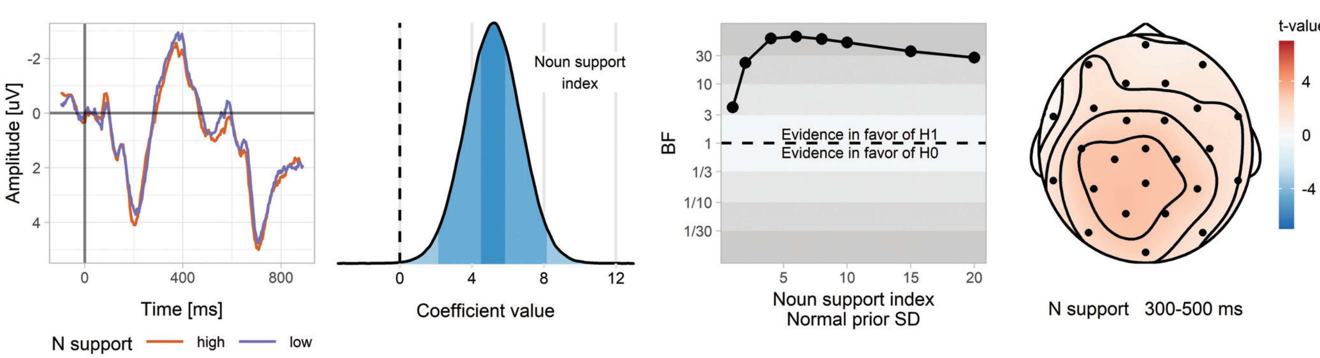
To sum up, there is weak evidence for facilitation of the adjective coming from the preceding semantic context and for facilitation of the adjective coming from (predictions for) associated strongly predictable nouns. At the same time, there is evidence against the hypothesis that N400 amplitudes at the adjective varied as a function of the amount of updating they induced in the probability distribution of the subsequent nouns. In other words, the N400 to the adjective varied depending on factors related to the ease of access for the adjective itself, but not with factors related to consequences the adjective might have for the probability of encountering upcoming nouns (as posited by surprisal and event prediction error theories).

Discussion

In this study we asked if comprehenders can rapidly make use of information (specifically, an adjective) that changes the probability of likely upcoming nouns. In particular, we sought to determine if the mechanisms involved in updating cannot only bring new information online but can also decrease support for words that had already been made predictable before the adjective. To that end, we presented sentences that varied in their constraint for the sentence-final target noun and preceded the noun with an adjective that either increased or decreased the predictability of that noun, as assessed by off-line norming. We examined whether the changes in predictability measured offline are accompanied by corresponding changes in the amplitude of the N400, an index of a word's current state of activation in long-term semantic memory (Kutas & Federmeier, 2011). We adopted a single-item-level

Figure 8

Results of a Bayesian Model Testing if the N400 to the Adjective Depends on the Amount of Support From the Most Predictable Head Noun



Note. Left: Event related potentials (ERPs) to the adjective split at the median index of support from the head noun (data-based). Middle-left: Posterior distribution of the effect. Fill color corresponds to 66% and 95% credible intervals. Middle-right: Sensitivity analysis showing Bayes factors in favor of the model including the effect of support from a predictable head noun for different values of standard deviation of the normally distributed prior centered at zero. Right: Scalp map of the noun support effect in the 300–500 ms time-window at the adjective estimated by linear mixed-effects models with the same structure as the main Bayesian model. See the online article for the color version of this figure.

regression approach, which enabled us to separate the change in the noun's predictability brought about by the adjective from the predictability of the noun driven by the context that preceded the adjective. As a result, we were able to determine whether updating can both increase and decrease support for particular nouns and to assess the relative weight of the local and global contextual information. We also probed responses at the adjective to look for neural activity associated with the updating processing itself and thereby gain a better understanding about the nature of the processing mechanisms involved.

Updating Works in Both Positive and Negative Directions but is Not Symmetrical

Our main finding concerned the effect the adjective had on the processing of the noun. We found that the adjectives could both facilitate and hamper access to the noun. This result shows that the adjective's impact on activation states in semantic memory must involve mechanisms that go beyond, for example, simple spreading activation. An updating mechanism relying on mere spreading activation would remain yoked to word probability only in sentences that contain no surprising or even ambiguous information. In that case, accruing semantic information would always build upon the previously activated features but never contradict them. Instead, the results show that information from the adjective cannot only facilitate noun processing but also make it more difficult to access nouns that become less likely in the context. We found that although the influence of negative updating may be weaker than that of positive updating, the effect is nevertheless bidirectional. The comprehension system's ability to immediately use information from a word to guide access to upcoming words is presumably part of the reason why, with a few exceptions (i.e., unlicensed negations, quantifiers, or counterfactuals: Ferguson et al., 2008; Fischler et al., 1983; Nieuwland & Kuperberg, 2008; Urbach & Kutas, 2010), cloze probabilities correlate so well with the amplitude of the N400.

Since the updating of the noun is unlikely to be explained using simple spreading activation, what is the mechanism responsible? To try to address this question, we turn to the outcome of the measures obtained at the cue inducing the update—the adjective. This will also allow us to make a fuller interpretation of the asymmetry between the effects of positive and negative updating at the noun.

Relating the (Null) Findings at the Adjective to Prior Literature

The analysis at the adjective showed no discernible modulation of ERPs associated with the amount of updating that a given adjective was likely to induce, with mean differences in amplitudes between adjectives associated with high versus low levels of updating being smaller than 0.5 μV at all electrodes and timepoints. This stands in contrast to the impact of updating, measured at the noun, which altered the N400 amplitude on the scale of a few microvolts. In looking for an explanation of this null finding at the adjective, we first want to stress that the effect at the noun *must* have resulted from the semantic information apprehended at the adjective, as the adjective was the only point of divergence across otherwise identical sentence contexts that ultimately yielded very different brain responses to the same nouns. In other words, we must exclude the possibility that the adjective did not alter the processing of the sentence because eventually (within less than a second after being apprehended) it did.

Before we speculate about the possible reasons this pattern of results, we note that it strikingly resembles the findings in an earlier experiment that also tested adjective-driven updating (Szewczyk & Wodniecka, 2020). As here, that study used sentences in which the critical words were an adjective and a head noun. The adjective's grammatical form (in Polish) matched or mismatched (via gender and number agreement, while the meaning of the adjective remained the same) the set or a subset of nouns predictable at that point in the sentence. The results showed that adjective-driven updating led to two different patterns of results at the adjective, depending on whether the adjective mismatched ALL predictable nouns or mismatched only some of them (matching others). In both cases, the updating led to a change in the amplitude of the N400 at the following noun. The differences in patterns of results concerned the adjective. The baseline condition for both types of updating were adjectives whose form was uninformative about the upcoming noun; that is, adjectives whose grammatical form matched all potentially predictable nouns (and did not rule out any of them). Adjectives mismatching all nouns led to a larger N400 than uninformative adjectives. This N400 effect stood in contrast to the pattern observed for adjectives that matched only a subset of the predictable nouns. In this latter case, although the adjective was informative about the noun (changed the probability of likely continuations), there was no discernable ERP effect associated with this update in the noun probability distribution. This was true despite the fact that the manipulation did later modulate the N400 to the noun.

The pattern of results in the present study resembles the outcome seen for adjectives that matched some predictable nouns in the study by Szewczyk and Wodniecka: Neither study found an effect at the adjective despite the fact that, in both, the adjective induced a modulation of the N400 to the noun. The similarity between the two studies also extends to the nature of the manipulation. Our adjectives always matched at least one noun from the set of nouns that had been made predictable by the context before the adjective. As noted, updating in the Szewczyk and Wodniecka (2020) study was triggered by the (mis)match between the grammatical form of the adjective and some of the predictable nouns, whereas in the current study, the updating was triggered by semantic information from the adjective. The similarity in results across the studies shows that the dimension along which the adjective and nouns are being aligned—semantic or grammatical—is not the critical factor for determining whether the updating adjective leads to an N400 effect or no effect at all. Instead, what seems to matter is whether the adjective matches some of the predictable nouns. In line with this conclusion, other studies in which the adjective (or other prenominal word) did not match any predictable noun found N400 effects for the words introducing the update (Boudewyn et al., 2015; Freunberger & Roehm, 2017; Maess et al., 2016; Wicha, Bates, et al., 2003; Wicha, Moreno, et al., 2003).

Adjectives Get Semantically Accessed but Do Not Update the Situation Model

What, then, are possible mechanisms by which, even though some adjectives do not elicit any ERP effect, they still alter the

N400 at the following noun? Of course, one possibility is simply that the kind of processing that underlies noun updating at the adjective creates neural activity that happens not to be detectable by EEG (e.g., because it occurs in a cortical or subcortical area that is not well reflected on the scalp; cf. Koessler et al., 2009). However, a more interesting possibility is that the lack of effect highlights how the construction of word meaning may be influenced by the larger information structure of a sentence.

Note that, as we pointed out in the preceding subsection, the core difference between adjectives that in the prior literature did and did not elicit an N400 effect is whether they matched or did not match any predictable nouns. Adjectives that match at least some of the predictable head nouns usually have features that cohere with the event being described (e.g., "sweet" in the context of a birthday party and a cake), whereas adjectives mismatching all predictable nouns are likely to activate some features that are not already in the discourse (e.g., "healthy" in the birthday party context; Boudewyn et al., 2015). In the present study we did not find any effects of Kullback-Leibler divergence at the adjective probably because the amount of new information provided by the adjectives was not different across different degrees of noun updating. On the other hand, we did find that the N400 to the adjectives varied depending on their own activation level. First, we found N400 effects at the adjectives yoked to their cloze probability in the sentence context, as estimated by the GPT-2 model. This index captures the amount of context-based facilitation for the adjective itself (see Szewczyk & Federmeier, 2021). Second, we also found that N400 amplitudes to the adjectives varied depending on the amount of support they received from the most predictable noun. This latter effect is analogous to findings from the related anomaly paradigm, which has been used to demonstrate that comprehenders can preactivate highly predictable words, which, in turn, prime other related words (Federmeier & Kutas, 1999; Ito et al., 2016; Kutas, 1993) or shapes (Rommers et al., 2013).[7] In the current study, the same priming was presumably responsible for activating the features of adjectives that were associated with a strongly predictable (and predicted) noun. Because this effect occurred only when the noun was highly predictable and when the association between the adjective and the noun was strong, we think that this effect reflects priming of the adjective by the noun and not some kind of updating of the noun driven by the adjective. In summary, we found that the N400 to the adjectives varied with respect to the degree to which they would be expected to bring online relatively more new semantic information. In other words, these two effects show that the adjectives were semantically accessed immediately after being encountered.

The fact that the adjective elicited semantic information, however, does not necessarily mean that this information is immediately integrated with the preadjectival context and used to update the situation model (that also involves updating the activation of nouns that potentially might follow). Indeed, it seems plausible that, to be useful for building or updating a message, adjectival information must first be linked with its head noun, when it is eventually encountered; for example, to make sense of the adjective "unbalanced," one would need to know if it is referring to a person, news coverage, the tower of Pisa, or next year's budget. We hypothesize that when an adjective is encountered before its head noun, the semantic representations it evokes are first accessed but then temporarily buffered, likely using semantic short-term memory (STM). The idea that adjective information is stored in STM until the head noun becomes available is supported by neuropsychological research. Case studies of patients with STM deficits find that individuals with deficits in semantic (but not phonological) STM have problems with comprehending (and producing) noun phrases in which a noun is preceded by two or more adjectives, which suggests that the interpretation of the adjectives in those sentences relies on STM (Hanten & Martin, 2000; Martin & He, 2004; Martin & Romani, 1994). Strikingly, comprehension is unaffected in these patients when the adjectives are presented postnominally ("big brown dog" vs. "the dog that is brown and big"), indicating that the use of adjective information depends on STM only when the adjectives occur before the noun.

On the surface, this conclusion may seem to conflict with the findings of visual world paradigm studies, which have attested that information provided by adjectives can immediately guide eye-movements to relevant objects displayed on the screen. For example, in an experiment by Sedivy et al. (1999), subjects made faster looks to a tall glass, given the instruction "touch the tall glass," as long as the displayed scene also included a short glass. However, the crucial thing to note is that in the visual world paradigm the candidate visual referents of the not-yet-presented head noun are *already available* when the participants hear the adjective. This enables relating (or coordinating, see Knoeferle & Crocker, 2006) the adjective with those nouns/objects to guide the next saccade (note that the listeners/viewers must know the crude identity and location of each object to be able to shift their gaze to an object based on its relevance). Thus, results from the visual world paradigm are similar to what we predict might happen for postnominal adjectives. However, our data, taken together with the patient data from Martin and colleagues, suggest that there may be important differences in how prenominal adjectives are processed and used when these are encountered in the absence of the kind of preestablished set of referents available in the visual world paradigm (for additional discussion of why the visual world paradigm may not always generalize to sentence comprehension under other task conditions, see Huettig et al., 2011; Knoeferle & Crocker, 2006).

Finally, we will note that the findings at the adjective are incompatible with models linking N400 amplitude directly to incremental updates of the situation model (surprisal theory, Levy, 2008;

[7] A brief terminological note is in order here. Throughout the manuscript, we were careful to refer to general predictability effects that result in N400 amplitude reductions simply as "facilitation," without specifically trying to locate them under the umbrella of "prediction" as the term might sometimes be used (e.g., Kuperberg & Jaeger, 2016). We prefer to reserve the term prediction for effects that can be demonstrated to go beyond basic context-based facilitation (e.g., Federmeier & Kutas, 1999). This is because, in contrast to basic effects of context-based facilitation, which are nearly universally attested, more specific effects of prediction are nonobligatory, as they are not observed in all comprehenders (Federmeier et al., 2002; Huettig & Pickering, 2019; Ng et al., 2017) or all processing circumstances (Wlotko & Federmeier, 2015), and they seem to be more dependent on left-hemispheric processing (Federmeier, 2007), and appear have a different neural underpinning from more general, context-based facilitation (that is attested bilaterally; see review in Federmeier, in press. Thus, brain and behavioral correlates of word predictability may be a sum of different mechanisms (see also Szewczyk & Federmeier, 2021, for an additional dissociation within context-based facilitation). The present results at the adjective provide a clear example in which context-based facilitation and prediction combine in modulating the N400 to the same word.

event prediction error theory, Rabovsky et al., 2018). According to these theories, if the situation model has been updated by the adjective such that it led to N400 effects at the noun, then such updating should be reflected by an N400 effect at the adjective. The finding in which we see N400 effects at the noun but not at the adjective is clearly not in line with these theories, at least as long as they assume that the situation model is incrementally updated at each word and the N400 amplitude always reflects the amount of updating. In contrast, more compatible with these results are theories assuming that the N400 reflects incremental processing at the level of semantic features and not the situation model (Brouwer et al., 2012; Kutas & Federmeier, 2000; Lau et al., 2008). In particular, these latter theories are compatible with the finding that the N400 at the adjective was modulated by the adjective's contextual predictability, as well as by its association with a strongly predicted noun (both increasing the activation of adjective's semantic features), but not by the degree to which the adjective changed the upcoming noun's probability distribution (updating of the situation model). To accommodate the findings of the present study, the event prediction error/surprisal theories would have to incorporate the assumption that the situation model is not updated at each word and, at least at the adjective, that N400 modulations do not reflect updating of the situation model. The N400 seems to tell us when there is new semantic information but not everything about how that information is used.

How Do Adjectives Modulate Access to the Noun?

Given that, according to our proposal, all adjectives get buffered until the head noun is encountered, how do they end up altering the N400 to the noun, once it is encountered? We speculate that the N400 gets modulated because the adjectives shape the very process of semantic access to the noun. For example, Huang et al. (2010, 2012) observed what they called "compositional concreteness effects"—effects of the concreteness of an adjective that are observed at the subsequent noun. The same noun (e.g., "book") elicited ERP responses associated either with concrete or with abstract words, depending on whether or not the preceding adjective, when combined with the noun, pointed to a physical reading of the referent (e.g., "green book") or a more abstract conceptualization of it (e.g., "engaging book"). Thus, adjective information can change how semantics is derived from a given noun. In some ways, this view may be seen as related to the compound cue theory, which Ratcliff and McKoon (1988) proposed as an alternative to spreading activation accounts of lexical priming. They posited that the prime and the target word constitute a compound cue that is used to retrieve concepts from long-term memory. In the present study, the N400 pattern observed at the noun similarly suggests that information about the adjective and the noun, once both have been obtained, are used together to access meaning.

Moreover, our pattern of results reveals that meaning access is shaped, in parallel, by the adjective information and the representation of the sentence. To illustrate this point, consider an example item from our study: "His skin was red from spending the day at the . . ." With no adjective, "beach" is the HiCP completion (and "pool" is a plausible LoCP alternative). When "beach" is preceded with the pro-HiCP adjective "sandy", the adjective facilitates the noun, presumably because the features of "sandy" overlap with those that would generally be elicited by "beach", reducing the amount of new information that is being activated when "beach" is eventually encountered. In contrast, when "beach" is preceded with the pro-LoCP adjective "neighborhood," the meaning of the noun gets accessed and constructed in a context that does not only consist of the preceding sentence (describing sunburns on the skin) but also of that modifier, "neighborhood". Certain features of a "beach" are important for understanding it as a place where a sunburn might occur. However, those are not the same semantic features as the ones necessary for understanding what would make something a "neighborhood beach." We argue that the word "beach" must be simultaneously understood in conjunction with the representation of the sentence and of the buffered adjective information. Thus, when the semantic features drawn out by a given combination of adjective with its noun are not the same as those drawn out by the sentence context, there is a net increase in the amount of new semantic information that is accessed in response to that noun—and hence, a larger N400 response to "beach" following "neighborhood" than without the adjective. It is worth noting that, whereas cloze probability seems to be a robust predictor of the degree to which the sentence context and adjective jointly facilitate access to features of a given noun, CP is not as strongly correlated with the increase in semantic feature activation entailed when, as just discussed, the meaning of the noun that was presented must be accessed in the context of heterogeneous information induced by the sentence and the adjective—hence, the shallower slope we observed for negative updating.

Our explanation entails that it is not the same information that is looked up every time a noun is encountered (e.g., "beach" after different modifiers and/or in different sentence contexts). Rather, which meaning features are activated and to what degree is contextually dependent, and context itself is multifaceted (here, consisting of both the sentence frame and the adjective information). Thus, this view entails that activating meaning is a constructive process (see Elman, 2009, for an overview). Moreover, on this view, the N400 effects entailed by understanding a "neighborhood beach" obtain at the noun because this is the first point when the adjective information can be used. We predict that if the two variants of the adjective were instead presented postnominally, N400 modulations would be instead seen on the adjective itself, because all the information necessary to integrate the adjective with the context—crucially, including the head noun—would have already been processed (in analogy to the aforementioned case studies by Martin and colleagues).

To conclude, even though sentence comprehension is a strongly incremental process—that is, the parser generally does not wait to integrate each upcoming word, as has been shown by rich behavioral, eye-tracking, and EEG literatures (Boland et al., 1995; Eberhard et al., 1995; Marslen-Wilson & Tyler, 1975; Tanenhaus et al., 1995; Van Petten et al., 1999)—there are times when information must get buffered instead of being immediately integrated with the sentence. In the case of prenominal adjectives (and, perhaps, all such modifiers) this happens because the adjective is linked to the rest of the sentence *through* the head noun. Other examples of incomplete incrementality can be found in filler-gap constructions, in which case the filler does not seem to be integrated with the sentence until the gap is encountered (King & Kutas, 1995), as well as in constructions such as "before X" (Münte et al., 1998), and object-relative clauses (Weckerly & Kutas, 1999). More generally, language comprehension tries to be maximally incremental because it helps comprehenders manage working memory load. But working memory loads are not the only pressure shaping how language can unfold: A hierarchical thought cannot always be linearized into a sentence in a fully incremental way, and speakers sometimes sacrifice incrementality to use the full range of devices afforded by the syntax, such as clefts, preposed or extraposed elements, and commonplace passives, in order emphasize something or to maintain a higher-order information structure (e.g., to start the sentence with a concept introduced in the preceding sentence). Thus, online comprehension necessarily must include strategies that accommodate cases where fully incremental interpretation is not possible.

Summary

In this study we have shown that adjectives can decrease, as well as increase, the ease of semantic access for upcoming nouns. However, taking into account the pattern of responses to the adjectives themselves, our results further indicate that it is unlikely that this updating happens through a direct integration of the adjective information into the sentence context representation, before the noun's apprehension. Instead, the results point to the idea that, even though semantic information linked to the adjective gets accessed when that adjective is encountered, the semantics of the adjective is buffered until the head noun is encountered, at which point access of the noun's semantics is shaped in parallel by both the adjective and the sentence-level representation. Meaning construction, while highly incremental, is nevertheless guided and constrained by the discourse and the syntactic structure.

References

Alday, P. M. (2019). How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. *Psychophysiology*, *56*(12), e13451. https://doi.org/10.1111/psyp.13451

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

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. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, *22*(2), 245–268. https://doi.org/10.1016/0021-9991(76)90078-4

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*, *34*(6), 774–806. https://doi.org/10.1006/jmla.1995.1034

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, *59*(1), 55–73. https://doi.org/10.1016/j.brainresrev.2008.05.003

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436. https://doi.org/10.1016/j.jecp.2012.01.005

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective & Behavioral Neuroscience*, *15*(3), 607–624. https://doi.org/10.3758/s13415-015-0340-0

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143. https://doi.org/10.1016/j.brainres.2012.01.055

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Chan, K.-S., & Ripley, B. (2020). *TSA: Time series analysis (R package version 1.3)* [Computer software]. https://CRAN.R-project.org/package=TSA

Chow, W.-Y., & Chen, D. (2020). Predicting (in)correctly: Listeners rapidly use unexpected information to revise their predictions. *Language, Cognition and Neuroscience*, *35*(9), 1149–1161. https://doi.org/10.1080/23273798.2020.1733627

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Craddock, M. (2019). *eegUtils: A collection of utilities for EEG analysis*. https://github.com/craddm/eegUtils

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, *1084*(1), 89–103. https://doi.org/10.1016/j.brainres.2006.02.010

Delogu, F., Brouwer, H., & Crocker, M. W. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain and Cognition*, *135*, 103569. https://doi.org/10.1016/j.bandc.2019.05.007

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. https://doi.org/10.1038/nn1504

DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, *121*(3), 226–239. https://doi.org/10.1016/j.bandl.2012.02.006

Dowle, M., & Srinivasan, A. (2021). *data.table: Extension of ©data.framè (R package version 1.14.0)* [Computer software]. https://CRAN.R-project.org/package=data.table

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. N. (Eds.). (2007). *Bayesian brain: Probabilistic approaches to neural coding*. MIT Press.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*(6), 409–436. https://doi.org/10.1007/BF02143160

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*(4), 547–582. https://doi.org/10.1111/j.1551-6709.2009.01023.x

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

Federmeier, K. D. (in press). Connecting and considering: Electrophysiology provides insights into comprehension. *Psychophysiology*.

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. https://doi.org/10.1006/jmla.1999.2660

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*(3), 149–161. https://doi.org/10.1016/j.bandl.2010.07.006

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. https://doi.org/10.1016/j.brainres.2006.06.101

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133–146. https://doi.org/10.1111/1469-8986.3920133

Ferguson, H. J., Sanford, A. J., & Leuthold, H. (2008). Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds. *Brain Research*, *1236*, 113–125. https://doi.org/10.1016/j.brainres.2008.07.099

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W., Jr. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*(4), 400–409. https://doi.org/10.1111/j.1469-8986.1983.tb00920.x

Fleur, D. S., Flecken, M., Rommers, J., & Nieuwland, M. S. (2020). Definitely saw it coming? The dual nature of the pre-nominal prediction effect. *Cognition*, *204*, 104335. https://doi.org/10.1016/j.cognition.2020.104335

Freunberger, D., & Roehm, D. (2017). The costs of being certain: Brain potential evidence for linguistic preactivation in sentence processing. *Psychophysiology*, *54*(6), 824–832. https://doi.org/10.1111/psyp.12848

Fruchter, J., Linzen, T., Westerlund, M., & Marantz, A. (2015). Lexical preactivation in basic linguistic phrases. *Journal of Cognitive Neuroscience*, *27*(10), 1912–1935. https://doi.org/10.1162/jocn_a_00822

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2018). *bridgesampling: An R package for estimating normalizing constants*. ArXiv. http://arxiv.org/abs/1710.08162

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., & Dale, A. M. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage*, *17*(3), 1101–1116. https://doi.org/10.1006/nimg.2002.1268

Hanten, G., & Martin, R. C. (2000). Contributions of phonological and semantic short-term memory to sentence processing: Evidence from two cases of closed head injury in children. *Journal of Memory and Language*, *43*(2), 335–361. https://doi.org/10.1006/jmla.2000.2731

Huang, H.-W., Lee, C.-L., & Federmeier, K. D. (2010). Imagine that! ERPs provide evidence for distinct hemispheric contributions to the processing of concrete and abstract concepts. *NeuroImage*, *49*(1), 1116–1123. https://doi.org/10.1016/j.neuroimage.2009.07.031

Huang, H.-W., Meyer, A. M., & Federmeier, K. D. (2012). A "concrete view" of aging: Event related potentials reveal age-related changes in basic integrative processes in language. *Neuropsychologia*, *50*(1), 26–35. https://doi.org/10.1016/j.neuropsychologia.2011.10.018

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

Huettig, F., & Pickering, M. J. (2019). Literacy advantages beyond reading: Prediction of spoken language. *Trends in Cognitive Sciences*, *23*(6), 464–475. https://doi.org/10.1016/j.tics.2019.03.008

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

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

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*(10), 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007

Jeffreys, H. (1939). *The theory of probability*. Oxford University Press.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55. https://doi.org/10.1023/A:1021933015362

King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*(3), 376–395. https://doi.org/10.1162/jocn.1995.7.3.376

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*(3), 481–529. https://doi.org/10.1207/s15516709cog0000_65

Koessler, L., Maillard, L., Benhadid, A., Vignal, J. P., Felblinger, J., Vespignani, H., & Braun, M. (2009). Automated cortical projection of EEG sensors: Anatomical correlation via the international 10-10 system. *NeuroImage*, *46*(1), 64–72. https://doi.org/10.1016/j.neuroimage.2009.02.006

Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and Language*, *100*(3), 257–261. https://doi.org/10.1016/j.bandl.2006.07.006

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

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? Language., *Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/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. https://doi.org/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*(4), 533–572. https://doi.org/10.1080/01690969308407587

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470. https://doi.org/10.1016/S1364-6613(00)01560-6

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*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

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

Lau, E. F., Namyst, A., Fogel, A., & Delgado, T. (2016). A direct comparison of N400 effects of predictability and incongruity in adjective-noun combination. *Collabra. Psychology*, *2*(1), 13. https://doi.org/10.1525/collabra.40

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Maess, B., Mamashli, F., Obleser, J., Helle, L., & Friederici, A. D. (2016). Prediction signatures in the brain: semantic pre-activation during language comprehension. *Frontiers in Human Neuroscience*, *10*, 591. https://doi.org/10.3389/fnhum.2016.00591

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Marslen-Wilson, W., & Tyler, L. K. (1975). Processing structure of sentence perception. *Nature*, *257*(5529), 784–786. https://doi.org/10.1038/257784a0

Martin, R. C., & He, T. (2004). Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language*, *89*(1), 76–82. https://doi.org/10.1016/S0093-934X(03)00300-6

Martin, R. C., & Romani, C. (1994). Verbal working memory and sentence comprehension: A multiple-components view. *Neuropsychology*, *8*(4), 506–523. https://doi.org/10.1037/0894-4105.8.4.506

Meng, X.-L., & Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, *6*(4), 831–860.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. ArXiv. http://arxiv.org/abs/1301.3781

Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, *395*(6697), 71–73. https://doi.org/10.1038/25731

Ness, T., & Meltzer-Asscher, A. (2018). Predictive pre-updating and working memory capacity: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *30*(12), 1916–1938. https://doi.org/10.1162/jocn_a_01322

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. https://doi.org/10.1080/10888438.2017.1310213

Nicenboim, B., Vasishth, S., & Rösler, F. (2020). Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a Bayesian random-effects meta-analysis using publicly available data. *Neuropsychologia*, *142*, 107427. https://doi.org/10.1016/j.neuropsychologia.2020.107427

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, *19*(12), 1213–1218. https://doi.org/10.1111/j.1467-9280.2008.02226.x

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Wolfsthurn, S. V. G. Z., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D. J., Rousselet, G. A., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E. M., ... Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife*, *7*, e33468. https://doi.org/10.7554/eLife.33468

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Palmer, J. A., Kreutz-Delgado, K., & Makeig, S. (2011). *AMICA: An adaptive mixture of independent component analyzers with shared components* [Technical Report]. Swartz Center for Computational Neuroscience, University of California San Diego.

Payne, B. R., Lee, C.-L., & Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. *Psychophysiology*, *52*(11), 1456–1469. https://doi.org/10.1111/psyp.12515

Pedersen, T. L. (2020). *patchwork: The composer of plots*. https://CRAN.R-project.org/package=patchwork

Peelle, J. E., Miller, R. L., Rogers, C. S., Spehar, B., Sommers, M. S., & Van Engen, K. J. (2020). Completion norms for 3085 English sentence contexts. *Behavior Research Methods*, *52*, 1795–1799. https://doi.org/10.3758/s13428-020-01351-1

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, *12*(5), 410–430. https://doi.org/10.1002/bs.3830120511

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, *2*(9), 693–705. https://doi.org/10.1038/s41562-018-0406-4

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training*. Preprint. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, *95*(3), 385–408. https://doi.org/10.1037/0033-295X.95.3.385

Rommers, J., & Federmeier, K. D. (2018). Lingering expectations: A pseudo-repetition effect for words previously expected but not presented. *NeuroImage*, *183*, 263–272. https://doi.org/10.1016/j.neuroimage.2018.08.023

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, *51*(3), 437–447. https://doi.org/10.1016/j.neuropsychologia.2012.12.002

Rouder, J. N., Haaf, J. M., & Vandekerckhove, J. (2018). Bayesian inference for psychology, part IV: Parameter estimation and Bayes factors. *Psychonomic Bulletin & Review*, *25*(1), 102–113. https://doi.org/10.3758/s13423-017-1420-7

Schad, D. J., Nicenboim, B., Bürkner, P.-C., Betancourt, M., & Vasishth, S. (2021). *Workflow techniques for the robust use of Bayes factors*. ArXiv. http://arxiv.org/abs/2103.08744

Sedivy, J. C. K., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147. https://doi.org/10.1016/S0010-0277(99)00025-6

Signal Developers. (2014). *signal: Signal processing*. http://r-forge.r-project.org/projects/signal/

Szewczyk, J. M., & Federmeier, K. D. (2021). *Context-Based Facilitation of Semantic Access Follows Both Logarithmic and Linear Functions of Stimulus Probability*. PsyArXiv. https://doi.org/10.31234/osf.io/rbych

Szewczyk, J. M., & Wodniecka, Z. (2020). The mechanisms of prediction updating that impact the processing of upcoming word: An event-related potential study on sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(9), 1714–1734. https://doi.org/10.1037/xlm0000835

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, *30*(4), 415–433. https://doi.org/10.1177/107769905303000401

Urbach, T. P., & Kutas, M. (2010). Quantifiers more or less quantify online: ERP evidence for partial incremental interpretation. *Journal of Memory and Language*, *63*(2), 158–179. https://doi.org/10.1016/j.jml.2010.03.008

Van De Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. T. W. M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, *3*(5), 1211–1224. https://doi.org/10.1111/j.1749-818X.2009.00163.x

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, *18*(4), 380–393. https://doi.org/10.3758/BF03197127

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 394–417. https://doi.org/10.1037/0278-7393.25.2.394

Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience*, *3*(2), 131–150. https://doi.org/10.1162/jocn.1991.3.2.131

Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology*, *36*(5), 559–570. https://doi.org/10.1111/1469-8986.3650559

Weston, S., & Microsoft Corporation. (2020). doParallel: Foreach parallel adaptor for the “parallel” package (R package version 1.0.16) [Computer software.] https://CRAN.R-project.org/package=doParallel

Wicha, N. Y. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, *346*(3), 165–168. https://doi.org/10.1016/S0304-3940(03)00599-8

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex*, *39*(3), 483–508. https://doi.org/10.1016/S0010-9452(08)70260-0

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Wickham, H. (2019). stringr: Simple, consistent wrappers for common string operations (R package version 1.4.0) [Computer software]. https://CRAN.R-project.org/package=stringr

Wickham, H., & Bryan, J. (2019). readxl: Read Excel files (R package version 1.3.1) [Computer software]. https://CRAN.R-project.org/package=readxl

Wickham, H., & Seidel, D. (2020). *scales: Scale functions for visualization*. https://CRAN.R-project.org/package=scales

Wilke, C. O. (2021). *ggridges: Ridgeline plots in “ggplot2.”* 1 https://CRAN.R-project.org/package=ggridges

Wlotko, E. W., & Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology*, *49*(6), 770–785. https://doi.org/10.1111/j.1469-8986.2012.01366.x

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

Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: Age-related differences in the use of sentential context. *Psychology and Aging*, *27*(4), 975–988. https://doi.org/10.1037/a0029206

Yan, S., Kuperberg, G. R., Jaeger, T. F. (2017). *Prediction (or not) during language processing. commentary on Nieuwland et al. (2017) and DeLong et al. (2005).* BioRxiv. https://doi.org/10.1101/143750

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart’s N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–979. https://doi.org/10.3758/PBR.15.5.971

Received March 22, 2021
Revision received August 19, 2021
Accepted August 26, 2021 ■