Title
An ERP index of real-time error correction within a noisy-channel framework of human communication

Authors
Ryskin, Rachel
Stearns, Laura
Bergen, Leon
et al.

Publication Date
2021-07-01

DOI
10.1016/j.neuropsychologia.2021.107855

Peer reviewed

An ERP index of real-time error correction within a noisy-channel framework of human communication

Short title: ERP index noisy-channel error correction

Rachel Ryskin*[1], Laura Stearns*[2], Leon Bergen[3], Marianna Eddy[4], Evelina Fedorenko[4,5] and Edward Gibson[4]

* The first two authors contributed equally.

[1]University of California, Merced; [2]Wellesley College; [3]University of California, San Diego; [4]Massachusetts Institute of Technology; [5]McGovern Institute for Brain Research

Corresponding author:
Rachel Ryskin: rryskin@ucmerced.edu

Abstract

Recent evidence suggests that language processing is well-adapted to noise in the input (e.g., spelling or speech errors, misreading or mishearing) and that comprehenders readily correct the input via rational inference over possible intended sentences given probable noise corruptions. In the current study, we probed the processing of noisy linguistic input, asking whether well-studied ERP components may serve as useful indices of this inferential process. In particular, we examined sentences where semantic violations could be attributed to noise—for example, in "The storyteller could turn any incident into an amusing <u>antidote</u>", where the implausible word "antidote" is orthographically and phonologically close to the intended "anecdote". We found that the processing of such sentences—where the probability that the message was corrupted by noise exceeds the probability that it was produced intentionally and perceived accurately—was associated with a reduced (less negative) N400 effect and an increased P600 effect, compared to semantic violations which are unlikely to be attributed to noise ("The storyteller could turn any incident into an amusing <u>hearse</u>"). Further, the magnitudes of these ERP effects were correlated with the probability that the comprehender retrieved a plausible alternative. This work thus adds to the growing body of literature that suggests that many aspects of language processing are optimized for dealing with noise in the input, and opens the door to electrophysiologic investigations of the computations that support the processing of imperfect input.

Communication between biological agents and/or artificial systems involves the transmission of a signal over a channel, in the course of which the signal may get corrupted by noise (Shannon, 1948). Because the input to the human comprehension system often contains errors (e.g., Bond, 1999; Bortfeld et al., 2001; Ferreira & Patson, 2007; Fromkin, 1971), examination of language processing through a noisy-channel lens can yield critical insights about human language comprehension mechanisms (Levy, 2008).

Noise in the linguistic input can result from a) production errors (speech errors, typographical errors, etc.), and b) perception errors (due to sub-optimal listening/viewing conditions, noise in the environment, etc.). The fact that communication typically proceeds smoothly suggests that comprehension mechanisms are well-adapted to this noise. A rational comprehender's guess of what was intended in a noise-corrupted linguistic exchange can be expressed as the probability of the speaker's intended sentence, s_i, given the perceptual input, s_p: $P(s_i \mid s_p)$. By Bayes' rule, this value is proportional to the product of the prior (what is likely to be communicated), $P(s_i)$, and the likelihood that a noise process would generate s_p from s_i, $P(s_p \mid s_i)$ (e.g., Gibson et al., 2013).

Prior behavioral studies using offline comprehension questions to probe interpretation suggest that readers often take the meaning of a sentence to differ from that of the literal string when the literal compositional meaning of the string has low prior probability, $P(s_i)$ (Christianson et al., 2001; Ferreira, 2003), and/or the potential noise corruption that might have generated that string has high probability, $P(s_p \mid s_i)$ (Gibson et al., 2013; Poppels & Levy, 2016). For example, readers often infer that the meaning of the sentence "The mother gave the candle the daughter" corresponds to a more plausible alternative (e.g., "The mother gave the candle *to* the daughter"). According to the noisy-channel account, $P(s_i =$ "The mother gave the candle the daughter") is low whereas $P(s_i =$ "The mother gave the candle to the daughter") is higher, and the probability that the more plausible sentence was intended by the producer but corrupted (e.g., by the deletion of "to") into the implausible version that was perceived, $P(s_p =$ "The mother gave the candle the daughter"$\mid s_i =$ "The mother gave the candle to the daughter"), is relatively high. As a result, $P(s_i =$ "The mother gave the candle to the daughter" $\mid s_p =$ "The mother gave the candle the daughter") is higher than $P(s_i =$ "The mother gave the candle the daughter" $\mid s_p$ = "The mother gave the candle the daughter") and readers interpret the sentence accordingly. Further, readers maintain uncertainty about the preceding input as they process a sentence and can revise their initial parse in real time, as needed. For example, in an eye-tracking study, Levy et al. (2009) showed that, when a later portion of a sentence (e.g., "The coach smiled at the player tossed the ball") renders $P(s_i)$ low, readers look back to previous locations in the sentence (e.g., "at") which are probable loci of noise corruptions (e.g., because P("at" | "as") is high).

Thus, according to the noisy-channel framework, comprehenders consider a range of alternatives, in proportion to their probability, as a sentence unfolds. Capturing the full distribution of reader/listener expectations is crucial for future progress in developing a noisy-channel model that quantitatively predicts human language comprehension.

However, behavioral measures such as reading time and comprehension accuracy may lack the sensitivity to reveal this fine-grained, rapidly changing probabilistic signal.

In the present work, we take a first step towards probing how noisy-channel inference unfolds in the moment of processing by leveraging the temporal resolution of the electroencephalogram (EEG) signal. Two event-related potential (ERP) components have been consistently linked to sentence comprehension in electrophysiological investigations of language processing: the N400 and the P600. In what follows, we briefly review current and former accounts of these components with an eye towards their interpretation in light of a noisy-channel lens on human communication. We then propose, and provide empirical evidence, that the relative magnitudes of the N400 and P600 may constitute useful indices of the relative probabilities of literal interpretation versus noisy-channel correction of the input.

ERP signatures of language processing

The N400—a negativity peaking 400ms after word onset—is hypothesized to index the ease of accessing the semantic representation of a word given the preceding input (e.g., after "I take my coffee with cream and...", "dog" elicits a more negative deflection than "sugar"; Kutas & Hillyard, 1984; Kutas & Federmeier, 2011). Recent computational models construe the N400 as indexing the lexico-semantic prediction error or the update in network activation elicited by a word as it is integrated into the preceding context (e.g., Fitz & Chang, 2019; Rabovsky et al., 2018; see also Cheyette & Plaut, 2017 for a model of the timecourse of the N400). Consistent with a noisy-channel view of language processing, the N400 is reduced when an incongruous completion is orthographically related to a plausible continuation, such that a plausible noise corruption might be inferred (e.g., "Before lunch he has to deposit his paycheck at the bark [vs. bank]"; Laszlo & Federmeier, 2009; Ito et al., 2016).

The P600—a positivity most pronounced 600-900ms after word onset—is less well understood. It was originally hypothesized to reflect syntactic integration difficulty (e.g., after "Every Monday he...", "mow" elicits a larger positivity than "mows"; Osterhout & Holcomb, 1992; Friederici, 1995; Hagoort et al., 1993). However, this interpretation has faced numerous challenges. First, a number of non-syntactic manipulations elicit a P600 (e.g., spelling errors - "fone" instead of "phone"; Münte, Heinze, Matzke, Wieringa, & Johannes, 1998; van de Meerendonk, Indefrey, Chwilla, & Kolk, 2011; Vissers, Chwilla, & Kolk, 2006). Second, sentences like "The hearty meal was devouring…" elicit a P600 in spite of being syntactically well-formed (e.g., Kim & Osterhout, 2005; Kuperberg, 2007; Kuperberg et al., 2003; van Herten et al., 2005). According to traditional interpretations of these components, because these sentences are semantically anomalous, an N400 should ensue in place of these "semantic P600's" (Brouwer et al., 2012).

Consequently, alternative accounts of the P600 have been put forward in the literature. Some appeal to parallel streams of (syntactic and semantic) processing in constructing the representation for an input string (e.g., Kim & Sikos, 2011; Kos et al., 2010; Kuperberg, 2007). Others argue that, given its scalp distribution and tight time-locking to responses, the P600 belongs to the P300 family of domain-general components (Coulson et al., 1998; Sassenhagen et al., 2014; Sassenhagen & Fiebach, 2019; for a review, see Leckey & Federmeier, 2019), which are thought to index the process of updating one's model of the world when one encounters low-probability ("oddball") events (Donchin, 1981; Sutton et al., 1965). Consistent with a connection to the P300, Kolk and colleagues proposed an account of the P600 as indexing our continuous monitoring of the linguistic (or other) input for possible errors (Kolk et al., 2003; Kolk & Chwilla, 2007; van de Meerendonk et al., 2011; Vissers et al., 2006). Recent computational accounts take different approaches: Brouwer et al. (2017) propose a single-stream model of N400 and P600 effects, and argue that the P600 indexes semantic integration[1] into the unfolding utterance, and Fitz and Chang (2019) model the P600 as the prediction error at the sequencing layer of a neural network.

An ERP signature of noisy-channel inference

Building on the general error-monitoring idea (van de Meerendonk et al., 2011), we propose that ***the relative magnitudes of the N400 and P600 may provide a useful index of rational inference in the noisy-channel framework of sentence comprehension***. When the input is anomalous but unlikely to have been an error, a large N400 ensues and no P600 is typically observed. In contrast, if the input is anomalous but can be explained by a plausible noise process, readers infer that a more probable intended sentence was corrupted, and a P600 ensues while the N400 is reduced. We do not here aim to provide a mechanistic account of either component (but see discussion); instead, we aim to relate well-known patterns in the EEG signal to a computational-level account of sentence comprehension (Gibson et al., 2013; Levy, 2008) in order to illuminate the noisy-channel inference process.

More precisely, Equation 1 describes the proposed relationship between the N400 and P600 effects and noisy-channel inference. Given (i) a preceding sentence context C and its most probable parse[2]; (ii) an expected completion word, $w_{expected}$; (iii) the incoming (target) word: $w_{received}$; and (iv) $s_{expected}$ and $s_{received}$ —the sentences that correspond to connecting $w_{expected}$ and $w_{received}$ to C respectively, there is a smaller (less negative) N400 and larger (more positive) P600 effect whenever $P(s_i=s_{received} \mid s_p=s_{received})$ is lower than $P(s_i=s_{expected} \mid s_p=s_{received})$:

$$N400\ and\ P600\ \propto \frac{P(s_i = s_{expected} | s_p = s_{received})}{P(s_i = s_{received} | s_p = s_{received})} = \frac{P(s_p = s_{received} | s_i = s_{expected}) P(s_i = s_{expected})}{P(s_p = s_{received} | s_i = s_{received}) P(s_i = s_{received})} \quad (1)$$

[1] Note that this is conceptually similar to the N400 in Rabovsky et al.'s model and is somewhat at odds with many classic findings (outside of the "semantic P600" cases) showing that semantic integration difficulty leads primarily to negativity in the N400 time window and not positivity in the P600 time window. In fact, recent large-scale investigations observe negative voltages in the P600 time window for highly semantically implausible continuations (Fleur et al., 2020; Nieuwland et al., 2020).

[2] For the current purposes, we set aside the possibility of multiple parallel parses of the preceding context, C, and how their relative probabilities can be re-weighted given new input but see Levy et al. (2009) for discussion.

For example, when C = "The storyteller could turn any incident into an amusing…", $w_{expected}$ may be "anecdote." If $w_{received}$ = "antidote", $P(s_i = s_{expected})$ is higher than $P(s_i = s_{received})$ and the probability of a noise corruption (e.g., a typographical error) transforming "anecdote" into "antidote" $P(s_p = s_{received} \mid s_i = s_{expected})$ is relatively high (see Figure 1). As a result, $P(s_i{=}s_{expected} \mid s_p{=}s_{received})$ is higher than $P(s_i{=}s_{received} \mid s_p{=}s_{received})$. If $w_{received}$ = "hearse", $P(s_i = s_{expected})$ is higher than $P(s_i = s_{received})$ but the probability of a noise corruption transforming "anecdote" into "hearse" $P(s_p = s_{received} \mid s_i = s_{expected})$ is much lower than the probability of no corruption occurring, $P(s_p = s_{received} \mid s_i = s_{received})$. As a result, $P(s_i{=}s_{expected} \mid s_p{=}s_{received})$ is lower than $P(s_i{=}s_{received} \mid s_p{=}s_{received})$. According to Equation 1, both the N400 and P600 amplitudes will be more positive when $w_{received}$ = "antidote" compared to when $w_{received}$ = "hearse" (see Figure 1 for additional examples).

Several previously observed empirical phenomena in the ERP literature can be reinterpreted through the lens of noisy-channel comprehension. The following non-exhaustive list of findings lends further support to the proposal that the relative component magnitudes reflect the probability of a noisy-channel inference:

1) Number, gender, and case agreement errors elicit no N400 effect (or a much smaller N400 relative to those elicited by semantically incongruous words; Guajardo & Wicha, 2014; Wicha et al., 2004) but a robust P600 effect, because a close alternative exists in these cases, which the comprehender can correct to. For example, the probability of the meaning/structure resulting from completing "Every Monday he…" with "mow", $P(s_i$="Every Monday he mow"), is low, whereas $P(s_i$="Every Monday he mows") is relatively high. Critically, the probability of a noise process changing "mows" to "mow," $P(s_p$="Every Monday he mow"$|s_i$="Every Monday he mows"), is relatively high; "mow" involves only a single character/morpheme deletion from "mows".
2) "Semantic P600s" exist when a close alternative exists that the producer plausibly intended. For example, $P(s_i$=" The hearty meal was devouring…"), is low, while $P(s_i$=" The hearty meal was devoured…") is relatively high, and critically $P(s_p$= "The hearty meal was devouring…" $| s_i$= "The hearty meal was devoured…") is high.
3) Semantic violations that have been typically found in the literature elicit a large N400 effect and little to no P600 effect[3] because the prior probability of the received sentence (or word given the preceding sentence context) is low (e.g., $P(s_p$= "I take my coffee with cream and dog" is low) but noise corruption is implausible (e.g., P(s_p= "I take my coffee with cream and dog" $| s_i$ = " I take my coffee with cream and <u>sugar</u>") is low).

[3] In some studies, a P600 is reported after an N400 for canonical semantic violations (see Brouwer, Fitz, & Hoeks, 2012; Van Petten & Luka, 2012). It is noteworthy that this is more likely to occur when an unnatural secondary task (e.g., acceptability judgments) is included. Kolk et al. (2003) directly compared judgment and naturalistic comprehension tasks and found a P600 in the semantic violation cases only for the former. When the task is to find errors, participants plausibly assume an increased likelihood of errors across the board. Other aspects of the task, e.g., the proportion of errors in the fillers or the proportion of incongruous sentences in the environment, also affect the prior and likelihood and, therefore, the probability of the N400 and P600 on the current account (e.g., Delaney-Busch et al., 2019; but see Nieuwland, 2020). We return to this issue in the discussion.

4) Semantic violations which involve words that are orthographically close to a highly plausible completion elicit smaller N400 effects (compared to words that are incongruous and not orthographically close to congruous word), which are followed by P600 effects. In such cases, the probability of the received sentence (e.g., P(s_p= “Before lunch he has to deposit his paycheck at the bark”)) is low, but a noise corruption (e.g., that a typographical error caused “n” to be replaced by “r”) is plausible (e.g., P(s_p= “Before lunch he has to deposit his paycheck at the bark” | s_i= “Before lunch he has to deposit his paycheck at the bank”) is high).
5) Further, semantic violations involving words that are orthographically close to a more plausible alternative elicit smaller N400 and larger P600 effects when the cloze probability of the most likely completion is higher (Ito et al., 2016, 2017). All else being equal, P(s_i= “…bank” | s_p= “…bark”) is higher when the prior probability of the most likely completion, P(s_i= “...bank”) is higher.
6) Similarly, spelling errors which create pseudo-words (e.g., “pank”) that are close to a plausible word elicit no N400 effect and a larger P600 effect (compared to real words that are incongruous, such as “bark”). In these cases, a noise corruption (e.g., that a typo caused “b” to be replaced by “p”) is also plausible but the prior probability of the pseudo-word, P(s_i= “...pank”), is very low. Therefore the likelihood of the noisy-channel inference is even higher.
7) Syntactic errors in “Jabberwocky” sentences, i.e., sentences that include function words/morphemes but cannot be interpreted with respect to world knowledge lead to reduced or no P600 effects (Münte et al., 1997; Yamada & Neville, 2007). In such cases, it is difficult to infer plausibly intended meanings because the materials are, by design, devoid of meaning.
8) Finally, a P600 has been observed in studies with semantic violations in extended discourse contexts. For example, in a study by Nieuwland and Van Berkum (2005), participants read a story (e.g., about a tourist and his suitcase; both entities were mentioned several times). In critical sentences like “Next, the woman told the tourist/suitcase…”, a P600 was observed for “suitcase” (not an N400, as in a null context), plausibly because a word substitution error, when both lexical entries are highly probable in the discourse, is a probable production error. Similarly, code switches, which are probable in bilingual speech, elicit a P600 (Moreno et al., 2002).

Here, we directly evaluate whether the ERP components track noisy-channel inferences by measuring the probability that a close alternative would be inferred for each target word. We first replicate several existing effects using an experimental design with four conditions (Figure 1): (1) a control condition with no violations, (2) a condition with a canonical semantic violation, (3) a condition with a canonical syntactic violation (number agreement error), and critically, (4) a condition where the target word is semantically inappropriate but orthographically and phonologically close (e.g., in terms of Levenshtein distance) to a semantically plausible neighbor. Behavioral norming indicates that the proximity of such a neighbor makes the plausibly intended word recoverable. As a result, the critical condition is expected to elicit a noisy-channel inference and, by hypothesis, a P600 (similar to the syntactic condition and in contrast to the semantic condition). We then look at the relationship between the probability of a noisy-channel inference (determined behaviorally on an independent set of participants) and the magnitude of the N400 and P600 effects. The noisy-channel framework predicts that words that are more likely to be interpreted as a close alternative will elicit a smaller N400 and larger P600 effect.

Figure 1. Example materials and predictions for N400 and P600 ERP effects according to the noisy-channel framework for human sentence comprehension.

Methods

Participants

Twenty-nine right-handed native English speakers participated in the ERP experiment, 24 of whom were included in the final analysis (10 males; age 18-40 years). Participants were recruited from the MIT Brain and Cognitive Sciences subject pool and the Wellesley College student community. Informed consent was obtained in accordance with the MIT Committee on the Use of Humans as Experimental Subjects. Participants were compensated with cash for their participation. Five subjects were excluded from final analysis due to an excessive number of artifacts in the EEG signal. An additional 475 participants were recruited for behavioral tasks on Amazon's Mechanical Turk. Data from 32 of those participants were excluded before analysis because they did not meet inclusion criteria (located in the US, self-reported native English speaker, answered at least 75% of fillers correctly, provided responses on at least 90% of trials).

Materials & Behavioral Norming

160 ten-word-long sentences were constructed (with four conditions each, as described above; Figure 1) and distributed across four presentation lists following a Latin Square design, so that each list contained only one version of an item (and 40 trials per condition). The target word (always a noun) was the last word in the sentence. The target words in the semantic violation and critical conditions were target words in the control condition for other items (e.g., "hearse" in the example above was the target word in the control condition of another item); the target words were thus identical across these conditions (and only differed in the number feature between these conditions and the syntactic violation condition). In addition, 320 10-word-long filler items were constructed. These contained no semantic or syntactic violations.

Experimental materials were evaluated for cloze probability of the final word, acceptability, perceived likelihood of the error, and recoverability of the intended word, each on an independent set of participants to ensure that a) the target words were judged less likely to be errors in the semantic violation condition than in the critical and syntactic violation conditions, and b) the intended words were more recoverable in the critical and syntactic violation conditions than in the semantic violation condition.

In the cloze task, participants were presented with each item with the last word missing and asked to complete it. The final words of experimental items in the control condition had a mean cloze probability of 0.40 (± 0.31 SD), suggesting that the sentence contexts were effective in setting up expectations for a particular word but were not overly constraining.

The results of the remaining norming tasks are summarized in Table 1 and the full set of materials is available at https://osf.io/vcsfb/?view_only=ba0079719cfa4118be5cc99714135acf. In the acceptability task, participants were asked to judge a sentence for how natural it sounds on a 7-point scale. To assess participants' perceived likelihood of the error in different conditions, we had participants first read the control (no violation) condition and answer a yes/no question about it. They were then asked the following question, "How likely would it be for someone to produce the following speech error for the last word, when they intended the above sentence?" Participants judged this likelihood on a 7-point scale. Finally, to assess how recoverable the intended word was in different conditions, we presented participants with a complete sentence in all but the control condition and the following instructions, "The final word in each of the following sentences is wrong: someone typed the wrong word. Please type in a different word, the one that was probably intended."

	Control	Syntactic error	Semantic error	Recoverable semantic error (with a plausible neighbor)
Acceptability rating (N=102)	6.0 (0.7)	3.9 (0.7)	2.3 (0.6)	2.9 (0.8)
Error likelihood rating (N=106)	N/A	4.4 (0.4)	1.2 (0.1)	3.3 (0.7)
Error recoverability rating (N=235)	N/A	0.92 (0.11)	0.38 (0.28)	0.84 (0.17)

Table 1. Results of the norming studies. The acceptability and error likelihood ratings are on a 7-point scale. The error recoverability is a proportion (i.e., the number of times the sentences were completed with the final word from the control condition out of the total number sentence completions). Standard deviations over items are given in parentheses.

EEG recording

EEG was recorded from 32 scalp sites (10-20 system positioning), a vertical eye channel for detecting blinks, a horizontal eye channel to monitor for saccades, and two additional electrodes affixed to the skin above the mastoid bone. EEG was acquired with the Active Two Biosemi system using active Ag-AgCl electrodes mounted on an elastic cap (Electro-Cap Inc.). All channels were referenced offline to an average of the mastoids. The EEG was recorded at 512 Hz sampling rate and filtered offline (bandpass 0.1-40 Hz). Trials with blinks, eye movements, muscle artifact, and skin potentials were rejected prior to averaging and analysis. An average of 15.6% of trials were rejected per participant (range: min = 0.6%, max = 26.3%).

Procedure

Participants were tested individually in a sound-attenuated booth where stimuli were presented on a computer monitor. Stimuli appeared in the center of the screen word-by-word, time-locked to the vertical refresh rate of the monitor (75 Hz). The sentences were displayed word-by-word in white on a black background. Each trial began with a pre-trial fixation (1,000 ms), followed by 500 ms of a blank screen. Then, the sentence was presented for 5,800 ms (400 ms per word and 100 ms ISI, with an ISI of 900 ms after the last word). The order of trials was randomized separately for each participant. Each list was pseudo-randomly divided into ten "runs", in order to give participants breaks as needed. Each run contained 4 trials of each condition and 32 fillers.

To ensure that participants read the sentences for meaning, yes/no comprehension questions appeared after 60 of the 480 trials (experimental and filler), constrained such that there were no more than three consecutive trials with a question, and no more than 20 consecutive trials without a question. The correct answer was "yes" half of the time. Comprehension questions were displayed all at once (for 3,500 ms + 100 ISI) in aqua on a black background, and participants responded "yes" or "no" by pressing buttons on a gamepad. At the beginning of the experiment, participants were shown a set of 12 practice items to familiarize them with the procedure. The experiment took ~1 hour.

Analysis

Eight centro-parietal electrode sites (C3, Cz, C4, CP1, CP2, P3, Pz, and P4) were included in the analysis. These sites reflect the typical distribution of N400 and P600 effects reported in the literature (Kutas & Federmeier, 2011; Tanner, 2019)[4]. ERP signals were time-locked to the onset of the sentence-final (target) word and individual trial epochs from 100 ms prior to the onset of this stimulus until 1,000 ms after onset were extracted. The time window from -100 ms to word onset was used as the pre-stimulus baseline. Mean amplitude measurements were computed in two time windows – 300-500 ms and 600-800 ms – to quantify the N400 and P600 components, respectively. Time windows were chosen to match standard time windows used in the literature (Gouvea et al., 2010; Kutas & Federmeier, 2011) and to be equal in duration with a 100 ms gap in between to reduce dependence between the windows.

For each of the two time windows of interest (300-500 ms and 600-800 ms), the mean amplitude was entered as the dependent variable in a linear mixed-effects regression model, with condition (control, semantic violation, syntactic violation, recoverable) as a dummy-coded fixed effect (with control as the reference level). The models included random intercepts for participants, items, and electrodes, and random condition slopes for each grouping variable. Analyses were performed using the "brms" package for Bayesian regression modeling in R (Bürkner, 2017), which interfaces with the Stan probabilistic programming language (Carpenter et al., 2017). Moderately regularizing priors were chosen based on prior literature. In particular, a normal distribution with mean 0 and standard deviation 2.5 was chosen for the beta coefficients based on the reasoning that most ERP effects fall between +/- 5μV. Sampling for all models was done with four chains for 2,000 iterations (1,000 for warmup). R-hat values were 1.00 for all parameters in all models reported. Data and analysis code are available at https://osf.io/vcsfb/?view_only=ba0079719cfa4118be5cc99714135acf.

Results

Participants mostly answered the comprehension questions accurately (mean = 0.88, bootstrapped 95% confidence interval = [0.85, 0.91]), which suggests that they were engaged in the task.

N400 and P600 components.

[4] Visual inspection of Appendix Figure A1 indicates that both the N400 and P600 were indeed centro-parietally distributed, consistent with the literature. Crucially, the later positivity was not larger anteriorly, which would suggest a qualitatively different component (see Van Petten & Luka, 2012).

The average timecourse by condition of the ERP amplitude for the eight electrodes included in the analysis is shown in Figure 2a. As expected, and replicating many previous studies, in the ***N400 window***, the ERP amplitude decreased by -4.09 µV (95% Credible Interval (CI) = [-5.06, -3.02]) in the semantic condition relative to the control condition. The amplitude was also somewhat more negative (Estimate = -1.37, 95% CI = [-2.51, -0.17]) in the recoverable condition relative to the control condition but more positive than in the semantic condition (Estimate = -2.73, 95% CI = [-3.88, -1.53]). An N400 effect is expected for the recoverable condition target word because it is not strongly facilitated by the semantic context, unlike the control condition target word. The N400 effect in the syntactic condition was not different from the control condition (Estimate = -0.48, 95% CI = [-1.66, 0.72]).

In the ***P600 window***, the ERP amplitude did not differ between the control condition and the semantic condition (Estimate = -0.85, 95% CI = [-2.08, 0.35]). However, P600 amplitude was more positive both in the syntactic (Estimate = 2.10, 95% CI = [0.91, 3.22]) and in the recoverable condition (Estimate = 1.34, 95% CI = [0.11, 2.52]). In other words, as predicted by the noisy-channel inference account, the critical recoverable condition, where the target word was semantically inappropriate but phonologically and orthographically close to a plausible neighbor, elicited a P600 effect, similar to the syntactic condition. See Figures 2b and 2c for summaries and Table 2 for full model estimates.

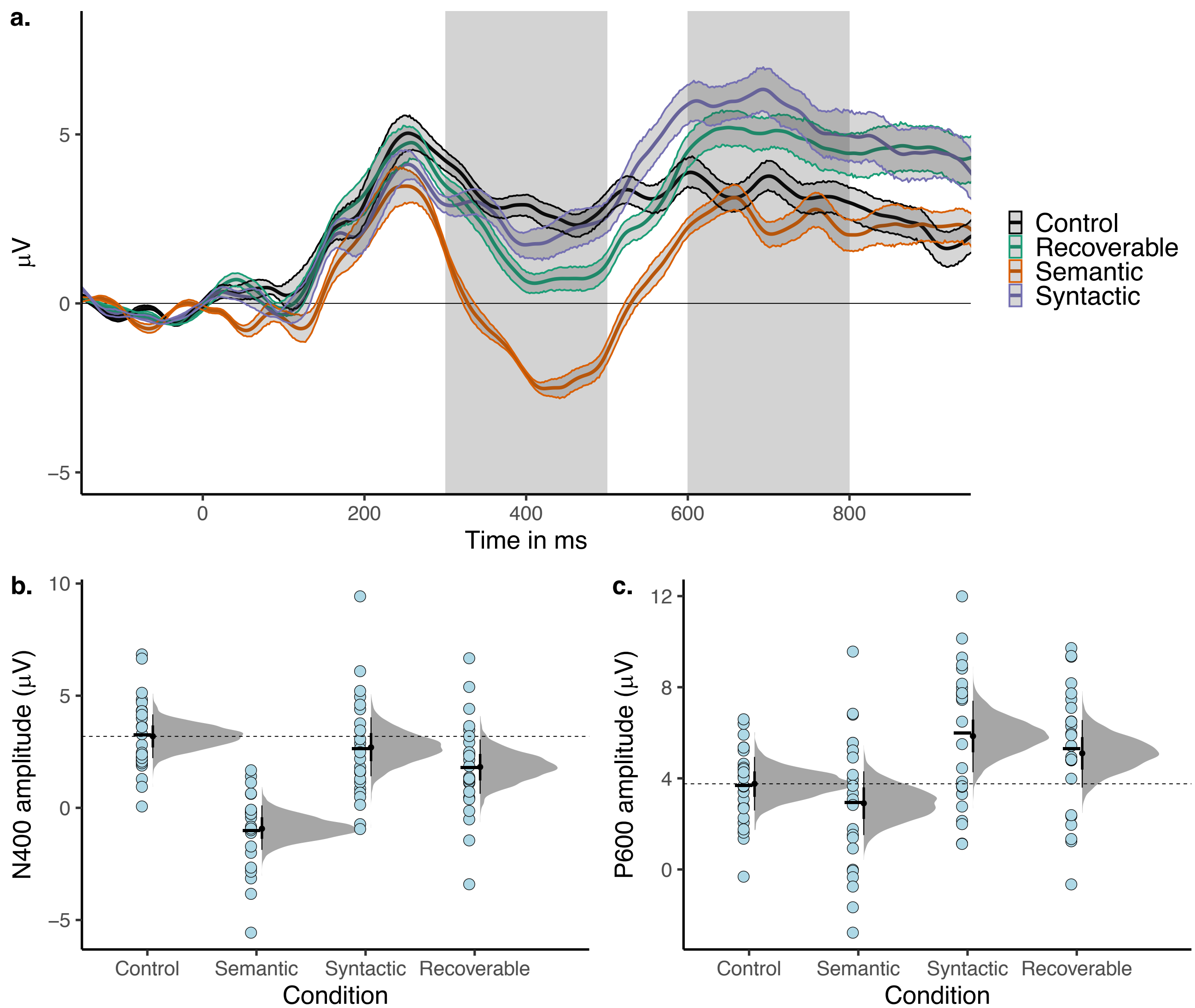


Figure 2. a. Timecourse of grand average ERP for each condition from eight centro-parietal channels used in the analysis (see Appendix A for all channels). The x-axis shows time from the onset of the presentation of the critical word, and the y-axis shows voltage (negative plotted down), as compared to the mean voltage of the baseline 100 ms pre-stimulus interval. Ribbons indicate bootstrapped 95% confidence intervals over the channel means. The two gray rectangles in each plot indicate the time windows of interest: 300-500ms (N400 window) and 600-800ms (P600 window). b-c. Mean amplitudes of (b) the N400 and (c) P600 components. Light blue points represent individual participant means and the black horizontal bar represents the overall mean for each condition. Densities and point intervals represent the distribution of fitted conditional means from Bayesian linear mixed-effects model posteriors. Dashed horizontal line indicates the mean amplitude in the control condition.

	N400		**P600**	
Predictors	*Estimates*	*CI (95%)*	*Estimates*	*CI (95%)*

Intercept	3.19	2.22 – 4.17	3.76	2.59 – 4.94
Semantic	-4.11	-5.06 – -3.02	-0.85	-2.08 – 0.35
Recoverable	-1.36	-2.51 – -0.17	1.35	0.11 – 2.52
Syntactic	-0.48	-1.66 – 0.72	2.11	0.91 – 3.22
Random Effects Variances				
σ^2	67.98		84.09	
Electrode (n=8)	0.34		0.94	
Item (n=160)	16.81		12.31	
Subject (n=24)	3.06		3.83	
Electrode:Semantic	0.10		0.10	
Electrode:Recoverable	0.05		0.05	
Electrode:Syntactic	0.10		0.11	
Item:Semantic	20.93		22.98	
Item:Recoverable	25.01		28.67	
Item:Syntactic	22.88		21.50	
Subject:Semantic	3.29		6.76	
Subject:Recoverable	4.71		5.52	
Subject:Syntactic	5.23		4.90	
ICC: 0.22, *Observations:* 25928				

Table 2. Summary of estimates from Bayesian mixed-effects regression models predicting N400 and P600 amplitudes by condition (formula: `Amplitude ~ condition +(1+condition|subject)+ (1+condition|item) + (1+condition|electrode)`).

Magnitude of N400 and P600 and recoverability of the plausible alternative.

To further explore these effects, we assessed whether the magnitudes of the N400 and P600 are linearly related to the recoverability of the word. We computed two measures of recoverability. The first is the Levenshtein distance between each target word (e.g., antidote) and its control condition counterpart (e.g., anecdote). Levenshtein distance was computed using the adist() function in R. The second measure was taken from the norming data (summary available at https://osf.io/vcsfb/?view_only=ba0079719cfa4118be5cc99714135acf): the percentage of correct guesses about which word was intended. The relationships between the magnitude of the ERP effects for an item (averaging over participants and electrodes and subtracting the amplitude for the control condition from the amplitudes in the other three conditions within each time window) and the two measures of recoverability are shown in Figure 3. Five simple linear regression models were fitted using brms (Table 3), with the same priors as in the above models where applicable (see further analysis details at https://osf.io/vcsfb/?view_only=ba0079719cfa4118be5cc99714135acf). Items which were less likely to elicit successful recovery of the control version had a larger Levenshtein distance from their control version (Estimate = -6.93, 95% CI = [-7.54, -6.34]), confirming the validity of operationalizing recoverability as Levenshtein distance from the nearest neighbor. Items for which participants were more likely to recover the control word elicited smaller N400 effects (Estimate = 3.15, 95% CI = [1.67, 4.65]) and larger P600 effects (Estimate = 3.29, 95% CI = [1.76, 4.85]). Similarly, items with a larger Levenshtein distance from their control elicited larger N400 effects (Estimate = -0.52, 95% CI = [-0.68,-0.37]) and smaller P600 effects (Estimate = -0.42, 95% CI = [-0.58,-0.26]). Note that these bivariate relationships are somewhat expected given that the 3 conditions were designed to be differentially recoverable. Models which include condition as an additional covariate indicate that these two predictors (condition and Levenshtein distance or Percent recovered) explain largely redundant variance (i.e., neither predictor is estimated to have a non-zero independent contribution).

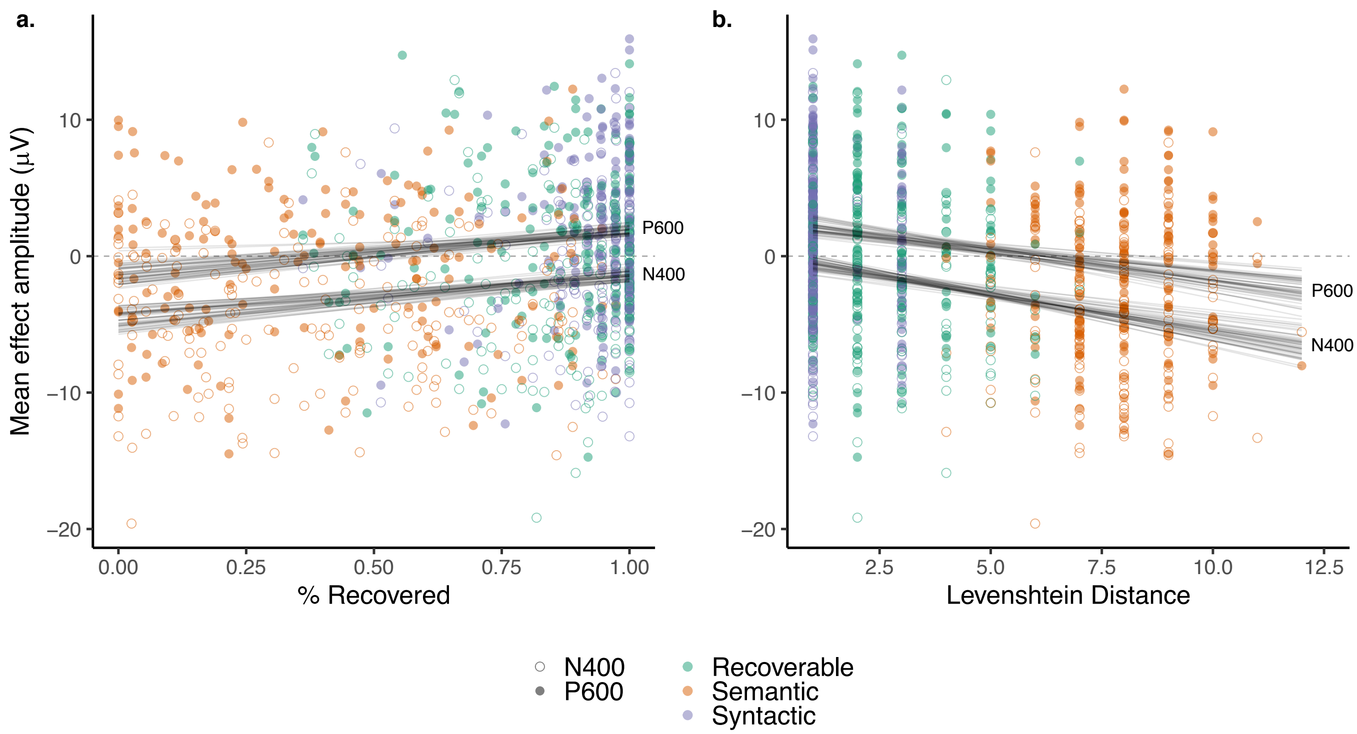


Figure 3. Relationships between the average N400 (unfilled points) and P600 effect (filled points) for each item in each experimental condition (after subtraction of corresponding amplitude in the Control condition) and two measures of recoverability: Percent of correctly recovered completions (a) and Levenshtein distance (b). Gray solid lines represent 50 fitted regression lines (randomly sampled from model posteriors) from models predicting either the size of the N400 effect or the P600 effect (separately) from percent recovered (a) and Levenshtein distance (b).

	Levenshtein Distance		N400		P600	
Predictors	*Estimates*	*CI (95%)*	*Estimates*	*CI (95%)*	*Estimates*	*CI (95%)*
Intercept	8.90	8.44 – 9.37	-4.50	-5.63 – -3.34	-1.42	-2.61 – -0.19
Percent recovered	-6.93	-7.54 – -6.34	3.14	1.67 – 4.65	3.31	1.76 – 4.85
	N400		P600			
Predictors	*Estimates*	*CI (95%)*	*Estimates*	*CI (95%)*		
Intercept	-0.17	-0.95 – 0.60	2.60	1.80 – 3.42		
Levenshtein Distance	-0.52	-0.68 – -0.37	-0.42	-0.58 – -0.26		

Table 3. Bivariate Bayesian linear regressions relating average recoverability of each target word to average ERP effects in both N400 and P600 time windows. Observations = 480.

Discussion

We observed a reduced (more positive) N400 effect and a P600 effect when participants read sentences where the target word was semantically inappropriate but had a close orthographic and phonological neighbor, allowing for the possibility that the received message was corrupted by noise (replicating Ito et al., 2016; Laszlo & Federmeier, 2009). The intended (plausible) word was thus recoverable, and comprehenders could ‘correct’ the signal. This effect was similar to that observed for the canonical syntactic violation condition. A large N400 and no P600 were observed for the canonical semantic violation, where the intended meaning could not be recovered. Further, the amplitudes of the N400 and P600 effects were linearly related to the likelihood of recovering the plausible alternative. Thus, the reduced N400 (relative to the semantic condition) and large P600 plausibly index the presence of a noisy-channel inference. Although we cannot rule out the possibility that readers are simultaneously attempting to correct the preceding sentence context which they hold in memory (Futrell et al., 2020; Levy et al., 2009), these data suggest that readers can also correct a word to a more plausible alternative as its semantic representation is being accessed.

Though similar in spirit, the predictions of the noisy-channel account can be dissociated from those of a purely error-monitoring based account (van de Meerendonk et al., 2011). For instance, it is typically assumed, on an error-monitoring view, that the larger the error, the more likely it is that a P600 effect will be observed. Holding other probabilities constant, the noisy-channel account would predict the opposite pattern, as larger noise corruptions are less plausible: an insertion of one letter is a more plausible typographical error than an insertion of five letters. Whereas the latter may be easier to detect as an error, the former is more likely to lead to recovery of the intended word. Indeed, the relationship between recoverability and ERP effects points to a process that goes beyond the "mere" detection of errors. However, a more fine-grained investigation of this relationship, in which both prior probabilities and error probabilities are carefully controlled and parametrically manipulated, will be needed in order to conclusively distinguish between the error-monitoring and the noisy-channel inference accounts.

The noisy-channel framework, as a computational-level account, concerns the *probability of an intended sentence* given the perceived form, no matter *how* the most likely representation for an input string might have been computed. Recent implemented models of the N400 and P600 (e.g., Brouwer et al., 2017; Fitz & Chang, 2019; Rabovsky, 2020; Rabovsky et al., 2018) have focused on addressing how the comprehension mechanisms compute sentence meanings or learn their relative probabilities. Both levels of analysis are important for making progress on understanding language processing, and to the extent that they generate accurate predictions regarding the probability of perceived sentences and their alternatives, the existing mechanistic models may be compatible with the noisy-channel framework. In fact, several connectionist models of the N400 make the same (partial) predictions as the current proposal in that they capture the absence or reduction of an N400 effect when an anomalous word is orthographically close to a plausible word relative to when an anomalous word has no plausible neighbors (Laszlo & Plaut, 2012; Rabovsky et al., 2018). Among models that explicitly include a P600 component, neither addresses this particular case. Brouwer et al. focus on semantic P600 effects, whereas Fitz and Chang focus on morpho-syntactic errors only. It is beyond the scope of the current paper to determine whether either or both of these models can be straightforwardly extended to account for the effects explored here.

Underlying mechanisms

The current proposal does not aim to provide a mechanistic model of how the N400 and P600 are generated but rather build a bridge between the growing noisy-channel sentence comprehension literature and the psycholinguistic ERP literature. As a result, we do not make strong claims regarding the underlying processes that may give rise to the observed ERP patterns. For example, despite the fact that we discuss two components with different timecourses, we do not know whether computing the plausibility of a noise corruption is part of the same process underlying the semantic update that takes place when a new word is encountered or whether the P600 reflects a later, attention-demanding reanalysis (see Rabovsky & McClelland, 2020). Similarly, the current data do not address whether the correlation between the P600 effect and the probability of noise inference only holds for sentences in which readers explicitly noticed the error and chose to interpret the sentence as the more plausible alternative[5]. In a recent study, Qian, Garnsey, and Christianson (2018) found that the magnitude of the P600 effect to an

[5] In the present experiment, though a few critical sentences were followed by comprehension questions, they were not always directly aimed at determining whether the reader had made a noisy-channel inference and they were too few in number to conduct any meaningful analyses.

ambiguous verb[6] in a garden-path sentence (e.g., While the man hunted the deer ran into the woods) was not related to accuracy on the subsequent comprehension question (e.g., Did the man hunt the deer?) (see also Sanford et al., 2009). However, these garden-path sentences are particularly complex and infrequent, so extending this work to the stimuli used in the present paradigm may yield different results.

Limitations

Further work is needed to determine whether there exist examples of N400 and P600 effects which cannot be readily tied to a noisy-channel inference process. Though it is beyond the scope of this paper to review the entirety of the N400 and P600 literatures, we discuss several such examples here and speculate about the ways in which they may be interpreted within the current framework. As mentioned above, "canonical" N400 effects are sometimes followed by P600s (Van Petten & Luka, 2012). The addition of an explicit task (e.g., grammaticality judgment) may provide a partial explanation (e.g., Osterhout & Mobley, 1995) – the task may increase the expected likelihood of noise across the board, which increases the rate of noisy-channel inference (Gibson et al., 2013). More importantly, for a word that is low probability in context it will often be possible to treat it as either a faithfully represented word that is unexpected, leading to an N400, or a corrupted version of an expected word, leading to a P600-like response. As noted by Van Petten and Luka[7], the majority of semantic violation tasks were not designed with noise-correction (or reanalysis) questions in mind and thus are unlikely to have thoroughly controlled how likely the materials are to have been corrected to an alternative. In the current dataset, where we have carefully chosen the semantic violations to not be the plausible outcome of a noise corruption, there is no hint of a P600 effect in the semantic violation condition. A re-analysis of existing datasets showing biphasic N400/P600 responses with this variable (plausibility of a noisy-channel inference) in mind may be a fruitful avenue for future progress on this topic. Further, different participants may have different relative weightings of the two options (literal interpretation or noisy-channel inference) given their own idiosyncratic language experience, so a blended response could reflect averaging across participants or, more intriguingly, high uncertainty within the same individual.

In addition, pragmatic processing has been linked to P600 effects (see Hoeks & Brouwer, 2014 for a review), for example in experiments looking at comprehension of figurative language (Regel et al., 2011) and jokes (Du et al., 2013). It is possible that jokes, for instance, violate the reader's expectation (i.e., their literal meaning has lower prior probability) and lead them to consider the alternative that would have been said, if the

---

[6] Levy (2008) proposed that readers may infer that "it" was deleted during transmission and interpret the sentence as a more (syntactically) plausible alternative (e.g., While the man hunted it the deer ran into the woods).

[7] "Assuming a continuity between the parietal post-N400 positivity and the semantic P600 implies that the parietal PNP also reflects attempted re-analysis or checking of bad sentences. This process may be only variably invoked by incongruent sentence completions depending on a host of difficult-to-quantify factors: the exact construction of incongruent sentences by different experimenters (i.e., whether there is any hint that a sentence could be re-interpreted in a way that makes sense) and/or the verbal abilities or motivation of individual subjects."

sentence were meant to be serious. We leave it to future work to investigate this intriguing speculation. Critically, this puzzle doesn't undermine the utility of the P600 as an index of noisy-channel correction in future experiments, provided experimenters carefully control pragmatics in their materials.

Future directions

The focus of this paper has been on the probability of noise corruptions at the orthographic/phonological level within a word during reading, but we expect that the same principles would apply more generally: a P600 is predicted to ensue (on average) whenever the probability that some portion of the linguistic input was corrupted by noise exceeds the probability that it was transmitted faithfully (according to Equation 1). Thus, ERP evidence could be brought to bear on a variety of timely questions in sentence processing. For example, offline responses indicate that readers commonly draw noisy-channel inferences when reading sentences such as, "The mother gave the candle the daughter." Yet it is unknown at what point this inference takes place—in the moment of first-pass processing, or during reanalysis, or only after reading the comprehension question—or what kind of noise corruption readers assume (e.g., an exchange of "candle" and "daughter," or a deletion of "to"). ERPs recorded during word-by-word sentence reading have the potential to shed light on these questions (e.g., if the noise inference process takes place in real time and readers assume a "to" deletion, then a P600 effect would be expected on the third "the") as well as other questions related to the reader's (or listener's) implicit model of the noise which have been left open by the use of behavioral methods.

A key prediction of the noisy-channel proposal is that the N400 and P600 should be modulated by the distribution of errors in the input because a rational comprehender will tune their noise model to the observed distribution of errors in the environment (Gibson et al., 2013; Ryskin et al., 2018). Indeed, increasing the number of sentences that contain syntactic violations leads to a reduction of the P600 magnitude (Coulson et al., 1998; Hahne & Friederici, 1999). Similarly, Hanulíková et al. (2012) showed reduced P600s to syntactic errors in foreign-accented speech, where an agreement error is more expected (compared to native-sounding speech), suggesting that listeners take speaker-specific information into account, in addition to the overall proportion of errors in the input (see also Gibson et al., 2017). Moreover, Zhou, Garnsey, and Christianson (2019) replicated Hanulíková et al.'s finding and additionally showed that the P600 reduction was absent for pronoun errors, which are less common than subject-verb agreement errors in the speech of the population in question, namely native Chinese, L2 English speakers. This finding is consistent with the fact that listeners not only adapt the overall rate of errors in their noise model to the particular speaker-context (e.g., high error-rate vs. low error-rate), but also tune their model of the noise to reflect the relative probabilities of different errors in a context-specific way (e.g., speaker A and B make errors at the same rate overall but A makes mostly deletion errors whereas B makes mostly insertion errors; Ryskin et al., 2018). We speculate that these noise models are likely continuously updated based on experience with the environment such that listeners bring a set of expectations about the noise model to the lab and then update it throughout the experiment (similar to current views of how syntactic processing reflects both lifelong and recent language statistics via the same mechanism, e.g., Chang et al., 2012; Fine et al., 2013; Ryskin et al., 2017). Future work is needed to provide a systematic test of the quantity and nature of input that will shift the noise model and what consequence this will have for the N400 and the P600.

Conclusion

To conclude, the relative magnitudes of the N400 and P600 ERP components seem to be well approximated by the probability of the comprehender inferring that the input has been corrupted by noise (e.g., perceptual or production error) during transmission. This work contributes to a growing literature suggesting that the human language system is well-adapted to potential corruption of the linguistic signal. Though we can only speculate about the underlying mechanisms and future studies are needed to generalize this proposal to a wider set of scenarios, the P600 effect is promising as a signal of noisy-channel error correction taking place in real-time. The current framework is particularly useful in that it allows predictions to be generated for any arbitrary set of sentences for which it is possible to estimate relative prior probabilities and noise likelihoods and it opens the door to electrophysiological investigations of the comprehender's implicit noise model.

Acknowledgments:
We thank Roger Levy, Peter Hagoort, Gina Kuperberg, Milena Rabovsky, and the audiences at the Neurobiology of Language 2012 conference, and the CUNY 2013 Sentence Processing conference for feedback on this work. We are also grateful to Steve Piantadosi for comments on a draft of the manuscript.

Conflict of Interest:
Authors report no conflict of interest.

Funding sources:
This work was supported by National Science Foundation Grant 0844472 from the Linguistics Program to EG, by the K99/R00 grant 057522 from NICHD to EF, and F32 015163 from NIDCD to RR.

References

Bond, Z. (Dzintra). (1999). Morphological Errors in Casual Conversation. *Brain and Language*, *68*, 144–150. https://doi.org/10.1006/brln.1999.2071

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, *44*(2), 123–147. https://doi.org/10.1177/00238309010440020101

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A Neurocomputational Model of the N400 and the P600 in Language Processing. *Cognitive Science*, *41*(S6), 1318–1352. https://doi.org/10.1111/cogs.12461

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

Bürkner, P.-C. (2017). **brms**: An *R* Package for Bayesian Multilevel Models Using *Stan*. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Chang, F., Janciauskas, M., & Fitz, H. (2012). Language adaptation and learning: Getting explicit about implicit learning: Language Adaptation. *Language and Linguistics Compass*, *6*(5), 259–278. https://doi.org/10.1002/lnc3.337

Cheyette, S. J., & Plaut, D. C. (2017). Modeling the N400 ERP component as transient semantic over-activation within a neural network model of word comprehension. *Cognition*, *162*, 153–166. https://doi.org/10.1016/j.cognition.2016.10.016

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes 13*, 21–58.

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

Donchin, E. (1981). Surprise!… Surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Du, X., Qin, Y., Tu, S., Yin, H., Wang, T., Yu, C., & Qiu, J. (2013). Differentiation of stages in joke comprehension: Evidence from an ERP study. *International Journal of Psychology: Journal International De Psychologie*, *48*(2), 149–157. https://doi.org/10.1080/00207594.2012.665162

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7

Ferreira, F., & Patson, N. D. (2007). The "Good Enough" Approach to Language Comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid Expectation Adaptation during Syntactic Comprehension. *PLoS ONE*, *8*(10), e77661. https://doi.org/10.1371/journal.pone.0077661

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

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

Friederici, A. D. (1995). The Time Course of Syntactic Activation During Language Processing: A Model Based on Neuropsychological and Neurophysiological Data. *Brain and Language*, *50*(3), 259–281. https://doi.org/10.1006/brln.1995.1048

Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*.

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-Context Surprisal: An Information-Theoretic Model of Memory Effects in Sentence Processing. *Cognitive Science*, *44*(3), e12814. https://doi.org/10.1111/cogs.12814

Gibson, E., Bergen, L., & Piantadosi, S. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056. https://doi.org/10.1073/pnas.1216438110

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't Underestimate the Benefits of Being Misunderstood.

Psychological Science, *28*(6), 703–712. https://doi.org/10.1177/0956797617690277

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, *25*(2), 149–188. https://doi.org/10.1080/01690960902965951

Guajardo, L. F., & Wicha, N. Y. Y. (2014). Morphosyntax can modulate the N400 component: Event related potentials to gender-marked post-nominal adjectives. *NeuroImage*, *91*, 262–272. https://doi.org/10.1016/j.neuroimage.2013.09.077

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (sps) as an erp measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439–483. https://doi.org/10.1080/01690969308407585

Hahne, A., & Friederici, A. D. (1999). Electrophysiological Evidence for Two Steps in Syntactic Analysis: Early Automatic and Late Controlled Processes. *Journal of Cognitive Neuroscience*, *11*(2), 194–205. https://doi.org/10.1162/089892999563328

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When One Person's Mistake Is Another's Standard Usage: The Effect of Foreign Accent on Syntactic Processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887. https://doi.org/10.1162/jocn_a_00103

Hoeks, J. C. J., & Brouwer, H. (2014, September 1). *Electrophysiological Research on Conversation and Discourse Processing*. The Oxford Handbook of Language and Social Psychology. https://doi.org/10.1093/oxfordhb/9780199838639.013.024

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*, 157–171. https://doi.org/10.1016/j.jml.2015.10.007

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 635–652. https://doi.org/10.1037/xlm0000315

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225. https://doi.org/10.1016/j.jml.2004.10.002

Kim, A., & Sikos, L. (2011). Conflict and surrender during sentence processing: An ERP study of syntax-semantics interaction. *Brain and Language*, *118*(1), 15–22. https://doi.org/10.1016/j.bandl.2011.03.002

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

Kolk, H., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*(1), 1–36. https://doi.org/10.1016/S0093-934X(02)00548-5

Kos, M., Vosse, T. G., Van Den Brink, D., & Hagoort, P. (2010). About Edible Restaurants: Conflicts between Syntax and Semantics as Revealed by ERPs. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00222

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., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct Patterns of Neural Modulation during the Processing of Conceptual and Syntactic Anomalies. *Journal of Cognitive Neuroscience*, *15*(2), 272–293. https://doi.org/10.1162/089892903321208204

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. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161. https://doi.org/10.1038/307161a0

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326–338. https://doi.org/10.1016/j.jml.2009.06.004

Laszlo, S., & Plaut, D. C. (2012). A Neurally Plausible Parallel Distributed Processing Model of Event-Related Potential Word Reading Data. *Brain and Language*, *120*(3), 271–281. https://doi.org/10.1016/j.bandl.2011.09.001

Leckey, M., & Federmeier, K. D. (2019). The P3b and P600(s): Positive contributions to language comprehension. *Psychophysiology*, *0*(0), e13351. https://doi.org/10.1111/psyp.13351

Levy, R. (2008). *A noisy-channel model of rational human sentence comprehension under uncertain input*. 234. https://doi.org/10.3115/1613715.1613749

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086–21090. https://doi.org/10.1073/pnas.0907664106

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching Languages, Switching Palabras (Words): An Electrophysiological Study of Code Switching. *Brain and Language*, *80*(2), 188–207. https://doi.org/10.1006/brln.2001.2588

Münte, T. F., Heinze, H.-J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia*, *36*(3), 217–226. https://doi.org/10.1016/S0028-3932(97)00119-X

Münte, T. F., Matzke, M., & Johannes, S. (1997). Brain Activity Associated with Syntactic Incongruencies in Words and Pseudo-Words. *Journal of Cognitive Neuroscience*, *9*(3), 318–329. https://doi.org/10.1162/jocn.1997.9.3.318

Nieuwland, M. S. (2020). *How 'rational' is semantic prediction? A discussion and reanalysis of Delaney-Busch, Morgan, Lau & Kuperberg (2019)*. PsyArXiv. https://doi.org/10.31234/osf.io/sm7fq

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., Matthew Husband, E., Ito, A., Kazanina, N., Kogan, V., Kohút, Z., Kulakova, E., Mézière, D., Politzer-Ahles, S., Rousselet, G., … Von Grebmer Zu Wolfsthurn, S. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the*

Royal Society B: Biological Sciences, *375*(1791), 20180522. https://doi.org/10.1098/rstb.2018.0522

Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, *24*(3), 691–701. https://doi.org/10.1016/j.cogbrainres.2005.04.003

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*(6), 739–773. https://doi.org/10.1006/jmla.1995.1033

Poppels, T., & Levy, R. P. (2016). Structure-sensitive Noise Inference: Comprehenders Expect Exchange Errors. *Proceedings of the 38th Annual Conference of the Cognitive Science Society.*, 378–383.

Qian, Z., Garnsey, S., & Christianson, K. (2018). A comparison of online and offline measures of good-enough processing in garden-path sentences. *Language, Cognition and Neuroscience*, *33*(2), 227–254. https://doi.org/10.1080/23273798.2017.1379606

Rabovsky, M. (2020). Change in a probabilistic representation of meaning can account for N400 effects on articles: A neural network model. *Neuropsychologia*, *143*, 107466. https://doi.org/10.1016/j.neuropsychologia.2020.107466

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. https://doi.org/10.1038/s41562-018-0406-4

Rabovsky, M., & McClelland, J. L. (2020). Quasi-compositional mapping from form to meaning: A neural network-based approach to capturing neural responses during human language comprehension. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1791), 20190313. https://doi.org/10.1098/rstb.2019.0313

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). *Isn't It Ironic? An Electrophysiological Exploration of Figurative Language Processing*. *23*(2), 277–293.

Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition*, *181*, 141–150. https://doi.org/10.1016/j.cognition.2018.08.018

Ryskin, R., Qi, Z., Duff, M. C., & Brown-Schmidt, S. (2017). Verb biases are shaped through lifelong learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 781–794. https://doi.org/10.1037/xlm0000341

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2009). Anomalies at the Borderline of Awareness: An ERP Study. *Journal of Cognitive Neuroscience*, *23*(3), 514–523. https://doi.org/10.1162/jocn.2009.21370

Sassenhagen, J., & Fiebach, C. J. (2019). Finding the P3 in the P600: Decoding shared neural mechanisms of responses to syntactic violations and oddball targets. *NeuroImage*, *200*, 425–436. https://doi.org/10.1016/j.neuroimage.2019.06.048

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, *137*, 29–39. https://doi.org/10.1016/j.bandl.2014.07.010

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*, 379–423.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-Potential Correlates of Stimulus Uncertainty. *Science*, *150*(3700), 1187–1188. https://doi.org/10.1126/science.150.3700.1187

Tanner, D. (2019). Robust neurocognitive individual differences in grammatical agreement processing: A latent variable approach. *Cortex*, *111*, 210–237. https://doi.org/10.1016/j.cortex.2018.10.011

van de Meerendonk, N., Indefrey, P., Chwilla, D. J., & Kolk, H. (2011). Monitoring in language perception: Electrophysiological and hemodynamic responses to spelling violations. *NeuroImage*, *54*(3), 2350–2363. https://doi.org/10.1016/j.neuroimage.2010.10.022

van Herten, M., Kolk, H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, *22*(2), 241–255. https://doi.org/10.1016/j.cogbrainres.2004.09.002

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

Vissers, C. Th. W. M., Chwilla, D. J., & Kolk, H. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, *1106*(1), 150–163. https://doi.org/10.1016/j.brainres.2006.05.012

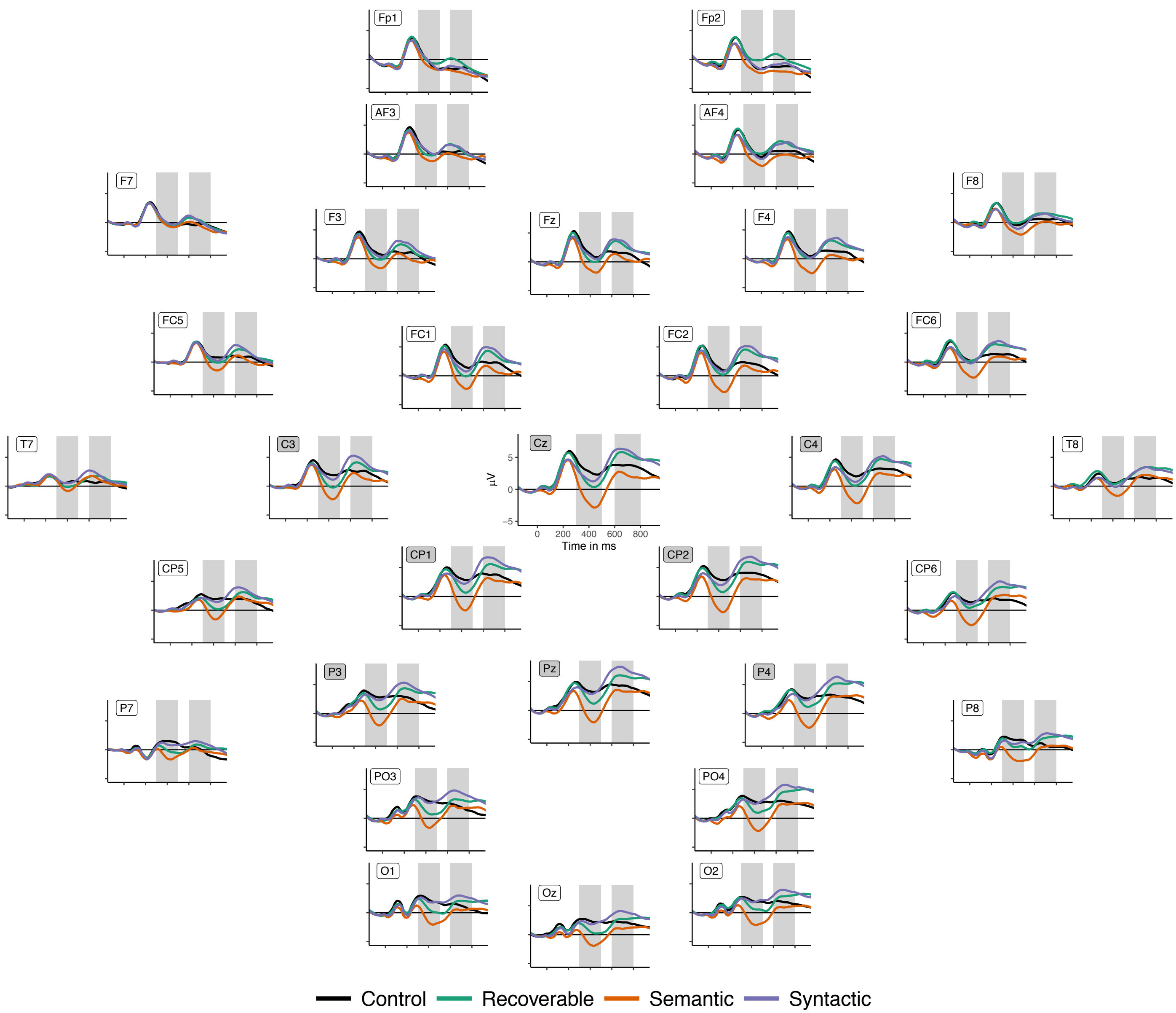
Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender Agreement in Spanish Sentence Reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288. https://doi.org/10.1162/0898929041920487

Yamada, Y., & Neville, H. J. (2007). An ERP study of syntactic processing in English and nonsense sentences. *Brain Research*, *1130*, 167–180. https://doi.org/10.1016/j.brainres.2006.10.052

Zhou, P., Garnsey, S., & Christianson, K. (2019). Is imagining a voice like listening to it? Evidence from ERPs. *Cognition*, *182*, 227–241. https://doi.org/10.1016/j.cognition.2018.10.014

Appendix

Figure A1. Grand average ERPs for each condition at every recorded electrode. The x-axis shows time from the onset of the presentation of the final word, and the y-axis shows loess-smoothed (span=0.2) voltage (negative plotted down), as compared to the mean voltage of the baseline 100 ms pre-stimulus interval. (The subset of channels used in the statistical analyses is indicated by the gray labels and the two gray rectangles in each plot indicate the time windows of interest: 300-500ms and 600-800ms.)


Fp1
Fp2
AF3
AF4
F7
F3
Fz
F4
F8
FC5
FC1
FC2
FC6
T7
C3
Cz
C4
T8
µV
5
0
−5
0
200
400
600
800
Time in ms
CP5
CP1
CP2
CP6
P7
P3
Pz
P4
P8
PO3
PO4
O1
Oz
O2
Control
Recoverable
Semantic
Syntactic