# The P600 ERP component as an index of rational error correction within a noisy-channel framework of human communication

Short title: P600 indexes rational error correction

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

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

Corresponding authors:
Rachel Ryskin: ryskin@mit.edu, Massachusetts Institute of Technology, 43 Vassar Street 46-3037, Cambridge, MA 02139

* The first two authors contributed equally.

**Abstract**

Models that aim to explain electrophysiological responses to linguistic processing (e.g., the N400 and P600 components) typically assume that the representation of linguistic input is error-free. Recent evidence suggests that language processing mechanisms are well-adapted to input corrupted by noise (e.g., speech errors, mishearing) and readily correct for it via rational inference over possible intended sentences and probable noise corruptions. We hypothesize that a P600 ensues whenever the probability that the message was corrupted by noise exceeds the probability that it was produced intentionally and perceived accurately. We show that semantic violations that are attributable to noise—for example, in "The storyteller could turn any incident into an amusing antidote", the implausible word "antidote" is orthographically and phonologically close to the intended "anecdote"—elicit a P600 (rather than N400). Further, the magnitude of this P600 is shown to relate to the probability that the comprehender will retrieve a plausible alternative.

Two event-related potential (ERP) components have been consistently linked to sentence comprehension in electrophysiological investigations of language processing. The N400—a negativity peaking 400ms after word onset—is hypothesized to index the ease of semantic retrieval (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, Hansen, & McClelland, 2018; cf. Cheyette & Plaut, 2017).

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, Brown, & Groothusen, 1993). However, this interpretation has faced numerous challenges. First, the latency of the effect may be inconsistent with evidence that syntactic analysis is fast (McElree & Griffith, 1998). Second, 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). And third, 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, Kolk, & Chwilla, 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, Fitz, & Hoeks, 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, Vosse, Van Den Brink, & Hagoort, 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, King, & Kutas, 1998; Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 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, Braren, Zubin, & John, 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 & Chwilla, 2007; Kolk, Chwilla, van Herten, & Oor, 2003; 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 into the unfolding utterance (conceptually similar to the N400 in Rabovsky et al.'s model). And Fitz and Chang (2019) model the P600 as the prediction error at the sequencing layer of a neural network.

Most of these accounts of the P600 share an assumption that the *received* linguistic input is treated as the *intended* linguistic input. A recent key insight in the sentence processing literature is that the input to our comprehension system is often noisy (Levy, 2008; see also Ferreira & Patson, 2007). This noise stems from a) production errors (speech errors, typos, etc.), and b) perception errors (due to sub-optimal listening/viewing conditions, noise in the environment, etc.). However, communication typically proceeds smoothly, suggesting 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 $s_i$, $P(s_p \mid s_i)$. Indeed, behavioral and eye-tracking studies suggest that readers often take the meaning of a sentence to differ from that of the literal string when that literal string has low prior probability, $P(s_i)$, and/or the potential noise corruption that might have generated that string has high probability, $P(s_p \mid s_i)$ (Gibson, Bergen, & Piantadosi, 2013; Levy, Bicknell, Slattery, & Rayner, 2009).

In line with the error-monitoring perspective (van de Meerendonk et al., 2011), we propose that the P600 may reflect the *noise correction process* in this "noisy-channel" framework. When 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. In contrast, if the input is anomalous but unlikely to have been an error, no P600 ensues, and an N400 might be observed. Critically, though multiple P600 accounts propose a role for computing the mismatch between the expected and received inputs, we argue that the *noise model*—the relative probabilities of different noise corruptions—plays an important role in the noise correction process and will affect the magnitude of the P600 (i.e., all else being equal, the P600 is expected to be smaller/less likely if the noise corruption is implausible).

More precisely (Eq.1), given (i) a preceding sentence context and its most probable parse, $C$[1]; (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 larger P600 noise correction signal whenever $P(s_i{=}s_{received} \mid s_p{=}s_{received})$ is lower than $P(s_i{=}s_{expected} \mid s_p{=}s_{received})$:

$$P600\ signal \propto \frac{P(s_i = s_{expected} \mid s_p = s_{received})}{P(s_i = s_{received} \mid s_p = s_{received})} = \frac{P(s_p = s_{received} \mid s_i = s_{expected})P(s_i = s_{expected})}{P(s_p = s_{received} \mid s_i = s_{received})P(s_i = s_{received})} \quad (1)$$

This *noisy-channel account* can straightforwardly explain a wide range of empirical phenomena. First, a P600 occurs for the "traditional" syntactic violations (number/gender/case agreement errors) and for other minor deviations from the target utterance (e.g., spelling errors), 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

[1] 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.

mow"), is low, while 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".

Second and similarly, this account explains "semantic P600s" because a close alternative exists that the producer plausibly intended.

Third, no P600 occurs for "traditional" semantic violations[2] because no noise corruption is plausible in those cases (e.g., P($s_p$="I take my coffee with cream and dog"|$s_i$="I take my coffee with cream and sugar") is low).

Fourth, a much reduced or no P600 is observed for syntactic errors in "Jabberwocky" sentences, i.e., sentences that include function words/morphemes but cannot be interpreted with respect to world knowledge (Münte, Matzke, & Johannes, 1997; Yamada & Neville, 2007). In such cases, it is difficult to infer plausibly intended meanings because the materials are, by design, devoid of meaning.

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 production errors in bilingual speech, elicit a P600 (Moreno, Federmeier, & Kutas, 2002).

Here, we directly evaluate this proposal for the P600 component using an experimental design with four conditions (Table 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 was semantically inappropriate but orthographically and phonologically close to a semantically plausible neighbor. The proximity of such a neighbor makes the plausibly intended word recoverable. As a result, the critical condition is expected to elicit a P600 (similar to the syntactic condition and in contrast to the semantic condition).

[2] In some studies, a P600 is reported after an N400 for canonical semantic violations (see Brouwer, Fitz, & Hoeks, 2012; Van Petten & Luka, 2012). However, this typically occurs when an unnatural secondary task (e.g., acceptability judgments) is included. Kolk et al. (2003) directly compared judgment and comprehension tasks and found a P600 in the semantic violation cases only for the former. When the task is to find errors, participants plausibly engage their error-correction mechanism for diverse anomalies. The current proposal focuses on more communicative scenarios where comprehenders are simply trying to interpret the linguistic signal (e.g., passive reading, answering comprehension questions).

| The storyteller could turn any incident into an amusing … | | | |
|---|---|---|---|
| **Completion** | **Condition** | **N400 prediction** | **P600 prediction** |
| anecdote | Control | - | - |
| hearse | Semantic violation | + | - |
| anecdotes | Syntactic violation | - | + |
| antidote | Critical | +/- | + |

**Table 1.** Example materials and predictions.

## Methods

### Participants

Twenty-nine right-handed native English speakers participated in this study, 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.

### Materials

160 10-word-long items were constructed (with 4 conditions each, as described above) 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 were only different in the number feature between these conditions and the syntactic violation condition). Materials were normed 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 (see OSF repository for details: https://osf.io/vcsfb/?view_only=ba0079719cfa4118be5cc99714135acf). In addition, 320 10-word-long filler items were constructed. These contained no semantic or syntactic violations.

### 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%).

**Testing 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). 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, Phillips, Kazanina, & Poeppel, 2010; Kutas & Federmeier, 2011) and to be equal in duration with a 100ms gap in between to reduce dependence between the windows.
For each of the two time windows of interest (300-500ms and 600-800ms), the mean amplitude was entered as the dependent variable in a linear mixed-effects regression model, with condition (control, semantic violation, syntactic violation, critical) 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 an ERP effect of +/- 5μV is fairly common. 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**.

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 critical condition relative to the control condition but not in the syntactic condition (Estimate = -0.48, 95% CI = [-1.66, 0.72]). (An N400 effect is expected for the critical condition target word because it is not strongly facilitated by the semantic context, unlike the control condition target word.) 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 critical condition (Estimate = 1.34, 95% CI = [0.11, 2.52]). In other words, as predicted by the rational error correction account, the critical 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 1 and 2 for summaries and https://osf.io/vcsfb/?view_only=ba0079719cfa4118be5cc99714135acf for full model estimates.

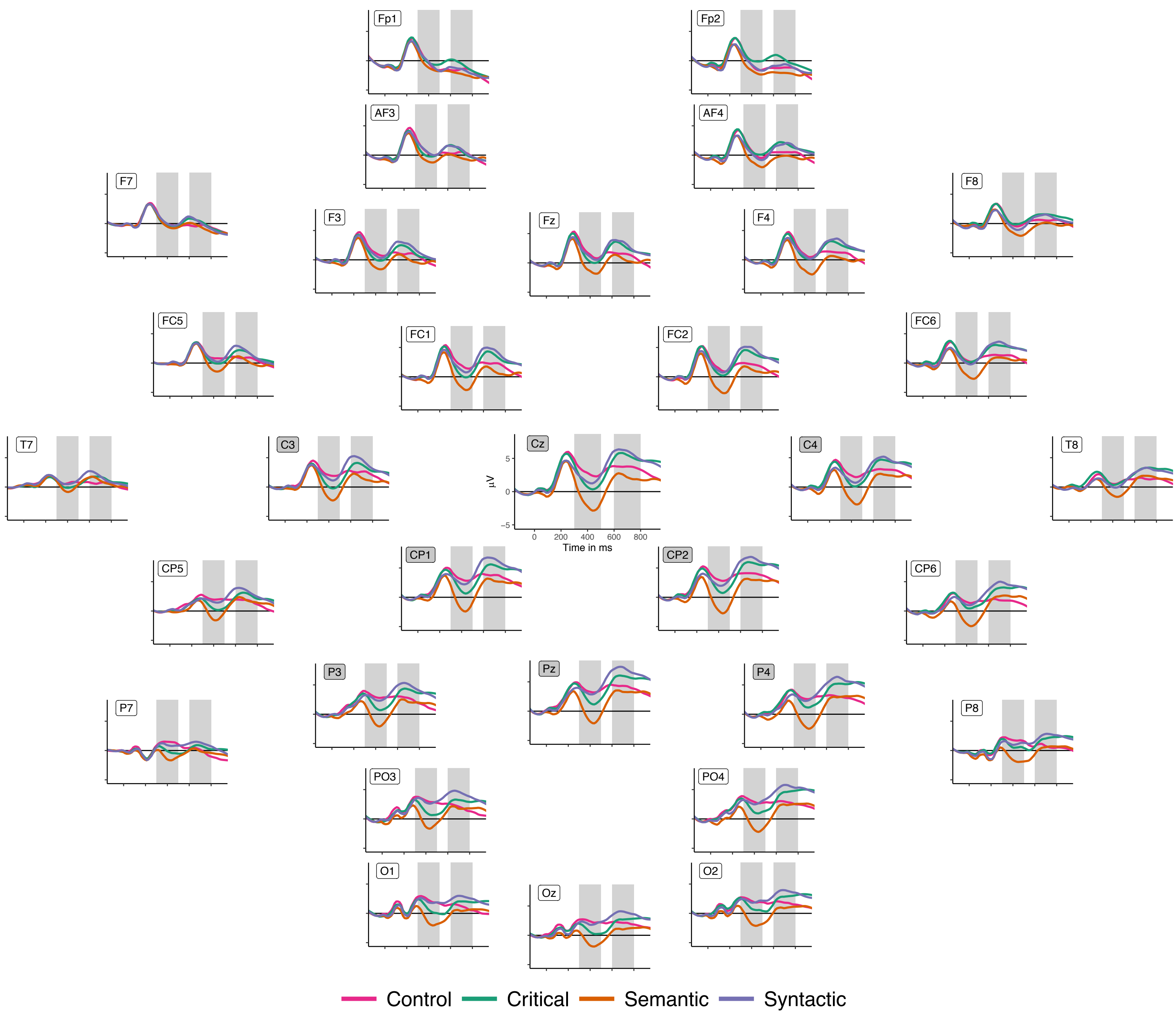
Fp1
Fp2
AF3
AF4
F7
F3
Fz
F4
F8
FC5
FC1
FC2
FC6
T7
C3
Cz
C4
T8
μV
5
0
−5
0
200
400
600
800
Time in ms
CP5
CP1
CP2
CP6
P7
P3
Pz
P4
P8
PO3
PO4
O1
Oz
O2
Control
Critical
Semantic
Syntactic

**Figure 1.** 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 voltage, 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.)

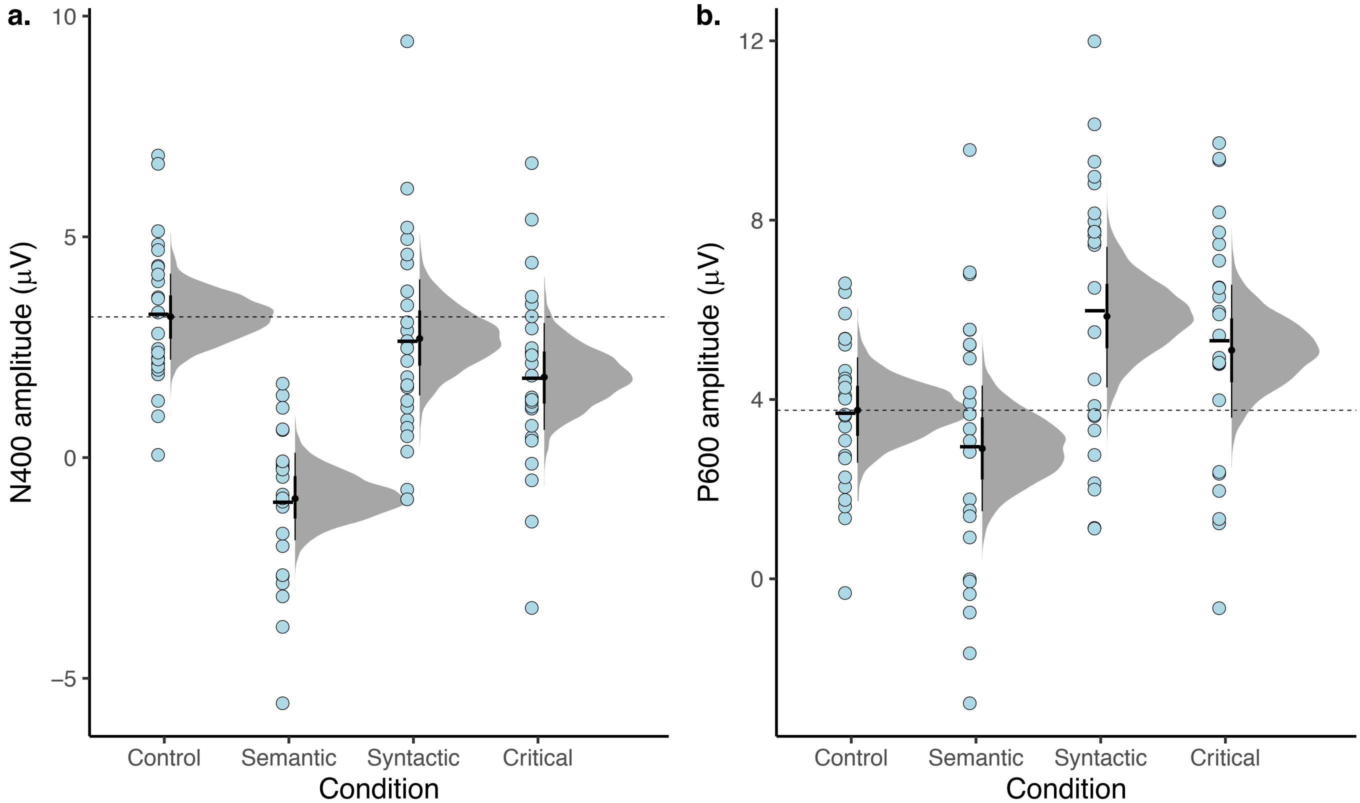


**Figure 2.** Mean amplitudes of (a) the N400 and (b) P600 components. Light blue points represent individual participant means and the black horizontal bar represents the overall mean for each condition. Densities and point intervals represent the distribution of fitted conditional means from Bayesian linear mixed-effects model posteriors. Dashed line indicates the mean amplitude in the control condition.

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

To further explore these effects, we assessed whether the magnitude of the P600 is 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 P600 effect for an item (averaging over participants and electrodes and subtracting the P600 amplitude for the control condition from the amplitudes in the other three conditions) and the two measures of recoverability are shown in Figure 3. Three simple linear regression models were fitted using brms, with the same priors as in the above models where applicable (see further analysis details at https://osf.io/vcsfb/?view_only=ba0079719cfa4118be5cc99714135acf). Items with a larger Levenshtein distance from their control version were less likely to elicit successful recovery of the 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 with a larger Levenshtein distance from their control also elicited smaller P600 effects (Estimate = -0.42, 95% CI = [-0.58,-0.26]). Similarly, items for which participants were more likely to recover the control elicited larger P600 effects (Estimate = 3.29, 95% CI = [1.76, 4.85]). 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).

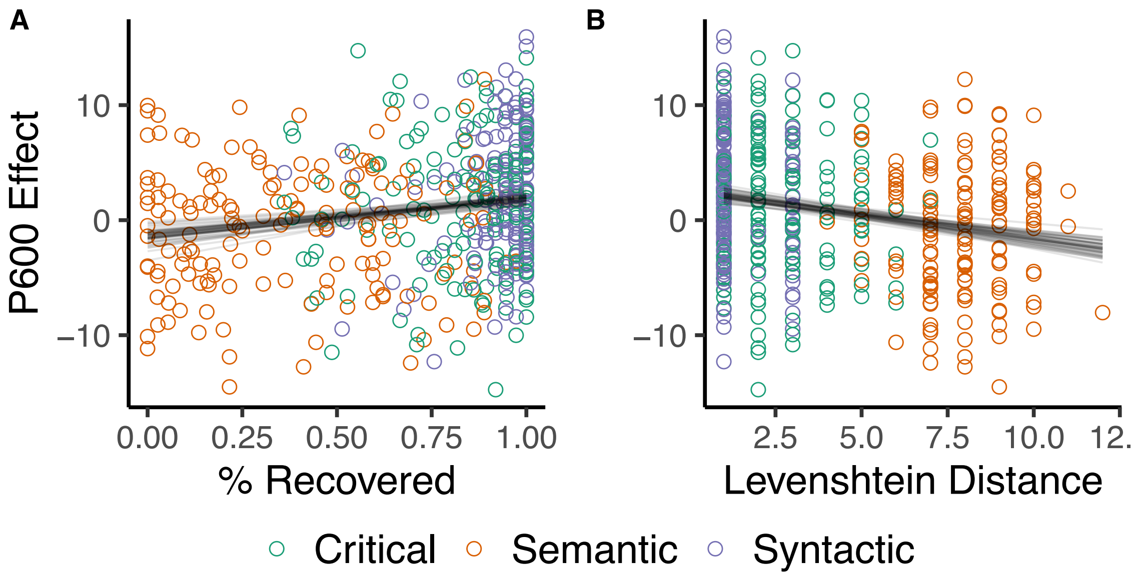


**Figure 3.** Relationships between the average P600 effect for each item in each experimental condition (after subtraction of P600 amplitude in the Control condition) and two measures of recoverability: Percent of correctly recovered completions (A) and Levenshtein distance (B). Gray lines represent 50 fitted regression lines (randomly sampled from model posteriors).

## Discussion

As predicted by the rational error-correction proposal, we observed a P600 when participants read sentences where the target word was semantically inappropriate but had an appropriate orthographic and phonological neighbor, allowing for the possibility that the received message was corrupted by noise. 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. No P600 was observed for the canonical semantic violation, where the intended meaning could not be recovered. Further, the size of the P600 was linearly related to the likelihood of recovering the plausible alternative, suggesting that this component could be leveraged to probe the reader's implicit inferences about the noise process.

In contrast to other recent accounts of the P600 (e.g., Brouwer et al., 2017; Fitz & Chang, 2019), the noisy-channel framework abstracts over questions of how the comprehension mechanisms compute sentence meanings or learn their relative probabilities. The focus is on the *probability* of an intended sentence given the perceived form, no matter how the most likely representations for an input string might have been computed. Nonetheless, the robustness of the human comprehension system to noise is largely compatible with these other accounts and suggests that existing models may improve their fit to human data by training on data with (plausible) noise.

A key prediction of the current account is that the 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, Futrell, Kiran, & Gibson, 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. Future work is needed to provide a systematic test of the quantity and nature of input that will shift the noise model and, consequently, the P600.

To conclude, the rational error correction account of the P600 within a noisy-channel framework i) explains a wide range of prior empirical evidence, ii) is supported by evidence from a new ERP experiment designed to directly evaluate the account, and iii) makes testable predictions for any new scenario: a P600 is predicted whenever the received input can be explained as a perceptual or production error. This finding contributes to a growing literature suggesting that the human language system is well-adapted to potential corruption of the linguistic signal and opens the door to investigation of the comprehender's implicit noise model.

## References

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., … Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

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

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.

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

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

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

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

Gibson, E., Bergen, L., & Piantadosi, S. T. (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

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

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

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

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

McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 432–460. http://dx.doi.org.libproxy.mit.edu/10.1037/0278-7393.24.2.432

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., & 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

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

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

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

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

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

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

## Acknowledgments

We thank Roger Levy, Peter Hagoort, Gina Kuperberg, 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 the draft of the manuscript. This work was supported by National Science Foundation Grant 0844472 from the Linguistics Program (to EG) and by the K99/R00 grant from NICHD (to EF).

## Author contributions

EF, and TG developed the study concept and design. Testing and data collection were performed by LS, ME, and TG. RR, LS, ME, and LB performed the data analysis and interpretation under the supervision of EF and TG. RR and LS drafted the manuscript, and EF and TG provided critical revisions. All authors approved the final version of the manuscript for submission.