Expecting the unexpected: A reanalysis of a multi-laboratory study provides evidence for prediction in language

Sophie Jano, Zachariah Cross, Alex Chatburn, Matthias Schlesewsky, Ina Bornkessel-Schlesewsky

University of South Australia

Words: 4906 (excluding methods, figure and table legends, and references)

Abstract

The present study aimed to investigate the extent of prediction in language, by reanalysing Nieuwland and colleagues' (2018) replication of DeLong et al. (2005). Participants (n = 356) viewed sentences containing articles and nouns of varying predictability, while their electroencephalogram (EEG) was recorded. We measured pre-stimulus and N400 event-related activity in the Nieuwland et al. (2018) data and calculated lexical surprisal using Generative Pre-trained Transformer-2 (GPT-2), to capture word predictability. Our results demonstrate increases in N400 amplitude as article surprisal increased, supporting DeLong et al.'s (2005) findings. N400 amplitudes following surprising articles were also reduced when prior word surprisal was high, suggesting that surprising information may impair the brain's ability to form precise predictions about upcoming words. These findings have important implications for existing understandings of prediction, as they provide evidence for prediction during language comprehension and support the proposal that prediction is an adaptive and unified mechanism of cognition.

Introduction

The question of how the brain converts sensory information into abstract meaning has led to the proposal of prediction as a core mechanism underlying perception (Bar, 2011; Friston, 2005, 2010; Friston & Kiebel, 2009; Press et al., 2020). This is the major assumption of the predictive coding theory, which suggests that prior experiences are used to form generative models (Friston, 2005; Friston & Kiebel, 2009). These models are applied to predict incoming sensory inputs, thus reducing processing effort and facilitating environmental adaptation (Clark, 2013; Friston, 2010; Hohwy, 2013). Although prediction is largely accepted as an explanatory mechanism within numerous cognitive domains (Apps & Tsakiris, 2014; Horga et al., 2014; Koster-Hale & Saxe, 2013), a longstanding debate remains within the cognitive neuroscience of language literature between prediction-based and integrationist theories of language processing (Ferreira & Chantavarin, 2018). Prediction theories propose that the brain anticipates upcoming linguistic content in a top-down manner (Kuperberg & Jaeger, 2016; Mantegna et al., 2019), whilst integrationist theories argue that language comprehension proceeds via the bottom-up integration of linguistic units with the prior context (Mantegna et al., 2019; Van Petten & Luka, 2012). Although seemingly specific to language, this debate could have important implications for existing understandings of prediction as a ubiquitous cognitive mechanism (Friston, 2010), particularly if linguistic prediction does not occur. As such, clarifying the role of prediction in language processing could provide insight into whether predictive processing is indeed fundamental to general cognition as has been argued (Friston, 2010).

As prediction and integration in language are notoriously difficult to disentangle empirically, the debate surrounding their role in linguistic processing has hinged prominently on an influential study by DeLong et al. (2005). DeLong and colleagues investigated the N400 event-related potential (ERP), which has been linked to prediction errors in language (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2019; Hodapp & Rabovsky, 2021; Kuperberg & Jaeger, 2016). By the nature of its design, their experiment sought to exclude an integration-based explanation of purported prediction effects, by capitalising on the English regularity whereby the indefinite articles 'a' and 'an' precede words beginning with a consonant, and beginning with a vowel, respectively (e.g., "a kite" or "an airplane"). DeLong et al. (2005) observed larger N400 amplitudes time-locked to the article preceding unexpected (vs expected) nouns (e.g., "The day was breezy, so the boy went outside to fly *an* …"; the article "an" rules out the expected continuation "kite"). As the indefinite articles 'a/an' have the same meaning, this observation is difficult to reconcile with integrationist views of the N400, which postulate that N400 differences reflect the ease with which a word's meaning can be integrated with its preceding context (Van Berkum et al., 1999; Brown & Hagoort, 1993; Chwilla & Kolk, 2000; Holcomb, 1993). Moreover, the N400 effect was graded in accordance with how predictable the article word was, suggesting that language comprehension involves the prediction and pre-activation of words down to the highly detailed level of speech sounds. This assumption is highly compatible with the assumed prediction of expected sensory input in predictive coding approaches (Rao & Ballard, 1999).

The investigation by DeLong and colleagues (2005) suggests that prediction serves as a likely mechanism for linguistic processing. This is supported by research which has observed similar ERP effects time-locked to adjectives (Van Berkum et al., 2005) and gender-marked determiners (Wicha et al., 2004) preceding a noun. However, other studies have observed differential patterns (for a review see Ito et al., 2017; Schoknecht et al., 2022), including a large-scale direct replication study by Nieuwland et al. (2018), who analysed data from 356

participants using the same stimuli as in DeLong et al. (2005). In contrast to DeLong and colleagues, Nieuwland and colleagues did not observe a more pronounced N400 following unexpected articles. Accordingly, they suggested that, although the brain may predict upcoming contextual or semantic information, this may not extend to form-based linguistic characteristics such as speech sounds (Nieuwland et al., 2018).

The differences in findings between Nieuwland and colleagues and DeLong and colleagues have previously been attributed to Nieuwland and colleagues' high statistical power (Nieuwland et al., 2018), differential baseline correction procedures, and lack of filler sentences (DeLong et al., 2017; Yan et al., 2017). However, further explanation for the divergence in findings reported in these two landmark studies may arise from an investigation of variability at the item and the individual level, with such variability capturing effects over and above group-level contrasts in previous studies (e.g., Roehm et al., 2013). As predictions occur prior to the input in question and likely depend on the specific sentence context, sources of item variability in the pre-stimulus window may explain differences in predictive processing. Further, intrinsic neural characteristics may be associated with changes in prediction at the inter-individual level (Kurthen et al., 2020); a factor that, in conjunction with pre-stimulus variability, was considered neither by DeLong et al. (2005) nor by Nieuwland et al. (2018). We propose that an analysis of pre-stimulus and inter-subject variability in the Nieuwland et al. (2018) data may provide insight into the depth of predictive processing in language and the ubiquity of prediction more generally.

Pre-stimulus ERPs have been proposed to reflect the transient, time-locked activity associated with prediction (Grisoni et al., 2017, 2021). Such components are generally negative-going deflections with frontal topographies and larger amplitudes preceding predictable information (Grisoni et al., 2021; León-Cabrera et al., 2019). This morphology has been demonstrated by several studies (Grisoni et al; 2017, 2021, León-Cabrera et al; 2017, 2019) that presented participants with predictable and unpredictable sentences. The semantic prediction potential (SPP) exhibited greater amplitudes in the 1000ms interval preceding expected vs unexpected critical words and demonstrated topographical variance in accordance with the meaning of each critical word (Grisoni et al., 2016, 2017, 2019, 2021). This provides evidence not only for prediction but for the pre-activation of the semantic characteristics of a word before it arises (Grisoni et al., 2016, 2017, 2019, 2021). Such correlates also strongly resemble the stimulus preceding negativity (SPN) ERP (León-Cabrera et al., 2019), a component associated with the subjective appraisal of prediction accuracy (Ono et al., 2021), and with feedback anticipation during reinforcement learning (Brunia et al., 2011; Morís et al., 2013). Therefore, while ERP activity in the pre-stimulus window may be associated with prediction, the extent to which such activity indexes prediction for specific stimulus properties or more general predictive processing is unclear. The importance of examining pre-stimulus variability is further highlighted by the somewhat confounded nature of post-stimulus (e.g., N400) activity, where prediction error signals and semantic integration are difficult to tease apart. Although DeLong et al. (2005) mitigated this in their experimental design, an investigation of item variability in the pre-stimulus window may provide an alternative way to capture the extent of prediction in language more directly, thus informing the nature of prediction as a unified perceptual mechanism (Friston, 2010).

In addition to item variability in the pre-stimulus window, sources of inter-subject variability may account for differences in findings between DeLong et al. (2005) and Nieuwland et al. (2018). A potential neurobiological marker of individual differences in prediction is the individual alpha frequency (IAF), the individual peak frequency within the alpha rhythm

(Corcoran et al., 2018; Grandy, Werkle-Bergner, et al., 2013a; Klimesch, 1999; Surwillo, 1961). The link between IAF and prediction primarily stems from IAF's role as a trait-like marker of information processing, whereby those with a high IAF were shown to have faster reaction times (Klimesch et al., 1996; Surwillo, 1961, 1963) and improved discrimination of near-threshold stimuli (Samaha & Postle, 2015). Enhanced information processing in high vs low IAF individuals is proposed to translate to a stronger reliance on bottom-up perceptual information as opposed to a top-down model, resulting in more frequent model updating (Kurthen et al., 2020; however, see Bornkessel-Schlesewsky et al., 2022 for stronger model adaptation effects in low IAF individuals). As such, high IAF individuals may exhibit weaker pre-stimulus ERP correlates, reflecting the attenuation of predictions during processing.

Despite periodic oscillatory components such as IAF traditionally being linked to cognitive processing, the importance of considering the aperiodic noise-like activity (particularly the 1/f slope of the distribution) that underlies the EEG spectra has been recently emphasised in the literature (Cross et al., 2022; Demuru & Fraschini, 2020; Immink et al., 2021; Merkin et al., 2023; Ouyang et al., 2020). Although such activity is often removed from oscillatory analyses to directly probe periodic (i.e., transient oscillatory) patterns (Donoghue et al., 2020), the 1/f component has been shown to predict between-participant variance in information processing speeds (Ouyang et al., 2020; Pathania et al., 2022). Regarding prediction, steeper 1/f slopes have been related to prediction error processing in language (Cross et al., 2022; Dave et al., 2018), and to stronger model adaptation effects (Bornkessel-Schlesewsky et al., 2022). As prediction errors are driven by bottom-up information, those with steeper 1/f slopes may exhibit smaller pre-stimulus ERP amplitudes, reflecting weaker reliance on a top-down prediction. As such, an investigation of IAF and the 1/f slope in the Nieuwland et al. (2018) dataset could capture the individual variability relating to prediction and provide insight into the extent to which prediction facilitates linguistic processing.

Although linguistic accounts may challenge the pervasiveness of prediction in higher-order cognition, an investigation of pre-stimulus ERPs and individual neural factors may shed light upon this longstanding debate, whilst reconciling existing understandings of the role of intrinsic neural activity in cognition. The present investigation reanalysed the Nieuwland et al. (2018) dataset to examine the role of individual neural factors and pre-stimulus neural activity in predictive processing. It was hypothesised that increases in article and noun predictability would be associated with greater pre-stimulus, frontal negative ERP amplitudes, and that this would be modulated by IAF and the 1/f slope.

Methods

Original study

The present study involved a reanalysis of Nieuwland et al. (2018), which included 356 participants (females = 222, mean age = 19.8 years) who were all right-handed native English speakers. Cloze probability (the probability that a word will be chosen as the likely continuation based on the prior sentence context; Taylor, 1953) was used as an index of stimulus predictability. Participants viewed 80 sentences with two cloze probability continuations: a more-expected article/noun combination (e.g., "…the boy went outside to fly *a kite*") and a less-expected article/noun combination (e.g., "…the boy went outside to fly *an airplane*"). Words within each sentence were presented for 200ms sequentially on a computer screen and separated by a 300ms interstimulus interval during which participants viewed a blank screen. Following 25% of sentences, participants answered a yes/no comprehension question. EEG was recorded throughout the study to enable an analysis of the event-related activity in response to articles and nouns. For a detailed description of the study protocol and materials, the reader is directed to Nieuwland et al. (2018).

Participants

Data for 356 participants were originally obtained from the Open Science Framework (OSF) repository (see https://osf.io/eyzaq/). Five participants were removed prior to the BIDS conversion (see *Data processing* section) due to incomplete files, answering questions randomly or due to being a non-native English speaker. As such, the BIDS conversion was completed for 351 participants. Subsequent pre-processing and epoching led to the removal of 18 more participants (ten due to trigger issues that could not be resolved, and eight participants due to excessive signal noise). As such, data for 333 participants were epoched, with four participants being excluded due to a lack of trials.

Seventeen participants were also identified as having an insufficient number of epochs for the calculation of the aperiodic (1/f) slope and IAF (less than 16 trials). However, given that there were sufficient epochs around the critical words, these participants were included in the main ERP analyses, taking the total number of participants to 329 (females = 206, mean age = 22.36 years), and the total number of participants with measures of intrinsic neural activity (e.g., IAF, 1/f slope) to 312. See *Figure 1* for a schematic of participants at each stage of the data analysis pipeline.

Data Processing

Data were first converted to the Brain Imaging Data Structure (BIDS) using the MNE-BIDS package (Appelhoff et al., 2019) in *MNE-python* version 3.8.8 (Gramfort et al., 2013), to account for file structure variations as a result of the differing EEG systems used across the laboratories. As files with the .cnt extension were unsupported by the BIDS package in python (*MNE-BIDS 0.10 Documentation*, 2022), such files were converted to the Brainvision Core Data format (Brain Products Gmbh) prior to the BIDS conversion using MATLAB (MathWorks, Natick, Massachusetts, USA). As most raw files encompassed both the experimental session and the control session, each file was then cropped to exclude the control session, as this was not of interest to the present analysis.

Data were then re-sampled to either 250Hz or 256Hz, depending on the original sampling rate, which ranged from 250Hz to 2048Hz between laboratories. Resampling occurred at an equal interval to avoid data loss when epoching (Gramfort et al., 2013), hence the two sampling frequencies. Therefore, the final sample consisted of 184 participants with data sampled at

250Hz, and 145 participants with sampling rates of 256Hz. This is not expected to have affected the single-trial ERP analyses, as samples were averaged across time intervals of interest and slight discrepancies in the number of samples per window should not introduce significant variability. Following resampling, the EEG data were re-referenced to linked mastoids, while EOGs were set as Fp1 and Fp2 across laboratories. For laboratories that used the alphabetized Biosemi system (BioSemi, 2019), EOG channels were those closest to Fp1 and Fp2 (i.e., C29 and C16, respectively). The data were filtered using a bandpass filter that ranged from 0.1 to 30Hz.

To conduct an independent component analysis (ICA), temporary epochs around events of interest were first generated. Rejection thresholds were then computed from the epochs using the *Autoreject* package in MNE-python (Jas et al, 2016), to be included in the ICA for the detection and removal of ocular artefacts. This procedure is argued to lead to more precision in the ICA, as thresholds are tailored specifically depending on the activity in segments of interest (Jas et al., 2016). The ICA was then fit to a 1Hz high pass filtered copy of the temporary epochs, to minimise the influence of noise from the inter-trial intervals. Subsequently, the ICA was applied to the raw data (See *Table 1* for average number of trials per condition). Following this, channels marked as bad were interpolated using spherical splines and the raw data were epoched around events of interest into segments ranging from -800ms to 1000ms (1.8 seconds total). The *Autoreject* function (Jas et al, 2016) was then applied again to repair or remove bad epochs.

ERP analysis

Event-related activity was extracted within the pre-stimulus and N400 time windows on a trial-by-trial basis. The pre-stimulus window ranged from -300 to 0ms prior to the article and noun words, to encompass the length of the interstimulus intervals. The N400 time window ranged from 200 to 500ms post-stimulus onset, in line with Nieuwland et al. (2018). To control for differences in neural activity unrelated to the experimental manipulation, epochs were also generated around the word preceding the article. The inter-stimulus interval (-300 to 0ms) preceding the pre-article word was chosen as a baseline period due to its temporal proximity to the trials of interest, and because the pre-stimulus activity preceding the articles and nouns could not be used since it was the focus of the present investigation. This pre-article "baseline" window was not used for traditional baseline correction, but rather the mean amplitude across this window was entered as a predictor into the statistical models of pre-stimulus effects involving both the article and the noun that followed (in line with Alday, 2019). For N400 analyses, the pre-stimulus activity itself was used as the "baseline" period that was similarly included as a predictor for statistical modelling.

To generate ERPs, *MNE-Python* (Gramfort et al., 2013) was used to calculate the mean ERP amplitude across time windows of interest. Due to differing EEG systems, one laboratory lacked channels that directly corresponded to a common 10-20 channel shared by the other laboratories. As such, a region-based approach was taken whereby pre-stimulus event-related activity was analysed over frontal regions, in line with previous research (León-Cabrera et al., 2017, 2019; Morís et al., 2013). For N400 analyses, we included the channels that were used in the original study (Cz, C3, C4, Pz, P3, and P4). However, the laboratory that used the alphabetized system did not have channels that directly corresponded to P3 and P4, and thus the closest channels were used (A7 and B4, respectively). ERP grand average plots were created using *ggplot2* in *R version 4.1.1* (Wickham, 2016). Due to the differing sampling intervals, the timing of the samples for four laboratories was adjusted by a maximum of four milliseconds, to match the timing of the other laboratories. This was done for visualisation

purposes only. Additionally, for ease of visualisation, channel-specific ERP plots include data from the alphabetized Biosemi channel that was closest to the shared 10-20 channel nomenclature.

Individual alpha frequency and 1/f exponent estimation

Individual alpha frequency (IAF) and the 1/f exponent were calculated using the *Fitting Oscillations & one-over f (FOOOF)* package (version 1.0.0) in *MNE-python* (Gramfort et al., 2013). Epochs were first generated around the between sentence intervals as during this time, participants should not have been actively engaged in the task (in exception is the 25% of trials that included a comprehension question, which were removed), and thus task-related effects were likely reduced. Epochs ranged from 200ms after the final word (when the presentation of the final word had ended), to 2.2 seconds, to ensure a sufficient number of samples to obtain reliable estimates. As the between-sentence intervals were self-paced, trials in which the intervals were shorter than 2.2 seconds were removed from the analysis. Similarly to the treatment of task-related epochs, the *Autoreject* function (Jas et al, 2017) was then applied to the between-sentence epochs to repair or remove bad trials.

Following epoching, the power spectral density was computed for each epoch across all EEG channels (Donoghue et al., 2020). Channels Oz, O1, and O2 were selected, as alpha activity is prominent over occipital regions (Kirschfeld, 2005), and the *FOOOF* algorithm was fit to PSD estimates derived from each channel, per epoch. For the laboratory with the alphabetized system, the channels closest to Oz, O1 and O2 were used (A23, A15 and A28, respectively). Parameterization of the power spectra occurred across a frequency range of 1-30Hz. Settings for the *FOOOF* algorithm included a peak width limit of 2 to 8, a maximum peak number of four, the minimum peak height as 0.0, a peak threshold of 2.0, and a fixed aperiodic mode. To obtain individual metrics, activity across the three channels was averaged across all epochs, resulting in single-subject slope and offset values. IAF estimates were calculated by selecting peaks within the 7-13Hz range and averaging over the activity across each channel and epoch to obtain a single IAF value per participant.

Statistical analysis

To examine effects of interest, four linear mixed-effects models (LMMs) were computed using the *Lme4* package in *R version 4.1.2* (Bates et al., 2015). P-values were calculated using the *summary* function from the *lmerTest* package, which is based on Satterthwaite's degrees of freedom (Kuznetsova et al., 2017). The distributions of the data and of the model residuals were plotted using the *ggplot2* package (Wickham, 2016), to ensure a normal distribution of the outcome measures, and to confirm that model assumptions were met. Model effects were also plotted using *ggplot2* (Wickham, 2016), with 83% confidence intervals to enable visual identification of statistical effects. For plots in which variables are presented dichotomously (e.g., high/low, flat/steep), this distinction was categorised according to the first and third quartiles.

When investigating the effect of word predictability (operationalized as cloze probability) on pre-stimulus ERPs time-locked to articles (Model 1), cloze probability was contrasted between-sentences, as opposed to between expected and unexpected sentence conditions. This is because, prior to the article, the sentence contexts were identical. The likely continuation (the expected article cloze value for that sentence) was taken as the cloze metric for both versions of the sentence (See *Table 2* for an example). In contrast, when measuring pre-stimulus effects time-locked to the nouns, the original noun cloze values for both expected and unexpected

continuations were applied because the sentences arguably differed once participants had been exposed to the article variation.

Statistical model structures are presented in *Table 3*. Note that IAF and the 1/f slope are included in the same models, specifying different interaction terms, as the relationship between the two factors was not of interest. These combined models also exhibited either significantly better model fit, or the model fit did not significantly differ from separate models for IAF and 1/f, as measured by the Akaike criterion (See *supplementary material A* for Akaike criterion values).

Results

Statistical model outputs are presented in the supplementary materials.

Pre-stimulus ERP effects

Pre-stimulus ERPs time-locked to articles are shown in *Figure 2,* and ERP effects for nouns are shown in *Figure 3.* Model 1 addressed the hypothesis that as cloze probability increases, pre-stimulus ERP amplitude will increase, and that this will be modulated by IAF or 1/f slope, for articles specifically. The effect of article cloze probability and IAF on pre-stimulus ERP amplitude was nonsignificant ($p = 0.395$, $\beta = -0.0004$, $SE = 0.0005$). The effect of article cloze and 1/f slope was also nonsignificant ($p = 0.136$, $\beta = -0.0008$, $SE = 0.0005$). Therefore, there were no detectable effects of cloze probability and individual factors on pre-stimulus ERP activity following the articles.

Model 2 addressed the hypothesis that increases in noun cloze probability would lead to greater pre-stimulus ERP amplitudes, and that this would be modulated by IAF or the aperiodic 1/f slope. The model demonstrated a significant two-way interaction between cloze and IAF on pre-stimulus ERP activity ($p < .001$, $\beta = -0.001$, $SE = 0.0002$), whereby pre-stimulus ERP amplitude decreased as cloze increased, with this pattern being strongest for those with a low IAF as compared to those with a high IAF (See *Figure 4*). The effect of 1/f slope and cloze probability on pre-stimulus ERP amplitude was nonsignificant ($p = 0.387$, $\beta = 0.0002$, $SE = 0.0002$).

N400 ERP effects

N400 grand average ERPs time-locked to articles are presented in *Figure 5.* Model 3 was fitted to examine the effects of cloze probability, IAF and the 1/f slope on N400 amplitude time-locked to articles. The model revealed a significant effect of cloze probability on N400 amplitude ($p < .001$, $\beta = -0.003$, $SE = 0.0004$), such that as cloze probability increased, N400 amplitude increased.

The effect of cloze probability, IAF and the 1/f slope on N400 amplitude time-locked to nouns was explored in Model 4, which revealed a significant two-way interaction between 1/f slope and cloze probability on N400 amplitude ($p < .001$, $\beta = 0.002$, $SE = 0.0004$), whereby those with steeper 1/f slopes exhibited greater N400 amplitudes to low-cloze nouns than those with flatter 1/f slopes, whilst the effect was reversed following high-cloze nouns (See *Figure 8a*). A significant interaction between IAF and cloze probability on N400 amplitude was also observed ($p = 0.020$, $\beta = -0.0009$, $SE = 0.0004$), such that high IAF individuals exhibited lower N400 amplitudes than those with a low IAF, with this difference appearing greatest following low-cloze nouns (See *Figure 8b*). Therefore, whilst cloze probability influenced N400

amplitudes following the articles, individual factors interacted with cloze to affect N400 amplitudes to nouns.

Exploratory N400 surprisal analyses

Early ERP grand average plots revealed event-related differences according to between-sentence cloze probability at earlier stages of the sentence (See *Figure 3*). This sparked interest regarding the accumulation of predictions over time, and of how the predictability of the prior words in the sentence may influence later predictions. Two additional exploratory LMMs were included to analyse whether the surprisal of the preceding words interacted with the surprisal of the articles and nouns respectively, to influence N400 amplitude. The raw surprisal values of the two preceding words (article-2 and article-1 in the case of article analyses, and article-1 and the article in the case of noun analyses), and of the critical articles and nouns, were included in the statistical models. To maintain consistency with Nieuwland and colleagues' original analysis, and as a sanity check of Yan et al. (2017; who also reanalysed the same data with log-transformed cloze as a metric of surprisal), we computed two more LMMs predicting N400 amplitude from critical word surprisal alone (See *Table 4* for model structures). Finally, Pearson correlations were computed to examine the relationship between cloze probability and surprisal for the two respective word types.

Lexical surprisal was calculated using an open artificial intelligence model known as Generative Pre-Trained Transformer-2 (GPT-2; Radford et al., 2019) using the *Transformers* package in Python (Wolf et al., 2020). Surprisal in a linguistic context is calculated based on the probability of a given word in a linguistic corpus or database, whereby a higher surprisal value corresponds to lower predictability of a word within, for example, a sentence (Frank et al., 2013). GPT-2 is trained on a large corpus of English texts to traditionally predict upcoming words in a sentence, hence its efficacy in calculating surprisal (Radford et al., 2019). This measure was specifically used as the original experiment lacked cloze probability values of the prior words (Nieuwland et al., 2018).

Results of N400 models with surprisal

Grand average N400 ERPs according to article surprisal are presented in *Figure 6,* whilst grand averages according to noun surprisal are shown in *Figure 7.* Model 5 examined the effect of article surprisal on N400 amplitude. The effect was significant ($p < .001$, $\beta = -0.048$, $SE = 0.009$), such that increases in article surprisal were associated with greater N400 amplitudes. To extend upon this, Model 6 examined whether the surprisal of the a-1, a-2, and the article words interacted to influence N400 amplitude. A significant three-way interaction effect between a-2 surprisal, a-1 surprisal and article surprisal was revealed ($p = .0004$, $\beta = 0.003$, $SE = 0.0009$). Although the N400 amplitude became larger with increases in article surprisal, high a-2 and a-1 surprisal values led to smaller N400 amplitudes to highly surprising articles (See *Figure 8c*).

To investigate the effect of noun surprisal on N400 amplitude, Model 7 was fitted. A significant main effect of noun surprisal was observed ($p < .001$, $\beta = -0.234$, $SE = 0.006$), such that greater noun surprisal values were associated with larger N400 amplitudes. Model 8 was fitted to measure the effect of a-1, article, and noun surprisal on N400 amplitudes time-locked to nouns. A significant three-way interaction between a-1, article, and noun surprisal was observed ($p = 0.012$, $\beta = 0.003$, $SE = 0.001$). Increases in noun surprisal were associated with larger N400 amplitudes, with low article and a-1 surprisal values leading to smaller N400 amplitudes to highly surprising nouns (See *Figure 8d*).

Correlational models

A Pearson correlation was computed to examine the relationship between article cloze probability and article surprisal, revealing a significant but weak correlation ($r(158) = -.2$, $p = 0.013$), whereby increases in article cloze were associated with decreases in article surprisal. A second Pearson correlation was calculated to determine the relationship between noun surprisal and noun cloze probability. The two variables were significantly negatively correlated ($r(157) = -.58$, $p < .001$), with increases in noun cloze associated with decreases in noun surprisal (see *Figure 9* for correlation plots).

Discussion

The present study sought to investigate the degree to which prediction underlies linguistic processing. Our exploratory analysis revealed larger N400 amplitudes following surprising articles, in support of DeLong et al.'s (2005) suggestion that the brain pre-activates words when interpreting language. N400 amplitudes were also influenced by prior word surprisal, illuminating the complex interplay between predictions and prediction errors across sentences. The role of IAF and the 1/f aperiodic slope in modulating the N400 component suggests that prediction tendencies may differ between subjects according to intrinsic neural factors, which may account for discrepancies in findings between DeLong et al. (2005) and Nieuwland et al. (2018). Broadly, our reanalysis contributes novel insights to the longstanding debate surrounding prediction in language by suggesting that the brain predicts form-based characteristics during linguistic processing.

Article-N400 models with surprisal suggest that prediction underlies language processing
Our article N400 results differed according to word predictability metrics, such that analyses with surprisal revealed a role for prediction in language processing. Initial article models with cloze probability showed increases in N400 amplitude as cloze increased. This result opposes both DeLong et al. (2005), who found the reversed effect, and Nieuwland et al. (2018), whose N400-article analysis did not reach significance (although the trend was detected in some laboratories in Nieuwland and colleagues' replication analysis). Differences in the data analysis pipeline between our analysis and Nieuwland et al.'s (2018) study may have contributed to the differential results. Variations in pre-processing steps and our inclusion of pre-stimulus activity as a model covariate (which is shown to improve statistical power over baseline correction; Alday, 2019), may explain this outcome. Although unexpected, this finding may also stem from the use of cloze probability, which may be limited in its ability to gauge the predictability of function words such as articles.

When surprisal was used to operationalize word predictability as opposed to cloze probability, the N400 effect reversed, such that surprising articles were followed by significantly larger N400 amplitudes than articles with low surprisal. As the articles 'a' and 'an' have the same meaning, the observed N400 difference is unlikely to result from difficulties relating to semantic integration, as has been previously proposed (Van Berkum et al., 1999; Brown & Hagoort, 1993; Chwilla & Kolk, 2000; Holcomb, 1993). Rather, the increased N400 for surprising articles may suggest that participants were predicting not only the upcoming noun but its phonological characteristics, such that a surprising article violated a prediction for the subsequent noun. This finding supports DeLong et al.'s (2005) original study, where articles with low cloze probabilities evoked greater N400 amplitudes than high-cloze articles. It is also compatible with Yan et al.'s (2017) reanalysis of Nieuwland et al.'s (2018) data using the log-transformation of cloze as a measure of surprisal, which revealed the same effect as in our analysis. Our result thus speaks to the discrepancies between surprisal and cloze probability, with the two methods yielding different article predictability outcomes in the current context. Article surprisal and article cloze probability were also weakly correlated, whilst noun cloze and noun surprisal were strongly related, with all noun analyses revealing a consistent effect in the expected direction (DeLong et al., 2005; Nieuwland et al., 2018). This discrepancy between the two metrics for articles may indicate that cloze probability is a limited measure of function word predictability, potentially because such words are not as semantically rich as content words (e.g., nouns). Whilst semantic information accumulates across a sentence to constrain subsequent word choices, early speech sounds do not strongly limit later word possibilities (Huettig et al., 2022). As such, article predictability differences may be more

subtle and difficult to detect using cloze probability (for a discussion of the limitations of cloze, see Lopukhina et al., 2021; Smith & Levy, 2011; Yan et al., 2017). Broadly, the modulation of prediction error signals according to article surprisal suggests that the brain predicts linguistic information at the level of speech sounds but may use semantic cues to strengthen predictions.

Predictions are influenced by the surprisal of the prior words in a sentence

The N400 amplitude was modulated by prior and critical word surprisal, suggesting that predictive processing is influenced by the accumulation of contextual information. High surprisal in the two words preceding the articles was associated with smaller N400 amplitudes to surprising articles. This result is consistent with the proposal that the N400 reflects a precision-weighted prediction error (Bornkessel-Schlesewsky & Schlesewsky, 2019), with precision reflecting the amount of certainty (or variance) in the sensory data and in the prediction (Feldman & Friston, 2010; Friston, 2005). According to Bayesian frameworks, precision influences the relative weightings of the sensory input and the prior belief, such that high sensory precision and low prior precision bias the posterior belief (i.e., the predictive model) toward incorporating the sensory information (Adams et al., 2013). In the present study, when the two prior words were surprising, it may have been difficult to generate a specific prediction for the upcoming word, thus reducing model precision. An imprecise model may be less easily violated by the sensory information, leading to a weaker error signal. This suggests that after exposure to surprising information, the brain may have begun to 'expect the unexpected,' such that it was unable to generate specific predictions for upcoming articles, and N400 amplitudes (i.e., prediction errors) decreased. This may explain how internal language models cope with change as an inherent property of language, as prior uncertainty may inform the extent to which subsequent information is used for model updating. It also highlights the complex interplay between predictions and sensory data, by suggesting that the brain's reliance on either-or depends strongly on prior contextual information. However, further research would be beneficial to understand how changes in early surprisal may directly influence precision, potentially by investigating correlates that index prediction errors for precision weightings themselves (e.g., the P300 ERP, as suggested by Bornkessel-Schlesewsky, Alday, et al., 2022). Nevertheless, this result supports the predictive coding notion that the weighting of one's predictive model over the sensory data emerges as a function of prior learning (Hohwy, 2013).

The 1/f slope and IAF affect prediction error processing

Steeper 1/f slopes may promote the identification of relevant errors for model updating, as steeper slopes led to stronger N400 amplitudes to low-cloze (unexpected) nouns, in contrast to flatter slopes. This finding is consistent with prior research whereby steeper 1/f slopes were associated with greater N400 amplitudes (Dave et al., 2018), and with enhanced model adaptation effects (Bornkessel-Schlesewsky et al., 2022). Coupled with our observation that those with steeper slopes also showed smaller N400 amplitudes to high-cloze (expected) words, such individuals may have developed a more precise predictive model that was attuned towards updating only the most relevant erroneous information. Mechanistically, steeper slopes are associated with increased synchronous neuronal spiking and a shift towards a greater inhibition to excitation ratio within the underlying neuronal populations (Gao et al., 2017; Voytek et al., 2015; Voytek & Knight, 2015). This is thought to translate to lower neural noise and improved neural gain control (i.e., enhanced regulatory activity within the norepinephrine system; Pertermann et al., 2019; Voytek & Knight, 2015). Regarding prediction, increasing the amount of neural gain is proposed to lead to enhanced attention and to improvements in the precision of prediction error signals (Bastos et al., 2012; Feldman & Friston, 2010; Shipp, 2016). This suggests that the heightened neural gain control accompanying steeper 1/f slopes

may promote receptivity to relevant prediction errors via increases in precision, hence the greater N400 amplitudes to low-cloze words observed in the present study. As such, steeper 1/f aperiodic slopes may be linked to optimised error processing, due to their association with low neural noise and better neural gain control.

The observed 1/f findings suggest that individual factors may influence prediction, which is strengthened by the observation that the N400 was modulated by individual differences in processing speed (as measured via IAF; Klimesch et al., 1996; Samaha & Postle, 2015; Surwillo, 1961, 1963). Our analysis revealed greater N400 amplitudes following low-cloze nouns for low versus high IAF individuals. As prediction errors prompt model adaptation (Clark, 2013), this finding supports research by Bornkessel-Schlesewsky et al. (2022) where low IAFs were associated with improved adaptation to surprisal during a language task as compared to high IAFs. As such, low IAF individuals may exhibit stronger model updating tendencies, despite their proposed slower information processing windows (Samaha & Postle, 2015). Although this may appear counterintuitive, this finding may be linked to the nature of linguistic paradigms, which present information at the word level that must be combined to form a broader sentence (Zuidema & de Boer, 2018). When information must be integrated across longer perceptual windows (such as in language), slower processing windows that capture information spanning longer periods may be ideal, hence the improved error signalling in those with a low IAF (Cecere et al., 2015). This is also supported by prior research by Howard et al. (2017), who observed enhanced performance for low IAF individuals on a spatial task requiring integration. In contrast, high IAF individuals exhibit superior performance in tasks that require the discrimination of rapidly-presented stimuli (Samaha & Postle, 2015; Zhang et al., 2021), highlighting the differential benefits of IAF according to task demands. Collectively, the influence of IAF and the 1/f aperiodic slope on the N400 component emphasises the role of individual neural factors in predictive processing and may account for the differences in findings between Nieuwland et al. (2019) and DeLong et al. (2005).

Pre-stimulus ERPs may not index prediction in natural language

Our pre-stimulus analyses did not reveal a significant effect of article cloze probability on pre-stimulus ERP amplitude, whilst pre-stimulus amplitudes preceding the nouns decreased as cloze probability increased. These findings are inconsistent with prior literature, which demonstrates amplitude increases in the semantic prediction potential (SPP) and the stimulus-preceding negativity (SPN) as word predictability increases (Grisoni et al., 2019, 2021; León-Cabrera et al., 2017, 2019; Pulvermüller & Grisoni, 2020). The observed null article finding may be due to the overlap in meaning between the articles ‘a’ and ‘an,’ resulting in a similar semantic prediction between the two. However, the noun effect suggests that both article and noun analyses are likely limited by the temporal overlap between the pre-stimulus and N400 windows following the prior words. Most studies observing SPP/SPN effects employed intervals preceding the critical words of 1000ms or longer to capture activity independent of ERP components following the prior word (Grisoni et al., 2017, 2021; León-Cabrera et al., 2017, 2019), whilst the present study included interstimulus intervals of 300ms. In our analysis, the N400 may have been probed, rather than a separate component, highlighting the difficulty in gauging pre-stimulus correlates with the use of more naturalistic sentence designs. This raises the question of whether the longer intervals in other studies encourage prediction to a greater extent, or whether they give rise to a process that differs from natural language comprehension. Therefore, the more robust nature of the observed N400 effects may suggest that prediction error components have more utility in the analysis of linguistic prediction, with pre-stimulus correlates failing to gauge prediction in a more naturalistic language setting.

Conclusion

Our results strongly suggest that the brain pre-activates information during linguistic processing, supporting the proposal that prediction underlies brain functioning (Friston, 2010). Prediction error signals differed according to critical and prior word surprisal, providing evidence for the word-by-word adaptation of predictive models in language. Further, the effect of IAF and the 1/f aperiodic slope on the N400 component highlights the modulation of predictive processing according to intrinsic neural factors, which may influence prediction error weightings and model updating tendencies. Whilst this study provides insight into the neural components relating to prediction, extending this research toward behavioural outcomes may provide insight into the consequences of such individual differences for learning and information processing. Furthermore, whilst language has a rich contextual architecture that draws largely on previously learnt rules (Zuidema & de Boer, 2018), understanding how predictions accumulate in more linear, as opposed to hierarchical sequences, may contribute to existing perspectives that posit prediction as a unified cognitive mechanism. In summary, the current investigation suggests that the brain may predict language, such that prior experience informs the pre-activation of phonological linguistic characteristics.

References

Adams, R., Stephan, K., Brown, H., Frith, C., & Friston, K. (2013). The Computational Anatomy of Psychosis. *Frontiers in Psychiatry*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyt.2013.00047

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

Appelhoff, S., Sanderson, M., Brooks, T. L., Vliet, M. van, Quentin, R., Holdgraf, C., Chaumon, M., Mikulan, E., Tavabi, K., Höchenberger, R., Welke, D., Brunner, C., Rockhill, A. P., Larson, E., Gramfort, A., & Jas, M. (2019). MNE-BIDS: Organizing electrophysiological data into the BIDS format and facilitating their analysis. *Journal of Open Source Software*, *4*(44), 1896. https://doi.org/10.21105/joss.01896

Apps, M. A. J., & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, *41*, 85–97. https://doi.org/10.1016/j.neubiorev.2013.01.029

Bar, M. (2011). *Predictions in the Brain*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195395518.001.0001

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

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

Bornkessel-Schlesewsky, I., Alday, P. M., Corcoran, A. W., Wilkinson, E. M., Sharrad, I., Kliegl, R., Lewis, R. L., Small, S. L., & Schlesewsky, M. (2022). *Effects of neural noise on predictive model updating across the adult lifespan* [Preprint]. Neuroscience. https://doi.org/10.1101/2022.12.14.520501

Bornkessel-Schlesewsky, I., Sharrad, I., Howlett, C. A., Alday, P. M., Corcoran, A. W., Bellan, V., Wilkinson, E., Kliegl, R., Lewis, R. L., Small, S. L., & Schlesewsky, M. (2022). *Rapid adaptation of predictive models during language comprehension: Aperiodic EEG slope, individual alpha frequency and idea density modulate individual differences in real-time model updating.* https://doi.org/10.3389/fpsyg.2022.817516

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*, *10*(298), 1–17. https://doi.org/10.3389/fpsyg.2019.00298

Brown, C., & Hagoort, P. (1993). The Processing Nature of the N400: Evidence from Masked Priming. *Journal of Cognitive Neuroscience*, *5*(1), 34–44. https://doi.org/10.1162/jocn.1993.5.1.34

Brunia, C. H. M., Hackley, S. A., van Boxtel, G. J. M., Kotani, Y., & Ohgami, Y. (2011). Waiting to perceive: Reward or punishment? *Clinical Neurophysiology*, *122*(5), 858–868. https://doi.org/10.1016/j.clinph.2010.12.039

Cecere, R., Rees, G., & Romei, V. (2015). Individual Differences in Alpha Frequency Drive Crossmodal Illusory Perception. *Current Biology*, *25*(2), 231–235. https://doi.org/10.1016/j.cub.2014.11.034

Chwilla, D. J., & Kolk, H. H. J. (2000). *Mediated Priming in the Lexical Decision Task: Evidence from Event-Related Potentials and Reaction Time*. *42*(3), 314-341. https://doi.org/10.1006/jmla.1999.2680

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Corcoran, A. W., Alday, P. M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2018). Toward a reliable, automated method of individual alpha frequency (IAF) quantification. *Psychophysiology*, *55*(7), e13064. https://doi.org/10.1111/psyp.13064

Cross, Z. R., Corcoran, A. W., Schlesewsky, M., Kohler, M. J., & Bornkessel-Schlesewsky, I. (2022). Oscillatory and Aperiodic Neural Activity Jointly Predict Language Learning. *Journal of Cognitive Neuroscience*, 1–20. https://doi.org/10.1162/jocn_a_01878

Dave, S., Brothers, T. A., & Swaab, T. Y. (2018). 1/f neural noise and electrophysiological indices of contextual prediction in aging. *Brain Research*, *1691*, 34–43. https://doi.org/10.1016/j.brainres.2018.04.007

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

DeLong, K., Urbach, T., & Kutas, M. (2017). *Concerns with Nieuwland et al. Multi-lab study (2017)*. https://doi.org/10.13140/RG.2.2.19318.60486

Demuru, M., & Fraschini, M. (2020). EEG fingerprinting: Subject-specific signature based on the aperiodic component of power spectrum. *Computers in Biology and Medicine*, *120*, 103748. https://doi.org/10.1016/j.compbiomed.2020.103748

Donoghue, T., Dominguez, J., & Voytek, B. (2020). Electrophysiological Frequency Band Ratio Measures Conflate Periodic and Aperiodic Neural Activity. *BioRxiv*, 2020.01.11.900977. https://doi.org/10.1101/2020.01.11.900977

Feldman, H., & Friston, K. J. (2010). Attention, Uncertainty, and Free-Energy. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00215

Ferreira, F., & Chantavarin, S. (2018). Integration and Prediction in Language Processing: A Synthesis of Old and New. *Current Directions in Psychological Science*, *27*(6), 443–448. https://doi.org/10.1177/0963721418794491

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2013). *Word surprisal predicts N400 amplitude during reading*. 7.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), Article 2. https://doi.org/10.1038/nrn2787

Friston, K., & Kiebel, S. (2009). Predictive Coding under the Free-Energy Principle. *Philosophical Transactions: Biological Sciences*, *364*(1521), 1211–1221.

Gao, R., Peterson, E. J., & Voytek, B. (2017). Inferring synaptic excitation/inhibition balance from field potentials. *NeuroImage*, *158*, 70–78. https://doi.org/10.1016/j.neuroimage.2017.06.078

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*(7), 267–267. https://doi.org/10.3389/fnins.2013.00267

Grandy, T. H., Werkle-Bergner, M., Chicherio, C., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2013). Individual alpha peak frequency is related to latent factors of general cognitive abilities. *NeuroImage*, *79*, 10–18. https://doi.org/10.1016/j.neuroimage.2013.04.059

Grisoni, L., Dreyer, F. R., & Pulvermüller, F. (2016). Somatotopic Semantic Priming and Prediction in the Motor System. *Cerebral Cortex*, *26*(5), 2353–2366. https://doi.org/10.1093/cercor/bhw026

Grisoni, L., Miller, T. M., & Pulvermüller, F. (2017). Neural Correlates of Semantic Prediction and Resolution in Sentence Processing. *The Journal of Neuroscience*, *37*(18), 4848–4858. https://doi.org/10.1523/JNEUROSCI.2800-16.2017

Grisoni, L., Mohr, B., & Pulvermüller, F. (2019). Prediction mechanisms in motor and auditory areas and their role in sound perception and language understanding. *NeuroImage*, *199*, 206–216. https://doi.org/10.1016/j.neuroimage.2019.05.071

Grisoni, L., Tomasello, R., & Pulvermüller, F. (2021). Correlated Brain Indexes of Semantic Prediction and Prediction Error: Brain Localization and Category Specificity. *Cerebral Cortex*, *31*(3), 1553–1568. https://doi.org/10.1093/cercor/bhaa308

Hodapp, A., & Rabovsky, M. (2021). The N400 ERP component reflects an error-based implicit learning signal during language comprehension. *European Journal of Neuroscience*, *54*(9), 7125–7140. https://doi.org/10.1111/ejn.15462

Hohwy, J. (2013). *The Predictive Mind*. OUP Oxford.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*(1), 47–61. https://doi.org/10.1111/j.1469-8986.1993.tb03204.x

Horga, G., Schatz, K. C., Abi-Dargham, A., & Peterson, B. S. (2014). Deficits in Predictive Coding Underlie Hallucinations in Schizophrenia. *Journal of Neuroscience*, *34*(24), 8072–8082. https://doi.org/10.1523/JNEUROSCI.0200-14.2014

Howard, C. J., Arnold, C. P. A., & Belmonte, M. K. (2017). Slower resting alpha frequency is associated with superior localisation of moving targets. *Brain and Cognition*, *117*, 97–107. https://doi.org/10.1016/j.bandc.2017.06.008

Huettig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition*, *224*, 105050. https://doi.org/10.1016/j.cognition.2022.105050

Immink, M. A., Cross, Z. R., Chatburn, A., Baumeister, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2021). Resting-state aperiodic neural dynamics predict individual differences in visuomotor performance and learning. *Human Movement Science*, *78*, 102829. https://doi.org/10.1016/j.humov.2021.102829

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). Why the A/AN prediction effect may be hard to replicate: A rebuttal to Delong, Urbach, and Kutas (2017). *Language, Cognition and Neuroscience*, *32*(8), 974–983. https://doi.org/10.1080/23273798.2017.1323112

Jas, M., Engemann, D., Raimondo, F., Bekhti, Y., & Gramfort, A. (2016). Automated rejection and repair of bad trials in MEG/EEG. *6th International Workshop on Pattern Recognition in Neuroimaging (PRNI)*. https://hal.archives-ouvertes.fr/hal-01313458

Kirschfeld, K. (2005). The physical basis of alpha waves in the electroencephalogram and the origin of the "Berger effect". *Biological Cybernetics*, *92*(3), 177–185. https://doi.org/10.1007/s00422-005-0547-1

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research. Brain Research Reviews*, *29*(2–3), 169–195. https://doi.org/10.1016/s0165-0173(98)00056-3

Klimesch, W., Doppelmayr, M., Schimke, H., & Pachinger, T. (1996). Alpha Frequency, Reaction Time, and the Speed of Processing Information. *Journal of Clinical Neurophysiology*, *13*(6), 511–518.

Koster-Hale, J., & Saxe, R. (2013). Theory of Mind: A Neural Prediction Problem. *Neuron*, *79*(5), 836–848. https://doi.org/10.1016/j.neuron.2013.08.020

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

Kurthen, I., Meyer, M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2020). Individual Differences in Peripheral Hearing and Cognition Reveal Sentence Processing Differences in Healthy Older Adults. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.573513

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

León-Cabrera, P., Flores, A., Rodríguez-Fornells, A., & Morís, J. (2019). Ahead of time: Early sentence slow cortical modulations associated to semantic prediction. *NeuroImage*, *189*, 192–201. https://doi.org/10.1016/j.neuroimage.2019.01.005

León-Cabrera, P., Rodríguez-Fornells, A., & Morís, J. (2017). Electrophysiological correlates of semantic anticipation during speech comprehension. *Neuropsychologia*, *99*, 326–334. https://doi.org/10.1016/j.neuropsychologia.2017.02.026

Lopukhina, A., Lopukhin, K., & Laurinavichyute, A. (2021). Morphosyntactic but not lexical corpus-based probabilities can substitute for cloze probabilities in reading

experiments. *PLOS ONE*, *16*(1), e0246133. https://doi.org/10.1371/journal.pone.0246133

Mantegna, F., Hintz, F., Ostarek, M., Alday, P. M., & Huettig, F. (2019). Distinguishing integration and prediction accounts of ERP N400 modulations in language processing through experimental design. *Neuropsychologia*, *134*, 107199. https://doi.org/10.1016/j.neuropsychologia.2019.107199

Merkin, A., Sghirripa, S., Graetz, L., Smith, A. E., Hordacre, B., Harris, R., Pitcher, J., Semmler, J., Rogasch, N. C., & Goldsworthy, M. (2023). Do age-related differences in aperiodic neural activity explain differences in resting EEG alpha? *Neurobiology of Aging*, *121*, 78–87. https://doi.org/10.1016/j.neurobiolaging.2022.09.003

Morís, J., Luque, D., & Rodríguez-Fornells, A. (2013). Learning-induced modulations of the stimulus-preceding negativity. *Psychophysiology*, *50*(9), 931–939. https://doi.org/10.1111/psyp.12073

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

Ono, K., Hashimoto, J., Hiramoto, R., Sasaoka, T., & Yamawaki, S. (2021). Modulatory Effects of Prediction Accuracy on Electroencephalographic Brain Activity During Prediction. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.630288

Ouyang, G., Hildebrandt, A., Schmitz, F., & Herrmann, C. S. (2020). Decomposing alpha and 1/f brain activities reveals their differential associations with cognitive processing speed. *NeuroImage*, *205*, 116304. https://doi.org/10.1016/j.neuroimage.2019.116304

Pathania, A., Euler, M. J., Clark, M., Cowan, R. L., Duff, K., & Lohse, K. R. (2022). Resting EEG spectral slopes are associated with age-related differences in information processing speed. *Biological Psychology*, *168*, 108261. https://doi.org/10.1016/j.biopsycho.2022.108261

Pertermann, M., Mückschel, M., Adelhöfer, N., Ziemssen, T., & Beste, C. (2019). On the interrelation of 1/f neural noise and norepinephrine system activity during motor response inhibition. *Journal of Neurophysiology*, *121*(5), 1633–1643. https://doi.org/10.1152/jn.00701.2018

Press, C., Kok, P., & Yon, D. (2020). The Perceptual Prediction Paradox. *Trends in Cognitive Sciences*, *24*(1), 13–24. https://doi.org/10.1016/j.tics.2019.11.003

Pulvermüller, F., & Grisoni, L. (2020). Semantic Prediction in Brain and Mind. *Trends in Cognitive Sciences*, *24*(10), 781–784. https://doi.org/10.1016/j.tics.2020.07.002

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language Models are Unsupervised Multitask Learners*. 24.

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

Roehm, D., Sorace, A., & Bornkessel-Schlesewsky, I. (2013). Processing flexible form-to-meaning mappings: Evidence for enriched composition as opposed to indeterminacy. *Language and Cognitive Processes*, *28*(8), 1244–1274. https://doi.org/10.1080/01690965.2012.712143

Samaha, J., & Postle, B. R. (2015). The Speed of Alpha-Band Oscillations Predicts the Temporal Resolution of Visual Perception. *Current Biology*, *25*(22), 2985–2990. https://doi.org/10.1016/j.cub.2015.10.007

Schoknecht, P., Roehm, D., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2022). The interaction of predictive processing and similarity-based retrieval interference: An ERP study. *Language, Cognition and Neuroscience*, *37*(7), 883–901. https://doi.org/10.1080/23273798.2022.2026421

Shipp, S. (2016). Neural Elements for Predictive Coding. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/article/10.3389/fpsyg.2016.01792

Smith, N. J., & Levy, R. (2011). *Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing*. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 33, No. 33).

Surwillo, W. W. (1961). Frequency of the 'Alpha' Rhythm, Reaction Time and Age. *Nature*, *191*(4790). https://doi.org/10.1038/191823a0

Surwillo, W. W. (1963). The relation of simple response time to brain-wave frequency and the effects of age. *Electroencephalography and Clinical Neurophysiology*, *15*(1), 105–114. https://doi.org/10.1016/0013-4694(63)90043-9

Taylor, W. L. (1953). "Cloze Procedure": A New Tool for Measuring Readability. *Sage Journals. 30*(4). https://doi.org/10.1177/107769905303000401

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading

Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443

Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic Integration in Sentences and Discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657–671. https://doi.org/10.1162/089892999563724

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

Voytek, B., & Knight, R. T. (2015). Dynamic network communication as a unifying neural basis for cognition, development, aging, and disease. *Biological Psychiatry*, *77*(12), 1089–1097. https://doi.org/10.1016/j.biopsych.2015.04.016

Voytek, B., Kramer, M. A., Case, J., Lepage, K. Q., Tempesta, Z. R., Knight, R. T., & Gazzaley, A. (2015). Age-Related Changes in 1/f Neural Electrophysiological Noise. *Journal of Neuroscience*, *35*(38), 13257–13265. https://doi.org/10.1523/JNEUROSCI.2332-14.2015

What was new in previous releases? —MNE-BIDS 0.10 documentation. (2022). Retrieved May 20, 2022, from https://mne.tools/mne-bids/stable/whats_new_previous_releases.html

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

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., … Rush, A. M. (2020). *HuggingFace's Transformers: State-of-the-art Natural Language Processing* (arXiv:1910.03771). arXiv. http://arxiv.org/abs/1910.03771

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

Zhang, Y., Lu, Y., Wang, D., Zhou, C., & Xu, C. (2021). Relationship between individual alpha peak frequency and attentional performance in a multiple object tracking task

among ice-hockey players. *PLOS ONE*, *16*(5), e0251443. https://doi.org/10.1371/journal.pone.0251443

Zuidema, W., & de Boer, B. (2018). The evolution of combinatorial structure in language. *Current Opinion in Behavioral Sciences*, *21*, 138–144. https://doi.org/10.1016/j.cobeha.2018.04.011

Figures and tables

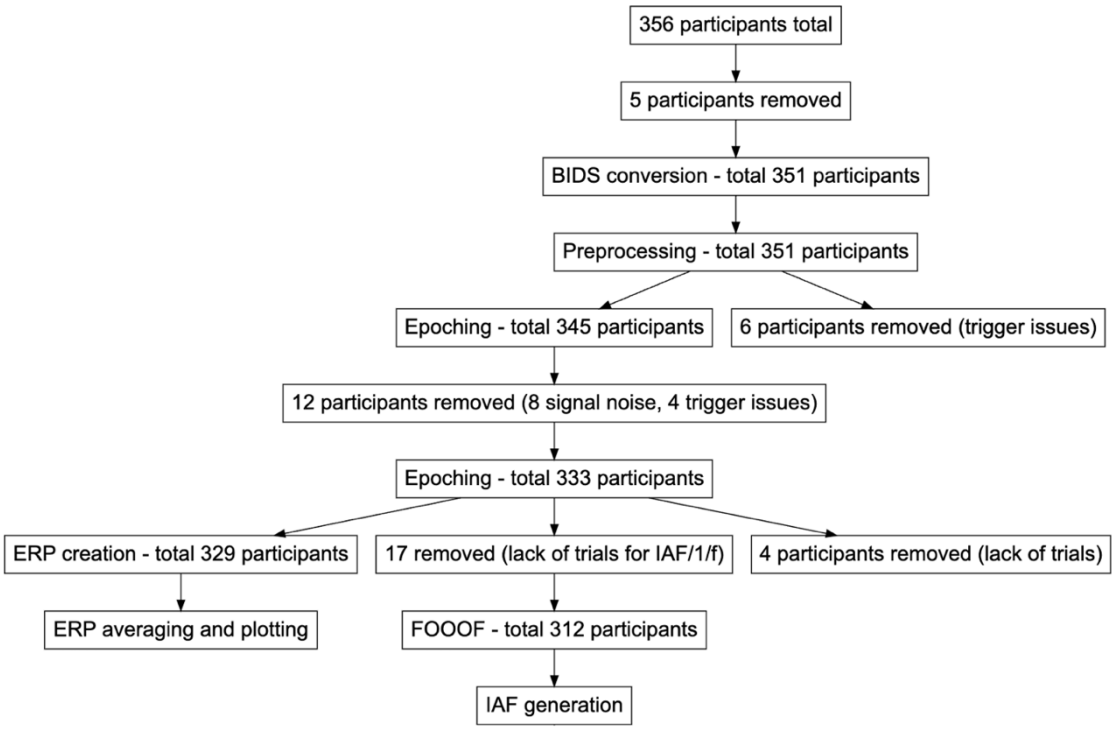

Figure 1. Participants at each stage of the data analysis
The number of participants at each stage of the data analysis, including the number removed and reasons for removal.

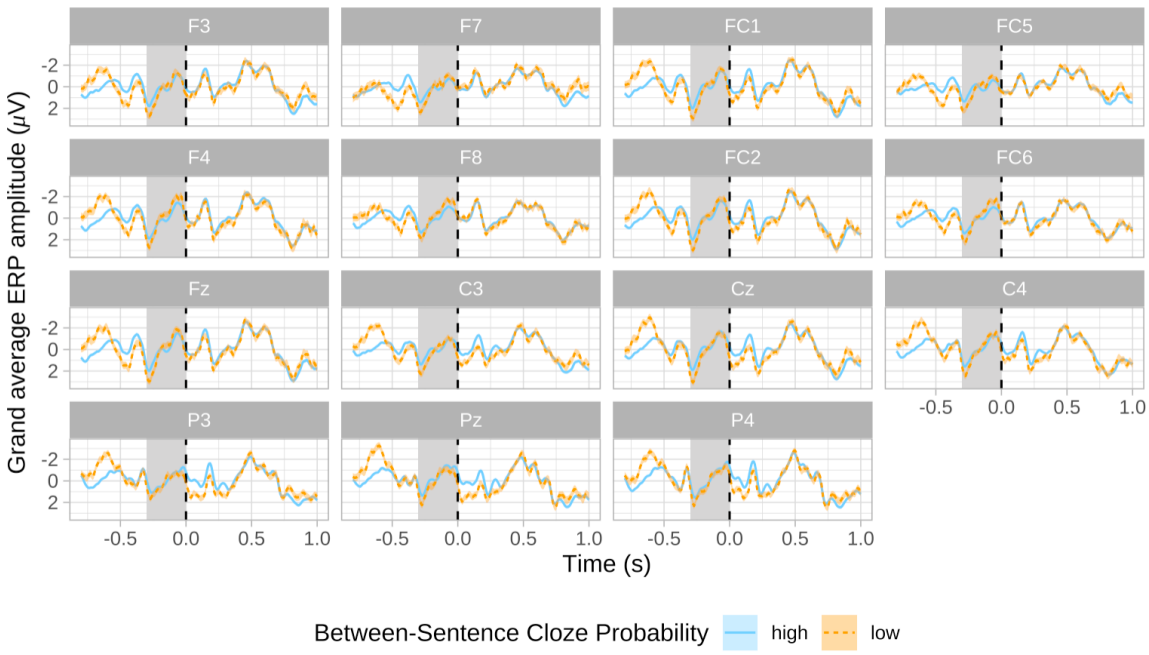

Figure 2. ERP correlates time-locked to the articles (between-sentence cloze)
ERPs for articles plotted over channels of interest. Blue lines represent high article cloze probability (>50%) whilst orange dashed lines reflect low article cloze probability (<50%). Note that between-sentence article cloze is plotted here to gauge pre-stimulus effects. Blue and orange shading reflects the standard errors for each condition. The pre-stimulus window is shaded grey, and the black dashed line depicts the onset of the article word.

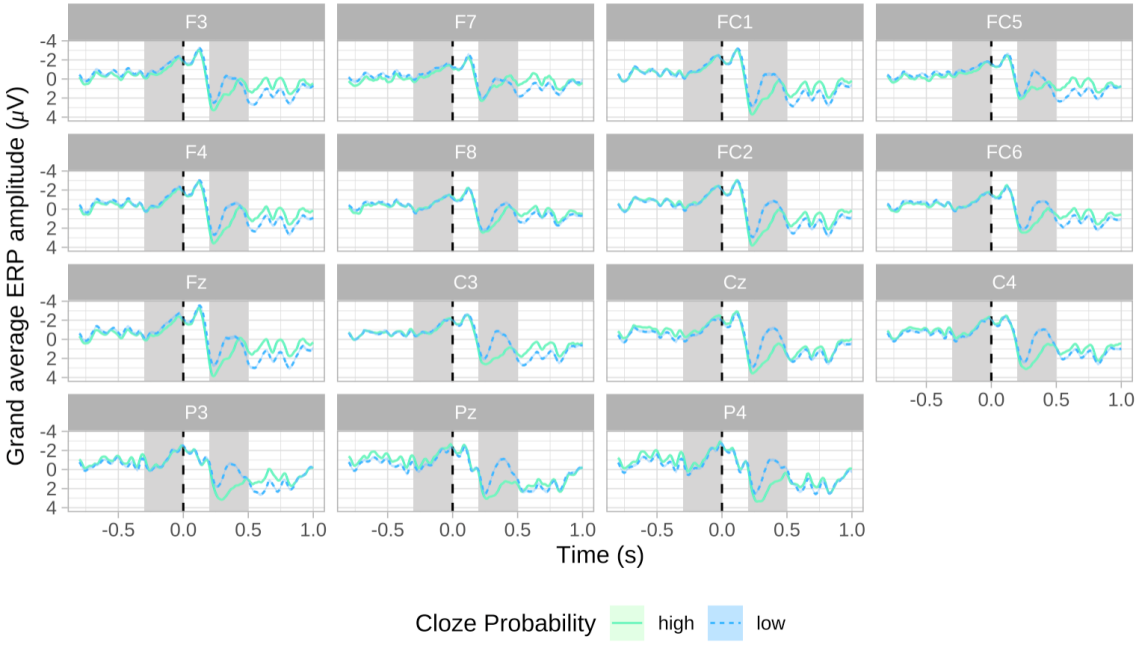

Figure 3. ERPs time-locked to nouns

Grand average ERP plots for noun words over channels of interest. Turquoise lines represent high noun cloze probability (>50) whilst dark blue dashed lines reflect low noun cloze probability (<50). Turquoise and blue shading reflects the standard errors for the respective conditions. The pre-stimulus ERP window and the N400 window are denoted via grey shading, and the black dashed line depicts the onset of the noun word.

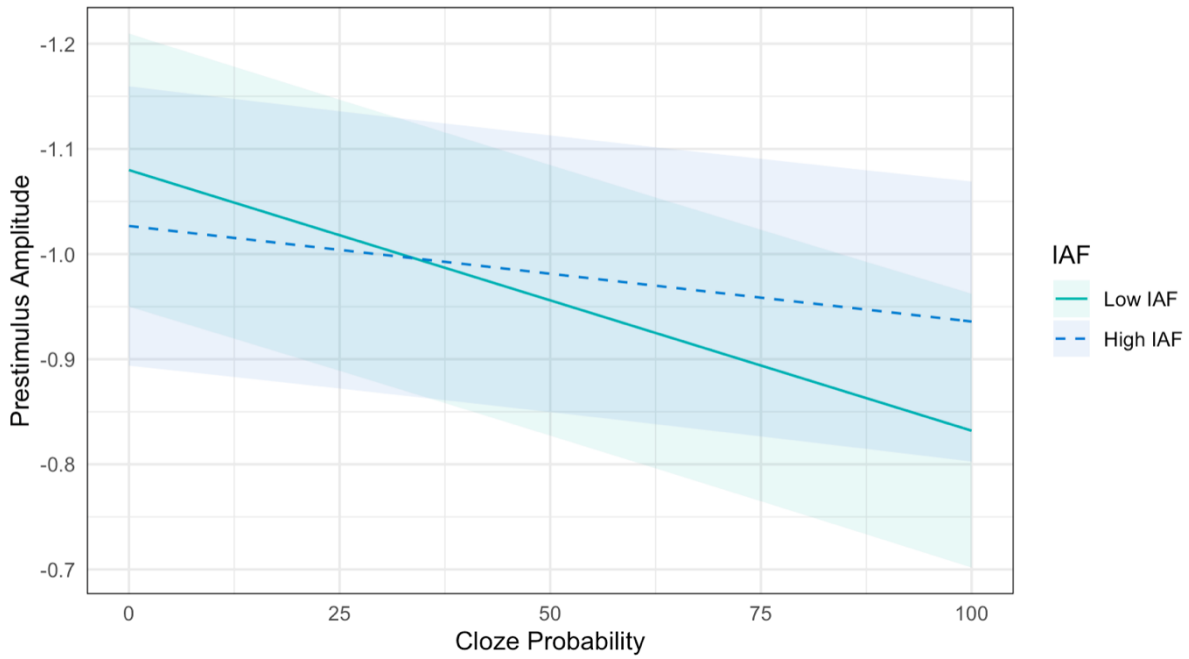

Figure 4. The interaction between cloze probability and IAF on pre-stimulus ERP activity, as modelled for nouns

The interaction between cloze probability and IAF on pre-stimulus ERP activity, as modelled for nouns. Lines represent model fit, with shading depicting the 83% confidence intervals. Dashed, dark blue lines represent a high IAF, and solid turquoise lines reflect a low IAF. The dichotomisation of IAF is for visualisation purposes only.

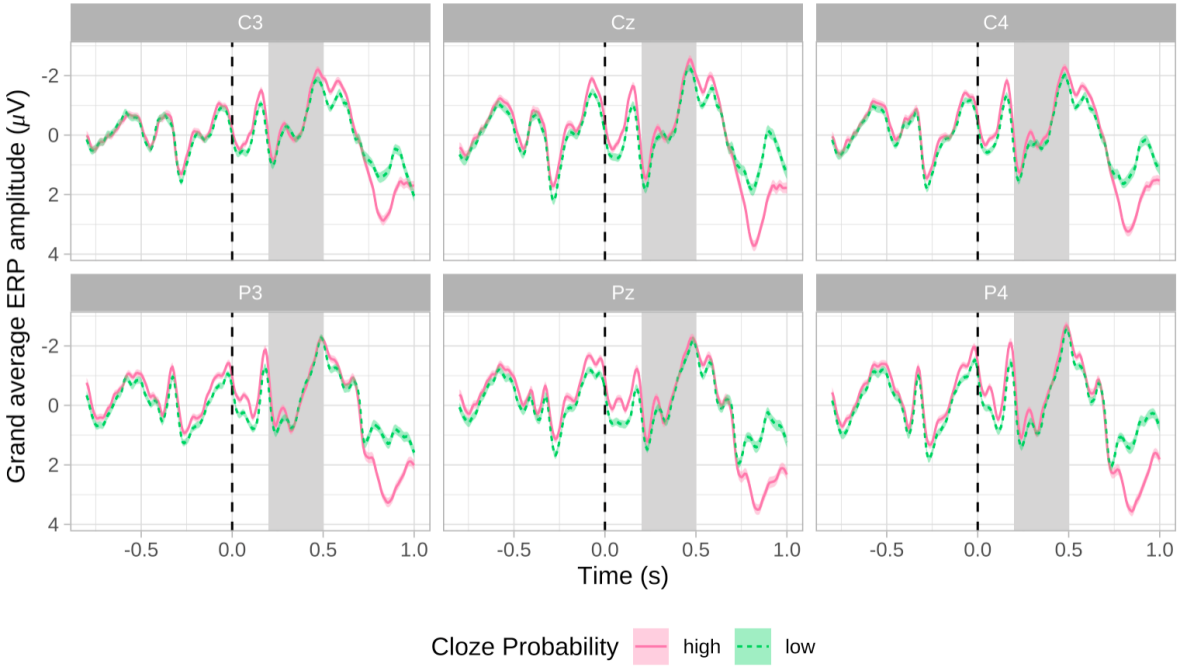

Figure 5. Grand average ERP plots of article-N400 effects
Grand average ERP plots for articles over channels of interest. Pink lines represent high cloze probability (>50) whilst green dashed lines reflect low cloze probability (<50). Pink and green shading depicts the standard errors for the two conditions. The N400 window is depicted via grey shading, and the black dashed line represents the onset of the article word.

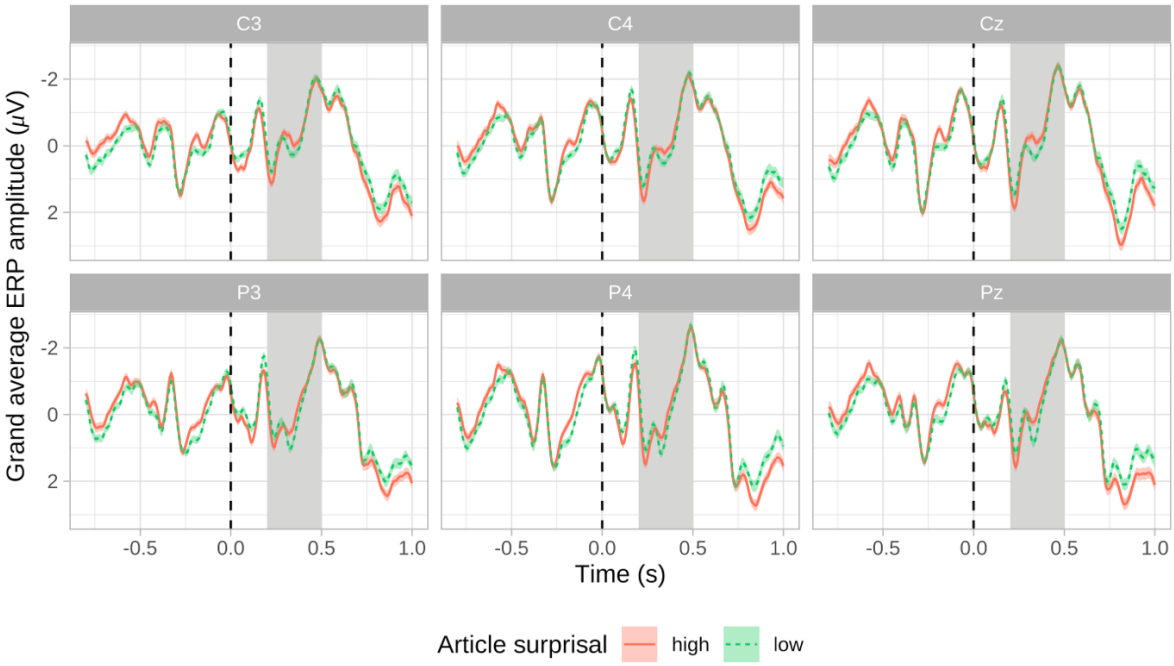

Figure 6. Grand average ERP plots of article-N400 effects according to surprisal
Grand average ERP plots for articles over channels of interest. Red lines represent high surprisal (>4.08; the mean article surprisal value) whilst green dashed lines reflect low surprisal (<4.08). Red and green shading depicts the standard errors for the two conditions. The N400 window is depicted via grey shading, and the black dashed line represents the onset of the article.

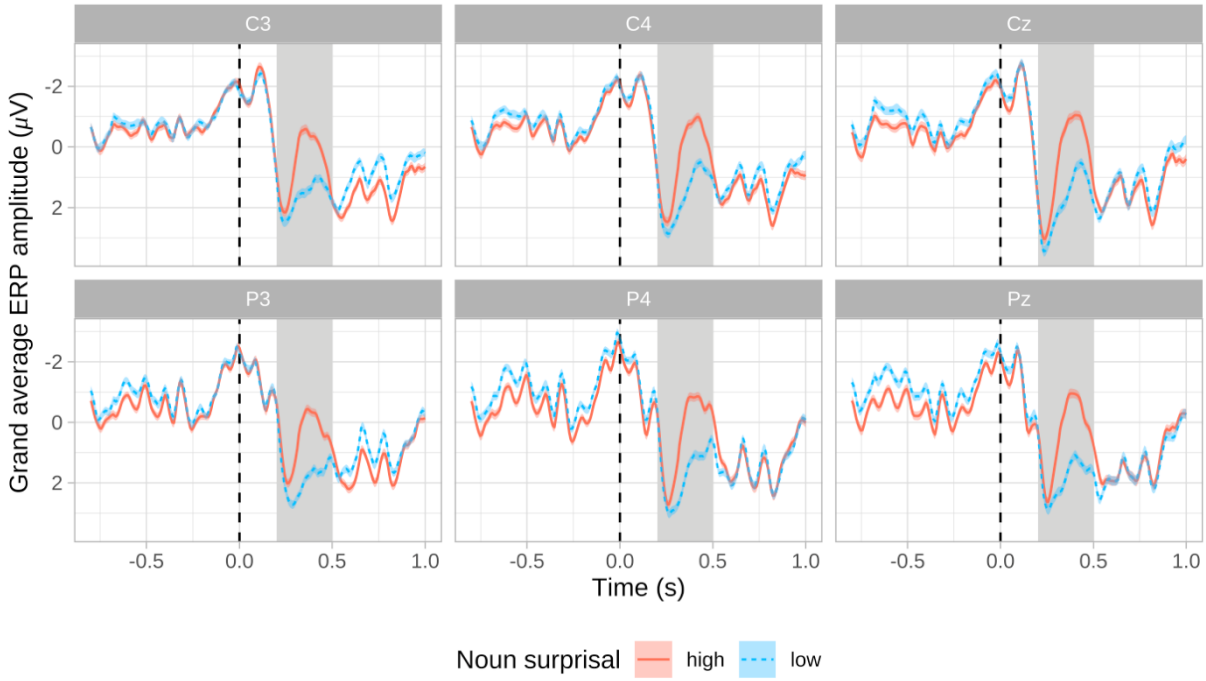

Figure 7. Grand average ERP plots of noun-N400 effects according to surprisal

Grand average ERP plots for nouns over channels of interest. Red lines represent high surprisal (>5.8; the mean noun surprisal value) whilst blue dashed lines reflect low surprisal (<5.8). Red and blue shading depicts the standard errors for the two conditions. The N400 window is depicted via grey shading, and the black dashed line represents the onset of the noun.

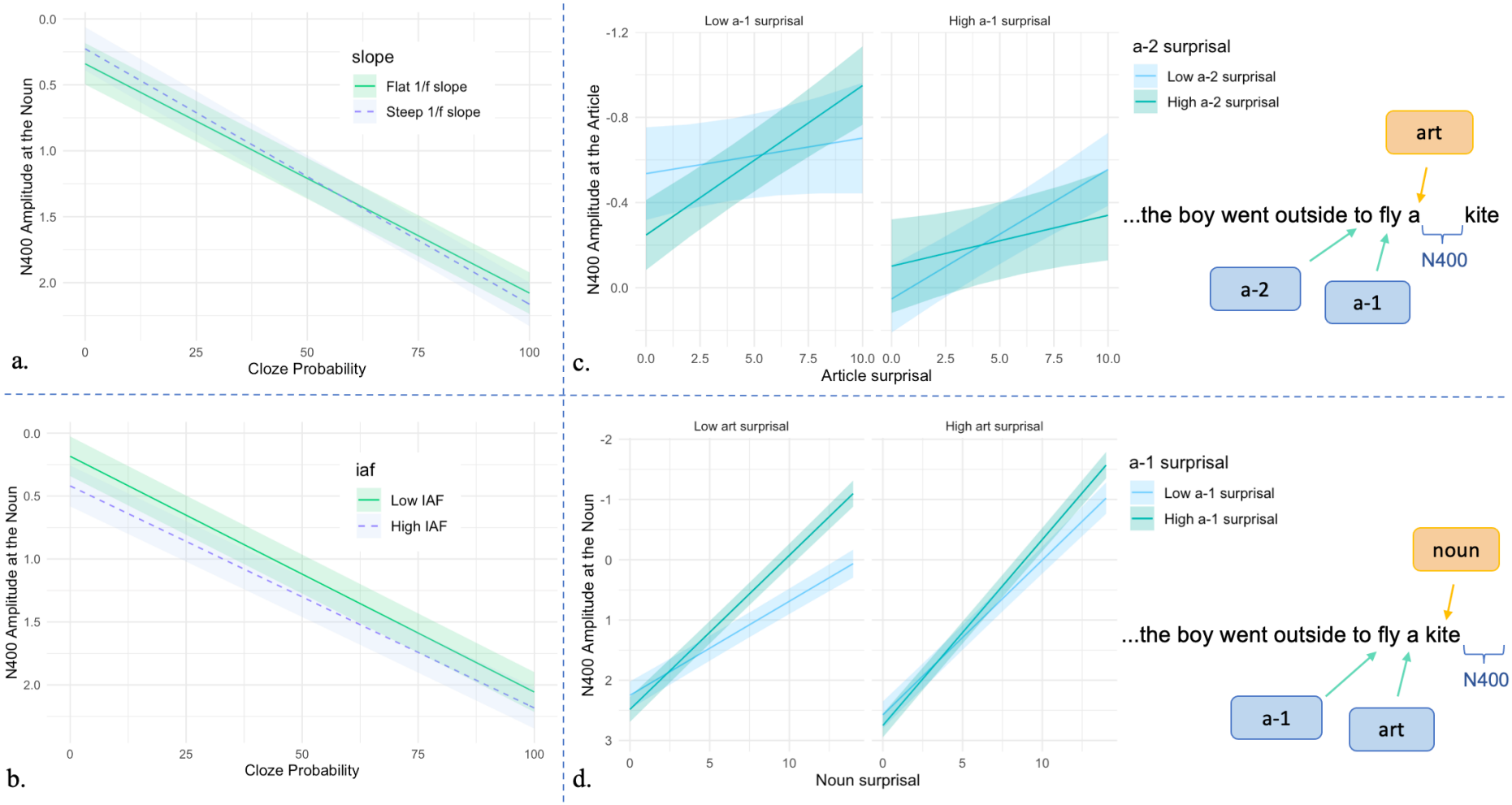

Figure 8. (a) N400 amplitude to nouns predicted by cloze and the aperiodic 1/f slope. **(b)** N400 amplitude to nouns, predicted by cloze and IAF. **(c)** N400 amplitude to articles, predicted by a-2, a-1, and article surprisal. **(d)** N400 amplitude time-locked to nouns, predicted by a-1, article, and noun surprisal. Shading in all plots reflects 83% confidence intervals, lines indicate model fit, and the dichotomisation of variables is for visualisation purposes only

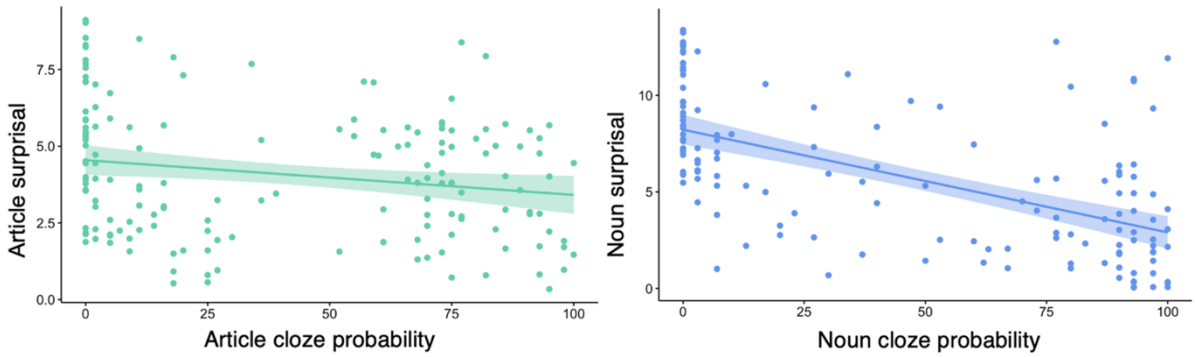

Figure 9. (a) The Pearson correlation between article cloze probability and article surprisal. **(b)** The Pearson correlation between noun cloze probability and noun surprisal.

Table 1.
Demonstrating the mean number of trials per condition that were entered into the ICA, with standard deviations (SD) in brackets.

Expected articles	Unexpected articles	Expected nouns	Unexpected nouns	Pre-article*
39.97 (0.29)	39.98 (0.20)	39.97 (0.30)	38.96 (1.01)	79.98 (0.15)

**The pre-article window (-300 to 0ms preceding the word before the article) was used as a baseline period.*

Table 2. Example sentence demonstrating the cloze probability values (in brackets) for both the unexpected and expected continuations.
Note that prior to the article, the sentence versions are identical, thus justifying the application of the expected article cloze value (82) to all versions of the sentence when analysing pre-stimulus effects time-locked to articles.

Sentence beginning	**Expected article**	**Unexpected article**	**Expected noun**	**Unexpected noun**
The old wives' tale says that if you want to keep the doctor away then you should eat	an (82)	a (0)	apple (97)	carrot (0)

Table 3.
*Linear mixed-effects models with outcome variables, covariates, fixed and random intercepts. Note that IAF and 1/f were included as separate interaction terms, such that in model 1 for example, the interaction of cloze*IAF was examined independently from the interaction of cloze*1/f slope.*

Model	Outcome	Covariates	Fixed effects	Separate interaction terms	Random intercepts
1	Pre-stimulus ERP amplitude time-locked to the **article**	Baseline (pre-article) ERP amplitude	Between-sentence cloze probability	1/f slope IAF	+ Subject + item + channel
2	Pre-stimulus ERP amplitude time-locked to the **noun**	Baseline (pre-article) ERP amplitude	Cloze probability	1/f slope IAF	+ Subject + item + channel
3	N400 amplitude time-locked to the **article**	Pre-stimulus ERP amplitude	Cloze probability	1/f slope IAF	+ Subject + item + channel
4	N400 amplitude time-locked to the **noun**	Pre-stimulus ERP amplitude	Cloze probability	1/f slope IAF	+ Subject + item + channel

Table 4.
Exploratory linear mixed-effects models with outcome variables, covariates, fixed and random effects.

Model	Outcome	Covariates	Fixed effects	Random effects
5	N400 amplitude time-locked to the **article**	Pre-stimulus ERP amplitude	Article surprisal	+ Subject + item + channel
6	N400 amplitude time-locked to the **article**	Pre-stimulus ERP amplitude	Article surprisal Article-1 surprisal Article-2 surprisal	+ Subject + item + channel
7	N400 amplitude time-locked to the **noun**	Pre-stimulus ERP amplitude	Noun surprisal	+ Subject + item + channel
8	N400 amplitude time-locked to the **noun**	Pre-stimulus ERP amplitude	Noun surprisal Article surprisal Article-1 surprisal	+ Subject + item + channel