eNeuro

Research Article: New Research

Cognition and Behavior

Strategic and Non-Strategic Semantic Expectations Hierarchically Modulate Neural Processing

Consuelo Vidal-Gran,[1,3] Rodika Sokoliuk,[1,3] Howard Bowman,[1,2,3] and Damian Cruse[1,3]

https://doi.org/10.1523/ENEURO.0229-20.2020

[1]School of Psychology, University of Birmingham, Birmingham B15 2TT, United Kingdom, [2]School of Computing, University of Kent, Canterbury, Kent CT2 7NF, United Kingdom, and [3]Centre for Human Brain Health, University of Birmingham, Birmingham B15 2TT, United Kingdom

Abstract

Perception is facilitated by a hierarchy of expectations generated from context and prior knowledge. In auditory processing, violations of local (within-trial) expectations elicit a mismatch negativity (MMN), while violations of global (across-trial) expectations elicit a later positive component (P300). This result is taken as evidence of prediction errors ascending through the expectation hierarchy. However, in language comprehension, there is no evidence that violations of semantic expectations across local-global levels similarly elicit a sequence of hierarchical error signals, thus drawing into question the putative link between event-related potentials (ERPs) and prediction errors. We investigated the neural basis of such hierarchical expectations of semantics in a word-pair priming paradigm. By manipulating the overall proportion of related or unrelated word-pairs across the task, we created two global contexts that differentially encouraged strategic use of primes. Across two experiments, we replicated behavioral evidence of greater priming in the high validity context, reflecting strategic expectations of upcoming targets based on "global" context. In our preregistered EEG analyses, we observed a "local" prediction error ERP effect (i.e., semantic priming) ~250 ms post-target, which, in exploratory analyses, was followed 100 ms later by a signal that interacted with the global context. However, the later effect behaved in an apredictive manner, i.e., was most extreme for fulfilled expectations, rather than violations. Our results are consistent with interpretations of early ERPs as reflections of prediction error and later ERPs as processes related to conscious access and in support of task demands.

Key words: ERP; predictive coding; relatedness proportion; semantic priming

Significance Statement

Semantic expectations have been associated with the event-related potential (ERP) N400 component, which is modulated by semantic prediction errors across levels of the hierarchy. However, there is no evidence of a two-stage profile that reflects violations of semantic expectations at a single level of the hierarchy, such as the mismatch negativity (MMN) and P3b observed in the local-global paradigm, which are elicited by violations of local and global expectations, respectively. In the present study, we provided evidence of an early ERP effect that reflects violations of local semantic expectations, followed by an apredictive signal that interacted with the global context. Thus, these results support the notion of early ERPs as prediction errors and later ERPs reflecting conscious access and strategic use of context.

Introduction

Predictive coding theory argues that the brain processes information in a hierarchical probabilistic Bayesian manner (Knill and Pouget, 2004; Friston, 2005) by contrasting sensory input with prior expectations generated

Received May 26, 2020; accepted September 11, 2020; First published October 6, 2020.
The authors declare no competing financial interests.

Author contributions: C.V.-G., H.B., and D.C. designed research; C.V.-G. and R.S. performed research; R.S., H.B., and D.C. contributed unpublished reagents/analytic tools; C.V.-G., R.S., and D.C. analyzed data; C.V.-G. and D.C. wrote the paper.

from context and the perceiver's knowledge (Clark, 2013; Heilbron and Chait, 2018). Expectations are sent down from higher levels of the hierarchy and any subsequent unexplained sensory input is sent back up the hierarchy as prediction error (Rao and Ballard, 1999; Heilbron and Chait, 2018; Friston and Kiebel, 2009; Bubic et al., 2010).

Some argue that evoked neural responses [e.g., event-related potentials (ERPs)] reflect prediction errors (Friston, 2005; Chennu et al., 2013). For example, the mismatch negativity (MMN) is larger in amplitude for stimuli that do not match short-term auditory expectations, relative to those that do (Wacongne et al., 2012; Heilbron and Chait, 2018). Prediction errors at higher levels of the hierarchy are investigated in paradigms that introduce violations of expectations formed from the global context in which stimuli occur. Indeed, generating such expectations involves complex cognition, including working memory and report of conscious expectation (Bekinschtein et al., 2009). The local-global paradigm (Bekinschtein et al., 2009) elegantly pits local expectation within each trial (i.e., standard vs deviant pitch tones) against a global expectation built from the context across blocks of trials. This paradigm elicits an initial MMN to local violations of expectation, and a subsequent centro-parietal positivity at ~300 ms poststimulus (P3b) to global violations of expectation (Faugeras et al., 2012; King et al., 2014; El Karoui et al., 2015); thereby, separating prediction error signals at two levels of an expectation hierarchy that unfold sequentially.

Within the realm of more ecologically valid stimulus processing, speech comprehension is similarly influenced by expectations at multiple levels of a hierarchy (Hutchison, 2007; Lau et al., 2013a; Lewis and Bastiaansen, 2015; Kuperberg and Jaeger, 2016; Ylinen et al., 2016). The N400, a negative deflection peaking around 400 ms poststimulus (Kutas and Federmeier, 2011), is a potential marker of errors of such semantic expectations (Rabovsky and McRae, 2014). On a local level, the N400 is larger to words that have not been primed relative to those that have (e.g., larger for DOG when preceded by Lamp than by Cat; Koivisto and Revonsuo, 2001; Lau et al., 2013a; Cruse et al., 2014), and at a more global level, the N400 is larger to words that are unexpected within a sentential context (Berkum et al., 1999; Thornhill and Van Petten, 2012; Boudewyn et al., 2015; Brothers et al., 2017). Interestingly, unlike the MMN/P3b in auditory processing, semantic prediction errors appear to be reflected in the magnitude of a single component, the N400, rather than in a series of components moving through the hierarchy of relative top-down involvement.

One approach to separate prediction error signals at two levels of a semantic expectation hierarchy is with a prime validity manipulation of a word-pair priming task. Specifically, we can pit the facilitation of target word processing that comes from presentation of a related prime against a global context in which it is not efficient for the comprehender to use the prime to predict the target, i.e., primes rarely followed by related targets (Keefe and Neely, 1990; Hutchison, 2007; Lau et al., 2013a,b). Therefore, as the proportion of related pairs increases within a context, the prime validity increases (i.e., the prime is more likely to predict the target). If individuals use the global context of prime validity to modulate their expectations, behavioral facilitation follows.

In ERP studies of prime validity, this hierarchy of local expectations (i.e., the prime relatedness) and global expectations (i.e., the prime validity) has not been reported to modulate the amplitudes of two sequential components (Lau et al., 2013a; Boudewyn et al., 2015); hence, there is no evidence of a two-stage profile to semantic expectation violation. Rather than reflecting error at one level, the N400 (or for N200 evidence, see Boudewyn et al., 2015) appears to account for a combination of errors across levels of the hierarchy. To disentangle these results, here we report a preregistered trial-by-trial manipulation of both local and global semantic expectations. First, we report a replication of the reaction time (RT) facilitation caused by global context as described by Hutchison (2007). Second, we report the associated electrophysiological markers of expectation and violation across levels of the hierarchy from a separate group of healthy participants performing the same task. In accordance with predictive coding, we hypothesized that ERP amplitudes would reflect violations of expectation at consecutive levels of the hierarchy, with local violations evident earlier than global violations.

Materials and Methods

Experiment 1, behavioral study

Participants

We recruited participants through the Research Participation Scheme website of the University of Birmingham, who received credits for their participation. A total of 64 participants were recruited, with the data of two participants excluded from analysis because of outlying data, as quantified by the non-recursive procedure for outlier elimination (detailed below, Behavioural data analyses section; Van Selst and Jolicoeur, 1994; Hutchison, 2007). Therefore, the final sample consisted of 62 participants (59 females, three males; median age: 19, range: 18–28). All participants reported to be monolingual native English speakers, right-handed, and with no history of neurologic conditions or diagnosis of dyslexia. All participants gave written informed consent before participation in this study, which was approved by the STEM Ethical Review Committee of the University of Birmingham.

Stimuli

Associated prime-target pairs were selected from the Semantic Priming Project database (Hutchison et al.,

This work was funded by the National Agency for Research and Development (ANID) from the Government of Chile, under the PhD programme "Doctorado en el extranjero BECAS CHILE 2016", the Medical Research Council Grant MR/P013228/1 (to D.C.), and the School of Psychology from the University of Birmingham.

Correspondence should be addressed to Damian Cruse at d.cruse@bham.ac.uk or Consuelo Vidal-Gran at cxv648@student.bham.ac.uk.

https://doi.org/10.1523/ENEURO.0229-20.2020

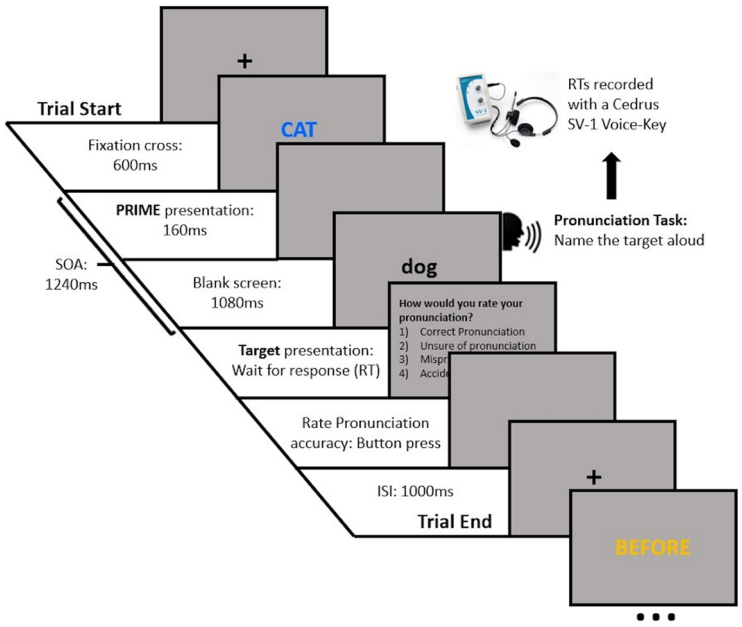

Figure 1. Semantic priming relatedness proportion task (Hutchison, 2007). Participants were required to name the target word aloud and as fast as possible, while their responses were recorded.

2013) and the experimental design was a replication of the paradigm implemented by Hutchison (2007). First, all word pairs available in the database (N = 1661) were ordered by Forward Associative Strength (i.e., the proportion of individuals who spontaneously name the same target after reading the prime word) and the 352 word-pairs with the highest strength were selected after removal of any specific American English associations (e.g., Clorox-Bleach; Slacks-Pants).

The first 156 word-pairs from this list of 352 word-pairs with the highest forward association were chosen to be the critical stimuli for statistical analysis. The remaining 196 word-pairs served as fillers to generate the global context and are not included in the statistical analysis. We divided all 156 critical word-pairs into two lists (N = 78 word-pairs per list) that were balanced according to the values from the database (Hutchison et al., 2013) for forward association, length, log HAL frequency, and orthographic neighborhood (all $p > 0.604$; all BF10 < 0.196). In the same way, we divided the 196 filler word-pairs into two balanced lists (N = 98 word-pairs per list; all $p > 0.284$, all BF10 < 0.267). Thus, we had created two critical related word-pair lists and two filler related word-pair lists. To create the unrelated word-pair lists, we manually re-paired (within list) all word-pairs in each of the four lists above, Stimuli section (two critical, two fillers) ensuring that unrelated targets were semantically unrelated to their prime. This resulted in a final set of eight lists: two critical related, two critical unrelated, two filler related, and two filler unrelated. Each participant was assigned two Critical sets of word-pairs (one related and one unrelated; 78 word-pairs per list) and two Filler sets (one related and one unrelated; 98 word-pairs per list). Hence, each participant saw all words within the full set of 352 word-pairs exactly once, composed of 176 related word-pairs and 176 unrelated word-pairs.

To create the prime-validity manipulation, first we assigned half of the critical word-pairs, including both related and unrelated items, to one color (yellow or blue), and the other half with the other color in an interleaved order. Next, the related filler set was assigned with one color (yellow or blue), and the unrelated filler set was assigned with the other color. Therefore, across all items seen by each participant, 77.8% of word-pairs presented in one of the two colors were related, thus giving that color high prime validity, and 77.8% of word-pairs presented in the other color were unrelated, thus giving that color low prime validity. Importantly, across the entire set of stimuli that each participant saw, exactly half were related (the other half unrelated) and half were presented in one color (the other half in the other color). However, the probability of a related target following a prime of one color was 77.8% and the probability of a related target following a prime of the other color was 22.2%. Across participants, the color assignment of the high validity primes was counterbalanced (i.e., half of participants saw high prime validity word-pairs in blue and low prime validity word-pairs in yellow; and the other half saw the opposite colors for each proportion), and all possible combinations of word lists were used, resulting in 32 permutations.

Procedure

The task was presented with Psychtoolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) in MATLAB (MathWorks). The vocal RTs were measured with a Cedrus SV-1 Voice Key (Cedrus Corporation), with all participants completing four practice trials under the experimenter¨s supervision to adjust the voice key threshold according to the participant¨s speech volume. The trial procedure is shown in Figure 1. Specifically, each trial started with a central fixation cross on a gray background

lasting 600 ms, then, the prime word was displayed in either yellow or blue, at the center of the screen for 160 ms, followed by a blank screen for 1080 ms, and subsequently the target was displayed on the screen; thus, the stimulus onset asynchrony (SOA) was 1240 ms. The target stayed on the screen until the participant pronounced the word; then the word disappeared from the screen, which remained blank for 300 ms. Afterwards, a rating for the quality of pronunciation was displayed on the screen with the following questions and potential responses: How would you rate your pronunciation? (1) Correct pronunciation; (2) unsure of pronunciation; (3) mispronunciation; (4) accidental voice-key triggering. Participants gave a button response on the keyboard (1–4) to rate their pronunciation (as per Hutchison, 2007). After the participant responded, the screen remained blank for 1000 ms, before the next trial began.

Each participant was tested individually and sat ~70 cm away from the computer screen. All participants received written information about the study, the instructions, and the consent form. In addition, the instructions were verbally repeated by the experimenter. We instructed all participants that a colored uppercase word (either blue or yellow) will be displayed on the screen and that they must read it silently to themselves; then, a black lower case word will be displayed on the screen, and they should pronounce the word aloud, as fast and accurately as possible. Participants were told that the color of the uppercase word will cue the probability of the lower case target being related or unrelated. Half of the participants received the following written instructions: "If the uppercase word is Blue, it is highly likely that the meaning of the lower case word will be related; and if the uppercase word is Yellow, it is highly likely that the meaning of the lower case word will be unrelated" (as per Hutchison, 2007). The other half of participants received the same instructions but with the colors flipped.

After the task, we asked participants to complete a self-report form about the use of strategy throughout the task, to determine whether they were using expectations strategically. The form was composed of three questions and a free text description of the strategy. The questions were the following: (1) which color was highly likely to be related? (responses: BLUE/YELLOW); (2) did you use the color of the UPPERCASE word (BLUE, YELLOW) as a cue for knowing whether the following word was related or unrelated? (responses: YES/NO); (3) did you engage in any strategy to speed up your responses using the color cue? (responses: YES/NO); (4) if YES, briefly describe. We considered participants to have used strategic expectation (i.e., those referred to as the strategy group) if they correctly identified the color that was assigned for the high validity condition (question 1), answered YES in questions 2 and 3, and described a strategy in question 4. All other participants were classified into the no strategy group.

Behavioral data analyses

To ensure the inclusion of trials pronounced correctly, we only included trials that were rated by the participants with a correct pronunciation (button press 1); moreover, we eliminated RTs that were longer than 2500 ms and shorter than 1 ms (i.e., not correctly triggered by the vocal onset). As raw RTs are skewed, some researchers opt to log transform the data, although this can result in other information about response speed being lost (Lo and Andrews, 2015). Here, we chose to follow the same procedure as in Hutchison (2007), namely, the non-recursive procedure for outlier elimination (Van Selst and Jolicoeur, 1994). Specifically, RTs that were more than X SDs from the mean were considered to be outliers and were removed, where the value of X decreases with decreasing sample size (i.e., number of trials in each condition for that participant) and is anchored at $X = 2.5$ for a sample size of 100. Next, across all participants we used the same procedure to determine outlier participants and rejected data from two participants that met the outlier criteria. For the remaining 62 participants, a median of 37 trials (range: 16–39) contributed to the high related condition; a median of 36 trials (range: 12–39) to the high unrelated condition; a median of 37 trials (range: 16–39) to the low related condition; and a median of 36 (range: 15–39) contributed to the low unrelated condition.

All behavioral analyses were conducted in Jasp 0.9.1.0 software (JASP Team, 2018). To test for an effect of global context on RTs, we conducted a two-way repeated measures ANOVA with factors of relatedness (i.e., related vs unrelated targets) and prime validity (i.e., high vs low prime validity). We also reported equivalent Bayesian repeated measures ANOVAs (Wagenmakers et al., 2018; Van Doorn et al., 2019). We expected individuals to show faster RTs for related (expected) in contrast with unrelated (unexpected) targets because of local level expectations, i.e., priming. Furthermore, we expected an interaction, with larger priming effects in a high validity context in contrast with a low validity context, reflecting the use of global level context to predict upcoming stimuli.

As a follow-up analysis, we conducted a three-way ANOVA, with its Bayesian equivalent, to test for the interaction and the report of strategy versus no strategy (self-report form) as a between-subjects factor.

Experiment 2, behavioral and electrophysiological study

This study was preregistered in the Open Science Framework website, details and all codes described in the paper can be found under the following link: https://osf.io/v35te/. Any deviations from the preregistered methods and analyses are specifically stated in the text.

Participants

We recruited participants through the Research Participation Scheme website and placed advertisement posters at the University of Birmingham; participants received a monetary compensation for their participation. We recruited 37 participants, however, since we only investigated those who reported using a strategy, the final sample only included 22 participants (15 female, seven males; median age: 21, range: 18–30; classified by the same report form as experiment 1). The inclusion criteria were the same as those for experiment 1; however, participants were also required to attend for a structural T1-

weighted MRI scan at the University of Birmingham; therefore, participants who had any metal parts in their body, were claustrophobic, or women who were pregnant were excluded from the study, as the scan was mandatory for participation. All participants gave written informed consent before participation in this study, which was approved by the STEM Ethical Review Committee of the University of Birmingham.

We aimed to detect a RT interaction of the same magnitude as seen in the strategy group of experiment 1; therefore, we conducted a power analysis to select an appropriate sample size for this goal. We performed non-parametric power calculations using the data of all participants of the strategy group from experiment 1. Specifically, from the pool of participants of the strategy group, we selected with replacement *N* participants and conducted the same two-way repeated measures ANOVA 1000 times to test for the RT interaction effect. With an *N* of 22 participants in the strategy group, we achieved 80% power at $p < 0.05$ (i.e., 80% of ANOVAs included a significant interaction).

As we did not know whether a participant was in the Strategy group until their self-report form was completed at the end of the study, we recruited participants until 22 of them were classified as being in the strategy group (median age: 21, range: 18–30; 12 in the no-strategy group, median age: 22, range: 19–33). After removal of trials rated as mispronunciations and those considered outliers according to the non-recursive outlier elimination procedure of Van Selst and Jolicoeur (1994; as experiment 1), a median of 28 trials (range: 11–38) contributed to the high related condition; a median of 29.5 trials (range: 13–38) to the high unrelated condition; a median of 29 trials (range: 12–39) to the low related condition; and a median of 28 (range: 14–37) contributed to the low unrelated condition.

Stimuli and procedure

Stimuli and procedure were the same as in experiment 1, except for the duration of the fixation cross (increased from 600 to 750 ms to provide more time for an EEG time-frequency baseline). Additionally, we checked each unrelated prime-target pair across all lists and re-paired 55 unrelated targets within list to ensure that each unrelated target shared no overlapping phonemes with their respective related target.

EEG recording

The EEG signal was continuously recorded with a 125 channel AntNeuro EEG system (AntNeuro b.v.) at a sampling rate of 500 Hz, with impedances kept below 20 kΩ. We placed the ground electrode on the left mastoid bone and referenced online to CPz. As participants were required to pronounce words aloud, we also recorded a bipolar EMG signal with one EMG electrode above the upper lip and the other below the lower lip on the left side of the mouth, approximately over the superior and inferior Orbicularis Oris muscles (Lapatki et al., 2003; Drake et al., 2009).

EMG preprocessing

As this task involved participants speaking, there were considerable artefacts in the EEG data around the vocal RT that were challenging to remove adequately. We therefore chose to analyze only the EEG data up to the point of vocal artifact. To minimize artefacts from additional preparatory muscular activity before vocal onset, in our preregistered methods, we planned to choose the latest time point for analysis post-target by identifying when the mouth EMG signal began to significantly differ between prime validity conditions in a temporal cluster mass randomization test, as implemented in FieldTrip (Oostenveld et al., 2011). However, this approach revealed no significant clusters (smallest cluster $p = 0.513$), and so did not provide a suitable cutoff time point for our analyses. Therefore, in a deviation from the preregistered plan, we chose our latest time point of EEG data to analyze as 150 ms before the fastest mean RT across conditions (in this instance, high validity-related = 532 ms; for a similar approach, see Kuperberg et al., 2018). Our post-target time window therefore continued to 382 ms post-target. From all the trials included for the statistical analysis, only 5.76% of trials had RTs earlier than this time point, comparable with previous studies (Kuperberg et al., 2018).

EEG preprocessing pipeline

We low-pass filtered the continuous EEG data at 40 Hz using the finite impulse response filter implemented in EEGLAB (Delorme and Makeig, 2004). Because of our interest in analyzing slow-waves (see below, Prime slow wave linear fit analyses), we performed no high-pass filtering. Next, we segmented the filtered EEG signals into epochs from 750 ms before the onset of the prime up to 382 ms post-target (for details, see above, EMG preprocessing). Subsequent artifact rejection proceeded in the following steps based on a combination of methods described by Nolan et al. (2010) and Mognon et al. (2011).

First, as in the behavioral data analysis, we excluded all trials in which the participant rated their response as incorrect (i.e., 2, 3, 4 button press) and those that had RTs that were classified as outliers in the non-recursive procedure for outlier elimination (Van Selst and Jolicoeur, 1994). Next, bad channels were identified and removed from the data. We considered a channel to be bad if its absolute *z* score across channels exceeded three on any of the following metrics: (1) variance of the EEG signal across all time points, (2) mean of the correlations between the channel in question and all other channels, and (3) the Hurst exponent of the EEG signal (a measure of the predictability of a time series (Nolan et al., 2010), estimated with the discrete second order derivative from the MATLAB function wfbmesti. After removal of bad channels, we identified and removed trials containing non-stationary artefacts. Specifically, we considered a trial to be bad if its absolute *z* score across trials exceeded three on any of the following metrics: (1) the mean across channels of the voltage range within the trial, (2) the mean across channels of the variance of the voltages within the trial, and (3) the mean across channels of the difference between the mean voltage at that channel in the trial in question and the mean voltage at that channel across all trials. After removal of these individual trials, we conducted an additional check for bad channels, and removed them, by interrogating the average of the channels across all trials (i.e., the ERP,

averaged across all conditions). Specifically, we considered a channel to be bad in this step if its absolute *z* score across channels exceeds three on any of the following metrics: (1) the variance of voltages across time within the ERP, (2) the median gradient of the signal across time within the ERP, and (3) the range of voltages across time within the ERP.

To remove stationary artefacts, such as blinks and eye movements, the pruned EEG data were subjected to an independent component analysis with the runica function of EEGLAB. The MATLAB toolbox ADJUST (Mognon et al., 2011) subsequently identified which components reflect artefacts on the basis of their similarity to stereotypical spatiotemporal patterns associated with blinks, eye movements, and data discontinuities, and the contribution of these artifact components was then subtracted from the data. Next, we interpolated the data of any previously removed channels via the spherical interpolation method of EEGLAB and re-referenced the data to the average of the whole head. We chose to use the average reference as this is common practice in high-density EEG recordings and allows for clearer comparison of ERPs with other relevant paradigms (Bekinschtein et al., 2009; Faugeras et al., 2012).

Before proceeding to group-level analyses, single-subject averages for the ERP analysis were finalized in the following way. First, a robust average was generated for each condition separately, using the default parameters of SPM12. Robust averaging iteratively down-weights outlier values by time point to improve estimation of the mean across trials. As recommended by SPM12, the resulting ERP was low-pass filtered below 20 Hz using a FIR filter (again, with EEGLAB's pop_neweegfilt), and the mean of the baseline window (−200–0 ms) was subtracted.

Single-subject data for the time-frequency analysis were preprocessed in a similar way. However, first, we concatenated the individual trials into a matrix of channels x all time points, and filtered each channel in two-steps (high-pass then low-pass) to retain the frequency bands of interest (i.e., 8–12 Hz α and 13–30 Hz β), using EEGLAB's finite impulse response filter (function: pop_eegnewfilt). Next, we extracted the squared envelope of the signal (i.e., the squared complex magnitude of the Hilbert-transformed signal) to provide a time-varying estimate of power within that frequency band. The resulting time course was re-segmented into its original epochs and averaged within each condition separately using SPM12's robust averaging procedure. As with the ERP analyses, we low-pass filtered the resulting average time series below 20 Hz (EEGLAB's pop_neweegfilt). Finally, we converted the power estimates to decibels relative to the mean of the baseline window (−200–0 ms).

EEG/MRI co-registration

We recorded the electrode locations of each participant relative to the surface of the head using a Xensor Electrode Digitizer device and the Visor2 software (AntNeuro b.v.). Furthermore, on a separate day, we acquired a T1-weighted anatomic scan of the head (nose included) of each participant with a 1 mm resolution using a 3T Philips Achieva MRI scanner (32 channel head coil). This T1-weighted anatomic scan was then co-registered with the digitized electrode locations using FieldTrip.

Analyses

Behavioral data analysis. The behavioral analyses are the same as for the strategy group in experiment 1.

EEG analysis.

Target ERP, prime ERP and prime time frequency analyses.Time courses (ERPs/time frequency) within the time window of interest (0–1240 ms for primes; 0–382 ms for targets) were compared with the cluster mass method of the open-source MATLAB toolbox FieldTrip (Oostenveld et al., 2011). This procedure involves an initial parametric step followed by a non-parametric control of multiple comparisons (Maris and Oostenveld, 2007). Specifically, we conducted two-tailed dependent samples *t* tests at each spatiotemporal data point within our time window of interest. Spatiotemporally adjacent electrodes (*t* values) with $p < 0.05$ were then clustered based on their proximity, with the requirement that a cluster must span more than one time point and at least four neighboring electrodes, with an electrode's neighborhood containing all electrodes within an ~4-cm radius (median: 8, range:2–10). Finally, we summed the *t* values at each spatiotemporal point within each cluster. Next, we estimated the probability under the null hypothesis of observing cluster sum Ts more extreme than those in the experimental data, i.e., the *p* value of each cluster. Specifically, FieldTrip randomly shuffles the trial labels between conditions, performs the above spatiotemporal clustering procedure, and retains the largest cluster sum T. Consequently, the *p* value of each cluster observed in the data are the proportion of the largest clusters observed across 1000 such randomizations that contain larger cluster sum Ts. As our analyses were two-tailed, we set the family-wise error corrected cluster α to 0.025.

Prime slow wave linear fit analyses. To further test for ERP evidence of expectation formation in response to the prime, we analyzed whether a slow wave differentiates high validity and low validity conditions. For this comparison, we used a least-squares linear fit to the averaged ERPs of each condition (high and low validity primes) for each electrode and participant (as per Chennu et al., 2013). Next, the slope values were compared between conditions with the spatial cluster mass analysis in FieldTrip (Oostenveld et al., 2011).

Source estimation analysis. We constructed individual boundary element head models (BEM; four layers) from subject-specific T1-weighted anatomic scans, by using the "dipoli" method of the MATLAB toolbox FieldTrip (Oostenveld et al., 2011). Next, we aligned the electrode locations, that were recorded with Xensor Electrode Digitizer device, to the surface of the scalp layer that was segmented from the T1-weighted anatomic scan. For reference points, we used the fiducial points and electrode locations as head shape. We visually checked that the electrode positions and the scalp surface were aligned, and we manually fixed imperfections. We prepared the EEG data before subjecting it to statistical analyses, where we balanced the number of trials in each condition, by taking the smallest condition N as a reference and randomly discarding trials from the other conditions

surpassing that N, resulting in equal datasets. This procedure is used to remove the potential influence of biases from unequal sample sizes.

ERPs whole brain. For the whole brain ERP source analysis, we used single-trial data that had not been subjected to robust averaging, and defined trials as time windows from −382 to 382 ms relative to target onset. This data were then bandpass filtered between 1 and 40 Hz using a firws filter as implemented in FieldTrip (Oostenveld et al., 2011). Subsequently, relative to the different conditions, data were divided into seven sets: one containing all trials, one containing only related trials, one only unrelated trials, one all high-validity related and one all low-validity related trials, one containing all high-validity unrelated and one all low-validity unrelated trials. The sensor covariance matrix was estimated for all these sets of data in the time window −382–382 ms relative to target onset. A common spatial filter was then computed on the dataset containing all trials using a linear constraint minimum variance (LCMV) beamformer (Van Drongelen et al., 1996; Van Veen et al., 1997; Robinson and Vrba, 1999). Beamformer parameters were chosen including a fixed dipole orientation, a weighted normalization (to reduce the center of head bias), as well as a regularization parameter of 5% to increase the signal-to-noise ratio (cf. Popov et al., 2018; Sokoliuk et al., 2019a). This common spatial filter served then for source estimation of the remaining six sets of trials. Subsequently, the dipole moments of the different source estimates were extracted within the poststimulus time windows of interest (time windows for source estimates of related vs unrelated trials: 226–280, 232–290, 306–382, 316–350 ms; time window to test interaction effect for source estimates of highly related and unrelated trials and low related and unrelated trials: 316–350 ms) and their absolute values averaged over time to obtain one average source estimation value per grid point ([dot]VE) and condition.

To test for significant differences between conditions we conducted five contrasts as mentioned above; first, an interaction between prime validity (high/low) and relatedness of the target (related/unrelated) in a time window from 316 to 350 ms; next, we tested the early and late main effects of relatedness of the target (related/unrelated) as observed in the sensor analyses results (four main effects), in their respective time windows for the early effect (226–280 and 232–290 ms); and the late effect (306–382 and 316–350 ms). Montecarlo cluster-based permutation tests were computed as implemented in FieldTrip (Oostenveld et al., 2011) by using averaged data over each time window; moreover, we used an α and a cluster α level of 0.025 and 1000 permutations.

Automated anatomical labeling (AAL) analysis. We tested for the post-target interaction, between the relatedness of the target (related/unrelated) and the validity of the prime (High prime validity/Low prime validity) in five specific anatomic regions of interest that are defined using the automated anatomic labeling (AAL) atlas (for similar analyses with MEG and EEG data, see Brookes et al., 2016; Sokoliuk et al., 2019b). The selected regions are the left inferior frontal gyrus (LIFG), including pars opercularis, pars triangularis and pars orbitralis; the posterior left middle temporal gyrus (LMTG); and posterior left superior temporal gyrus (LSTG), as Weber et al. (2016) reported a relatedness proportion interaction in these regions. In addition, we tested the post-target interaction in the anterior LMTG and anterior LSTG, as Lau et al. (2013b) found differences in the anterior left superior temporal region (LSTG) in related versus unrelated items in a high validity condition. Moreover, as a deviation from our preregistered analyses, we tested the main effects found in the Related – Unrelated contrast at the sensor level (ERPs) in the same anatomic regions (for more details, see Results). To determine both the anterior and posterior parts of the LMTG and LSTG, we calculated the center of mass of each AAL region and selected all virtual electrodes that were anterior or posterior to the center of mass.

We aggregated the AAL regions of interest to each participant's T1-weighted image. Next, for each participant individually, we extracted the average source estimation values of all VEs (from prior source estimation; cf. above, ERPs whole brain) within each AAL region, weighted them according to their Euclidian distance to the center of mass of the AAL region (Brookes et al., 2016) and averaged over VEs within each AAL region of interest. We then conducted paired-sample *t* tests between the post-target conditions (SP-high validity/SP-low validity) for all AAL regions; and another paired-sample *t* test between the relatedness conditions (related/unrelated) for each AAL region in four time windows (226–280, 232–290, 316–350, 306–382 ms) from the main effects obtained in the sensor level ERP analyses (Results). The *p* values that we obtained were corrected for multiple comparisons across AAL regions using false discovery rate (FDR; Yekutieli and Benjamini, 1999). Furthermore, to test for evidence for the null hypothesis, we calculated Bayes Factors using the Bayes equivalent *t* test, according to Rouder et al. (2009). A Bayes factor between 1 and 3 is considered to be weak/anecdotal evidence in support of the hypothesis being tested; from 3 to 10 is substantial evidence and 10 to 100 is strong evidence (Jeffreys, 1961). Note that, as the Bayes factor is the ratio of evidence for two hypotheses, the same category descriptions hold for the inverse (i.e., 1:3, 1:10, 1:100).

Results

Experiment 1, behavioral only

In a two-way repeated measures ANOVA, we found a significant interaction between prime validity and relatedness of the target ($F_{(1,61)} = 13.751$, $p < 0.001$, $\eta p^2 = 0.184$), which was also strongly supported by a Bayesian repeated measures ANOVA ($BF_{inclusion} = 19.25$). As shown in Table 1, this interaction stems from the larger semantic priming effect in the high prime validity context ($t_{(61)} = -6.525$, $p < 0.001$, Cohen's $d = -0.829$, CI = −1.115 −0.537) relative to the low prime validity context [$t_{(61)} = -5.169$, $p < 0.001$, Cohen's $d = -0.656$, confidence interval (CI) = −0.929 −0.380]. Furthermore, RTs to unrelated items were markedly similar across contexts ($t_{(61)} < 0.001$, $p = 0.999$, Cohen's $d < 0.001$, CI = −0.249 0.249), while the difference in semantic priming

Table 1: Descriptive statistics including mean RT (ms) and SD of related and unrelated word-pairs on each validity context, high prime validity and low prime validity

Condition	Low validity = 22.2% mean RTs (SD)	High validity = 77.8% mean RTs (SD)	Prime validity effect
Unrelated	508 ms (76 ms)	508 ms (75 ms)	
Related	493 ms (73 ms)	472 ms (76 ms)	
Priming effect	15 ms (32 ms)	36 ms (54 ms)	21 ms (60 ms)

Semantic priming effects and prime validity effect (relatedness proportion effect).

stems from significantly different RTs to related items ($t_{(61)}$ = −3.797, $p < 0.001$, Cohen's d = −0.482, CI = −0.744 −0.217).

Of 62 participants, 32 were classified in the "no-strategy" group and 30 were classified in the "strategy" group according to their responses on the self-report form (for description, see Materials and Methods). A *post hoc* mixed design ANOVA with two within factors (relatedness of target; validity of the prime) and one between subjects factor (strategy; no-strategy) revealed a significant target × prime validity × strategy interaction ($F_{(1,60)} = 7.537$, p =0.008, ηp^2 = 0.112, $BF_{inclusion}$ = 3.203), reflecting the apparent presence of a prime validity effect when participants reported using the prime strategically ($F_{(1,29)} = 20.388$, $p < 0.001$, ηp^2 = 0.413; $BF_{inclusion}$ = 34.67) but absence of a prime validity effect when participants reported no strategy ($F_{(1,31)} = 0.860$, $p = 0.361$, ηp^2 = 0.027; $BF_{inclusion}$ = 0.393; Fig. 2). The no strategy group, however, did exhibit a significant semantic priming effect by showing faster responses in the related relative to unrelated items ($F_{(1,31)} = 21.656$, $p < 0.001$, $\eta p^2 = 0.411$; inclusion $BF_{inclusion}$ = 4994.57).

Experiment 2

Behavioral results

These results were qualitatively consistent with those we observed in experiment 1. A two-way repeated measures ANOVA analysis showed a significant interaction between prime validity and relatedness of the target ($F_{(1,21)} = 9.071$, $p = 0.007$, $\eta p^2 = 0.302$), while the Bayesian repeated measures ANOVA analysis showed anecdotal evidence for the interaction ($BF_{inclusion}$ = 2.519). The interaction was driven by a larger semantic priming effect in the high prime validity context ($t_{(21)}$ = −4.254, $p < 0.001$, Cohen's d = −0.907, CI = −1.398 −0.400) than in the low prime validity context ($t_{(21)}$ = −2.046, $p = 0.054$, Cohen's d = −0.436, CI = −0.869 0.007; Table 2). There was no significant difference between the RTs to unrelated items across contexts ($t_{(21)} = 0.731$, $p = 0.473$, Cohen's $d = 0.156$, CI = −0.266 0.575) as opposed to a significant difference between related items across contexts ($t_{(21)}$ = −2.719, $p = 0.013$, Cohen's d = −0.580, CI = −1.027 −0.121).

EEG results, sensor level

Prime analyses: ERPs, time frequency and slow wave linear fit analyses

As the global context was instantiated by the prime words, we sought to also investigate potential electrophysiological markers of expectation setting (rather than post-target prediction errors). However, none of our preregistered analyses in the prime time window (0–1240 ms after prime onset) revealed evidence of markers of expectation in response to the prime. Specifically, there were no effects in analysis of the ERPs (smallest cluster $p = 0.233$), the slow wave linear fit analysis (no clusters formed), or the α-β time-frequency analysis (smallest cluster $p = 0.136$).

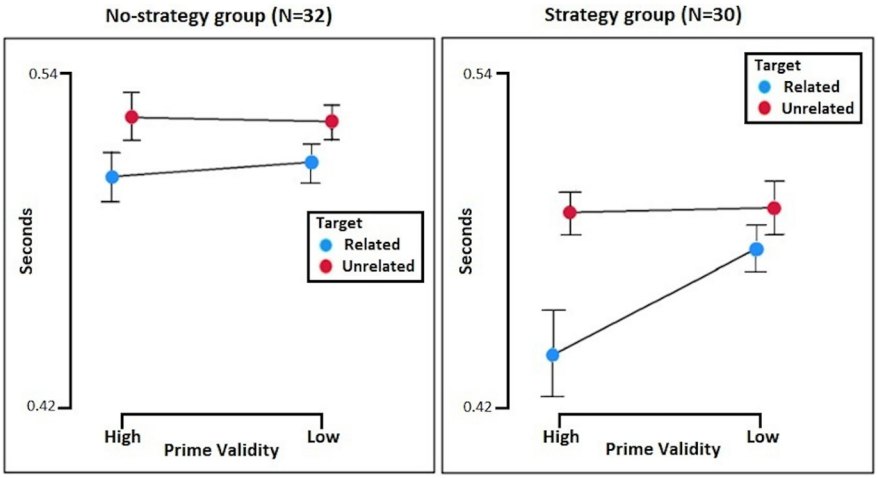

Figure 2. Mean RTs and confidence intervals (95%): prime validity (high/low), relatedness of the target (related/unrelated). Interaction ($p < 0.001$) between the validity of the prime and the relatedness of the target in the group of participants that reported the use of a conscious strategy (right), and no interaction ($p = 0.361$) in the group of participants that did not report a conscious strategy (left).

Table 2: Descriptive statistics including mean RT (ms) and SD of related and unrelated word pairs on each validity context, high prime validity and low prime validity

Condition	Low validity = 22.2% mean RTs (SD)	High validity = 77.8% mean RTs (SD)	Prime validity effect
Unrelated	576 ms (92 ms)	582 ms (87 ms)	
Related	560 ms (107ms)	532 ms (110 ms)	
Priming effect	16 ms (54 ms)	50 ms (69 ms)	34 ms (95 ms)

Semantic priming effects and prime validity effect (relatedness proportion effect).

Therefore, in exploratory analyses, we focused the time window of interest for the ERP analysis on the peak of the global field power Skrandies (1990) (530–1240 ms); however, this also revealed no significant difference between the high and low validity contexts (smallest cluster $p = 0.139$). Similarly, we used the window of interest for the α-β time-frequency analysis to the peak of the global field power Skrandies (1990) (602–1240 ms), which also yielded no significant difference between conditions (no clusters formed). Moreover, as α-β frequency bands include a wide range of frequencies we analyzed them separately. However, the time-frequency analysis in the α band (8–12 Hz) showed no significant differences between conditions in the 0- to 1240-ms time window (smallest cluster $p = 0.121$), nor in the 530- to 1240-ms time window (smallest cluster $p = 0.08$). The same was true for the β band (13–30 Hz; 0- to 1240-ms cluster $p = 0.312$; 530- to 1240-ms cluster $p = 0.197$). Together, these analyses suggested no apparent electrophysiological markers of pre-target expectation formation in our data.

Target results: ERPs. In our preregistered interaction contrast in the latency range from 0 to 382 ms poststimulus, the cluster-based permutation analysis yielded no clusters. However, in preregistered analyses of main effects in the same latency range, we found four significant main effects of relatedness of the target (i.e., unrelated vs related targets; Fig. 3). The clusters in our data occurred in two distinct periods within the time window as shown in Figure 3. Specifically, two clusters reflected a left frontotemporal dipolar effect of relatedness (Fig. 3*A*,*B*) at ~250 ms poststimulus (negative cluster: 226–280 ms, $p = 0.019$; positive cluster: 232–290 ms, $p = 0.009$), and two clusters reflected a later parieto-occipital dipolar effect of relatedness (Fig. 3*C*,*D*) at ~350 ms poststimulus (negative cluster: 316–350 ms, $p = 0.021$; positive cluster: 306–382 ms, $p = 0.004$). The early effects showed a predictive signal as in both clusters the voltage exhibited more extreme values for unrelated than related items. On the contrary, the later effects showed signs of an apredictive signal, especially in Figure 3*D*, as the voltage within the cluster had more extreme values for the related relative to the unrelated items.

As an exploratory analysis, and to increase power to detect a potential interaction effect, we tested for the interaction within each of the main effect clusters by averaging per condition and participant across all channels and time points within each main effect cluster. With this approach, the later negative cluster (Fig. 3*C*) showed a significant interaction ($F_{(1,21)} = 6.679$, $p = 0.017$, $\eta p^2 = 0.241$), reflecting a larger voltage difference between the related and unrelated targets in a high validity context with respect to a low validity context (other clusters $p = 0.396$, $p = 0.110$, $p = 0.273$). Bayesian equivalent analyses considered this to be anecdotal evidence for the alternative hypothesis ($BF_{inclusion} = 1.505$; Fig. 4).

Source estimate analyses

Our preregistered analyses included whole-brain interaction and main effect contrasts within the time windows of significant clusters at the sensor level. However, this approach returned no significant clusters at the source level (interaction smallest cluster $p = 0.147$; main effect smallest cluster $p = 0.067$). Furthermore, our preregistered source analyses included regions of interest from the following AAL regions: LIFG, LMTG, and LSTG. However, none of these regions exhibited significant interaction effects or main effects (all FDR corrected $p > 0.05$).

Consequently, for a qualitative visualization of the source estimates, here we plot the whole-brain thresholded *t* values ($p < 0.05$) of the source estimate contrasts, uncorrected for multiple comparisons. Specifically, we plot these *t* values for the early main effect (Fig. 3*A*,*B*) and the late main effect (Fig. 3*C*,*D*) in time windows selected to be entirely within the significant dipolar sensor level clusters (early: 232–280 ms; late: 316–350 ms; Fig. 5). The thresholded *t* values showed the peak of activity at the right middle and superior fontal gyri for the early effect; and the activity peak at the right supplementary motor area for the late effect, as shown in Figure 5.

Discussion

Predictive coding theory posits that the brain generates expectations about upcoming stimuli at varying levels of complexity, from low-level expectations about stimulus properties through to higher-level conceptual expectations. Here, we investigated the behavioral and electrophysiological correlates of such expectations and their violations at two levels of a semantic expectation hierarchy (local and global). First, on the behavioral level, participants of two separate experiments showed evidence of speeded RTs in related trials relative to unrelated trials, consistent with a local expectation generated about target word identity on the basis of the prime identity. Furthermore, participants generated a more conceptually complex expectation based on the global context (i.e., prime validity) to exhibit greater behavioral facilitation in the high prime validity context than the low prime validity context (Boudewyn et al., 2015). Importantly, only those individuals who reported conscious strategic expectation showed evidence of behavioral facilitation given by the

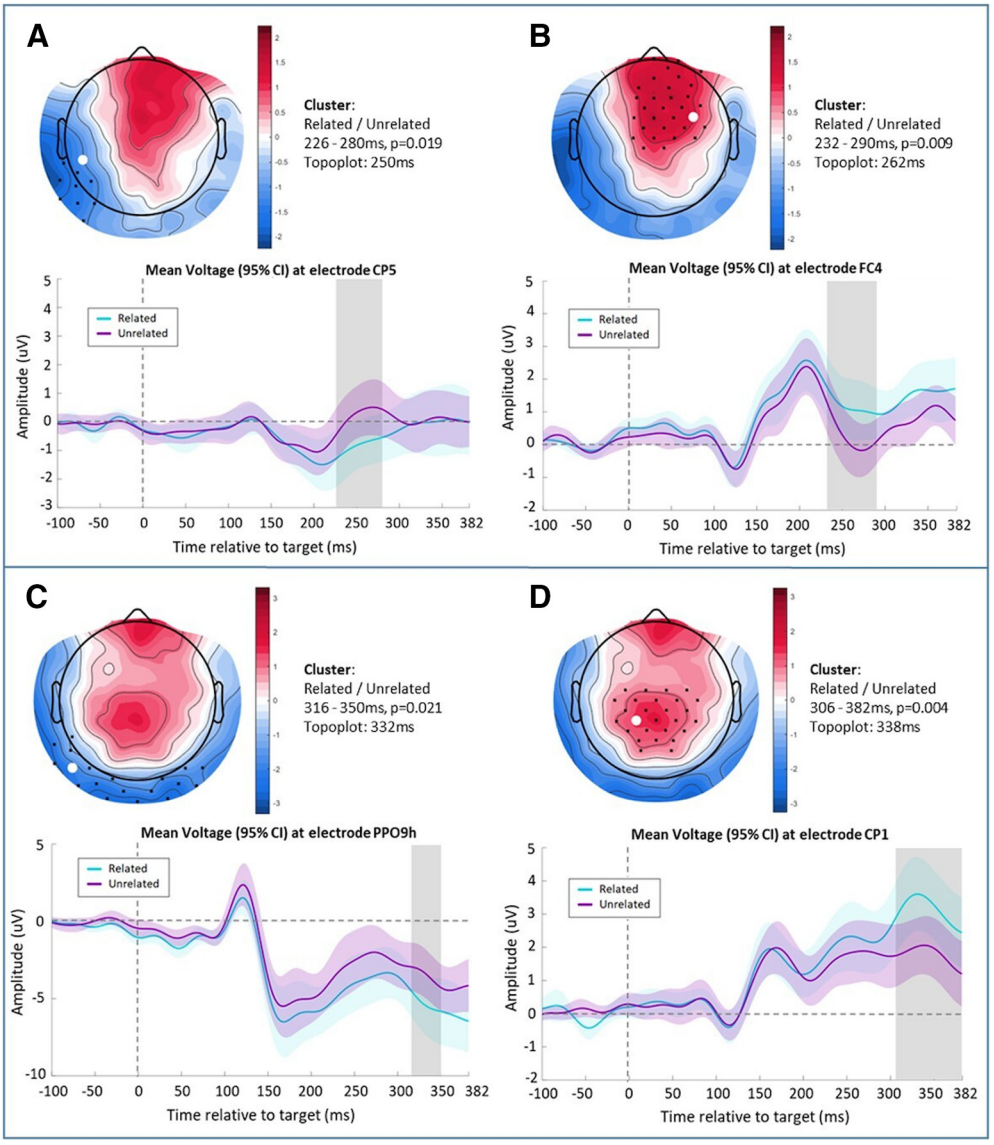

Figure 3. Four main effects from the cluster-based permutation analyses, which contrasted the voltage difference between related and unrelated word-pairs from 0 to 382 ms poststimulus. ERP scalp topographies revealed two dipolar effects; first, an early fronto-temporal effect at ~250 ms (***A***, ***B***); then, a later parieto-occipital effect at around 340 ms. Electrodes contributing to the clusters are marked with black dots. ***C***, ***D***, ERP plots show data (mean and shaded 95% confidence interval) from the electrode where the effect was maximal (highlighted with a white circle on the top plots), with the cluster period highlighted in gray.

global context, while those individuals who did not report a conscious strategy only exhibited facilitation as a result of the local context. Together, these behavioral data are consistent with a dissociation between a local expectation about the identity of the target generated by the prime, and a global expectation about the relatedness of the target that necessitates reportable, effortful, and strategic application of expectation. Moreover, the present data provides evidence for a successful replication of the behavioral effect elicited by the same paradigm as implemented by Hutchison (2007), who also found that the magnitude of the global facilitatory effect was modulated by the level of attentional control (i.e., weaker effect in individuals with lower attentional control; Hutchison, 2007). Similarly, our results suggested that only individuals that reported applying an effortful conscious strategy showed the global context effect as mentioned above.

Consistent with this two-stage expectation profile, the ERPs in response to the target words also exhibited a two-stage profile, with an early effect modulated by local expectation (around 250 ms) and a later effect modulated by global expectation (around 350 ms). These results are broadly consistent with the two-stage profile observed in the auditory oddball local, global paradigm (Bekinschtein et al., 2009), which includes an MMN in an early stage reflecting errors of the local context of the stimuli and a P3b response to errors of the global context given by blocks across the task.

Furthermore, the early effect in the present experiment showed more extreme amplitudes for unexpected targets

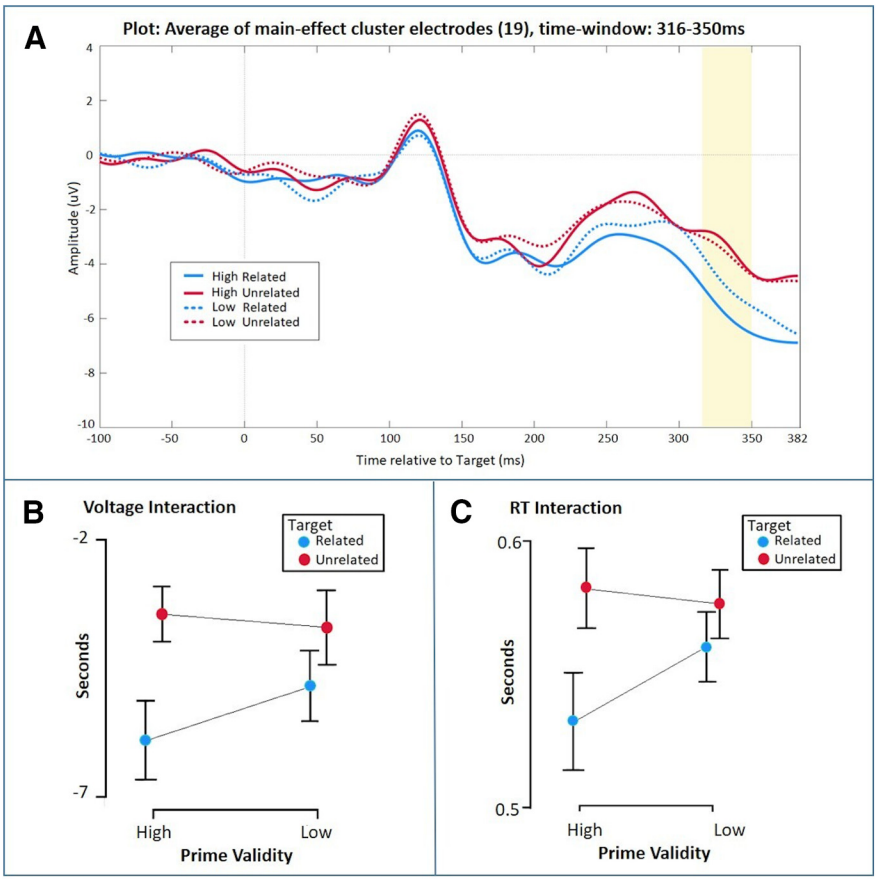

Figure 4. Exploratory analysis to test for the interaction between the four conditions [(HR – HU) – (LR – LU)]. The ERP plot in panel ***A*** shows the mean of electrodes (19 electrodes) within the 316- to 350-ms cluster found in the main effect analysis (Fig. 3*C*). ***B***, Mean and confidence intervals (95%) for each condition within the same time window that was analyzed with repeated measures ANOVA showing a significant voltage interaction ($p = 0.017$) with a larger difference in voltage between related and unrelated items in high validity context than low validity context. ***C***, Mean RTs and confidence intervals (95%) showing a significant interaction ($p = 0.007$) presented in Table 2. In this experiment, participant's behavior (RT; ***C***) showed the same pattern as their ERP responses (***B***).

relative to expected targets, consistent with a prediction error signal, such as the MMN to unexpected/deviant items observed across levels of stimulus awareness (Bekinschtein et al., 2009; Faugeras et al., 2012; Wacongne et al., 2012; Chennu et al., 2013; El Karoui et al., 2015). Moreover, the scalp topography of the early effect has a fronto-central peak, which is consistent with the MMN (Bekinschtein et al., 2009; Faugeras et al., 2012; Chennu et al., 2013); however, its latency is a little longer than seen in some of these previous papers and is not elicited here by a violation of auditory regularity. Additionally, in our source estimation analyses, the early effect was localized to the middle frontal gyrus (Fig. 5), whereas in another study the local MMN effect was localized to the temporal parietal junction and prefrontal cortex (Chennu et al., 2013), indicating not entirely overlapping neurocognitive processes. Nevertheless, as we observed behavioral semantic priming (as tracked by the early effect) even for participants who were not making strategic expectations, and because of the shared common features with the MMN (i.e., more extreme for errors and with a fronto-central focus), we consider the early effect to be consistent with an error of local expectation, i.e., expectation based on the identity of the prime, rather than the prime validity. Indeed, the MMN is elicited even by individuals who are not actively attending to the stimuli (Bekinschtein et al., 2009).

The late effect, however, was the opposite of what would be expected for a prediction error signal, i.e., its amplitude was more extreme for expected targets compared with unexpected targets. While the topography of this effect is similar to that of an N400 effect, with a maximum over midline parietal electrodes, it is evident from Figure 3*D* that the underlying waveforms are not consistent with an N400 effect. Indeed, a classical N400 effect is the difference between two negative-going ERPs that are more extreme (i.e., more negative-going) for semantically unexpected targets (Kutas and Federmeier, 2011). Conversely, the late effect here is evidently the difference between two positive-going ERPs (Fig. 3*D*) and is more extreme (i.e., more positive-going) for expected targets. This apredictive pattern is not readily explained by prediction error accounts without appeal to precision-weighting, in which a prediction error is weighted by the system's confidence in the signal (Friston, 2005; Wacongne et al., 2012; Chennu et al., 2013). Under precision-weighting, all

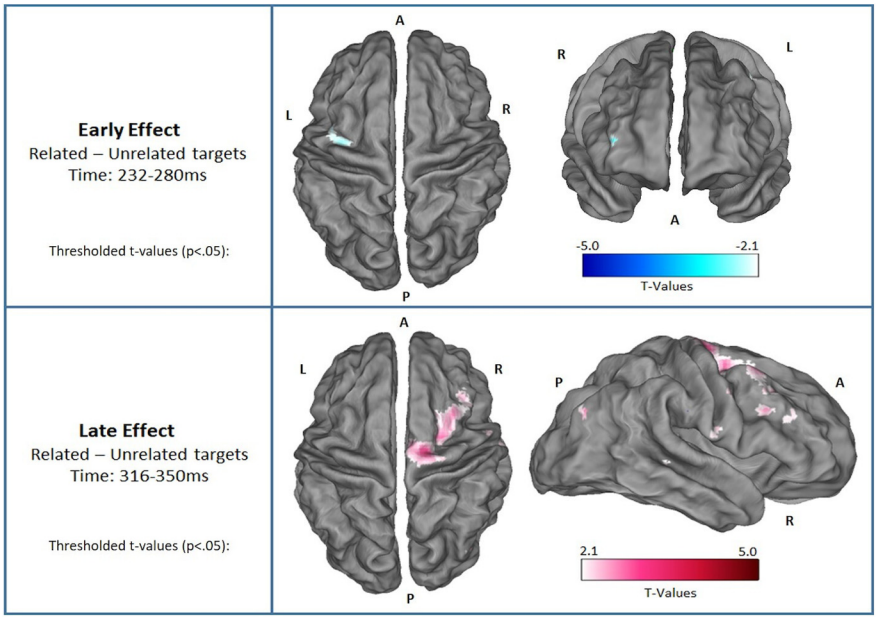

Figure 5. Thresholded *t* values ($p < 0.05$) of the ERP source estimates over two distinct time windows that corresponded to the early and late ERP effects reported above in Figure 3. Upper panel, Difference between related and unrelated targets in the early time window (232–280 ms). Lower panel, Same difference in a later time window (316–350 ms; thresholded *t* values, $p < 0.025$).

possible patterns of prediction error signals on the scalp are possible, including apredictive patterns as we observed here, as precision may vary freely across task conditions (Kok et al., 2012). For example, Barascud et al. (2016) reported a larger MEG signal for auditory stimuli that become predictable, relative to stimuli that are entirely unpredictable, i.e., an apredictive pattern, that they linked to up-weighting of the expected stimuli by precision (Heilbron and Chait, 2018). Within predictive coding, attention is one specific mechanism that is thought to increase precision (Hohwy, 2012). Therefore, under a predictive coding framework, one can appeal to varying levels of attention across task conditions. Therefore, we could *post hoc* theorize that our late apredictive effect reflects individuals paying greater attention to the high validity trials as they have a high level of predictability and paying greater attention to related targets than unrelated targets, as the former fulfil their expectations. Therefore, the relative levels of attention across conditions could interact to generate this apredictive effect. Indeed, consistent with this, 59% of our participants (13/22) self-reported that their strategy was to generate an expectation in the high validity condition only (i.e., "I was trying to guess next word if previous was blue," where blue was high validity condition).

An alternative interpretation stems from evaluation of our behavioral data. When comparing the behavioral RT interaction with the ERP voltage interaction (Table 2; Fig. 4, respectively), both show the same pattern: namely, that the interaction is driven by expected items in a high validity context, showing more extreme values with respect to the other three conditions. This similarity in behavior and ERP effects suggest that our late "error" effect may simply reflect processing in service of behavior, whereby sensory signals are routed to goal-driven analogous motor behavior (Zylberberg et al., 2010). One formulation of how this may occur is provided by the Brain's router model (Zylberberg et al., 2010), in which a set of neurons (the router) connects incoming sensory information to a set of possible responses, while a task-set specifies the response to be executed in response to a given stimulus. Within the paradigm described here, one might consider that the router links the lexical and semantic representations of the prime word input (e.g., CAT) with the motor commands for a set of related subsequent target words (e.g., DOG, KITTEN, etc.). The task-set is specified by the prime validity, such that under high prime validity, the links between words and motor responses for related words are foregrounded/preactivated, e.g., the response "DOG" is set. Under the router model, and the related global neuronal workspace model (Dehaene and Christen, 2011), the neural substrate of routing to action is considered to lie within the broad frontoparietal network. Having a goal, i.e., having the intention to anticipate and pronounce a target word, will form a task-set that prepares the motor representation of the target word for execution when sufficient evidence about the target identity has been accumulated. Therefore, having this goal will speed RTs by preparing the link to the appropriate motor plan. Indeed, consistent with the role of goal-based expectations about which appropriate motor plan to execute, participants in the no-strategy group, who do not report engaging in any form of conscious expectation of target identity, show no speeding of RTs under high validity. Our late apredictive ERP pattern may therefore not reflect a precision-weighted global prediction error, but more simply the result of the brain routing the incoming information into appropriate behavior. Under this interpretation, our results are therefore also consistent with interpretations of early ERPs as reflections of prediction error and later

ERPs as processes related to conscious access and in support of task demands (Dehaene and Christen, 2011; Rohaut et al., 2015).

It is possible that other later error signals were also evident in the neural response during our task, including those traditionally linked to the N400 (i.e., peaking ~400 ms post-target). However, we limited our analyses to the 0- to 382-ms time window post-target so as to avoid muscle artifact created by the pronunciation responses. We chose to use a pronunciation task as our aim was to observe the behavioral effect produced by the manipulation of both the local (relatedness) and global context (prime validity) as implemented by Hutchison (2007). Nevertheless, tasks that do not produce large muscular artefacts, such as a lexical decision task (LDT) in which individuals only produce motor responses on filler trials, would allow for analysis of the N400 time window. However, as argued by Hutchison (2007), participants can complete an LDT with a semantic-matching strategy, meaning that after seeing the target they can verify whether it is related to the prime, which could bias their responses as only words can be related and nonwords would be, by their nature, unrelated (Hutchison, 2007). Additionally, as we provided a global context by manipulating the proportion of related items across the task, individuals could bias their responses using the validity cue (Keefe and Neely, 1990); for example, primes that were presented in blue (high validity context) were more likely to be related (80%). Therefore, when seeing a blue prime, individuals could judge their response (word/non-word) solely based on the prime, in this case a "word" as most of the word-pairs are related. Instead, using a pronunciation task allows for a purer measure of expectation, with the caveat of limiting the time window of artifact-free EEG for analysis. While our limit of 382 ms for analysis post-target likely excludes the later portion of the N400 (generally peaking at ~400 ms post-target), it nevertheless captures the onset of any putative N400, generally considered to onset at 200 ms poststimulus and to reflect semantic processing (Kutas and Federmeier, 2011). Furthermore, we can be confident that the 382-ms time window allows us to capture semantic processes as this time window was defined on the basis of RTs (minus a motor preparation period) that are themselves modulated by the semantic associations of the words.

A recent prediction error view on language-related ERPs proposes that the N400 has similar properties to the MMN, as they both are modulated by the predictability of stimuli (i.e., increased ERP amplitude as a prediction-error response) but that their relative latencies indicate prediction-error processing at different levels of stimulus complexity (Bornkessel-Schlesewsky and Schlesewsky, 2019). In our findings, both consecutive effects could be similarly interpreted as reflecting different levels of complexity of precision-weighted prediction error processing across a semantic hierarchy. However, as noted above, appeal to precision-weighting problematically allows for *post hoc* explanations of all possible ERP patterns (Bowman et al., 2013).

Regarding the source estimation analyses, the early effect was localized to the middle frontal gyrus, which has been previously associated with semantic categorization when compared with passive listening (Noesselt et al., 2003). Furthermore, the ERP source estimation analysis for the late effect was localized to the supplementary motor area, consistent with the above interpretation that the late interaction reflects goal-driven routing toward action. Indeed, this area has been linked to speech motor control, verbal working memory, and predictive top-down mechanisms in speech perception (Hertrich et al., 2016). However, neither of these two regions were part of our preregistered hypotheses. Therefore, these source estimates should be interpreted with caution, and future studies with this paradigm will wish to replicate these sources.

In our preregistered analyses, we also hypothesized that we would observe electrophysiological markers of differential expectations generated by the high and low validity primes, before the onset of the target. Specifically, we expected these differential expectations to be reflected in the ERPs, including the slope of a putative slow wave, or contingent negative variation (CNV; Chennu et al., 2013), and/or in the power of the EEG in the α/β bands, as these have been previously associated with the precision of expectations (Bauer et al., 2014). However, we found no evidence of any differences in these measures between high and low validity primes before target onset. The CNV is typically observed in the preparatory period before a temporally-expected target and is considered to reflect priming of the neural circuits required for a task-appropriate response, whether that be motoric (Gómez et al., 2001) or cognitive (Chennu et al., 2013). Indeed, the magnitude of the CNV has been linked to the amount of top-down expectation instantiated by a stimulus (Chennu et al., 2013), thus leading to our hypothesis of differential amounts of expectation instantiated by high validity and low validity primes. One interpretation of our data's lack of support for this hypothesis is that our specific measures were simply not sensitive enough to detect the differential expectations in these conditions. Indeed, we powered our study to detect the post-target behavioral effect specifically. Furthermore, EEG may lack the spatial sensitivity to detect subtle differences in preparatory motor and/or semantic neural circuits. An alternative interpretation is that expectations were, in fact, not different between the two conditions. Indeed, under predictive coding, the brain is considered to optimize the difference between its expectations and sensory input by updating its internal model (Friston, 2010); hence, it is possible that the optimal means of minimizing prediction error in this task is to always predict the related target, regardless of the prime validity. This would result in preparatory priming of the same circuits in expectation of the upcoming target item, and therefore equivalent CNVs. For example, even if one were to consciously expect that an upcoming target will be unrelated (as in a low validity trial), it is simply not possible to accurately predict the identity of that target, as the range of possible unrelated target words is considerable. Therefore, although predicting the identity of a specific related target had only a ~22% probability of being correct in a low validity context, it was still more likely than predicting any one of the vast arrays of potential unrelated target words. An optimal expectation then would be to always predict DOG in response to CAT, thus resulting in CNVs that do not differ across prime validity.

Nevertheless, our speculation here is based on a null result, and future inspection of participants' meta-cognition in relation to their specific expectations following prime presentation will help speak to this interpretation.

Conclusions

In conclusion, we here reported ERP evidence of hierarchical matching of semantic expectations to incoming speech. Lower lever expectations based on the local context (i.e., the prime identity) elicited an early and predictive pattern that matches with prediction error accounts. Higher level expectations generated from the global context required awareness of the global rule and the use of a reportable strategy, and were associated with an apredictive pattern that can be interpreted within a precision-weighted prediction error account, or may reflect the routing of sensory signals and their expectations into task-directed behavior. This later effect was only evident in exploratory analyses (i.e., not our preregistered analyses) and therefore requires further future replication.

References

Barascud N, Pearce MT, Griffiths TD, Friston KJ, Chait M (2016) Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. Proceedings of the National Academy of Sciences 113:E616–E625.

Bauer M, Stenner MP, Friston KJ, Dolan RJ (2014) Attentional modulation of alpha/beta and gamma oscillations reflect functionally distinct processes. J Neurosci 34:16117–16125.

Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L (2009) Neural signature of the conscious processing of auditory regularities. Proc Natl Acad Sci USA 106:1672–1677.

Berkum JJV, Hagoort P, Brown CM (1999) Semantic integration in sentences and discourse: evidence from the N400. J Cogn Neurosci 11:657–671.

Boudewyn MA, Long DL, Swaab TY (2015) Graded expectations: predictive processing and the adjustment of expectations during spoken language comprehension. Cogn Affect Behav Neurosci 15:607–624.

Bornkessel-Schlesewsky I, Schlesewsky M (2019) Towards a neurobiologically plausible model of language-related, negative event-related potentials. Front Psychol 10:298.

Bowman H, Filetti M, Wyble B, Olivers C (2013) Attention is more than prediction precision [Commentary on target article]. Behav Brain Sci 36:206–208.

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436.

Brookes MJ, Tewarie PK, Hunt BAE, Robson SE, Gascoyne LE, Liddle EB, Liddle PF, Morris PG (2016) A multi-layer network approach to MEG connectivity analysis. Neuroimage 132:425–438.

Brothers T, Swaab TY, Traxler MJ (2017) Goals and strategies influence lexical prediction during sentence comprehension. J Mem Lang 93:203–216.

Bubic A, Von Cramon DY, Schubotz RI (2010) Prediction, cognition and the brain. Front Hum Neurosci 4:25.

Chennu S, Noreika V, Gueorguiev D, Blenkmann A, Kochen S, Ibáñez A, Owen AM, Bekinschtein TA (2013) Expectation and attention in hierarchical auditory prediction. J Neurosci 33:11194–11205.

Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 36:181–204.

Cruse D, Beukema S, Chennu S, Malins JG, Owen AM, McRae K (2014) The reliability of the N400 in single subjects: implications for patients with disorders of consciousness. Neuroimage Clin 4:788–799.

Dehaene S, Christen Y (2011) Characterizing consciousness: from cognition to the clinic? San Diego: Springer Science and Business Media.

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. J Neurosci Methods 134:9–21.

Drake R, Vogl AW, Mitchell AW (2009) Gray's anatomy for students E-book. Amsterdam: Elsevier Health Sciences.

El Karoui I, King JR, Sitt J, Meyniel F, Van Gaal S, Hasboun D, Adam C, Navarro V, Baulac M, Dehaene S, Cohen L, Naccache L (2015) Event-related potential, time-frequency, and functional connectivity facets of local and global auditory novelty processing: an intracranial study in humans. Cereb Cortex 25:4203–4212.

Faugeras F, Rohaut B, Weiss N, Bekinschtein T, Galanaud D, Puybasset L, Bolgert F, Sergent C, Cohen L, Dehaene S, Naccache L (2012) Event related potentials elicited by violations of auditory regularities in patients with impaired consciousness. Neuropsychologia 50:403–418.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815–836.

Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11:127–138.

Friston K, Kiebel S (2009) Predictive coding under the free-energy principle. Phil Trans R Soc B 364:1211–1221.

Gómez CM, Delinte A, Vaquero E, Cardoso MJ, Vázquez M, Crommelinck M, Roucoux A (2001) Current source density analysis of CNV during temporal gap paradigm. Brain Topogr 13:149–159.

Heilbron M, Chait M (2018) Great expectations: is there evidence for predictive coding in auditory cortex? Neuroscience 389:54–73.

Hertrich I, Dietrich S, Ackermann H (2016) The role of the supplementary motor area for speech and language processing. Neurosci Biobehav Rev 68:602–610.

Hohwy J (2012) Attention and conscious perception in the hypothesis testing brain. Front Psychol 3:96.

Hutchison KA (2007) Attentional control and the relatedness proportion effect in semantic priming. J Exp Psychol Learn Mem Cogn 33:645–662.

Hutchison KA, Balota DA, Neely JH, Cortese MJ, Cohen-Shikora ER, Tse CS, Yap MJ, Bengson JJ, Niemeyer D, Buchanan E (2013) The semantic priming project. Behav Res 45:1099–1114.

JASP Team (2018) JASP (Version 0.9. 0.1) [Computer software].

Jeffreys H (1961) Theory of probability. Oxford: Oxford University Press.

Keefe DE, Neely JH (1990) Semantic priming in the pronunciation task: the role of prospective prime-generated expectancies. Mem Cognit 18:289–298.

King JR, Gramfort A, Schurger A, Naccache L, Dehaene S (2014) Two distinct dynamic modes subtend the detection of unexpected sounds. PLoS One 9:e85791.

Kleiner M, Brainard D, Pelli D (2007) What's new in Psychtoolbox-3? Perception 36, ECVP Abstract Supplement.

Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci 27:712–719.

Koivisto M, Revonsuo A (2001) Cognitive representations underlying the N400 priming effect. Brain Res Cogn Brain Res 12:487–490.

Kok P, Rahnev D, Jehee JF, Lau HC, De Lange FP (2012) Attention reverses the effect of prediction in silencing sensory signals. Cereb Cortex 22:2197–2206.

Kuperberg GR, Jaeger TF (2016) What do we mean by prediction in language comprehension? Lang Cogn Neurosci 31:32–59.

Kuperberg GR, Delaney-Busch N, Fanucci K, Blackford T (2018) Priming production: neural evidence for enhanced automatic semantic activity preceding language production in schizophrenia. Neuroomage Clin 18:74–85.

Kutas M, Federmeier KD (2011) Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu Rev Psychol 62:621–647.

Lapatki BG, Stegeman DF, Jonas IE (2003) A surface EMG electrode for the simultaneous observation of multiple facial muscles. J Neurosci Methods 123:117–128.

Lau EF, Holcomb PJ, Kuperberg GR (2013a) Dissociating N400 effects of prediction from association in single-word contexts. J Cogn Neurosci 25:484–502.

Lau EF, Gramfort A, Hämäläinen MS, Kuperberg GR (2013b) Automatic semantic facilitation in anterior temporal cortex revealed through multimodal neuroimaging. J Neurosci 33:17174–17181.

Lewis AG, Bastiaansen M (2015) A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. Cortex 68:155–168.

Lo S, Andrews S (2015) To transform or not to transform: using generalized linear mixed models to analyse reaction time data. Front Psychol 6:1171.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG-and MEG-data. J Neurosci Methods 164:177–190.

Mognon A, Jovicich J, Bruzzone L, Buiatti M (2011) ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology 48:229–240.

Noesselt T, Shah NJ, Jäncke L (2003) Top-down and bottom-up modulation of language related areas–an fMRI study. BMC Neurosci 4:13.

Nolan H, Whelan R, Reilly RB (2010) FASTER: fully automated statistical thresholding for EEG artifact rejection. J Neurosci Methods 192:152–162.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:156869.

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442.

Popov T, Oostenveld R, Schoffelen JM (2018) FieldTrip made easy: an analysis protocol for group analysis of the auditory steady state brain response in time, frequency, and space. Front Neurosci 12:711.

Rabovsky M, McRae K (2014) Simulating the N400 ERP component as semantic network error: insights from a feature-based connectionist attractor model of word meaning. Cognition 132:68–89.

Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87.

Robinson SE, Vrba J (1999) Functional neuroimaging by synthetic aperture magnetometry (SAM). Recent Advances in Biomagnetism, pp 302–305. Sendai:Tohoku Univ. Press.

Rohaut B, Faugeras F, Chausson N, King JR, El Karoui I, Cohen L, Naccache L (2015) Probing ERP correlates of verbal semantic processing in patients with impaired consciousness. Neuropsychologia 66:279–292.

Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G (2009) Bayesian t tests for accepting and rejecting the null hypothesis. Psychon Bull Rev 16:225–237.

Skrandies W (1990) Global field power and topographic similarity. Brain Topogr 3:137–141.

Sokoliuk R, Mayhew SD, Aquino KM, Wilson R, Brookes MJ, Francis ST, Hanslmayr S, Mullinger KJ (2019a) Two spatially distinct posterior alpha sources fulfil different functional roles in attention. J Neurosci 39:7183–7194.

Sokoliuk R, Calzolari S, Cruse D (2019b) Dissociable electrophysiological correlates of semantic access of motor and non-motor concepts. Sci Rep 9:1–14.

Thornhill DE, Van Petten C (2012) Lexical versus conceptual anticipation during sentence processing: frontal positivity and N400 ERP components. Int J Psychophysiol 83:382–392.

Van Doorn J, van den Bergh D, Bohm U, Dablander F, Derks K, Draws T, Etz A, Evans NJ, Gronau QF, Haaf JM, Hinne M, Kucharsky S, Ly A, Marsman M, Matzke D, Gupta ARKN, Sarafoglou A, Stefan A, Voelkel JG, Wagenmakers EJ (2019) The JASP guidelines for conducting and reporting a Bayesian analysis. PsyArXiv. doi:10.31234/osf.io/yqxfr.

Van Drongelen W, Yuchtman M, Van Veen BD, Van Huffelen AC (1996) A spatial filtering technique to detect and localize multiple sources in the brain. Brain Topogr 9:39–49.

Van Selst M, Jolicoeur P (1994) A solution to the effect of sample size on outlier elimination. Quart J Exp Psychol A 47:631–650.

Van Veen BD, Van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng 44:867–880.

Wacongne C, Changeux JP, Dehaene S (2012) A neuronal model of predictive coding accounting for the mismatch negativity. J Neurosci 32:3665–3678.

Wagenmakers EJ, Love J, Marsman M, Jamil T, Ly A, Verhagen J, Selker R, Gronau QF, Dropmann D, Boutin B, Meerhoff F, Knight P, Raj A, van Kesteren EJ, van Doorn J, Šmíra M, Epskamp S, Etz A, Matzke D, de Jong T, et al. (2018) Bayesian inference for psychology. Part II: example applications with JASP. Psychon Bull Rev 25:58–76.

Weber K, Lau EF, Stillerman B, Kuperberg GR (2016) The yin and the yang of prediction: An fMRI study of semantic predictive processing. PLoS One 11:e0148637.

Yekutieli D, Benjamini Y (1999) Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. J Stat Plan Inf 82:171–196.

Ylinen S, Huuskonen M, Mikkola K, Saure E, Sinkkonen T, Paavilainen P (2016) Predictive coding of phonological rules in auditory cortex: a mismatch negativity study. Brain Lang 162:72–80.

Zylberberg A, Slezak DF, Roelfsema PR, Dehaene S, Sigman M (2010) The brain's router: a cortical network model of serial processing in the primate brain. PLoS Comput Biol 6:e1000765.