

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21

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}

¹School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.

²School of Computing, University of Kent, Canterbury, Kent, CT2 7NF, UK.

³Centre for Human Brain Health, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.

Corresponding authors email: Damian Cruse (d.cruse@bham.ac.uk); Consuelo Vidal-Gran (cxv648@student.bham.ac.uk).

Conflict of Interest: Authors report no conflict of interest.

Funding sources: This study was supported by a PhD scholarship given by the Chilean National Agency for Research and Development (Agencia Nacional de Investigación y Desarrollo, ANID), from the Government of Chile; the Medical Research Council (Damian Cruse PI: MR/P013228/1); and the School of Psychology from the University of Birmingham, Birmingham, UK.

22 **Abstract**

23

24 Perception is facilitated by a hierarchy of expectations generated from context and prior knowledge.

25 In auditory processing, violations of local (within-trial) expectations elicit a mismatch negativity,

26 while violations of global (across-trial) expectations elicit a later positive component (P300). This

27 result is taken as evidence of prediction errors ascending through the expectation hierarchy.

28 However, in language comprehension, there is no evidence that violations of semantic expectations

29 across local-global levels similarly elicit a sequence of hierarchical error signals – thus drawing into

30 question the putative link between event-related potentials and prediction errors. We investigated

31 the neural basis of such hierarchical expectations of semantics in a word-pair priming paradigm. By

32 manipulating the overall proportion of related or unrelated word-pairs across the task, we created

33 two global contexts that differentially encouraged strategic use of primes. Across two experiments,

34 we replicated behavioural evidence of greater priming in the high validity context, reflecting

35 strategic expectations of upcoming targets based on ‘global’ context. In our pre-registered EEG

36 analyses, we observed a ‘local’ prediction error ERP effect (i.e. semantic priming) approximately

37 250ms post-target, which, in exploratory analyses, was followed 100ms later by a signal that

38 interacted with the global context. However, the later effect behaved in an apredictive manner - i.e.

39 was most extreme for fulfilled expectations, rather than violations. Our results are consistent with

40 interpretations of early ERPs as reflections of prediction error and later ERPs as processes related to

41 conscious access and in support of task demands.

42

43

44

45

46 **Significance statement**

47

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

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70 **Introduction**

71

72 Predictive coding theory argues that the brain processes information in a hierarchical
73 probabilistic Bayesian manner (Friston 2005; Knill & Pouget, 2004) by contrasting sensory input with
74 prior expectations generated from context and the perceiver's knowledge (Heilbron and Chait,
75 2018; Clark, 2013). Expectations are sent down from higher levels of the hierarchy and any
76 subsequent unexplained sensory input is sent back up the hierarchy as prediction error (Heilbron
77 and Chait, 2018; Friston and Kiebel, 2009; Rao and Ballard, 1999).

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

91 Within the realm of more ecologically valid stimulus processing, speech comprehension is
92 similarly influenced by expectations at multiple levels of a hierarchy (e.g. Lewis, Bastiaansen, 2015;
93 Ylinen et al., 2016; Lau et al., 2013; Hutchison, 2007; Kuperberg, Jaeger, 2016). The N400 – a

94 negative deflection peaking around 400ms post-stimulus (Kutas, Federmeier, 2011) – is a potential
95 marker of errors of such semantic expectations (Rabovsky & McRae, 2014). On a local level, the
96 N400 is larger to words that have not been primed relative to those that have (e.g. larger for DOG
97 when preceded by Lamp than by Cat; Cruse et al., 2014; Lau et al., 2013; Koivisto & Revonsuo, 2001),
98 and at a more global level, the N400 is larger to words that are unexpected within a sentential
99 context (Brothers et al., 2017; Boudewyn, Long & Swaab, 2015; Thornhill, Van Petten 2012; Van
100 Berkum et al., 1999). Interestingly, unlike the MMN/P3b in auditory processing, semantic prediction
101 errors appear to be reflected in the magnitude of a single component –the N400– rather than in a
102 series of components moving through the hierarchy of relative top-down involvement.

103 One approach to separate prediction error signals at two levels of a semantic expectation
104 hierarchy is with a prime validity manipulation of a word-pair priming task. Specifically, we can pit
105 the facilitation of target word processing that comes from presentation of a related prime against a
106 global context in which it is not efficient for the comprehender to use the prime to predict the target
107 – i.e. primes rarely followed by related targets (Keefe and Neely, 1990; Hutchison, 2007; Lau et al.,
108 2013(a); Lau et al., 2013(b)). Therefore, as the proportion of related pairs increases within a context,
109 the prime validity increases (i.e. the prime is more likely to predict the target). If individuals use the
110 global context of prime validity to modulate their expectations, behavioural facilitation follows.

111 In ERP studies of prime validity, this hierarchy of local expectations (i.e. the prime
112 relatedness) and global expectations (i.e. the prime validity) has not been reported to modulate the
113 amplitudes of two sequential components (Boudewyn, Long & Swaab, 2015; Lau et al., 2013); hence,
114 there is no evidence of a two-stage profile to semantic expectation violation. Rather than reflecting
115 error at one level, the N400 (or see Boudewyn, Long & Swaab (2015) for N200 evidence) appears to
116 account for a combination of errors across levels of the hierarchy. To disentangle these results, here
117 we report a pre-registered trial-by-trial manipulation of both local and global semantic expectations.

118 First, we report a replication of the reaction time facilitation caused by global context as described
119 by Hutchison (2007). Second, we report the associated electrophysiological markers of expectation
120 and violation across levels of the hierarchy from a separate group of healthy participants performing
121 the same task. In accordance with predictive coding, we hypothesised that ERP amplitudes would
122 reflect violations of expectation at consecutive levels of the hierarchy, with local violations evident
123 earlier than global violations.

124

125 **Materials and Methods**

126

127 **Experiment 1 – Behavioural study**

128 Participants

129 We recruited participants through the Research Participation Scheme website of the
130 University of Birmingham, who received credits for their participation. A total of 64 participants
131 were recruited, with the data of two participants excluded from analysis due to outlying data, as
132 quantified by the non-recursive procedure for outlier elimination (detailed below; Van Selst,
133 Jolicoeur, 1994; Hutchison, 2007). Therefore, the final sample consisted of 62 participants (59
134 female, 3 male; median age: 19, range: 18 – 28). All participants reported to be mono-lingual native
135 English speakers, right-handed, and with no history of neurological conditions or diagnosis of
136 dyslexia. All participants gave written informed consent prior to participation in this study, which
137 was approved by the STEM Ethical Review Committee of the University of Birmingham.

138

139

140

141

142 Stimuli

143 Associated prime-target pairs were selected from the Semantic Priming Project database
144 (Hutchison et al., 2013) and the experimental design was a replication of the paradigm implemented
145 by Hutchison (2007). First, all word pairs available in the database (N: 1661) were ordered by For-
146 ward Associative Strength (i.e. the proportion of individuals who spontaneously name the same tar-
147 get after reading the prime word) and the 352 word-pairs with the highest strength were selected
148 after removal of any specific American English associations (e.g. Clorox-Bleach; Slacks-Pants).

149 The first 156 word-pairs from this list of 352 word-pairs with the highest forward association
150 were chosen to be the critical stimuli for statistical analysis. The remaining 196 word-pairs served
151 as fillers to generate the global context and are not included in the statistical analysis. We divided
152 all 156 critical word-pairs into two lists (N: 78 word-pairs per list) that were balanced according to
153 the values from the database (Hutchison et al., 2013) for forward association, length, log HAL fre-
154 quency, and orthographic neighbourhood (all $p > .604$; all $BF_{10} < .196$). In the same way, we divided
155 the 196 filler word-pairs into two balanced lists (N: 98 word-pairs per list; all $p > .284$, all $BF_{10} < .267$).
156 Thus, we had created two critical related word-pair lists and two filler related word-pair lists. To
157 create the unrelated word-pair lists, we manually re-paired (within list) all word-pairs in each of the
158 four lists above (two critical, two fillers) ensuring that unrelated targets were both semantically un-
159 related to their prime and shared no overlapping phonemes with their respective related target.
160 This resulted in a final set of eight lists: two critical related, two critical unrelated, two filler related,
161 and two filler unrelated. Each participant was assigned two Critical sets of word-pairs (one related
162 and one unrelated; 78 word-pairs per list) and two Filler sets (one related and one unrelated; 98
163 word-pairs per list). Hence, each participant saw all words within the full set of 352 word-pairs ex-
164 actly once, composed of 176 related word-pairs and 176 unrelated word-pairs.

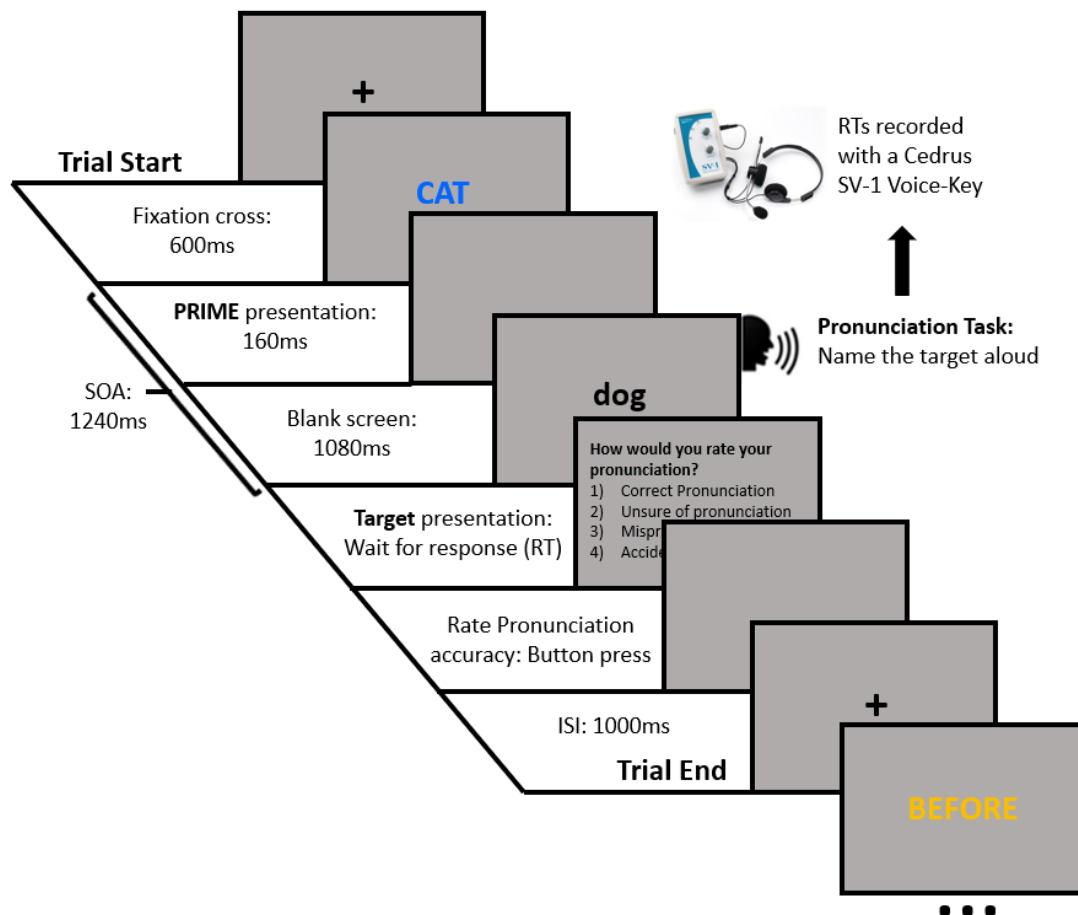
165 To create the prime-validity manipulation, first we assigned half of the Critical word-pairs,
166 including both related and unrelated items, to one colour (yellow or blue), and the other half with
167 the other colour in an interleaved order. Next, the related filler set was assigned with one colour
168 (yellow or blue), and the unrelated filler set was assigned with the other colour. Therefore, across
169 all items seen by each participant, 77.8% of word-pairs presented in one of the two colours were
170 related, thus giving that colour high prime validity, and 77.8% of word-pairs presented in the other
171 colour were unrelated, thus giving that colour low prime validity. Importantly, across the entire set
172 of stimuli that each participant saw, exactly half were related (the other half unrelated) and half
173 were presented in one colour (the other half in the other colour). However, the probability of a
174 related target following a prime of one colour was 77.8% and the probability of a related target
175 following a prime of the other colour was 22.2%. Across participants, the colour assignment of the
176 high validity primes was counterbalanced (i.e. half of participants saw high prime validity word-pairs
177 in blue and low prime validity word-pairs in yellow; and the other half saw the opposite colours for
178 each proportion), and all possible combinations of word lists were used, resulting in 32 permuta-
179 tions.

180

181 Procedure

182 The task was presented with Psychtoolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007)
183 in Matlab (Mathworks, Inc., Natick, Massachusetts). The vocal reaction times (RT) were measured
184 with a Cedrus SV-1 Voice Key (Cedrus Corporation), with all participants completing four practice
185 trials under the experimenter's supervision to adjust the voice key threshold according to the par-
186 ticipant's speech volume. The trial procedure is shown in Figure 1. Specifically, each trial started
187 with a central fixation cross on a grey background lasting 600 ms; then, the prime word was dis-
188 played in either yellow or blue, at the centre of the screen for 160 ms; followed by a blank screen

189 for 1080ms, and subsequently the target was displayed on the screen; thus, the stimulus onset asyn-
190 chrony (SOA) was 1240ms. The target stayed on the screen until the participant pronounced the
191 word; then the word disappeared from the screen, which remained blank for 300ms. Afterwards, a
192 rating for the quality of pronunciation was displayed on the screen with the following questions and
193 potential responses: How would you rate your pronunciation? 1) Correct pronunciation; 2) Unsure
194 of pronunciation; 3) Mispronunciation; 4) Accidental voice-key triggering. Participants gave a button
195 response on the keyboard (1-4) to rate their pronunciation (as per Hutchison, 2007). After the par-
196 ticipant responded, the screen remained blank for 1000ms, before the next trial began.



197

198 **Figure 1:** Semantic Priming Relatedness Proportion task (Hutchison, 2007). Participants were required to name
199 the target word aloud and as fast as possible, while their responses were recorded.

200 Each participant was tested individually and sat approximately 70 cm away from the com-
201 puter screen. All participants received written information about the study, the instructions and the
202 consent form. In addition, the instructions were verbally repeated by the experimenter. We in-
203 structed all participants that a coloured uppercase word (either blue or yellow) will be displayed on
204 the screen and that they must read it silently to themselves; then, a black lowercase word will be
205 displayed on the screen, and they should pronounce the word aloud, as fast and accurately as pos-
206 sible. Participants were told that the colour of the uppercase word will cue the probability of the
207 lowercase target being related or unrelated. Half of the participants received the following written
208 instructions: “If the uppercase word is Blue, it is highly likely that the meaning of the lowercase word
209 will be related; and if the uppercase word is Yellow, it is highly likely that the meaning of the lower-
210 case word will be unrelated” (as per Hutchison, 2007). The other half of participants received the
211 same instructions but with the colours flipped.

212 After the task, we asked participants to complete a self-report form about the use of strat-
213 egy throughout the task, to determine whether they were using expectations strategically. The form
214 was composed of three questions and a free text description of the strategy. The questions were
215 the following: 1) Which colour was highly likely to be related? (Responses: BLUE / YELLOW); 2) Did
216 you use the colour of the UPPERCASE word (BLUE, YELLOW) as a cue for knowing whether the fol-
217 lowing word was related or unrelated? (Responses: YES / NO); 3) Did you engage in any strategy to
218 speed up your responses using the colour cue? (Responses: YES / NO); 4) If YES, briefly describe. We
219 considered participants to have used strategic expectation (i.e. those referred to as the Strategy
220 group) if they correctly identified the colour that was assigned for the high validity condition (Ques-
221 tion 1), answered YES in questions 2 and 3, and described a strategy in question 4. All other partici-
222 pants were classified into the No Strategy group.

223

224 Behavioural Data Analyses

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

240 All behavioural analyses were conducted in Jasp 0.9.1.0 software (JASP Team, 2018). To test
241 for an effect of global context on reaction times, we conducted a two-way repeated measures
242 ANOVA with factors of relatedness (i.e. related vs unrelated targets) and prime validity (i.e. high vs
243 low prime validity). We also reported equivalent Bayesian Repeated Measures ANOVAs (Van Doorn
244 et al. 2019; Wagenmakers et al., 2018). We expected individuals to show faster RTs for related
245 (expected) in contrast with unrelated (unexpected) targets due to local level expectations – i.e.
246 priming. Furthermore, we expected an interaction, with larger priming effects in a high validity

247 context in contrast with a low validity context, reflecting the use of global level context to predict
248 upcoming stimuli.

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

252

253 **Experiment 2 – Behavioural and electrophysiological study**

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

257

258 Participants

259 We recruited participants through the Research Participation Scheme website and placed
260 advertisement posters at the University of Birmingham; participants received a monetary
261 compensation for their participation. We recruited 37 participants, however, since we only
262 investigated those who reported using a strategy, the final sample only included 22 participants (15
263 female, 7 male; median age: 21, range: 18 - 30; classified by the same report form as experiment 1).
264 The inclusion criteria were the same as those for Experiment 1; however, participants were also
265 required to attend for a structural T1-weighted MRI scan at the University of [name redacted for
266 double-blind review]; therefore, participants who had any metal parts in their body, were
267 claustrophobic, or women who were pregnant were excluded from the study, as the scan was
268 mandatory for participation. All participants gave written informed consent prior to participation in
269 this study, which was approved by the STEM Ethical Review Committee of the University of
270 Birmingham.

271 We aimed to detect a reaction time interaction of the same magnitude as seen in the
272 Strategy group of Experiment 1; therefore, we conducted a power analysis to select an appropriate
273 sample size for this goal. We performed non-parametric power calculations using the data of all
274 participants of the Strategy group from Experiment 1. Specifically, from the pool of participants of
275 the Strategy group, we selected with replacement N participants and conducted the same two-way
276 repeated measures ANOVA 1000 times to test for the reaction time interaction effect. With an N of
277 22 participants in the Strategy group we achieved 80% power at $p < .05$ (i.e. 80% of ANOVAs included
278 a significant interaction).

279 As we did not know if a participant was in the Strategy group until their self-report form was
280 completed at the end of the study, we recruited participants until 22 of them were classified as
281 being in the Strategy group (median age: 21, range: 18-30; 12 in the no-strategy group, median age:
282 22, range: 19-33). After removal of trials rated as mispronunciations and those considered outliers
283 according to the non-recursive outlier elimination procedure of Van Selst and Jolicoeur (1994; as
284 Experiment 1), a median of 28 trials (range: 11-38) contributed to the high related condition; a
285 median of 29.5 trials (range: 13-38) to the high unrelated condition; a median of 29 trials (range: 12-
286 39) to the low related condition; and a median of 28 (range: 14-37) contributed to the low unrelated
287 condition.

288

289 Stimuli and procedure

290 Stimuli and procedure were the same as in Experiment 1, except for the duration of the fixation
291 cross (increased from 600ms to 750ms to provide more time for an EEG time-frequency baseline).

292

293

294

295 EEG recording

296 The EEG signal was continuously recorded with a 125 channel AntNeuro EEG system (Ant-
297 Neuro b.v., Enschede, Netherlands) at a sampling rate of 500 Hz, with impedances kept below 20
298 k Ω . We placed the ground electrode on the left mastoid bone and referenced online to CPz. As
299 participants were required to pronounce words aloud, we also recorded a bipolar EMG signal with
300 one EMG electrode above the upper lip and the other below the lower lip on the left side of the
301 mouth; approximately over the superior and inferior Orbicularis Oris muscles (Lapatki, Stegeman &
302 Jonas, 2003; Drake, Vogl & Mitchell, 2009).

303

304 EMG Pre-processing

305 As this task involved participants speaking, there were considerable artefacts in the EEG
306 data around the vocal reaction time that were challenging to remove adequately. We therefore
307 chose to analyse only the EEG data up to the point of vocal artefact. To minimise artefacts from
308 additional preparatory muscular activity prior to vocal onset, in our pre-registered methods, we
309 planned to choose the latest time-point for analysis post-target by identifying when the mouth EMG
310 signal began to significantly differ between prime validity conditions in a temporal cluster mass ran-
311 domisation test, as implemented in Fieldtrip (Oostenveld et al., 2011). However, this approach re-
312 vealed no significant clusters (smallest cluster $p = 0.513$), and so did not provide a suitable cut-off
313 time-point for our analyses. Therefore, in a deviation from the pre-registered plan, we chose our
314 latest time-point of EEG data to analyse as 150ms prior to the fastest mean RT across conditions (in
315 this instance High Validity – Related = 532ms; see Kuperberg et al., 2018, for a similar approach).
316 Our post-target time-window therefore continued to 382ms post-target. From all the trials included
317 for the statistical analysis only 5.76% of trials had RTs earlier than this time-point, comparable with
318 previous studies (Kuperberg et al., 2018).

319 EEG Pre-Processing Pipeline

320 We low pass filtered the continuous EEG data at 40Hz using the finite impulse response filter
321 implemented in EEGLAB (Delorme & Makeig, 2004). Due to our interest in analysing slow-waves (see
322 below), we performed no high-pass filtering. Next, we segmented the filtered EEG signals into
323 epochs from 750ms before the onset of the prime up to 382ms post-target (see above for details).
324 Subsequent artefact rejection proceeded in the following steps based on a combination of methods
325 described by Nolan et al. (2010) and Mognon et al. (2011).

326 First, as in the behavioural data analysis, we excluded all trials in which the participant rated
327 their response as incorrect (i.e. 2, 3, 4 button press) and those that had reaction times that were
328 classified as outliers in the Non Recursive Procedure for outlier elimination (Selst & Jolicoeur, 1994).
329 Next, bad channels were identified and removed from the data. We considered a channel to be bad
330 if its absolute z-score across channels exceeded 3 on any of the following metrics: 1) variance of the
331 EEG signal across all time-points, 2) mean of the correlations between the channel in question and
332 all other channels, and 3) the Hurst exponent of the EEG signal (estimated with the discrete second
333 order derivative from the Matlab function `wfbmesti`). After removal of bad channels, we identified
334 and removed trials containing non-stationary artefacts. Specifically, we considered a trial to be bad
335 if its absolute z-score across trials exceeded 3 on any of the following metrics: 1) the mean across
336 channels of the voltage range within the trial, 2) the mean across channels of the variance of the
337 voltages within the trial, and 3) the mean across channels of the difference between the mean volt-
338 age at that channel in the trial in question and the mean voltage at that channel across all trials.
339 After removal of these individual trials, we conducted an additional check for bad channels, and
340 removed them, by interrogating the average of the channels across all trials (i.e. the ERP, averaged
341 across all conditions). Specifically, we considered a channel to be bad in this step if its absolute z-
342 score across channels exceeds 3 on any of the following metrics: 1) the variance of voltages across

343 time within the ERP, 2) the median gradient of the signal across time within the ERP, and 3) the
344 range of voltages across time within the ERP.

345 To remove stationary artefacts, such as blinks and eye-movements, the pruned EEG data
346 was subjected to an independent component analysis with the runica function of EEGLAB. The
347 Matlab toolbox ADJUST (Mognon et al., 2011) subsequently identified which components reflect
348 artefacts on the basis of their similarity to stereotypical spatio-temporal patterns associated with
349 blinks, eye-movements, and data discontinuities, and the contribution of these artefact components
350 was then subtracted from the data. Next, we interpolated the data of any previously removed chan-
351 nels via the spherical interpolation method of EEGLAB and re-referenced the data to the average of
352 the whole head.

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

359 Single-subject data for the time-frequency analysis were pre-processed in a similar way.
360 However, first, we concatenated the individual trials into a matrix of channels x all time-points, and
361 filtered each channel in two-steps (high-pass then low-pass) to retain the frequency bands of inter-
362 est (i.e. 8-12Hz alpha, and 13-30Hz beta), using EEGLAB's finite impulse response filter (function:
363 pop_eegnewfilt). Next, we extracted the squared envelope of the signal (i.e. the squared complex
364 magnitude of the Hilbert-transformed signal) to provide a time-varying estimate of power within
365 that frequency band. The resulting time-course was re-segmented into its original epochs and aver-
366 aged within each condition separately using SPM12's robust averaging procedure. As with the ERP

367 analyses, we low-pass filtered the resulting average time-series below 20Hz (EEGLAB's pop_newee-
368 gfilt). Finally, we converted the power estimates to decibels relative to the mean of the baseline
369 window (-200 – 0 ms.).

370

371 EEG / MRI co-registration

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

378

379 Analyses

380 Behavioural Data Analysis:

381 The behavioural analyses are the same as for the Strategy Group in Experiment 1.

382

383 EEG Analysis:

384 *Target ERP, Prime ERP and Prime time frequency analyses:*

385 Time-courses (ERPs / time-frequency) within the time-window of interest (0-1240ms for
386 primes; 0-382ms for targets) were compared with the cluster mass method of the open-source
387 Matlab toolbox FieldTrip (Oostenveld et al., 2011). This procedure involves an initial parametric step
388 followed by a non-parametric control of multiple comparisons (Maris, Oostenveld, 2007).
389 Specifically, we conducted two-tailed dependent samples t-tests at each spatio-temporal data-point
390 within our time-window of interest. Spatiotemporally adjacent electrodes (t-values) with p-values

391 < 0.05 were then clustered based on their proximity, with the requirement that a cluster must span
392 more than one time-point and at least 4 neighbouring electrodes, with an electrode's
393 neighbourhood containing all electrodes within an approximately 4-cm radius (median: 8, range:2-
394 10). Finally, we summed the t-values at each spatio-temporal point within each cluster. Next, we
395 estimated the probability under the null hypothesis of observing cluster sum Ts more extreme than
396 those in the experimental data - i.e. the p-value of each cluster. Specifically, Fieldtrip randomly
397 shuffles the trial labels between conditions, performs the above spatio-temporal clustering
398 procedure, and retains the largest cluster sum T. Consequently, the p-value of each cluster observed
399 in the data is the proportion of the largest clusters observed across 1000 such randomisations that
400 contain larger cluster sum Ts. As our analyses were two-tailed, we set the family-wise error
401 corrected cluster alpha to .025.

402

403 *Prime slow wave linear fit analyses:*

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

410

411 *Source estimation analysis:*

412 We constructed individual boundary element head models (BEM; four layers) from subject-
413 specific T1-weighted anatomical scans, by using the 'dipoli' method of the Matlab toolbox FieldTrip
414 (Oostenveld et al., 2011). Next, we aligned the electrode locations, that were recorded with Xensor

415 Electrode Digitizer device, to the surface of the scalp layer that was segmented from the T1-
416 weighted anatomical scan. For reference points, we used the fiducial points and electrode locations
417 as head shape. We visually checked that the electrode positions and the scalp surface were aligned,
418 and we manually fixed imperfections. We prepared the EEG data before subjecting it to statistical
419 analyses, where we balanced the number of trials in each condition, by taking the smallest condition
420 N as a reference and randomly discarding trials from the other conditions surpassing that N,
421 resulting in equal datasets.

422

423 *ERPs whole brain*

424 For the whole brain ERP source analysis, we used single-trial data that had not been sub-
425 jected to robust averaging, and defined trials as time windows from -382 to 382ms relative to target
426 onset. This data was then band-pass filtered between 1 and 40Hz using a firws filter as implemented
427 in Fieldtrip (Oostenveld et al., 2011). Subsequently, relative to the different conditions, data were
428 divided into seven sets: one containing all trials, one containing only related trials, one only unre-
429 lated trials, one all high-validity related and one all low-validity related trials, one containing all high-
430 validity unrelated and one all low-validity unrelated trials. The sensor covariance matrix was esti-
431 mated for all these sets of data in the time window -382 – 382ms relative to target onset. A common
432 spatial filter was then computed on the dataset containing all trials using a Linear Constraint Mini-
433 mum Variance (LCMV) beamformer (VanDrongelen, 1996; VanVeen, 1997; Robinson, 1999). Beam-
434 former parameters were chosen including a fixed dipole orientation, a weighted normalisation (to
435 reduce the center of head bias), as well as a regularisation parameter of 5% to increase the signal
436 to noise ratio (cf. Popov et al., 2018; Sokoliuk et al., 2019). This common spatial filter served then
437 for source estimation of the remaining six sets of trials. Subsequently, the dipole moments of the
438 different source estimates were extracted within the post-stimulus time windows of interest (time

439 windows for source estimates of related vs. unrelated trials: 226-280ms; 232-290ms; 306-382ms;
440 316-350ms; time window to test interaction effect for source estimates of highly related and unre-
441 lated trials and low related and unrelated trials: 316-350ms) and their absolute values averaged
442 over time to obtain one average source estimation value per grid point (VE) and condition.

443 To test for significant differences between conditions we conducted five contrasts as men-
444 tioned above; first, an interaction between prime validity (High/Low) and relatedness of the target
445 (Related/Unrelated) in a time-window from 316 to 350ms; next, we tested the early and late main
446 effects of relatedness of the target (Related/Unrelated) as observed in the sensor analyses results
447 (four main effects), in their respective time windows for the early effect (226-280ms and 232-
448 290ms); and the late effect (306-382ms and 316-350ms). Montecarlo Cluster-based permutation
449 tests were computed as implemented in Fieldtrip (Oostenveld et al., 2011) by using averaged data
450 over each time-window; moreover, we used an alpha and a cluster alpha level of 0.025 and 1000
451 permutations.

452

453 *Automated Anatomical Labelling (AAL) analysis:*

454 We tested for the post-target interaction, between the relatedness of the target
455 (related/unrelated) and the validity of the prime (High prime validity/Low prime validity) in five
456 specific anatomical regions of interest that are defined using the automated anatomical labelling
457 (AAL) atlas (see Brookes et al., 2016; Sokoliuk et al., 2019 for similar analyses with MEG and EEG
458 data). The selected regions are the Left inferior frontal gyrus (LIFG), including pars opercularis, pars
459 triangularis and pars orbitalis; the posterior Left middle temporal gyrus (LMTG); and posterior Left
460 superior temporal gyrus (LSTG), as Weber et al. (2016) reported a relatedness proportion interaction
461 in these regions. In addition, we tested the post-target interaction in the anterior LMTG and anterior
462 LSTG, as Lau et al. (2014) found differences in the anterior left superior temporal region (LSTG) in

463 related vs. unrelated items in a high validity condition. Moreover, as a deviation from our
464 preregistered analyses, we tested the main effects found in the Related – Unrelated contrast at the
465 sensor level (ERPs) in the same anatomical regions (more details in results section). To determine
466 both the anterior and posterior parts of the LMTG and LSTG, we calculated the centre of mass of
467 each AAL region and selected all virtual electrodes that were anterior or posterior to the centre of
468 mass.

469 We aggregated the AAL regions of interest to each participant’s T1-weighted image; next,
470 for each participant individually, we extracted the average source estimation values of all VEs (from
471 prior source estimation (cf. *ERPs whole brain*)) within each AAL region, weighted them according to
472 their Euclidian distance to the centre of mass of the AAL region (Brookes et al., 2016) and averaged
473 over VEs within each AAL region of interest. We then conducted paired-sample t-tests between the
474 post-target conditions (SP-High validity / SP-Low validity) for all AAL regions; and another paired-
475 sample t-test between the relatedness conditions (Related / Unrelated) for each AAL region in four
476 time windows (226-280ms; 232-290ms; 316-350ms; 306-382ms) from the main effects obtained in
477 the sensor level ERP analyses (results section). The p-values that we obtained were corrected for
478 multiple comparisons across AAL regions using False Discovery Rate, FDR (Yekutieli, Benjamini,
479 1999). Furthermore, to test for evidence for the null hypothesis, we calculated Bayes Factors using
480 the Bayes equivalent t-test, according to Rouder et al. (2009).

481

482

483

484

485

486

487 **Results**

488

489 **Experiment 1 – Behavioural only**

490 In a two-way repeated measures ANOVA, we found a significant interaction between prime
491 validity and relatedness of the target ($F(1, 61) = 13.751, p < 0.001, \eta_p^2 = 0.184$), which was also
492 supported by a Bayesian Repeated Measures ANOVA ($BF_{inclusion} = 19.25$). As shown in Table 1, this
493 interaction stems from the larger semantic priming effect in the high prime validity context
494 ($t(61) = -6.525, p < 0.001, \text{Cohen's } d = -0.829, \text{CI} = -1.115 -0.537$) relative to the low prime validity
495 context ($t(61) = -5.169, p < 0.001, \text{Cohen's } d = -0.656, \text{CI} = -0.929 -0.380$). Furthermore, reaction
496 times to unrelated items were markedly similar across contexts ($t(61) < .001, p = 0.999, \text{Cohen's } d$
497 $< 0.001, \text{CI} = -0.249 0.249$), while the difference in semantic priming stems from significantly
498 different reaction times to related items ($t(61) = -3.797, p < 0.001, \text{Cohen's } d = -0.482, \text{CI} = -0.744 -$
499 0.217).

500

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

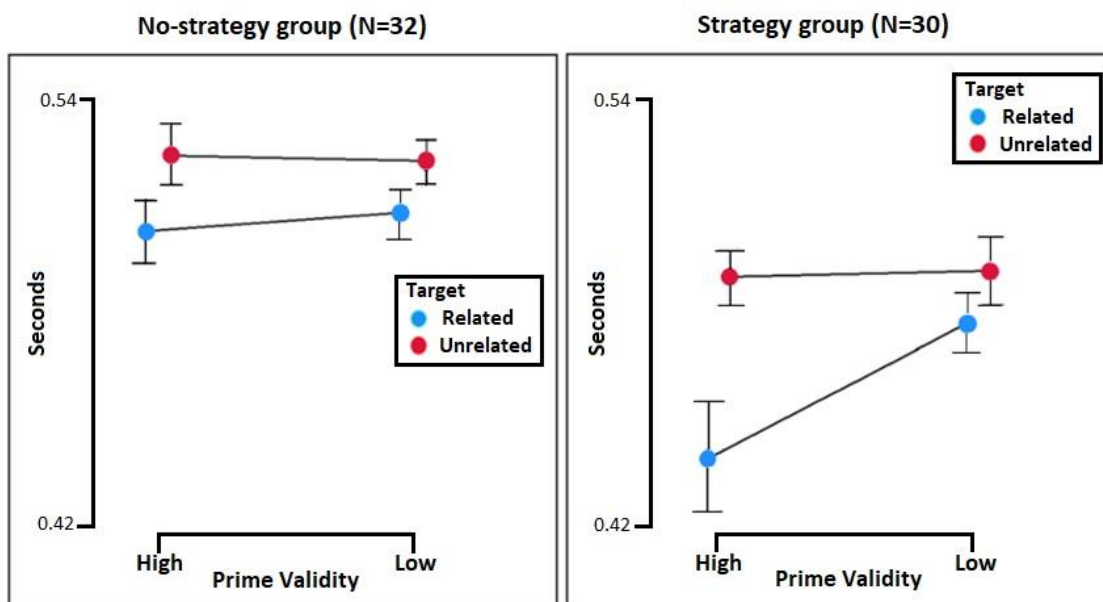
501

502 **Table 1:** Descriptive statistics including Mean RT (ms) and standard deviation of related and unrelated
503 word-pairs on each validity context, High Prime Validity and Low Prime Validity. Semantic priming effects and
504 prime validity effect (relatedness proportion effect).

505

506

507 Of 62 participants, 32 were classified in the “No-Strategy” group and 30 were classified in
508 the “Strategy” group. A post-hoc mixed design ANOVA with two within factors (Relatedness of
509 Target; Validity of the prime) and one between subjects factor (Strategy; No-strategy) revealed a
510 significant Target * Prime Validity * Strategy interaction ($F(1, 60) = 7.537, p = 0.008, \eta^2 = 0.112,$
511 $BF_{inclusion} = 3.203$), reflecting the apparent presence of a prime validity effect when participants
512 reported using the prime strategically ($F(1, 29) = 20.388, p < 0.001, \eta^2 = 0.413; BF_{inclusion} = 34.67$)
513 but absence of a prime validity effect when participants reported no strategy ($F(1, 31) = 0.860, p =$
514 $0.361, \eta^2 = 0.027; BF_{inclusion} = 0.393$; Figure 2). The No strategy group, however, did exhibit a
515 significant semantic priming effect by showing faster responses in the related relative to unrelated
516 items ($F(1, 31) = 21.656, p < 0.001, \eta^2 = 0.411; inclusion BF_{inclusion} = 4994.57$).



517 **Figure 2:** Mean RTs: Prime Validity (High / Low), Relatedness of the target (Related / Unrelated). Interaction
518 ($p < 0.001$) between the validity of the prime and the relatedness of the target in the group of participants that
519 reported the use of a conscious strategy (right), and no interaction ($p = 0.361$) in the group of participants that
520 did not report a conscious strategy (left).

521

522 **Experiment 2**

523

524 **Behavioural Results**

525 These results were qualitatively consistent with those we observed in Experiment 1. A two-
526 way repeated measures ANOVA analysis showed a significant interaction between prime validity
527 and relatedness of the target ($F(1, 21) = 9.071, p = 0.007, \eta p^2 = 0.302$), while the Bayesian Repeated
528 Measures ANOVA analysis showed anecdotal evidence for the interaction ($BF_{inclusion} = 2.519$). The
529 interaction was driven by a larger semantic priming effect in the high prime validity context
530 ($t(21) = -4.254, p < 0.001, \text{Cohen's } d = -0.907, CI = -1.398 -0.400$) than in the low prime validity
531 context ($t(21) = -2.046, p = 0.054, \text{Cohen's } d = -0.436, CI = -0.869 0.007$), see table 2. There was no
532 significant difference between the reaction times to unrelated items across contexts ($t(21) = 0.731,$
533 $p = 0.473, \text{Cohen's } d = 0.156, CI = -0.266 0.575$) as opposed to a significant difference between
534 related items across contexts ($t(21) = -2.719, p = 0.013, \text{Cohen's } d = -0.580, CI = -1.027 -0.121$).

535

536

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

537

538 **Table 2:** Descriptive statistics including Mean RT (ms) and standard deviation of related and unrelated word
539 pairs on each validity context, High Prime Validity and Low Prime Validity. Semantic priming effects and prime
540 validity effect (relatedness proportion effect).

541

542

543 EEG Results – Sensor Level

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

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

551 Therefore, in exploratory analyses, we focused the time-window of interest for the ERP
552 analysis on the peak of the global field power (530-1240ms), however this also revealed no
553 significant difference between the high and low validity contexts (smallest cluster $p = 0.139$).
554 Similarly, we used the window of interest for the alpha-beta time-frequency analysis to the peak of
555 the global field power (602-1240ms), which also yielded no significant difference between
556 conditions (no clusters formed). Moreover, as alpha-beta frequency bands include a wide range of
557 frequencies we analysed them separately. However, the time-frequency analysis in the Alpha band
558 (8-12Hz) showed no significant differences between conditions in the 0-1240ms time window
559 (smallest cluster $p = 0.121$), nor in the 530-1240ms time window (smallest cluster $p = 0.08$). The
560 same was true for the Beta band (13-30Hz; 0-1240ms cluster $p = 0.312$; 530-1240ms cluster $p =$
561 0.197). Together, these analyses suggested no apparent electrophysiological markers of pre-target
562 expectation formation in our data.

563

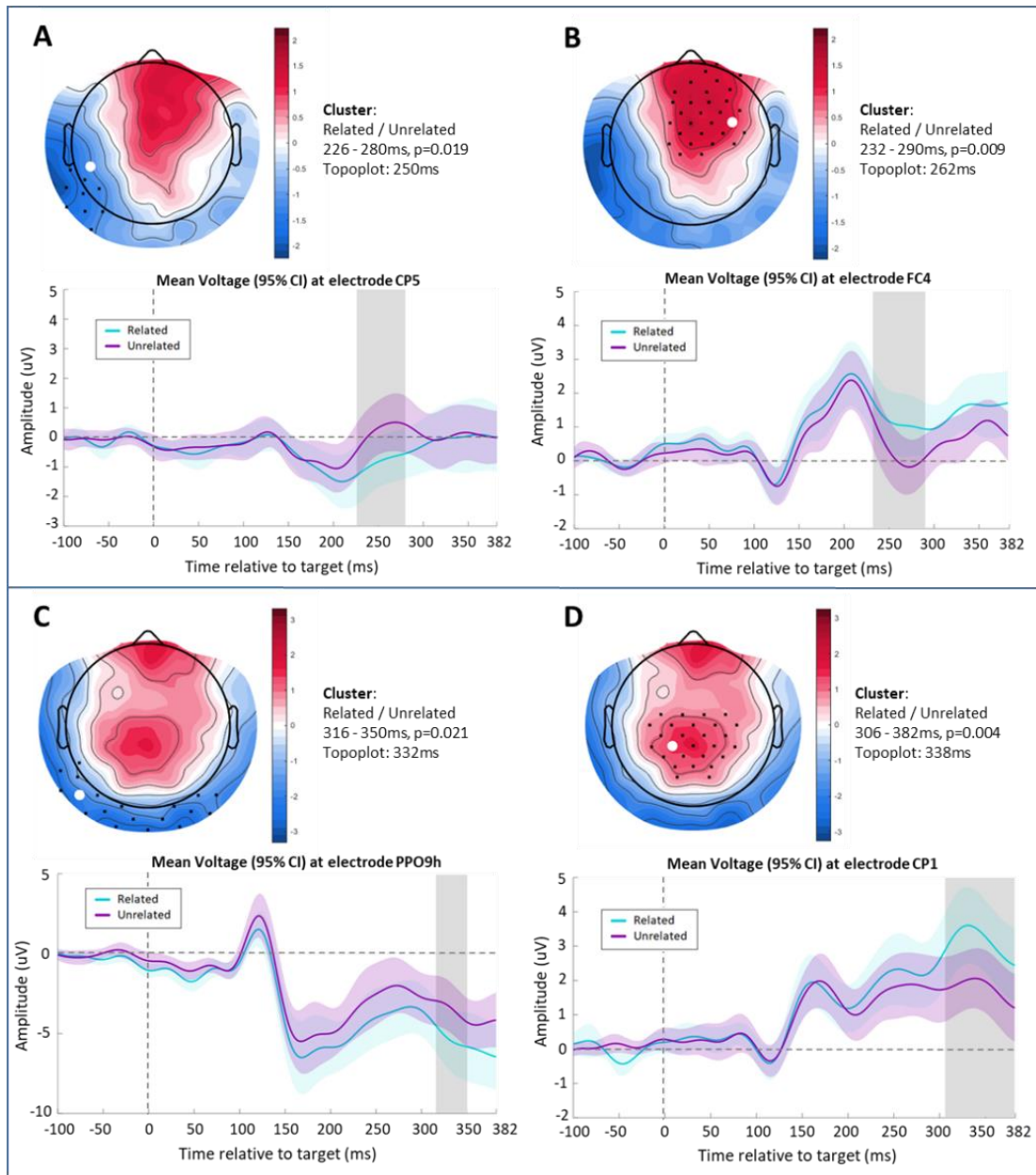
564

565

566

567 *Target Results: ERPs*

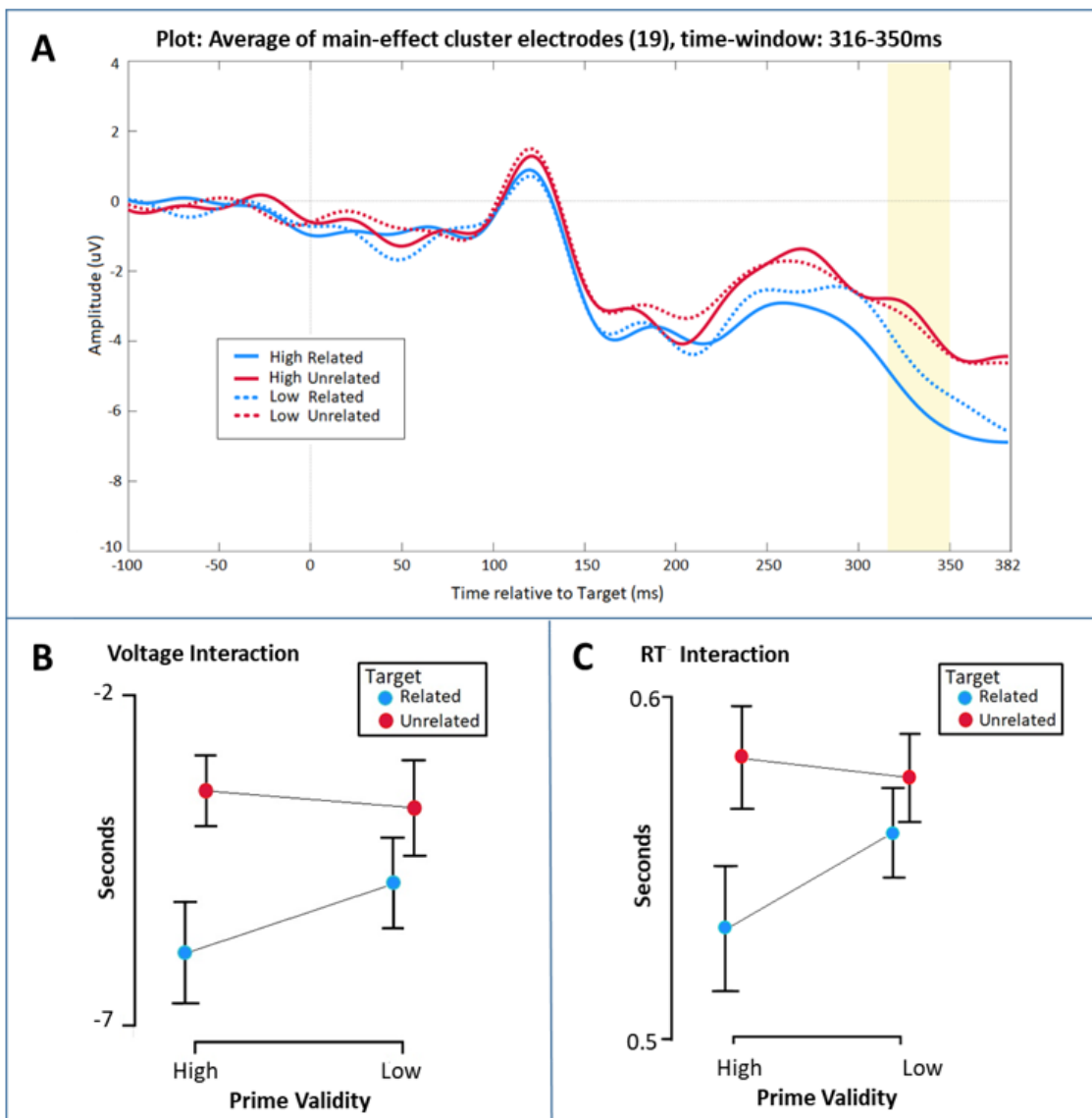
568 In our pre-registered interaction contrast in the latency range from 0 to 382ms post-
569 stimulus, the cluster-based permutation analysis yielded no clusters. However, in pre-registered
570 analyses of main effects in the same latency range, we found four significant main effects of
571 relatedness of the target (i.e. unrelated versus related targets; see Figure 3). The clusters in our data
572 occurred in two distinct periods within the time window as shown in Figure 3. Specifically, two
573 clusters reflected a left fronto-temporal dipolar effect of relatedness (Panels A & B in Figure 3) at
574 approximately 250ms post-stimulus (negative cluster: 226 – 280ms, $p = 0.019$; positive cluster: 232
575 – 290ms, $p = 0.009$), and two clusters reflected a later parieto-occipital dipolar effect of relatedness
576 (Panels C & D in Figure 3) at approximately 350ms post-stimulus (negative cluster: 316 – 350ms, $p =$
577 0.021 ; positive cluster: 306 – 382ms, $p = 0.004$). The early effects showed a predictive signal as in
578 both clusters the voltage exhibited more extreme values for unrelated than related items. On the
579 contrary, the later effects showed signs of an apredictive signal, especially in Panel D, as the voltage
580 within the cluster had more extreme values for the related relative to the unrelated items.



581

582 **Figure 3:** Four main effects from the cluster-based permutation analyses, which contrasted the voltage
583 difference between related and unrelated word-pairs from 0-382ms post-stimulus. ERP scalp topographies
584 revealed two dipolar effects; first, an early fronto-temporal effect at approximately 250ms (A and B); then, a
585 later parieto-occipital effect at around 340ms (C and D). ERP plots show data (mean and shaded 95%
586 confidence interval) from the electrode where the effect was maximal, with the cluster period highlighted in
587 grey.

588 As an exploratory analysis, and to increase power to detect a potential interaction effect,
589 we tested for the interaction within each of the main effect clusters by averaging per condition and
590 participant across all channels and time points within each main effect cluster. With this approach,
591 the later negative cluster (C in Figure 3) showed a significant interaction ($F(1, 21) = 6.679$, $p = 0.017$,
592 $\eta_p^2 = 0.241$), reflecting a larger voltage difference between the related and unrelated targets in a
593 high validity context with respect to a low validity context (other clusters $p = 0.396$; 0.110 ; 0.273).
594 Bayesian equivalent analyses considered this to be anecdotal evidence for the alternative
595 hypothesis ($BF_{inclusion} = 1.505$), see Figure 4.



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

603

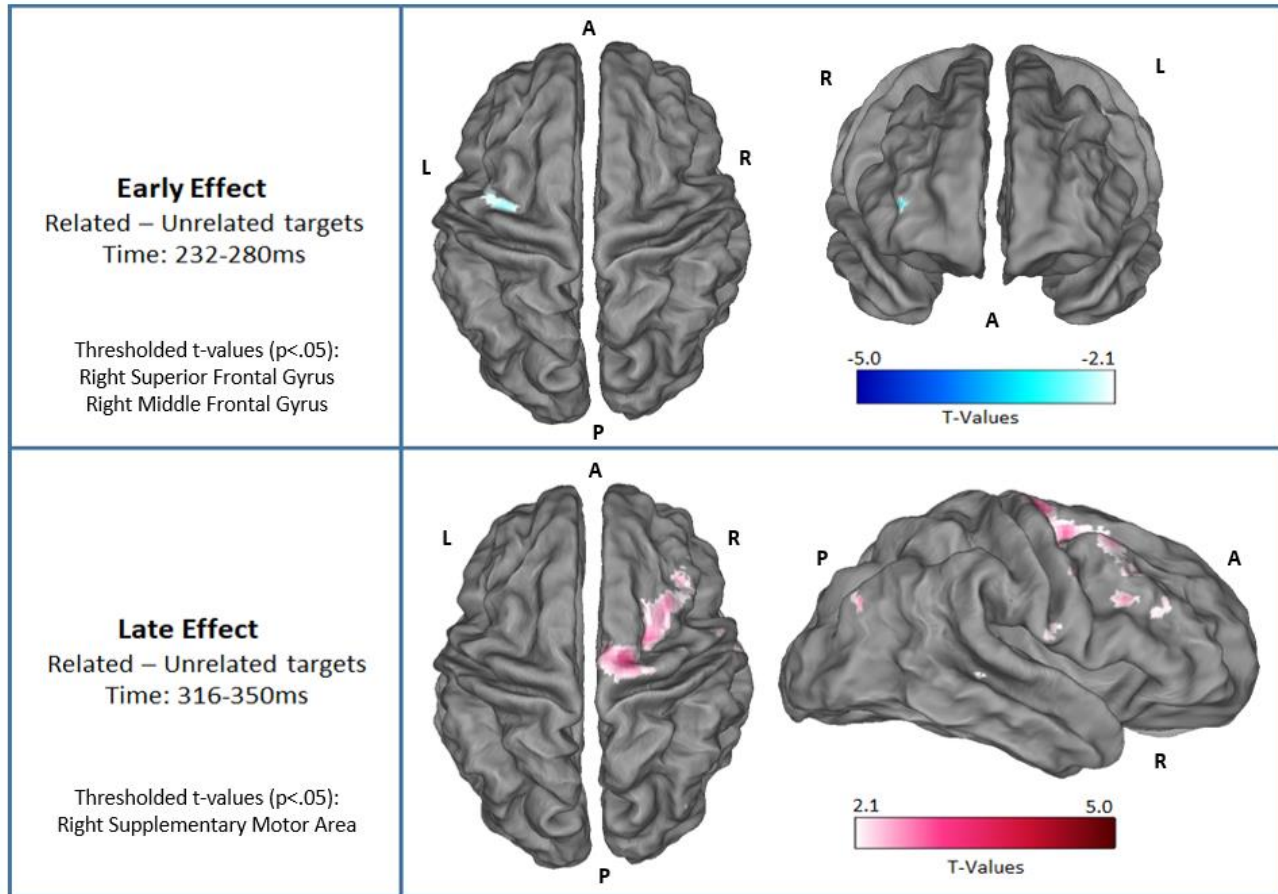
604 Source Estimate Analyses

605 Our pre-registered analyses included whole-brain interaction and main effect contrasts
606 within the time-windows of significant clusters at the sensor level. However, this approach returned
607 no significant clusters at the source level (interaction smallest cluster $p = 0.147$; main effect smallest
608 cluster $p = 0.067$). Furthermore, our preregistered source analyses included regions of interest from
609 the following AAL regions: Left inferior frontal gyrus (LIFG); Left middle temporal gyrus (LMTG); Left
610 superior temporal gyrus (LSTG). However, none of these regions exhibited significant interaction
611 effects or main effects (all FDR corrected p -values > 0.05).

612 Consequently, for a qualitative visualisation of the source estimates, here we plot the
613 whole-brain thresholded t -values ($p < 0.05$) of the source estimate contrasts, uncorrected for multiple
614 comparisons. Specifically, we plot these t -values for the early main effect (Figure 3A&B) and the late
615 main effect (Figure 3C&D) in time windows selected to be entirely within the significant dipolar
616 sensor level clusters (early: 232-280ms; late: 316-350), see Figure 5. The thresholded t -values
617 showed the peak of activity at the Right Middle and Superior Frontal Gyri for the early effect; and
618 the activity peak at the Right Supplementary Motor Area for the late effect, as shown in Figure 5.

619

620



621 **Figure 5:** Thresholded t -values ($p < .05$) of the ERP source estimates over two distinct time windows that
622 corresponded to the early and late ERP effects reported above in Figure 3. In the Figure, the upper panel shows
623 the difference between related and unrelated targets in the early time window (232-280ms), and the lower
624 panel indicates the same difference in a later time window (316-350ms) (thresholded t -values, $p < 0.025$).

625

626

627

628

629

630

631

632 **Discussion**

633 Predictive coding theory posits that the brain generates expectations about upcoming
634 stimuli at varying levels of complexity – from low-level expectations about stimulus properties
635 through to higher-level conceptual expectations. Here, we investigated the behavioural and
636 electrophysiological correlates of such expectations and their violations at two levels of a semantic
637 expectation hierarchy (local and global). First, on the behavioural level, participants of two separate
638 experiments showed evidence of speeded reaction times in related trials relative to unrelated trials,
639 consistent with a local expectation generated about target word identity on the basis of the prime
640 identity. Furthermore, participants generated a more conceptually complex expectation based on
641 the global context (i.e. prime validity) to exhibit greater behavioural facilitation in the high prime
642 validity context than the low prime validity context (Boudewyn, Long, and Swaab, 2015).
643 Importantly, only those individuals who reported conscious strategic expectation showed evidence
644 of behavioural facilitation given by the global context, while those individuals who did not report a
645 conscious strategy only exhibited facilitation as a result of the local context. Together, these
646 behavioural data are consistent with a dissociation between a local expectation about the identity
647 of the target generated by the prime, and a global expectation about the relatedness of the target
648 that necessitates reportable, effortful, and strategic application of expectation. Moreover, the
649 present data provides evidence for a successful replication of the behavioural effect elicited by the
650 same paradigm as implemented by Hutchison (2007), who also found that the magnitude of the
651 global facilitatory effect was modulated by the level of attentional control (i.e. weaker effect in
652 individuals with lower attentional control; Hutchison, 2007). Similarly, our results suggested that
653 only individuals that reported applying an effortful conscious strategy showed the global context
654 effect as mentioned above.

655

656 Consistent with this two-stage expectation profile, the ERPs in response to the target words
657 also exhibited a two-stage profile, with an early effect modulated by local expectation (around
658 250ms) and a later effect modulated by global expectation (around 350ms). These results are
659 broadly consistent with the two-stage profile observed in the auditory oddball local – global
660 paradigm (Bekinschtein et al., 2009), which includes an MMN in an early stage reflecting errors of
661 the local context of the stimuli and a P3b response to errors of the global context given by blocks
662 across the task.

663 Furthermore, the early effect in the present experiment showed more extreme amplitudes
664 for unexpected targets relative to expected targets, consistent with a prediction error signal, such
665 as the MMN to unexpected/deviant items observed across levels of stimulus awareness (Chennu et
666 al., 2013; Bekinschtein et al., 2009; Faugueras et al., 2012; El Karoui et al. 2014). Moreover, the scalp
667 topography of the early effect has a fronto-central peak, which is consistent with the MMN (Chennu
668 et al., 2013; Bekinschtein et al., 2009; Faugueras et al., 2012), although, its latency is a little longer
669 than seen in some of these previous papers. Additionally, in our source estimation analyses, the
670 early effect was localised to the middle frontal gyrus (Figure 5), whereas in another study the local
671 MMN effect was localised to the temporal parietal junction and prefrontal cortex (Chennu et al.,
672 2013), indicating not entirely overlapping neurocognitive processes. Nevertheless, as we observed
673 behavioural semantic priming (as tracked by the early effect) even for participants who were not
674 making strategic expectations, and due to the shared common features with the MMN (i.e. more
675 extreme for errors and with a fronto-central focus), we consider the early effect to be consistent
676 with an error of local expectation – i.e. expectation based on the identity of the prime, rather than
677 the prime validity. Indeed, the MMN is elicited even by individuals who are not actively attending
678 to the stimuli (Bekinschtein et al., 2009).

679

680 The late effect, however, was the opposite of what would be expected for a prediction error
681 signal – i.e. its amplitude was more extreme for expected targets compared to unexpected targets.
682 This *apredictive* pattern is not readily explained by prediction error accounts without appeal to
683 precision-weighting, in which a prediction error is weighted by the system’s confidence in the signal
684 (Chennu et al., 2013; Friston, 2005). Under precision-weighting, all possible patterns of prediction
685 error signals on the scalp are possible, including *apredictive* patterns as we observed here, as
686 precision may vary freely across task conditions (Kok et al., 2011). For example, Barascud et al.
687 (2016) reported a larger MEG signal for auditory stimuli that become predictable, relative to stimuli
688 that are entirely unpredictable – i.e. an *apredictive* pattern – that they linked to up-weighting of the
689 expected stimuli by precision (Heilbron, Chait, 2018). Within predictive coding, attention is one
690 specific mechanism that is thought to increase precision (Hohwy, 2012). Therefore, under a
691 predictive coding framework, one can appeal to varying levels of attention across task conditions.
692 Therefore, we could post-hoc theorise that our late *apredictive* effect reflects individuals paying
693 greater attention to the high validity trials as they have a high level of predictability and paying
694 greater attention to related targets than unrelated targets, as the former fulfil their expectations.
695 Therefore, the relative levels of attention across conditions could interact to generate this
696 *apredictive* effect. Indeed, consistent with this, 59% of our participants (13/22) self-reported that
697 their strategy was to generate an expectation in the high validity condition only (i.e. “I was trying to
698 guess next word if previous was blue”; where blue was high validity condition).

699 An alternative interpretation stems from evaluation of our behavioural data. When
700 comparing the behavioural reaction time interaction with the ERP voltage interaction (Table 2 and
701 Figure 4, respectively), both show the same pattern: namely, that the interaction is driven by
702 expected items in a high validity context, showing more extreme values with respect to the other
703 three conditions. This similarity in behaviour and ERP effects suggest that our late ‘error’ effect may

704 simply reflect processing in service of behaviour, whereby sensory signals are routed to goal-driven
705 analogous motor behaviour (Zylberberg et al., 2010). Our late apredictive ERP pattern may therefore
706 not reflect a precision-weighted global prediction error, but more simply the result of the brain
707 routing the incoming information into appropriate behaviour. Under this interpretation, our results
708 are therefore also consistent with interpretations of early ERPs as reflections of prediction error and
709 later ERPs as processes related to conscious access and in support of task demands (e.g. Dehaene &
710 Christen, 2011; Rohaut et al., 2015).

711 It is possible that other later error signals were also evident in the neural response during
712 our task, including those traditionally linked to the N400 (i.e. peaking approximately 400ms post-
713 target). However, we limited our analyses to the 0 to 382ms time-window post-target so as to avoid
714 muscle artefact created by the pronunciation responses. We chose to use a pronunciation task as
715 our aim was to observe the behavioural effect produced by the manipulation of both the local
716 (relatedness) and global context (prime validity) as implemented by Hutchison (2007). Nevertheless,
717 tasks that don't produce large muscular artefacts, such as a lexical decision task (LDT) in which
718 individuals only produce motor responses on filler trials, would allow for analysis of the N400 time-
719 window. However, as argued by Hutchison (2007), participants can complete an LDT with a
720 semantic-matching strategy, meaning that after seeing the target they can verify whether it is
721 related to the prime, which could bias their responses as only words can be related and non-words
722 would be, by their nature, unrelated (Hutchison, 2007). Additionally, as we provided a global context
723 by manipulating the proportion of related items across the task, individuals could bias their
724 responses using the validity cue (Keefe & Neely, 1990); for example, primes that were presented in
725 blue (high validity context) were more likely to be related (80%). Therefore, when seeing a blue
726 prime, individuals could judge their response (word/non-word) solely based on the prime, in this
727 case a 'word' as most of the word-pairs are related. Instead, using a pronunciation task allows for a

728 purer measure of expectation, with the caveat of limiting the time-window of artefact-free EEG for
729 analysis.

730 A recent prediction error view on language-related ERPs proposes that the N400 has similar
731 properties to the MMN, as they both are modulated by the predictability of stimuli (i.e. increased
732 ERP amplitude as a prediction-error response) but that their relative latencies indicate prediction-
733 error processing at different levels of stimulus complexity (Bornkessel-Schlesewsky & Schlewsky,
734 2019). In our findings, both consecutive effects could be similarly interpreted as reflecting different
735 levels of complexity of precision-weighted prediction error processing across a semantic hierarchy.
736 However, as noted above, appeal to precision-weighting problematically allows for post-hoc
737 explanations of all possible ERP patterns (Bowman, Filetti & Olivers, 2013).

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

748 In our pre-registered analyses, we also hypothesised that we would observe
749 electrophysiological markers of differential expectations generated by the high and low validity
750 primes, prior to the onset of the target. Specifically, we expected these differential expectations to
751 be reflected in the ERPs, including the slope of a putative slow wave (Chennu et al., 2013), and/or

752 in the power of the EEG in the alpha/beta bands, as these have been previously associated with the
753 precision of expectations (Bauer, Stenner, Friston & Dolan, 2014). However, we found no evidence
754 of any differences in these measures between high and low validity primes prior to target onset.
755 One interpretation is that our specific measures were simply not sensitive enough to detect the
756 differential expectations in these conditions. Indeed, we powered our study to detect the post-
757 target behavioural effect specifically. An alternative interpretation is that expectations were, in fact,
758 not different between the two conditions. Indeed, under predictive coding, the brain is considered
759 to optimize the difference between its expectations and sensory input by updating its internal model
760 (Friston, 2010); hence, it is possible that the optimal means of minimising prediction error in this
761 task is to always predict the related target, regardless of the prime validity. For example, even if one
762 were to consciously expect that an upcoming target will be unrelated (as in a low validity trial), it is
763 simply not possible to accurately predict the identity of that target, as the range of possible
764 unrelated target words is considerable. Therefore, even though predicting the identity of a specific
765 related target had only a ~22% probability of being correct in a low validity context, it was still more
766 likely than predicting any one of the vast arrays of potential unrelated target words. Future
767 inspection of participants' meta-cognition in relation to their specific expectations following prime
768 presentation will help speak to this interpretation.

769

770 *Conclusions*

771 In conclusion, we here reported ERP evidence of hierarchical matching of semantic
772 expectations to incoming speech. Lower level expectations based on the local context (i.e. the prime
773 identity) elicited an early and predictive pattern that matches with prediction error accounts. Higher
774 level expectations generated from the global context required awareness of the global rule and the
775 use of a reportable strategy, and were associated with an apredictive pattern that can be

776 interpreted within a precision-weighted prediction error account, or may reflect the routing of

777 sensory signals and their expectations into task-directed behaviour.

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796 **References**

- 797 1. Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual
798 word recognition of single-syllable words. *Journal of Experimental Psychology: Gen-*
799 *eral*, 133(2), 283.
- 800 2. Bauer, M., Stenner, M. P., Friston, K. J., & Dolan, R. J. (2014). Attentional modulation of
801 alpha/beta and gamma oscillations reflect functionally distinct processes. *Journal of Neuro-*
802 *science*, 34(48), 16117-16125.
- 803 3. Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neu-
804 ral signature of the conscious processing of auditory regularities. *Proceedings of the Na-*
805 *tional Academy of Sciences*, 106(5), 1672-1677.
- 806 4. Berkum, J. J. V., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and
807 discourse: Evidence from the N400. *Journal of cognitive neuroscience*, 11(6), 657-671.
- 808 5. Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: Predictive pro-
809 cessing and the adjustment of expectations during spoken language comprehension. *Cog-*
810 *nitive, Affective, & Behavioral Neuroscience*, 15(3), 607-624.
- 811 6. Bornkessel-Schlesewsky, I., & Schlewsky, M. (2019). Towards a neurobiologically plausible
812 model of language-related, negative event-related potentials. *Frontiers in psychology*, 10,
813 298.
- 814 7. Bowman, H., Filetti, M., Wyble, B., & Olivers, C. (2013). Attention is more than prediction
815 precision [Commentary on target article]. *Behavioral and Brain Sciences*, 36(3), 206-208.
- 816 8. Brainard, D. H. (1997) The Psychophysics Toolbox, *Spatial Vision* 10:433-436.
- 817 9. Brookes, M. J., Tewarie, P. K., Hunt, B. A., Robson, S. E., Gascoyne, L. E., Liddle, E. B., ... &
818 Morris, P. G. (2016). A multi-layer network approach to MEG connectivity analysis. *Neu-*
819 *roimage*, 132, 425-438.

- 820 10. Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical pre-
821 diction during sentence comprehension. *Journal of Memory and Language*, *93*, 203-216.
- 822 11. Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the
823 brain. *Frontiers in human neuroscience*, *4*, 25.
- 824 12. Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibáñez, A., ... & Bekin-
825 schtein, T. A. (2013). Expectation and attention in hierarchical auditory prediction. *Journal*
826 *of Neuroscience*, *33*(27), 11194-11205.
- 827 13. Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cog-
828 nitive science. *Behavioral and brain sciences*, *36*(3), 181-204.
- 829 14. Cruse, D., Beukema, S., Chennu, S., Malins, J. G., Owen, A. M., & McRae, K. (2014). The reli-
830 ability of the N400 in single subjects: implications for patients with disorders of conscious-
831 ness. *NeuroImage: Clinical*, *4*, 788-799.
- 832 15. Dehaene, S., & Christen, Y. (Eds.). (2011). *Characterizing consciousness: from cognition to*
833 *the clinic?* Springer Science & Business Media.
- 834 16. Delorme, A. & Makeig, S. (2004) EEGLAB: an open source toolbox for analysis of single-trial
835 EEG dynamics. *Journal of Neuroscience Methods*, *134*: 9-21.
- 836 17. Drake, R., Vogl, A. W., & Mitchell, A. W. (2009). *Gray's Anatomy for Students E-Book. Elsevier*
837 *Health Sciences*.
- 838 18. El Karoui, I., King, J. R., Sitt, J., Meyniel, F., Van Gaal, S., Hasboun, D., ... & Cohen, L. (2014).
839 Event-related potential, time-frequency, and functional connectivity facets of local and
840 global auditory novelty processing: an intracranial study in humans. *Cerebral cortex*, *25*(11),
841 4203-4212.

- 842 19. Faugeras, F., Rohaut, B., Weiss, N., Bekinschtein, T., Galanaud, D., Puybasset, L., ... & Nac-
843 cache, L. (2012). Event related potentials elicited by violations of auditory regularities in
844 patients with impaired consciousness. *Neuropsychologia*, 50(3), 403-418.
- 845 20. Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews neuro-*
846 *science*, 11(2), 127.
- 847 21. Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosoph-*
848 *ical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1211-1221.
- 849 22. Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal So-*
850 *ciety B: Biological sciences*, 360(1456), 815-836.
- 851 23. Heilbron, M., & Chait, M. (2017). Great expectations: is there evidence for predictive coding
852 in auditory cortex?. *Neuroscience*.
- 853 24. Hertrich, I., Dietrich, S., & Ackermann, H. (2016). The role of the supplementary motor area
854 for speech and language processing. *Neuroscience & Biobehavioral Reviews*, 68, 602-610.
- 855 25. Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Fron-*
856 *tiers in psychology*, 3, 96.
- 857 26. Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C. S., ...
858 & Buchanan, E. (2013). The semantic priming project. *Behavior Research Methods*, 45(4),
859 1099-1114.
- 860 27. Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in seman-
861 tic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4),
862 645.
- 863 28. Keefe, D. E., & Neely, J. H. (1990). Semantic priming in the pronunciation task: The role of
864 prospective prime-generated expectancies. *Memory & cognition*, 18(3), 289-298.

- 865 29. King, J. R., Gramfort, A., Schurger, A., Naccache, L., & Dehaene, S. (2014). Two distinct dy-
866 namic modes subtend the detection of unexpected sounds. *PLoS one*, 9(1), e85791.
- 867 30. Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3?.
- 868 31. Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding
869 and computation. *TRENDS in Neurosciences*, 27(12), 712-719.
- 870 32. Koivisto, M., & Revonsuo, A. (2001). Cognitive representations underlying the N400 priming
871 effect. *Cognitive Brain Research*, 12(3), 487-490.
- 872 33. Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., & De Lange, F. P. (2012). Attention reverses the
873 effect of prediction in silencing sensory signals. *Cerebral cortex*, 22(9), 2197-2206.
- 874 34. Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language com-
875 prehension? *Language, cognition and neuroscience*, 31(1), 32-59.
- 876 35. Kuperberg, G. R., Delaney-Busch, N., Fanucci, K., & Blackford, T. (2018). Priming production:
877 Neural evidence for enhanced automatic semantic activity preceding language production
878 in schizophrenia. *NeuroImage: Clinical*, 18, 74-85.
- 879 36. Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the
880 N400 component of the event-related brain potential (ERP). *Annual review of psychol-*
881 *ogy*, 62, 621-647.
- 882 37. Lapatki, B. G., Stegeman, D. F., & Jonas, I. E. (2003). A surface EMG electrode for the simul-
883 taneous observation of multiple facial muscles. *Journal of neuroscience methods*, 123(2),
884 117-128.
- 885 38. Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013) (a). Dissociating N400 effects of predic-
886 tion from association in single-word contexts. *Journal of cognitive neuroscience*, 25(3), 484-
887 502.

- 888 39. Lau, E. F., Gramfort, A., Hämäläinen, M. S., & Kuperberg, G. R. (2013) (b). Automatic seman-
889 tic facilitation in anterior temporal cortex revealed through multimodal neuroimaging. *Jour-*
890 *nal of Neuroscience*, 33(43), 17174-17181.
- 891 40. Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dy-
892 namics during sentence-level language comprehension. *Cortex*, 68, 155-168.
- 893 41. Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear
894 mixed models to analyse reaction time data. *Frontiers in psychology*, 6, 1171.
- 895 42. Lund, K., & Burgess, C. (1996, April). Hyperspace analogue to language (hal): A general
896 model semantic representation. In *Brain and Cognition* (Vol. 30, No. 3, pp. 5-5). 525 B ST,
897 STE 1900, SAN DIEGO, CA 92101-4495: ACADEMIC PRESS INC JNL-COMP SUBSCRIPTIONS.
- 898 43. Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-
899 data. *Journal of neuroscience methods*, 164(1), 177-190.
- 900 44. Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based
901 plausibility immediately influences on-line language comprehension. *Journal of Experi-*
902 *mental Psychology: Learning, Memory, and Cognition*, 37(4), 913.
- 903 45. McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the Word Frequency Effect: The Ne-
904 glected Role of Distributional Information in Lexical Processing. *Language and*
905 *Speech*, 44(3), 295–322.
- 906 46. Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG arti-
907 fact detector based on the joint use of spatial and temporal features. *Psychophysiol-*
908 *ogy*, 48(2), 229-240.
- 909 47. Noesselt, T., Shah, N. J., & Jäncke, L. (2003). Top-down and bottom-up modulation of lan-
910 guage related areas—an fMRI study. *BMC neuroscience*, 4(1), 13.

- 911 48. Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source soft-
912 ware for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computa-*
913 *tional intelligence and neuroscience*, 2011, 1.
- 914 49. Pelli, D. G. (1997) The VideoToolbox software for visual psychophysics: Transforming num-
915 bers into movies, *Spatial Vision* 10:437-442.
- 916 50. Popov, T., Oostenveld, R., & Schoffelen, J. M. (2018). FieldTrip made easy: An analysis pro-
917 tocol for group analysis of the auditory steady state brain response in time, frequency, and
918 space. *Frontiers in neuroscience*, 12, 711.
- 919 51. Rabovsky, M., & McRae, K. (2014). Simulating the N400 ERP component as semantic net-
920 work error: Insights from a feature-based connectionist attractor model of word mean-
921 ing. *Cognition*, 132(1), 68-89.
- 922 52. Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional inter-
923 pretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1), 79.
- 924 53. Robinson, S. (1999) Functional neuroimaging by synthetic aperture magnetometry (SAM).
925 Recent advances in biomagnetism.
- 926 54. Rohaut, B., Faugeras, F., Chausson, N., King, J. R., El Karoui, I., Cohen, L., & Naccache, L.
927 (2015). Probing ERP correlates of verbal semantic processing in patients with impaired con-
928 sciousness. *Neuropsychologia*, 66, 279-292.
- 929 55. Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests
930 for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review*, 16(2), 225-
931 237.
- 932 56. Skrandies, W. (1990). Global field power and topographic similarity. *Brain topography*, 3(1),
933 137-141.

- 934 57. Sokoliuk, R., Mayhew, S. D., Aquino, K. M., Wilson, R., Brookes, M. J., Francis, S. T., ... &
935 Mullinger, K. J. (2019). Two spatially distinct posterior alpha sources fulfil different func-
936 tional roles in attention. *Journal of Neuroscience*, 39(36), 7183-7194.
- 937 58. Sokoliuk, R., Calzolari, S., & Cruse, D. (2019). Dissociable electrophysiological correlates of
938 semantic access of motor and non-motor concepts. *Scientific reports*, 9(1), 1-14.
- 939 59. Team, J. A. S. P. (2018). JASP (Version 0.8. 6). *Computer software*.
- 940 60. Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sen-
941 tence processing: Frontal positivity and N400 ERP components. *International Journal of Psy-*
942 *chophysiology*, 83(3), 382-392.
- 943 61. Van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., ... & Ly, A.
944 (2019). The JASP guidelines for conducting and reporting a Bayesian analysis.
- 945 62. Van Drongelen, W., Yuchtman, M., Van Veen, B. D., & Van Huffelen, A. C. (1996). A spatial
946 filtering technique to detect and localize multiple sources in the brain. *Brain Topogra-*
947 *phy*, 9(1), 39-49.
- 948 63. Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimi-
949 nation. *The Quarterly Journal of Experimental Psychology Section A*, 47(3), 631-650.
- 950 64. Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain
951 electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transac-*
952 *tions on biomedical engineering*, 44(9), 867-880.
- 953 65. Wacongne, C., Changeux, J. P., & Dehaene, S. (2012). A neuronal model of predictive coding
954 accounting for the mismatch negativity. *Journal of Neuroscience*, 32(11), 3665-3678.
- 955 66. Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... & Meerhoff, F.
956 (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psycho-*
957 *nomic bulletin & review*, 25(1), 58-76.

- 958 67. Yekutieli, D., & Benjamini, Y. (1999). Resampling-based false discovery rate controlling mul-
959 tiple test procedures for correlated test statistics. *Journal of Statistical Planning and Infer-*
960 *ence*, 82(1-2), 171-196.
- 961 68. Ylinen, S., Huuskonen, M., Mikkola, K., Saure, E., Sinkkonen, T., & Paavilainen, P. (2016).
962 Predictive coding of phonological rules in auditory cortex: A mismatch negativity
963 study. *Brain and language*, 162, 72-80.
- 964 69. Zylberberg, A., Slezak, D. F., Roelfsema, P. R., Dehaene, S., & Sigman, M. (2010). The brain's
965 router: a cortical network model of serial processing in the primate brain. *PLoS computa-*
966 *tional biology*, 6(4).
- 967
- 968
- 969
- 970
- 971
- 972
- 973
- 974
- 975
- 976
- 977
- 978
- 979
- 980
- 981