


1 **Can Prediction Error Explain Predictability Effects on the N1 during Picture-Word**
2 **Verification?**

3 Jack E. Taylor^{1,2}, Guillaume A. Rousselet², and Sara C. Sereno²

4 ¹Department of Psychology, Goethe University Frankfurt

5 ²School of Psychology and Neuroscience, University of Glasgow

Author Note

Jack E. Taylor  <https://orcid.org/0000-0003-4765-0118>

Guillaume A. Rousselet  <https://orcid.org/0000-0003-0006-8729>

Sara C. Sereno  <https://orcid.org/0000-0001-7957-9542>

Correspondence concerning this article should be addressed to Jack E. Taylor
Department of Psychology, Goethe University Frankfurt; E-mail:
Taylor@psych.uni-frankfurt.de

Author Contributions: *J.E.T.* - Conceptualisation, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing, Visualisation. *S.C.S.* - Conceptualisation, Methodology, Writing, Supervision. *G.A.R.* - Methodology, Formal Analysis, Writing, Supervision.

Authors report no conflict of interest.

This research was supported in part by an Economic and Social Research Council (ESRC) postgraduate fellowship awarded to J. E. Taylor (reference: ES/P000681/I).

Abstract

24

25 Do early effects of predictability in visual word recognition reflect prediction error?

26 Electrophysiological research investigating word processing has demonstrated

27 predictability effects in the N1, or first negative component of the event-related potential

28 (ERP). However, findings regarding the magnitude of effects and potential interactions of

29 predictability with lexical variables have been inconsistent. Moreover, past studies have

30 typically used categorical designs with relatively small samples and relied on

31 by-participant analyses. Nevertheless, reports have generally shown that predicted

32 words elicit less negative-going (i.e., lower amplitude) N1s, a pattern consistent with a

33 simple predictive coding account. In our preregistered study, we tested this account via

34 the interaction between prediction magnitude and certainty. A picture-word verification

35 paradigm was implemented in which pictures were followed by tightly matched

36 picture-congruent or picture-incongruent written nouns. The predictability of target

37 (picture-congruent) nouns was manipulated continuously based on norms of association

38 between a picture and its name. ERPs from 68 participants revealed a pattern of effects

39 opposite to that expected under a simple predictive coding framework.

40 *Keywords:* N1, N170, Prediction, Predictive Coding, Word Recognition

41 **Can Prediction Error Explain Predictability Effects on the N1 during Picture-Word**
42 **Verification?**

43 **Introduction**

44 Readers and listeners routinely use context to predict upcoming semantic and
45 lexical content. Evidence for such predictive processes arises from both behavioural and
46 neural correlates of language comprehension (Kuperberg & Jaeger, 2016; Luke &
47 Christianson, 2016; Pickering & Gambi, 2018; Rayner et al., 2011; Van Petten & Luka,
48 2012), with demonstrated facilitation for the processing of predicted information
49 (Federmeier, 2007; Pickering & Garrod, 2013).

50 A key question in this area is, how early in the processing stream are predictive
51 processes able to modulate visual word recognition. One early stage in visual word
52 recognition, which may be sensitive to prediction, involves the processing of visual word
53 forms. A word form can be defined as the visual pattern of a single written word,
54 comprised of smaller orthographic components (e.g., letters, letter bigrams, graphemes,
55 strokes). While some electrophysiological evidence suggests sensitivity to orthographic
56 variables in an earlier posterior P1 component peaking at around 100 ms after word
57 presentation (e.g., Nobre et al., 1994; Segalowitz & Zheng, 2009; Sereno et al., 1998),
58 the event-related potential (ERP) component most identified as an index of orthographic
59 processing across different scripts is the first posterior negative-going wave, the N1
60 (Bentin et al., 1999; Lin et al., 2011; Ling et al., 2019; Maurer, Brandeis, & McCandliss,
61 2005; Maurer et al., 2008; Pleisch et al., 2019). The N1 is also sometimes referred to as
62 the N170 due to the timing of its peak in some studies, at around 170 ms. This typically
63 occipitotemporal, negative-going component shows reliable differences between
64 orthographic and non-orthographic stimuli (e.g., words elicit more negative-going N1s
65 than false-font strings do; Appelbaum et al., 2009; Bentin et al., 1999;
66 Eberhard-Moscicka et al., 2016; Maurer, Brandeis, & McCandliss, 2005; Maurer, Brem,
67 et al., 2005; Pleisch et al., 2019; Zhao et al., 2014).

68 Accounts of orthographic processing often stress the importance of top-down
69 predictions, and their interactions with bottom-up sensory input. For instance, the
70 interactive account of the ventral occipito-temporal cortex (vOT), a region which is a
71 likely generator of the N1 ERP component (Allison et al., 1994; Brem et al., 2009; Cohen
72 et al., 2000; Dale et al., 2000; Maurer, Brem, et al., 2005; Nobre et al., 1994; Taha et al.,
73 2013; Woolnough et al., 2021), suggests that sensitivity to orthography arises through
74 the synthesis of bottom-up visuospatial information and top-down predictions informed
75 by prior experience and knowledge (Price & Devlin, 2011). Such accounts exist within a
76 predictive coding framework, according to which the brain utilises higher-level
77 information to build, maintain, and continually update hierarchical series of estimators
78 that form generative models of sensory information (Friston, 2010; Rao & Ballard, 1999;
79 Rauss et al., 2011). Predictive coding accounts have been employed to explain
80 prediction effects observed in early evoked responses across a range of domains, such
81 as the mismatch negativity (Garrido et al., 2009) and sensory attenuation of
82 self-generated percepts (Knolle et al., 2013). A key feature of such accounts is that
83 higher-level predictions cause lower-level features to be preactivated, and that the
84 difference between the bottom-up sensory input and top-down predictions corresponds to
85 a prediction error, which the brain attempts to minimise (Clark, 2013; Walsh et al., 2020).

86 In a predictive coding framework, prediction errors are commonly determined by
87 two key attributes: the magnitude of the error, and the precision or certainty of the error
88 (Feldman & Friston, 2010; Kanai et al., 2015). Such variants of predictive coding models
89 are commonly referred to as *precision-weighted*. Feldman and Friston (2010) likened the
90 error signal to the calculation of the t statistic, where magnitude of an observation (i.e.,
91 mean, or mean difference) is divided by the inverse of its precision (i.e., standard error).
92 Prediction errors, weighted by precision in this manner, can be conceptualised as
93 representing the degree of “surprise” associated with a set of observations under a
94 specified hypothesis.

95 Firstly, the magnitude of the error should determine the size of the error signal,
96 with larger prediction errors resulting from greater mismatch between descending
97 (top-down) predictions and ascending (bottom-up) sensory input. In neutral
98 (non-biasing) contexts, a predictive coding account that includes learning of statistical
99 regularities over extended periods would assert that error signals should vary as a
100 function of stimulus regularity. More specifically, a predictive coding account of
101 orthographic processing would expect error signals to vary as a function of the size of
102 the difference between a general orthographic prior (e.g., an average word form) and a
103 presented word form. Some recent findings appear to support the notion that the N1
104 reflects a neutral-context error signal, with greater distance from an orthographic prior
105 eliciting greater amplitude (Gagl et al., 2020), while the profile of the N1's sensitivity to
106 word form regularity over experience matches that expected under a predictive coding
107 account (Huang et al., 2022; Zhao et al., 2019).

108 Secondly, the precision or certainty of the prediction error should influence the
109 response, with more certain descending predictions, and more certain ascending
110 sensory input, eliciting greater error signals when predictions are violated. In neutral
111 contexts, predictions, and certainty about them, may not be expected to vary much from
112 a context-general prior. Indeed, it is easier to envisage the expected role of prediction
113 precision for orthographic processing in biasing contexts, where precision is more
114 variable than it is in neutral contexts. A predictive coding model of orthographic
115 processing that allows for online, context-informed updating of orthographic priors would
116 expect that the predictability of word forms should influence error responses, with more
117 predictable contexts eliciting stronger prediction error effects. For instance, a sentential
118 context that elicits a clear and reliable prediction for an upcoming word (i.e., that has high
119 Cloze probability) should show a larger prediction error difference, between succeeding
120 prediction-congruent and -incongruent word forms, than should a more neutral sentential
121 context that is consistent with a large number of low-probability candidate words.

122 In this paper, we examine whether a simple predictive coding account that
123 includes online updating of context-biased predictions and expectations can explain
124 neural activity, captured in the N1, elicited by a word in context. Specifically, we examine
125 whether sensitivity to prediction error in the N1 is dependent on contextual predictability
126 in the manner that a predictive coding account would expect. This question is prompted
127 by (1) the emerging evidence that the N1 in neutral contexts is consistent with an
128 orthographic prediction error signal (Gagl et al., 2020; Huang et al., 2022; Zhao et al.,
129 2019), and (2) existing evidence that biasing semantic contexts can modulate the N1
130 ERP (outlined below). To address our question, we employ a paradigm informed directly
131 by predictive coding models, manipulating prediction congruency and precision
132 independently, to examine whether the N1 shows the pattern of amplitudes expected
133 under such a model, in biasing contexts. Moreover, we maximise our sensitivity to an
134 orthographic prediction error by presenting prediction-congruent and -incongruent words
135 that are carefully matched item-wise on possible confounders, with maximal orthographic
136 distance from one another. Importantly, evidence for a context-informed prediction error
137 signal at an early, likely orthographic, stage of processing, would not preclude the
138 existence of similar prediction error signals at later stages. Indeed, the hierarchically
139 composed generative model posited by a predictive coding account is fully compatible
140 with the production of prediction errors spanning a hierarchy of linguistic representations.

141 We hypothesise that according to a simple predictive coding model, the N1 should
142 be larger for prediction-incongruent than prediction-congruent word forms (i.e., prediction
143 error), in a manner dependent on the level of predictability (i.e., precision). We
144 hypothesise that as predictability increases, so too should the prediction error effect.

145 We begin by reviewing findings from prior studies. We make a distinction between
146 those studies that have biased expectations via *linguistic cues* (text preceding the target
147 word), and those that have employed *non-linguistic cues* (e.g., cross-modal contexts and
148 manipulation of task demands).

149 **Biasing Word Form Predictions via Linguistic Cues**

150 Readers' predictions of upcoming word forms are generally manipulated via
151 linguistic cues. In these studies, a target word's predictability is typically determined in a
152 pre-experiment norming study, operationalised via Cloze probability (i.e., the probability
153 that the target is correctly guessed given its preceding context). Such a measure of word
154 form predictability aligns closely with the concept of prediction precision or certainty in a
155 predictive coding account.

156 Recent ERP investigations that have manipulated sentential context have also
157 often varied word frequency, with the assumption that an interaction of predictability with
158 word frequency would provide evidence for top-down influences on lexical access. In
159 **Table 1**, we summarise N1 results reported from studies using sentential paradigms that
160 have employed such *Predictability × Frequency* designs. While effects often extend to
161 earlier and later components, we limit our summary to those involving predictability
162 within the N1 window. In Figure 1 we visualise the timing of N1 windows applied in these
163 studies and others cited in this introduction. Sentential studies using a Predictability ×
164 Frequency design have demonstrated effects in the N1, although the pattern of effects
165 observed across studies is varied (for a review, see Sereno et al., 2019). We also note
166 that studies using average reference showed more posterior effects, while effects
167 reported from studies using mastoid reference showed more centroparietal topography.

168 In addition to such studies that focused on the N1, some studies designed to
169 focus on N400 effects of predictability may also provide insight into early prediction
170 effects. For instance, Brothers et al. (2015) examined correlates of prediction accuracy
171 in the N400, in a sentential design with cloze probability of either medium (.5) or low
172 (<.01) cloze probability. Although they did not report effects in the N1, Brothers et al. did
173 show that accurate predictions of upcoming words were associated with more positive
174 amplitudes after 200 ms, in a P2 component immediately following the N1. In another
175 N400 study using a design related to that employed in the present study, Lau et al.

PREDICTION ERROR IN THE WORD N1

Study and Results Summary	SOA	Window	Effect Sites	Reference
Sereno et al. (2003) More negative amplitudes at lower predictability, but only for low frequency words.	450	132-192	Posterior & Anterior ^a	Average
Penolazzi et al. (2007) ^b More negative amplitudes at higher predictability, and no interaction with frequency.	700	170-190	Centroparietal	Mastoids
Dambacher et al. (2012) More negative amplitudes for low than high frequency words, but only at higher predictability.	280 ^c	135-155	Posterior	Average
More negative amplitudes for low than high frequency words, and no interaction with predictability.	280 ^c	190-260	Posterior	Average
Kretzschmar et al. (2015) More negative amplitudes at higher predictability, and no interaction with frequency.	- ^d	150-200	Centroparietal	Mastoids
Sereno et al. (2019) More negative amplitudes at higher predictability, but only for high frequency words.	300	160-200	Left	Average
More negative amplitudes at lower predictability, but only for high frequency words.	300	160-200	Right	Average

SOA: Stimulus Onset Asynchrony (ms).

Window: Analysed ERP Window (ms).

^a This topography describes the first factor in a spatial factor analysis.

^b This study additionally manipulated word length, finding no interaction in the N1.

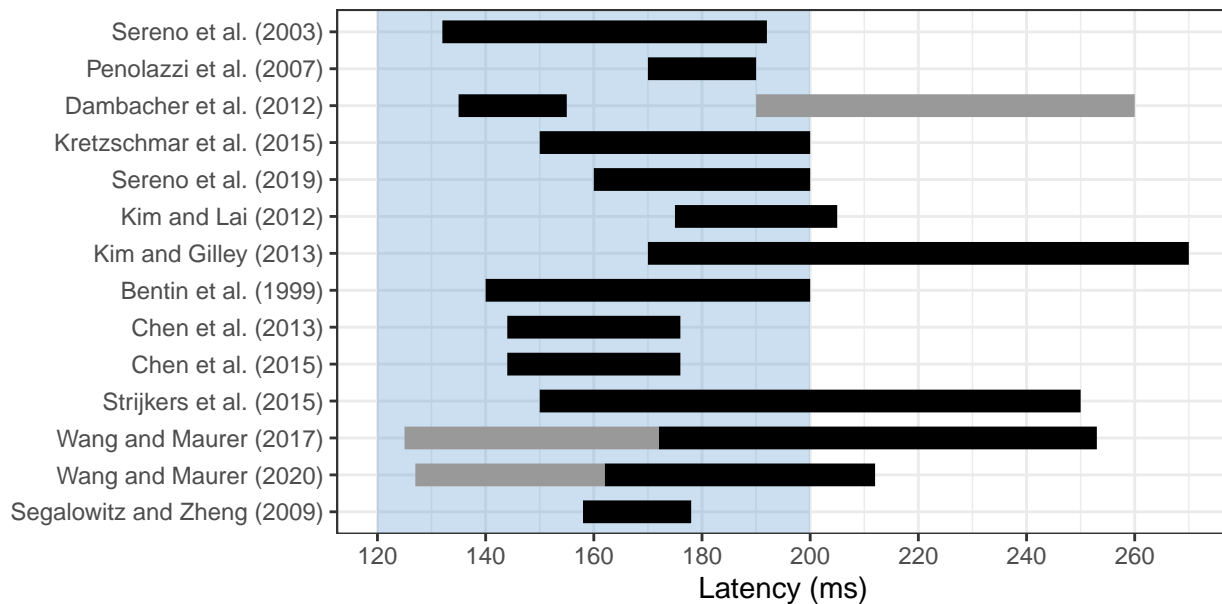
^c SOA varied (280, 490, & 700 ms), with N1 effects only observed at the SOA of 280 ms.

^d This study analysed Fixation-Related Potentials in self-paced reading.

Table 1

Summary of N1 effects reported in studies that biased word form predictions in sentential paradigms, using a Predictability (low, high) × Frequency (low, high) factorial design.

Figure 1
N1 windows in predictability studies.



Some studies analysed two N1 windows (e.g., onset and offset). N1 windows reported to show a predictability effect are highlighted in black, while N1 windows that failed to show a predictability effect are highlighted in grey. Studies are listed in order of their mention in our review. For reference, the blue region displays the N1 period that we pre-registered.

176 (2016) presented adjective-noun pairs to participants in which the effects of both
177 congruency and predictability were examined, showing small congruency, and large
178 predictability, effects in the N400. As with Brothers et al., Lau et al. report ERPs with no
179 robust differences prior to a P2 component.

180 Instead of manipulating error precision or certainty, as the above studies have by
181 varying predictability, Kim and Lai (2012) manipulated the orthographic error *magnitude*.
182 Using a 550 ms SOA, the target word or alternative orthographic versions of it were
183 presented in contexts that were acutely predictive of the target ($M_{Cloze} = .90$). Contexts
184 were followed by the predictable target word (e.g., *cake*), an orthographically similar
185 pseudoword (e.g., *ceke*), an orthographically dissimilar pseudoword (e.g., *tont*), or a
186 consonant-string nonword (e.g., *srdt*). Consistent with an orthographic explanation for
187 prediction effects in the N1, relative to targets, N1 (175-205 ms) amplitude was more

188 negative-going for orthographically dissimilar pseudowords and nonwords (i.e., when
189 orthographic prediction error magnitude was greater). Orthographically similar
190 pseudowords, while significantly different from all other conditions in the earlier P1,
191 elicited N1 components more similar in amplitude to target words.

192 Another linguistic cue that has been manipulated is grammaticality. Kim and
193 Gilley (2013) demonstrated effects of syntactic anomaly on the N1. Sentences leading to
194 a strong prediction for the determiner, *the*, were presented unchanged or with the
195 determiner replaced with an agrammatic preposition (e.g., *The thief was caught by*
196 *the/for police*). The left-lateralised occipitotemporal N1 (170-270 ms) was more
197 negative-going with the syntactically anomalous preposition than with the determiner. As
198 the authors point out, the N1 effect is unlikely to be evidence for sensitivity to syntax per
199 se. Rather, given evidence of the N1's sensitivity to orthographic features, it is probably
200 more accurate to posit that the high predictability of the determiner's orthographic
201 features elicited a less negative-going N1 when these predictions were confirmed.

202 Kim and Gilley's simultaneous manipulation of orthography and syntax highlights
203 a prevalent issue within the literature: namely, altering the visual word form necessitates
204 alteration of the semantics, syntax, and/or plausibility of the sentence or wider discourse.
205 Another limitation shared by studies using word-by-word presentation of sentences is
206 that ERPs elicited by the target word can become difficult to disentangle from ERPs
207 elicited by preceding or succeeding words, especially if the SOA is short or unjittered.
208 While fast presentation times of sentential contexts and targets are useful for
209 demonstrating that early modulation by predictive processes extends to realistic reading
210 rates, their application may not be necessary to demonstrate that such modulation can
211 occur. It is also of note that in a recent review of ERP studies using sentence- and
212 discourse-level contexts to examine early neural correlates of word form prediction,
213 Nieuwland (2019) concluded that findings thus far have been weak, inconsistent, and in
214 need of more replication attempts. Moreover, most studies to date were not

215 pre-registered and often used inappropriate analysis models that did not account for
216 measurement variability, raising questions about false positives in that literature.

217 **Biasing Word Forms via Non-Linguistic Cues**

218 Effects of prediction and expectation may alternatively be investigated using
219 paradigms that modulate non-linguistic features of tasks and stimuli. In one approach,
220 identical or suitably matched stimuli are presented under different task instructions (e.g.,
221 Compton et al., 1991). In that context, tasks requiring more explicit lexical and semantic
222 processing cause words to elicit more negative-going N1s (144-176 ms; Chen et al.,
223 2013). Tasks requiring more in-depth lexicosemantic processing may also increase
224 sensitivity to lexical variables such as word frequency in the N1 (144-176 ms, Chen et al.,
225 2015; 150-250 ms, Strijkers et al., 2015), and may increase the size of script familiarity
226 effects (more negative amplitudes for familiar scripts) (F. Wang & Maurer, 2017). This
227 may be especially in the N1's offset period (172-253 ms F. Wang & Maurer, 2017), where
228 onsets and offsets are defined respectively as the periods in the component's time
229 window which precede and succeed its peak. F. Wang and Maurer (2020) further
230 showed that biasing participants' word form predictions towards expecting a familiar
231 script increased the size of the script familiarity effect in the N1 offset (162-212 ms).

232 In addition to task manipulations, non-sentential semantic contexts, leading to
233 predictions for specific words or categories of words, have also been used to investigate
234 predictive processing. Segalowitz and Zheng (2009) reported an interaction between
235 stimulus type (word vs. pseudoword) and expectation (one vs. five categories) in the N1
236 (158-178 ms), wherein expectation affected N1 amplitudes for words but not for
237 pseudowords. Their finding suggested that the N1 was sensitive to the greater predictive
238 strength of a single semantic category. Using a similar paradigm, Hauk et al. (2012)
239 compared ERPs in lexical and semantic decision tasks, showing that effects of category
240 relevance were observed in the semantic decision task as early as 166 ms (data were
241 analysed continuously, with no N1 window definition). This finding suggests, consistent

242 with the findings of Segalowitz and Zheng, an early sensitivity to category relevance
243 during the N1 which, given the N1's robust sensitivity to orthography, is likely to reflect an
244 influence of semantic-level predictions on orthographic processing.

245 In another attempt to modulate top-down expectancy without linguistic context,
246 Dikker and Pylkkänen (2011) implemented a picture-noun phrase verification task. An
247 image of a target object alone or an image of objects related to the target object was
248 followed by a written noun phrase (article + noun) denoting the target object. They
249 manipulated congruency and predictability. For congruent trials, the noun phrase referred
250 to a food/drink or animal (e.g., *the apple* or *the monkey*) that matched the prior image of
251 the object presented on its own or 'contained' in a stylized image (e.g., a grocery bag or
252 Noah's Ark, respectively). In the incongruent condition, the noun phrase did not match
253 the prior image (single object or collection of objects). Predictability was considered high
254 when the target object appeared on its own, and was considered low when the target
255 object could be inferred to exist within the stylized images. Example conditions for the
256 noun phrase, *the apple*, are determined by its preceding image as follows: an apple
257 (congruent, high predictability), a banana (incongruent, high predictability), a bag of
258 groceries (congruent, low predictability), or Noah's Ark (incongruent, low predictability).
259 Noun phrases (40 food/drink, 40 animal) were repeated four times across conditions.
260 Although Dikker and Pylkkänen did not examine effects in the MEG equivalent of an N1
261 window, they did find effects of congruency only in the high predictive condition (i.e., the
262 apple preceded by an apple vs. a banana image) in temporal windows preceding (~100
263 ms) and succeeding (250-400 ms) the N1. Their stimuli were designed to minimise
264 orthographic similarity between congruent and incongruent pairs of noun phrases (i.e.,
265 maximising the magnitude of orthographic errors), suggesting that the authors
266 anticipated that any early sensory effect of predictability may be related to orthographic
267 processing. With only 7 participants, the study likely lacked the sample size necessary to
268 identify such an effect in an N1-like window. In a study using the same stimuli as Dikker

269 and Pylkkänen, Cheimariou et al. (2019) examined the effects of aging on lexical
270 prediction indexed by the N400 component. Cheimariou et al. also did not analyse an
271 N1 window, and used Mastoid reference sites, but did show that an early, though wide,
272 window from 125-348 ms showed topographically broad prediction effects in younger
273 adults, with more negative amplitudes for predictive content.

274 We note that related paradigms using fMRI often show orthography-semantics
275 interactions in the likely N1 generator, the left vOT (e.g., Branzi et al., 2022; Kherif et al.,
276 2011; J. Wang et al., 2019). However, fMRI prevents the interpretation of the timing of
277 such effects - its coarse temporal resolution means that mapping of semantic content to
278 representations in vOT could occur so late after word presentation as to be irrelevant to
279 initial orthographic word recognition processes.

280 One advantage of paradigms like picture-word verification tasks is that the
281 researcher can control and manipulate variables like predictability and specificity of the
282 picture-word relation. This was demonstrated in the design used by Dikker and
283 Pylkkänen (2011), where the picture preceding the target word unambiguously biased
284 participants' expectations to a single word form (with an image of one clearly identifiable
285 object), or instead biased a set of semantically related possible word forms (with an
286 image inducing multiple object candidates). Such a manipulation is comparable to the
287 use of Cloze probability in sentential contexts or single versus multiple category priming,
288 and similarly aligns with the concept of error precision or certainty.

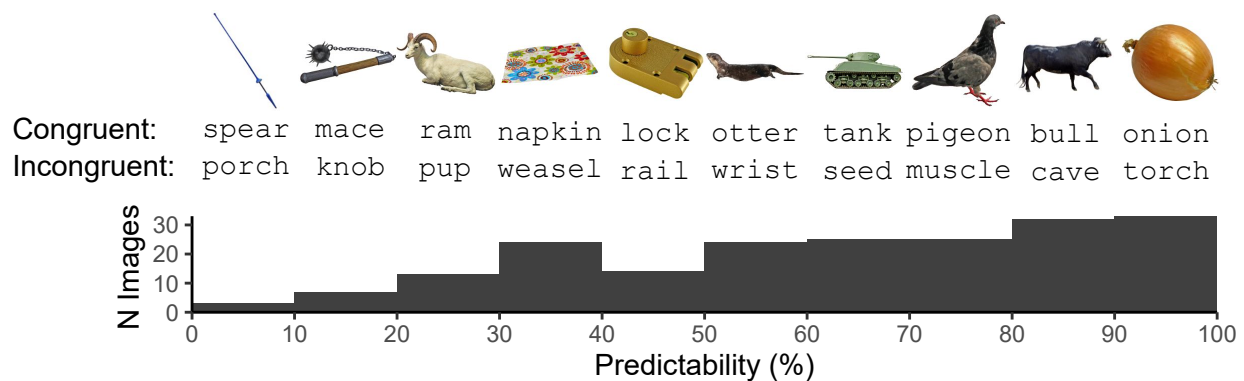
289 **The Present Study**

290 In the present study, we adapted the picture-word verification paradigm to
291 examine the Congruency-Predictability interaction in the N1. We presented participants
292 with PICTURE-word pairs that were congruent (e.g., ONION-onion) or incongruent (e.g.,
293 ONION-torch). Predictability of the congruent word, given the picture that precedes it,
294 was operationalised via a continuous variable drawn from picture naming norms
295 (Brodeur et al., 2014) reflecting the probability of the congruent word being given as a

296 name for the picture (**Figure 2**). Picture-congruent words with very low predictability
297 were always semantically appropriate names for their associated image, though they
298 were difficult to predict, often because several acceptable names exist. For example, the
299 image for *spear* in **Figure 2** could also be plausibly named with words like *lance*, *javelin*,
300 or *pole*. Incongruent words, meanwhile, were specifically selected to be semantically
301 incongruent with congruent words, but matched on relevant psycholinguistic dimensions.
302 By manipulating both Congruency and Predictability of word forms, we were able to
303 examine whether the effect of Congruency on the N1 (sensitivity to prediction error) is
304 contingent on Predictability (certainty or precision of prediction errors), in the manner
305 expected according to a simple predictive coding account of the N1 in which observed
306 N1 magnitude indexes prediction error.

Figure 2

Illustration of the experimental stimuli.



PICTURE-word pairs were either congruent (e.g., NAPKIN-*napkin*) or incongruent (e.g., NAPKIN-*weasel*), while predictability of congruent picture-word pairs varied continuously. Ten example picture-congruent and -incongruent pairs are presented, with their predictability corresponding to the histogram bin they appear above.

307 We hypothesised, consistent with such a predictive coding account, that that there
308 would be a Congruency-Predictability interaction in which at the highest levels of
309 Predictability, N1s elicited by picture-incongruent words would be more negative-going
310 than those elicited by picture-congruent words, while at the lowest level of Predictability
311 picture-congruent and -incongruent words should elicit N1s of similar magnitude. We

312 anticipated three patterns of results that would have been consistent with this hypothesis:

313 (1) higher levels of Predictability lead to a reduction in N1 magnitude only for
314 picture-congruent words, with no such effect for picture-incongruent words (**Figure 3a**);

315 (2) higher levels of Predictability lead to an increase in N1 magnitude only for
316 picture-incongruent words, with no such effect for picture-congruent words (**Figure 3b**);

317 or

318 (3) higher levels of Predictability lead to both a reduction in N1 magnitude for
319 picture-congruent words and an increase in N1 magnitude for picture-incongruent words
320 (**Figure 3c**).

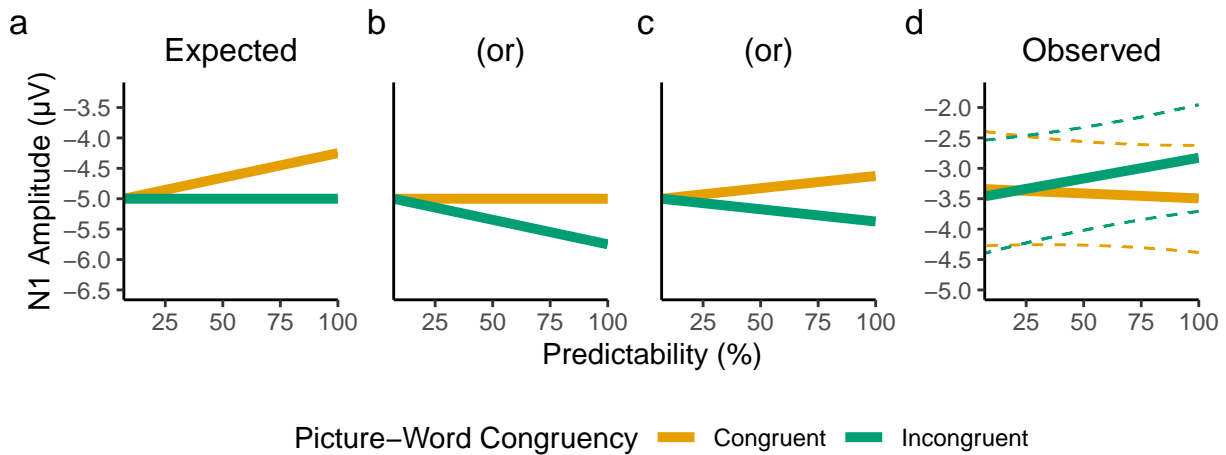
321 In our power analysis, we focused on the first of these possible patterns of results,
322 but importantly, the Congruency-Predictability interaction term that we pre-registered to
323 test our hypothesis would capture any of these patterns, as the interaction term's
324 coefficient would be in the same direction in all cases.

325 In our analysis, we found a pattern of effects counter to our pre-registered
326 hypothesis (**Figure 3d**), with a Congruency-Predictability interaction in the opposite
327 direction. An exploratory Bayesian analysis revealed that the observed interaction was
328 16.61 times more likely than our hypothesis. Based on these findings, we argue our
329 results suggest that such a simplistic predictive coding account is, at least on its own,
330 insufficient to explain the pattern of prediction effects observed in the N1 during a
331 picture-word verification task.

332 This study was pre-registered at <https://osf.io/jk3r4> and the reported methodology
333 and planned analysis conform to that specified in the pre-registration, except for two
334 changes: an accidental change to timing of stimuli, and a lowering of the EEG high-pass
335 filter cut-off. We explain these changes in the relevant sections, and demonstrate in
336 **Supplementary Materials F** that the change to the high-pass filter cut-off had minimal
337 effect on the results and conclusions. All data and code are available at
338 <https://osf.io/389ce/>.

Figure 3

A comparison between the predicted (a,b,c) and observed (d) patterns of results.



The predicted pattern of results was based on a predictive coding interpretation of the N1, according to which the magnitude of the N1 should be smaller for picture-congruent words relative to picture-incongruent words, and to a greater extent as Predictability increases. The observed pattern of results depicts the fixed effect predictions from the pre-registered linear mixed-effects model, with dashed lines depicting 95% bootstrapped prediction intervals (estimated from 5,000 bootstrap samples).

Method

339

340 The experiment included two separate tasks: The principal picture-word task was
341 preceded by a localiser task to account for between-participant variability in the N1's
342 timing and location. The details of stimulus selection and control as well as presentation
343 timing are provided in the following sections. For clarity, we first introduce the overall
344 Congruency-Predictability design of the picture-word task. In this task, pictures of single
345 objects are presented, followed by a noun, and participants decide whether the noun
346 corresponds to the object. The level of Predictability of the noun was determined from
347 norms of possible terms used to label a set of individual pictures (Brodeur et al., 2014).
348 The most frequent, modal name agreement varied across pictures. Thus, level of noun
349 Predictability was continuous and varied between 7% and 100%. The Congruency of the
350 noun was either congruent (matching the modal name of the picture) or incongruent (a
351 semantically unrelated noun matched across several lexical variables).

352 **Materials: Picture-Word Task**

353 A total of 400 words (200 per Congruency condition) were selected with LexOPS
354 (Taylor et al., 2020), a package for the generation and control of lexical variables in the R
355 programming language (R Core Team, 2021). Picture-congruent and -incongruent words
356 were matched precisely in an item-wise manner on a range of relevant psycholinguistic
357 variables, comprising word length, frequency, concreteness, OLD20, and character
358 bigram probability. To ensure that picture-incongruent words were not inadvertent
359 possible descriptors for images, we minimised the semantic relatedness between pairs
360 of words. Additionally, counterbalanced sets of stimuli were matched on distributions in
361 these variables using a measure of distributional similarity (Pastore & Calcagni, 2019). A
362 full description of the method by which stimuli were selected, and a full list of stimuli, is
363 available in **Supplementary Materials A**.

364 Before embarking on the electrophysiological picture-word experiment, we first
365 ran a proof-of-concept behavioural experiment using a different stimulus set generated
366 from a very similar pipeline. We anticipated that increased Predictability should cause
367 faster response time (RT) for congruent trials and have either no effect or a minimal
368 effect on performance for incongruent trials. The results from this behavioural validation
369 are presented in **Supplementary Materials B**. In short, we observed the pattern of
370 results consistent with our expectations, with Predictability leading to faster RTs for
371 congruent trials, but having almost no effect on incongruent trials.

372 **Materials: Localiser Task**

373 The precise location of the N1, and timing of its peak amplitude, is known to vary
374 across studies and among participants. As such, we did not specify a common N1
375 electrode or timepoint shared among all participants before data collection. Instead, we
376 employed a localiser task to identify, within an appropriate region and time period of
377 interest, the electrode and timepoint at which each participant's maximal sensitivity to
378 orthography emerges (i.e., more extreme amplitudes for words than false-font stimuli).

379 This data could then be used to extract N1 amplitudes in the picture-word task, while
380 accounting for variability among participants in timing and topography of orthographic
381 processes.

382 For the localiser task, three categories of stimuli were presented for 100 trials
383 each (**Figure 4**). These consisted of matched triplets of words (Courier New font),
384 false-font strings (BACS2serif font; Vidal et al., 2017), and phase-shuffled words. The
385 comparison between words and false-font strings is a standard measure of N1 sensitivity
386 to orthography, with previous evidence suggesting a more robust difference than exists
387 between nonwords and words (Brem et al., 2018; Maurer, Brandeis, & McCandliss,
388 2005; Pleisch et al., 2019). However, phase-shuffled words were employed as an
389 alternative comparison for exploratory analyses, with equal spatial-frequency amplitude
390 and permuted spatial-frequency phase. Similar phase-shuffled word stimuli have shown
391 robust differences to word forms in fMRI investigations of vOT activity (Rauschecker
392 et al., 2012; Rodrigues et al., 2019; White et al., 2019; Yeatman et al., 2013).

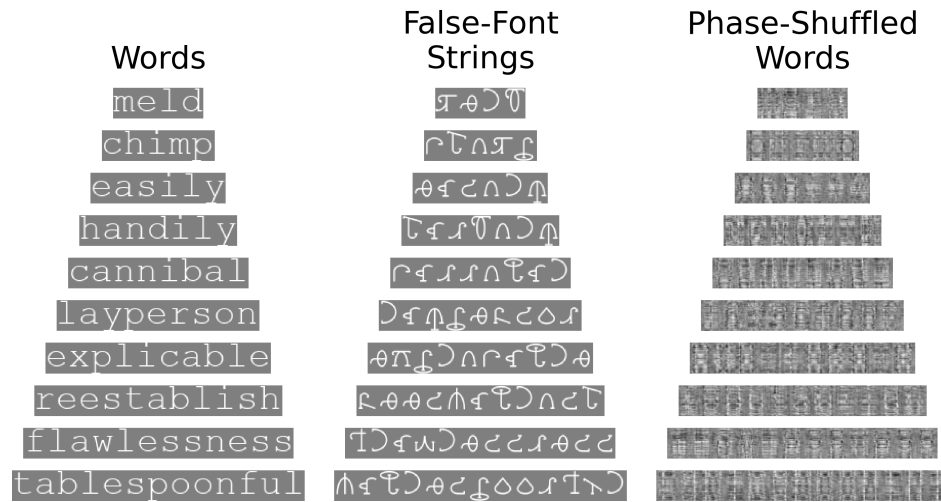
393 The word stimuli were selected to be widely known by participants (>90%
394 proportion known), and to be representative on a range of psycholinguistic variables
395 including length, frequency, part of speech, and prevalence. A full description of how the
396 Localiser Stimuli were selected, and a list of all word stimuli, is presented in
397 Supplementary Materials C.

398 **Participants**

399 The sample size of 68 participants was decided via a power analysis using
400 Monte-Carlo simulations of a realistic effect size (**Supplementary Materials D**). This
401 revealed that with ≥ 68 participants we could expect >80% statistical power in the long
402 run (**Figure 5**). All 68 participants (40 female, 27 male, 1 non-binary) were monolingual
403 native English speakers. Participants were randomly allocated into one of the four
404 combinations of stimulus set (Set 1, Set 2) and response group (i.e., the left-right
405 mapping of the two response buttons for affirmative and negative responses), such that

Figure 4

Ten example stimuli for each stimulus type in the localiser task.



Each row represents a matched triplet of word, false-font string, and phase-shuffled word stimuli. The phase-shuffled word images were generated uniquely for each trial.

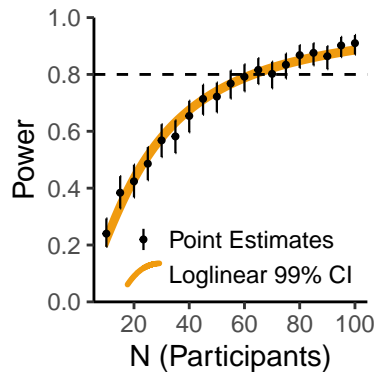
406 each combination of stimulus set and response group comprised 17 participants. No
407 participants reported diagnosis of any reading disorder. Ages ranged from 18 to 37 years
408 (M=22.69, SD=4.9), and all participants reported having normal or corrected-to-normal
409 vision. Participants' handedness was assessed via the revised short form of the
410 Edinburgh Handedness Inventory (Veale, 2014), with participants only permitted to take
411 part if they scored a laterality quotient of +40 indicating right handedness. Exclusion
412 criteria for participants were determined prior to data collection as follows: (1) if 10 or
413 more channels showed an offset more extreme than ± 25 mV (as measured on the
414 BioSemi acquisition software, ActiView), or (2) if more than 5% of the trials were lost due
415 to technical issues with the EEG system. As no participants satisfied these criteria, no
416 participants were excluded after data collection. Data collection was approved by the
417 Ethics Committee of the University of Glasgow College of Science and Engineering
418 (application number: 300200117), and all participants provided informed consent.

419 Procedure

420 Stimuli were presented on a VPixx Technologies VIEWPixx screen (resolution
421 1920×1080 pixels, diagonal length 23", model VPX-VPX-2004A). Participants completed

Figure 5

Estimated relationship between number of participants and statistical power.



Black points and error bars depict point estimates $\pm 99\%$ Binomial confidence intervals, each from 500 simulations. As 500 simulations provides a noisy estimate, we interpolated the relationship between N and power via a loglinear, logit-link Binomial model. The *orange* region depicts the 99% confidence intervals of this loglinear model.

422 the experiment on a chin rest positioned 48 cm from the centre of the screen. Stimuli
423 were presented on a grey background equal to 50% of the maximum intensity in each
424 colour channel, roughly 12.3 cd/m². The experiment was written using the Python library
425 PsychoPy (Peirce, 2007), and all code and materials are available in the repository
426 associated with the study. All stimuli were presented centrally (horizontally and
427 vertically). All trials in both tasks were presented in a pseudo-randomised order, such
428 that no more than five consecutive trials required the same response from the
429 participant. Trials were randomised across blocks, with the exception of the practice
430 block, for which trials were randomised within the one block.

431 A mistake in the lab setup, which we discovered after data collection, meant that
432 the display screen was running at 120 Hz rather than an expected 60 Hz. As we were
433 controlling stimulus presentation by screen refreshes, this meant that all our stimuli were
434 presented for half the expected durations. For this reason, the veridical stimulus
435 durations described here differ from those described in the pre-registration.

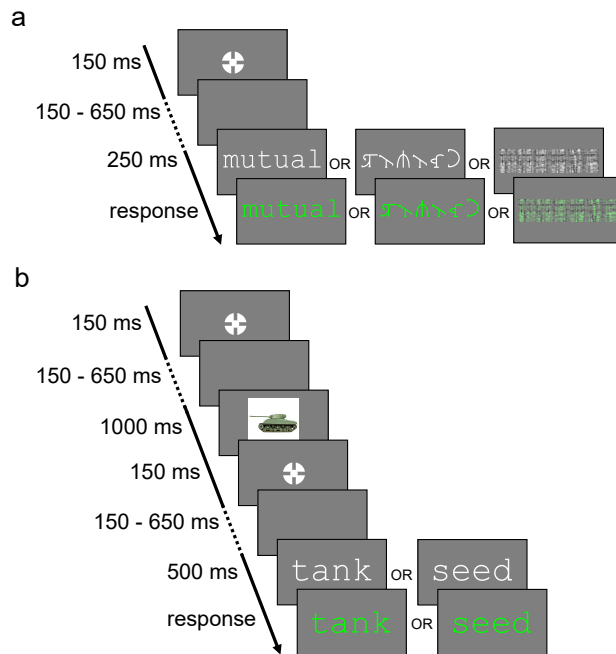
436 Participants started with the localiser task, in the form of a lexical decision task
437 (**Figure 6a**). The localiser task began with 30 practice trials, and was then followed by

438 300 trials split into 5 blocks of 60 trials. Each trial began with the bullseye fixation target
439 recommended by Thaler et al. (2013) (outer and inner circle diameters were 0.6° and
440 0.2° of visual angle), presented for 150 ms. This was followed by a jittered interval of
441 between 150 and 650 ms, during which the screen was blank. The stimulus (word,
442 false-font string, or phase-shuffled word image) was then presented at a height of 1.5°
443 (width of 1.07° for one character). Words and false-font strings were presented in white
444 (80 cd/m^2), in the respective fonts of non-proportional Courier New and BACS2serif font.
445 The stimulus was visible for 250 ms, after which the font colour changed to green to
446 signal participants to respond. Participants were requested to respond once after the
447 stimulus changed colour, quickly and accurately, to indicate whether the stimulus they
448 saw in each trial was either a word or not a word. The stimulus remained on screen until
449 the participant responded. Responses were given with the right and left control ('Ctrl')
450 keys of a QWERTY keyboard, with the mapping of affirmative and negative responses
451 counterbalanced across participants. After the participant had responded, there was a
452 delay of around 100 ms (variable as data was saved to disk during this interval), and
453 then the next trial began.

454 After the localiser task, participants completed the picture-word task (**Figure 6b**),
455 comprising an initial practice block of 20 trials, followed by 200 trials split into 5 blocks of
456 40 trials. As in the localiser task, each trial in the picture-word task began with the
457 bullseye fixation point, presented for 150 ms, after which there was a blank screen for a
458 jittered interval of between 150 and 650 ms. An image was then presented for 1000 ms,
459 at a size of $10 \times 10^\circ$. The bullseye fixation point was then presented again for 150 ms,
460 followed by another interval jittered between 150 and 650 ms. The word was then
461 presented in white Courier New font, at a height of 1.5° (width 1.07° for one character).
462 After 500 ms, the word turned green, and participants could provide their response to
463 indicate whether the word described the image they saw. The word remained on screen
464 until the participant responded. As in the localiser task, responses were given with the

Figure 6

Trial structure of the (a) localiser task and (b) picture-word task.



This figure is illustrative and the sizes are not to scale; in the experiment, images were in fact presented at a much larger scale than words.

465 right and left control ('Ctrl') keys of a QWERTY keyboard, with the mapping of affirmative
466 and negative responses counterbalanced across participants, but kept consistent within
467 participants across the two tasks. After participants had responded, there was a delay of
468 around 100 ms (again, variable as data was saved to disk during this interval), and then
469 the next trial began. There was no deadline for participants to respond. The instructions
470 given to participants for the picture-word task are presented in **Supplementary**
471 **Materials E**.

472 The first blocks of both tasks consisted of practice trials with 10 exemplars for
473 each stimulus type (word or false-font string or phase-shifted image, and congruent or
474 incongruent noun for the localiser and picture-word tasks, respectively), during which
475 participants were additionally given immediate feedback on their accuracy for each trial.
476 These practice trials were followed by green text reading "CORRECT!" if the participant
477 responded correctly, or else by red text reading "INCORRECT!", presented in Courier

478 New font with a height of 1.5°, for 1000 ms. Participants had self-paced breaks between
479 blocks for each task. Before the practice trials and at the start of every experimental
480 block, participants were presented with instructions for the task (available in
481 **Supplementary Materials E**), summarising what would occur in each trial, and
482 specifying that they should respond as quickly and accurately as possible once the
483 stimulus turned green. These instructions also specified which keys participants should
484 press to indicate their decision. After each experimental block, including the practice
485 trials, participants were presented with their average accuracy and median response
486 time. After the practice trials, participants were additionally given the option to run the
487 practice trials again. In the experimental blocks, no trial-level feedback was provided.

488 **Recording**

489 EEG data were recorded using a 64-channel BioSemi ActiveTwo system,
490 sampling at 512 Hz, with an online anti-aliasing low-pass filter cutoff at one fifth of the
491 sample rate (i.e., 102.4 Hz). Electrodes were positioned in the standard 10-20 system
492 locations. Four electro-oculography (EOG) electrodes were placed to record eye
493 movements and blinks: 2 were placed to the sides of eyes (on the right and left outer
494 canthi), and 2 below the eyes (on the infraorbital foramen). Electrode offset was kept
495 stable and low through the recording, within ± 25 mV, as measured by the BioSemi
496 ActiView EEG acquisition tool. Electrodes whose activity exceeded this threshold were
497 recorded but were removed (and interpolated) in data preprocessing.

498 **Preprocessing**

499 The following section details the procedure applied to EEG data from each
500 individual session, with the same pipeline being applied to both the localisation task and
501 picture-word task unless otherwise specified. EEG preprocessing was achieved using
502 functions from the EEGLAB (Delorme & Makeig, 2004) toolbox for MATLAB (MATLAB,
503 2022) or OCTAVE (Eaton et al., 2020). For both tasks, trials were excluded if responded
504 to incorrectly (total of $N_{trials}=368$, or .02%, in localiser task, and $N_{trials}=226$, or .02%, in

505 picture-word). Further trials were excluded if responded to later than 1500 ms after the
506 word (or nonword) changed colour (total of $N_{trials}=41$, or .002%, in localiser task,
507 $N_{trials}=42$, or .003%, in picture-word).

508 Channels recorded as having offsets ± 25 mV during data acquisition were
509 removed from the data (in both tasks, 56 channels, or 1.27%, were removed across all
510 participants), with their activity to be later interpolated. In addition, we found that even
511 when not identified as problematic during recording, the channel PO4 was consistently
512 noisy, and so we interpolated this channel for all participants. PO4 was not part of our
513 left occipitotemporal region of interest, but was interpolated for exploratory analyses of
514 the whole scalp, and to avoid affecting other preprocessing steps. Interpolating electrode
515 PO4 was not a preregistered step. However, we note that this change did not alter the
516 direction of any results, rather, only reducing the size of effects. After interpolation, the
517 EEG data were then re-referenced to the average activity across all electrodes and
518 filtered with a 4th order Butterworth filter between .1 and 40 Hz. To counteract the
519 distortion in signals' timing (phase) that is inherent to causal filters, the filter was applied
520 in both directions (i.e., two-pass), with the MATLAB function *filtfilt()*. In our
521 pre-registration, we specified that we would apply a Butterworth filter with a bandpass of
522 .5-40 Hz. However, after the pre-registration, we considered that, consistent with
523 research into the effects of high-pass filters (Rousselet, 2012; Tanner et al., 2015;
524 VanRullen, 2011), this could produce artefactually early effects. As a result, we lowered
525 the high-pass filter to a less problematic .1 Hz. For comparison, demonstrating that our
526 change to the pre-registered pipeline had minimal effect on the results or our
527 conclusions, the results using the original filter are presented in **Supplementary**
528 **Materials F**.

529 Segments of data outside of experimental blocks (i.e., in break periods) were
530 identified and removed so they did not impact the independent components analysis
531 (ICA) applied later in the pipeline. Blocks were identified as beginning 500 ms before

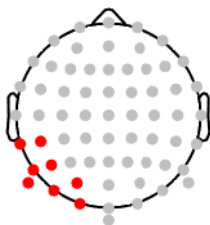
532 stimulus presentation in the first trial of each block, ending 500 ms after the end of the
533 last trial's epoch. To reduce the impact of occasional non-stationary artefacts with high
534 amplitude (such as infrequent muscle movements), artefact subspace reconstruction
535 (ASR; Chang et al., 2020) was used with a standard deviation cutoff of 20 to remove
536 non-stationary artefacts. Following this, an ICA was run on the data to identify more
537 stationary artefacts. The ICA was run using the FastICA algorithm (Hyvärinen & Oja,
538 1997), with a recorded random seed for reproducibility. The ICA was run on a copy of the
539 data with channel offsets removed to allow for better sensitivity to electro-oculogram
540 (EOG) artefacts (Groppe et al., 2009). The ICLabel classifier (Pion-Tonachini et al.,
541 2019) was used to automatically identify artefacts which were eye- or muscle-related.
542 Components classified by ICLabel as eye-related or muscle-related with a probability of
543 $\geq 85\%$ were removed from the data. Following eye movement artefact removal, activity
544 from channels which were removed was interpolated via spherical splines (Localiser:
545 $M=1.14$ per participant, $SD=1.58$; Picture-Word: $M=1.68$, $SD=2.03$), as implemented in
546 EEGLAB. Trials were then epoched and baseline-corrected to the 200 ms preceding
547 stimulus presentation. For the localiser task, stimulus presentation refers to the time
548 point at which words, false-font strings, or phase-shuffled images were presented; in the
549 picture-word task, stimulus presentation refers to the target word.

550 For the planned analysis, we pre-registered an approach to maximise sensitivity
551 to effects of Congruency and Predictability on the N1. To encompass the typical
552 topography and timing of the posterior left-lateralised N1, we selected eight
553 occipitotemporal electrodes (**Figure 7**; electrodes O1, PO3, PO7, P5, P7, P9, CP5, and
554 TP7) and a 120-200 ms window. In contrast to some previous studies whose N1
555 windows extended beyond 200 ms, we set 200 ms as an upper bound for the possible
556 maximal timepoint in the main analysis, to ensure effects were indeed restricted to the
557 N1, and not later components like the N400. For each participant, we identified the
558 electrode that showed maximal sensitivity to orthographic information in the N1 during

559 the localisation task. Specifically, each participant's "maximal electrode" (within the
560 region of interest and selected time window) was the one which showed the largest
561 mean amplitude difference, in the expected direction, across all localiser trials between
562 word and false-font string stimuli. The expected direction was a more negative-going N1
563 for words than for false-font strings, a pattern based on previous findings (Appelbaum
564 et al., 2009; Bentin et al., 1999; Eberhard-Moscicka et al., 2016; Pleisch et al., 2019;
565 Zhao et al., 2014). Each participant's "maximal timepoint" was the timepoint at which the
566 maximal electrode showed the greatest sensitivity to the word-versus-false-font
567 difference in the expected direction. Each participant's maximal electrode and maximal
568 timepoint were then used to extract their trial-level N1 amplitudes from the picture-word
569 task. To reduce the influence of noise on trial-level data, the trial-level N1 amplitudes in
570 the picture-word task were calculated as the maximal electrode's mean amplitude across
571 3 timepoints: the participant's maximal timepoint, and the timepoints immediately
572 preceding and following it. At the recorded sample rate of 512 Hz, this is equivalent to a
573 window of 5.85 ms (i.e., $1/512 \times 3$) centred on the maximal timepoint.

Figure 7

The left-lateralised occipitotemporal region of interest selected for the N1 (highlighted in red).



574 Planned Analysis

575 Our planned analysis tested the pre-registered hypothesis of a
576 Congruency-Predictability interaction in which N1 amplitudes are reduced (i.e., less
577 negative going) for picture-congruent trials than for picture-incongruent trials, and in

578 which this difference is greatest at the highest levels of predictability, and smallest at the
579 lowest levels of predictability. This was based on the notion that the N1 indexes
580 prediction error in biasing contexts.

581 The trial-level N1 amplitudes from the picture-word task were modelled using a
582 linear mixed-effects model fit with the R package *lme4* (Bates et al., 2015), estimating
583 the maximal random effects structure justified by the experiment's design (Barr et al.,
584 2013) as detailed in the section on the power analysis. The model was fit using the
585 *bobyqa* optimiser (Powell, 2009). In *lme4* syntax, the formula for the mixed-effect model
586 was specified as:

```
587 amplitude ~ 1 + congruency * predictability +  
588   (1 + congruency * predictability | participant_id) +  
589   (1 + congruency | image_id) +  
590   (1 | word_id)
```

591

592 In this formula, *amplitude* is the trial-level N1 amplitude in microvolts, while
593 *congruency* is a deviation-coded categorical variable indicating whether a given trial's
594 word was picture-congruent or -incongruent, and *predictability* refers to the proportion of
595 name agreement in the BOSS norms, normalised between 0 and 1. A consequence of
596 this coding method is that the model's intercept reflects the predicted amplitude at the
597 lowest level of Predictability, averaged across both levels of Congruency, while the
598 slopes' coefficients are standardised and directly comparable in their magnitude. The
599 variables of *participant_id*, *image_id*, and *word_id*, in the formula, identify each trial's
600 participant, image, and word, respectively.

601

Results

602 The planned, pre-registered analysis examined whether the hypothesised effect
603 of a Predictability-dependent reduction of N1 amplitudes for picture-congruent words
604 was observed at the electrode/timepoint in which each participant showed maximal
605 sensitivity to orthography. We then present exploratory analyses, which respectively

606 examine the Bayesian probability that our data are consistent with the hypothesis, and
607 delineate the time-course of the Congruency-Predictability interaction. We also
608 conducted further exploratory analyses, which we report in the supplementary materials,
609 examining behavioural results in the picture-word study (**Supplementary Materials G**),
610 and EEG and behavioural results from the localiser task (**Supplementary Materials H**)

611 **Planned Analysis**

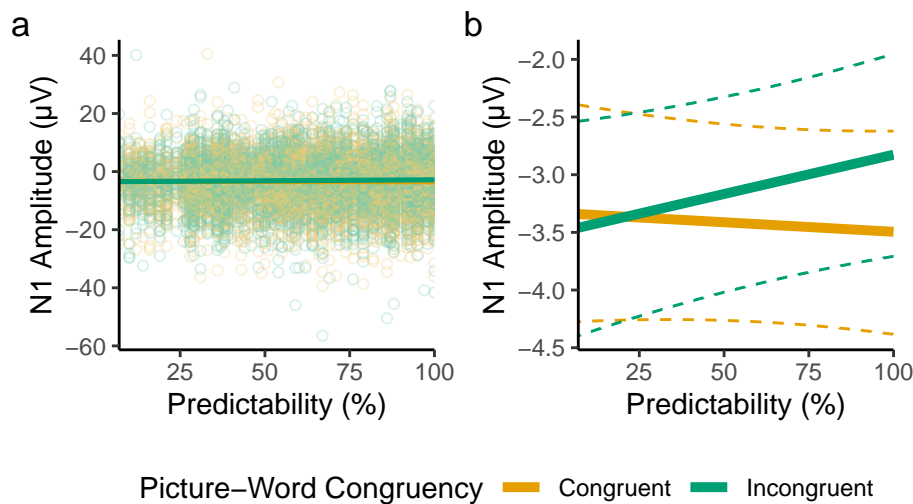
612 The fixed effect relationships estimated in the planned analysis are presented in
613 **Figure 8**. The model intercept, reflecting the average N1 amplitude at the lowest level of
614 Predictability, was estimated to be $\beta = -3.4 \mu\text{V}$ ($SE = .48$). The fixed effect of Congruency
615 from this model was estimated as $\beta = -.12 \mu\text{V}$ ($SE = .34$), which captures that, at the lowest
616 level of Predictability (7%), N1 components for picture-congruent and -incongruent words
617 were estimated to be quite similar (.12 μV difference). The main effect of Predictability
618 was estimated as $\beta = .25 \mu\text{V}$ ($SE = .27$), meaning that N1 amplitudes, averaged across
619 congruent and incongruent conditions, were .25 μV less negative-going at the highest
620 level (100%) than at the lowest level of Predictability (7%). The effect of interest, the
621 interaction between Congruency and Predictability, was in the opposite direction from
622 that hypothesised, estimated as $\beta = -.79 \mu\text{V}$ ($SE = .52$). As our hypothesis was directional,
623 with a prediction in the opposite direction, we interpret these results as a failure to find
624 evidence in favour of the hypothesis.

625 To describe the estimated interaction, for picture-incongruent words, the effect of
626 Predictability was estimated to be $\beta = .63 \mu\text{V}$ ($SE = .36$), while for picture-congruent words,
627 the effect of Predictability was estimated to be $\beta = -.15 \mu\text{V}$ ($SE = .4$). As such, the slopes for
628 the effect of Predictability in both Congruency conditions were of different magnitudes,
629 and were both in directions inconsistent with our predictive coding hypothesis.

630 For comparison, we also analysed the data altering aspects of our planned
631 analysis method: first using the maximal electrodes that would be identified from the
632 comparison between words and phase-shuffled words, and second using averages

Figure 8

Fixed effect predictions from the planned analysis of the picture-word task.



(a) Model-derived fixed-effect predictions, visualised over results from all trials (individual points). (b) Fixed-effect predictions visualised alone for visibility, with dashed lines depicting the bounds of 95% bootstrapped prediction intervals (estimated from 5,000 samples), where bootstrapped predictions were generated using the *bootMer()* function of *lme4*. For feasibility, bootstrapped predictions were generated from a version of the model that lacked random slopes.

633 within the occipitotemporal region of interest (**Supplementary Materials I**). These
634 exploratory analyses revealed very similar patterns of effects, with estimates of the
635 Congruency-Predictability interaction similarly inconsistent with our hypothesis, which we
636 derived from a simple predictive coding account of the N1.

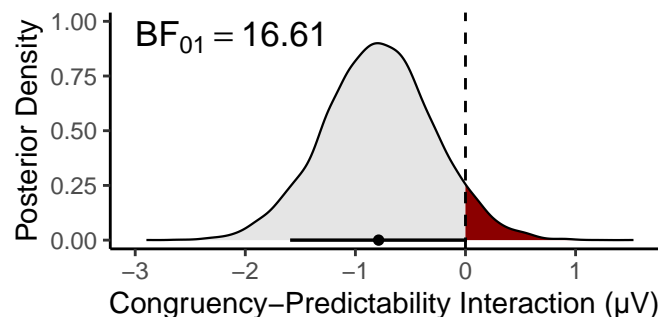
637 Exploratory Bayesian Analysis

638 We observed a Congruency-Predictability interaction in the opposite direction
639 (i.e., negative) to what we expected under our predictive coding hypothesis (i.e.,
640 positive). To explicitly quantify the probability of our predictive coding hypothesis, we fit a
641 Bayesian implementation of the model described in the planned analysis, in STAN
642 (STAN Development Team, 2023) via *brms* (Bürkner, 2017). This model was fit to the
643 same data, and estimated the same hierarchical formula, with the same Gaussian link
644 function as that described above, but was specified with weakly informative priors for the

645 fixed effects. Specifically, the prior for the fixed effect intercept was specified as a normal
646 distribution of mean -5 , and SD 10 , while all fixed effect slopes' priors were specified as
647 normal distributions centred on 0 , with SDs of 5 . Covariance matrices were assigned flat
648 priors, and default priors for brms were used for random effect SDs and the sigma
649 parameter of the normal distribution. The model was fit with 5 chains and 5000 iterations
650 per chain (split equally between warmup and sampling) such that there were a total of
651 12,500 posterior samples. Consistent with the linear mixed-effects model we fit via *lme4*,
652 this analysis revealed a median posterior estimate for the Congruency-Predictability
653 interaction of $\beta = -.79 \mu V$ (89% highest density interval = $[-1.59, .013]$; **Figure 9**). We
654 calculated, given this posterior distribution, that the Congruency-Predictability interaction
655 is 16.61 times more likely to be less than 0, than it is to be greater than zero (that is,
656 BF_{01}), which we consider to be strong evidence against our hypothesis.

Figure 9

Posterior density for the Congruency-Predictability interaction.



The region of the posterior distribution consistent with the predictive coding hypothesis (where $\beta > 0$) is highlighted in *red*. The point and horizontal line below the density plot depict respectively the median estimate and 89% highest density interval of the posterior distribution.

657 We considered that our use of a localiser task may have been an inappropriate
658 approach for identifying electrodes sensitive to orthographic information. Indeed, our
659 pre-registered approach for identifying maximal electrodes specified the direction of the
660 difference that should be used, with more negative N1s for words than for false-font
661 strings. However, exploratory ERP analyses of the localiser task showed that

662 left-lateralised occipitotemporal electrodes showed a more negative N1 peak overall for
663 false-font strings than for words (**Supplementary Materials H**). Our approach may
664 therefore have systematically selected electrodes that are not representative of the ROI.
665 As a result, we re-ran the Bayesian analysis as described above, but modelling average
666 amplitudes from all electrodes in the left occipitotemporal ROI (**Supplementary**
667 **Materials I**). This revealed even stronger evidence against the hypothesis, with a
668 Congruency-Predictability interaction for the average amplitude in the ROI of $\beta=-1.03 \mu\text{V}$
669 (89% highest density interval = [-1.52, -.058], estimated to be 2082 times more likely to
670 be less than 0, than it is to be greater than zero.

671 **Exploratory Time-Course Analysis**

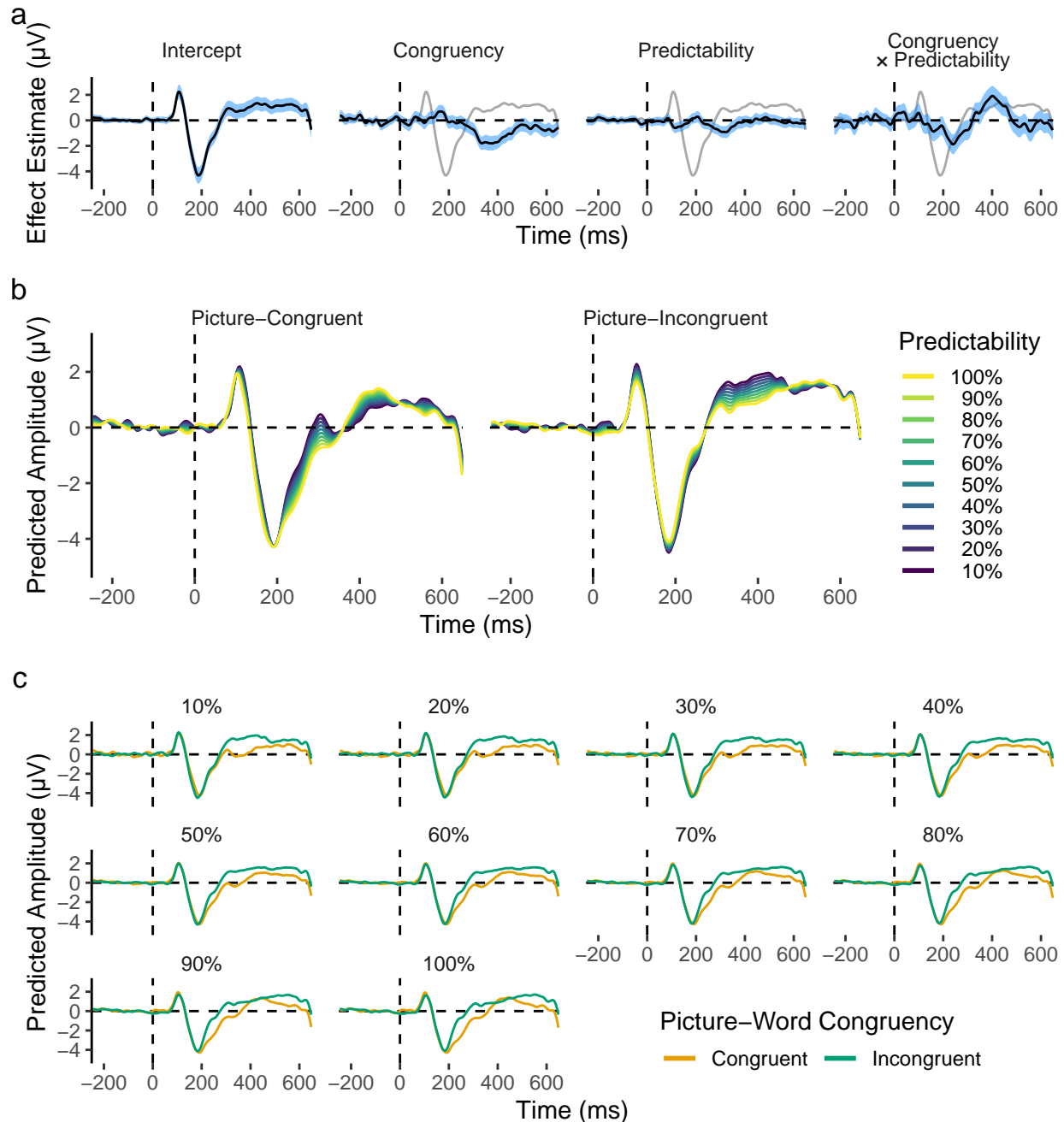
672 To examine the time-course of effects, we fit separate linear mixed-effects models
673 to sample level data for the left-lateralised occipitotemporal region of interest, with
674 variables coded as described for the planned analysis. For feasibility, data were
675 down-sampled to 256 Hz, and the models did not estimate random slopes. To account
676 for variability between electrodes, and for per-participant differences in topography,
677 random intercepts were estimated for each combination of participant and electrode. In
678 *lme4* syntax, the model formula was specified as follows:

```
679 amplitude ~ 1 + congruency * predictability +  
680 (1 | participant_id) +  
681 (1 | participant_id:electrode_id) +  
682 (1 | image_id) +  
683 (1 | word_id)
```

685 The results (**Figure 10**) reproduced findings from the planned analysis, with
686 increases in Predictability associated with more negative (larger) N1 amplitudes for
687 picture-congruent words, and with less negative (smaller) N1 amplitudes for
688 picture-incongruent words. The Congruency-Predictability interaction of interest
689 remained negative, and thus in the opposite direction to that hypothesised, throughout
690 the N1.

Figure 10

Time-course of fixed effects from the sample-level analysis of the left-lateralised occipitotemporal region of interest.

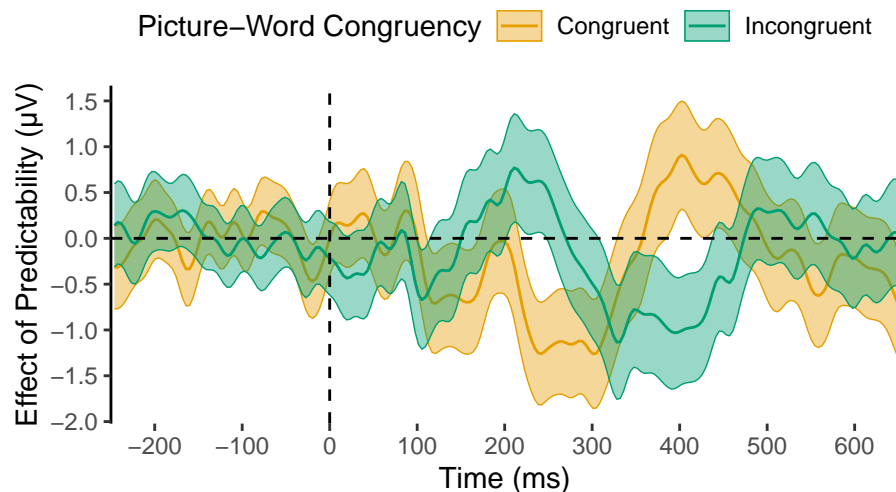


(a) Time-course of fixed effects estimates, with blue-shaded regions depicting 95% confidence intervals. The model intercept (reflecting average amplitudes at the lowest level of Predictability) is depicted as a grey line on each panel to provide a reference for timing and magnitude of effects. (b) Fixed-effect predictions for picture-congruent and -incongruent words at levels of Predictability from 10 to 100%, in steps of 10%. (c) Same data as (b), but split by Predictability rather than Congruency. The rapid change in amplitude after 650 ms was likely elicited by the stimulus colour change at 500 ms, as shown more clearly in **Figure 12**.

691 The sample-level analysis additionally suggested that the difference was largest
692 in the N1's offset period (succeeding the peak). A later Congruency-Predictability
693 interaction was also observed, peaking at around 400 ms (possibly resulting from effects
694 in the N400 component) in the opposite direction to that observed for the N1's offset. To
695 better understand the time-course of the Congruency-Predictability interaction, we
696 examined the time-course of the effect of Predictability for picture-congruent and
697 -incongruent words separately (i.e., simple effects; **Figure 11**). This showed more clearly
698 that Predictability reduced amplitudes in the N1 for picture-incongruent words, but
699 increased amplitudes for picture-congruent words. This difference peaked around 225
700 ms, but reversed in direction after 300 ms. It is of note that the peak of the observed
701 effects in the N1 was later than originally anticipated (the planned analysis was limited to
702 ≤ 200 ms). Nevertheless, the model intercept (**Figure 10a**) clearly shows that these
703 effects peaked during the N1's offset period.

Figure 11

Time-course of the effect of Predictability for picture-congruent and -incongruent words.



Central lines depict effect estimates, derived from sample-level models that were coded such that the model intercept lay at the respective levels of picture-word Congruency. Estimates reflect occipitotemporal ERPs for words at the maximum level of Predictability, minus those at the minimum level of Predictability. Shaded areas depict 95% confidence intervals of model estimates.

704 **Exploratory Scalp-Wide Analysis**

705 Finally, we examined how the full topography of effects changed over time
706 (**Figure 12**). Specifically, we fit a linear mixed-effects model to data from each time-point
707 and electrode separately, with variables coded as described in the section on the
708 time-course analysis of the region of interest. As in that analysis, we again excluded
709 random slopes for feasibility:

```
710 amplitude ~ 1 + congruency * predictability +  
711 (1 | participant_id) +  
712 (1 | image_id) +  
713 (1 | word_id)
```

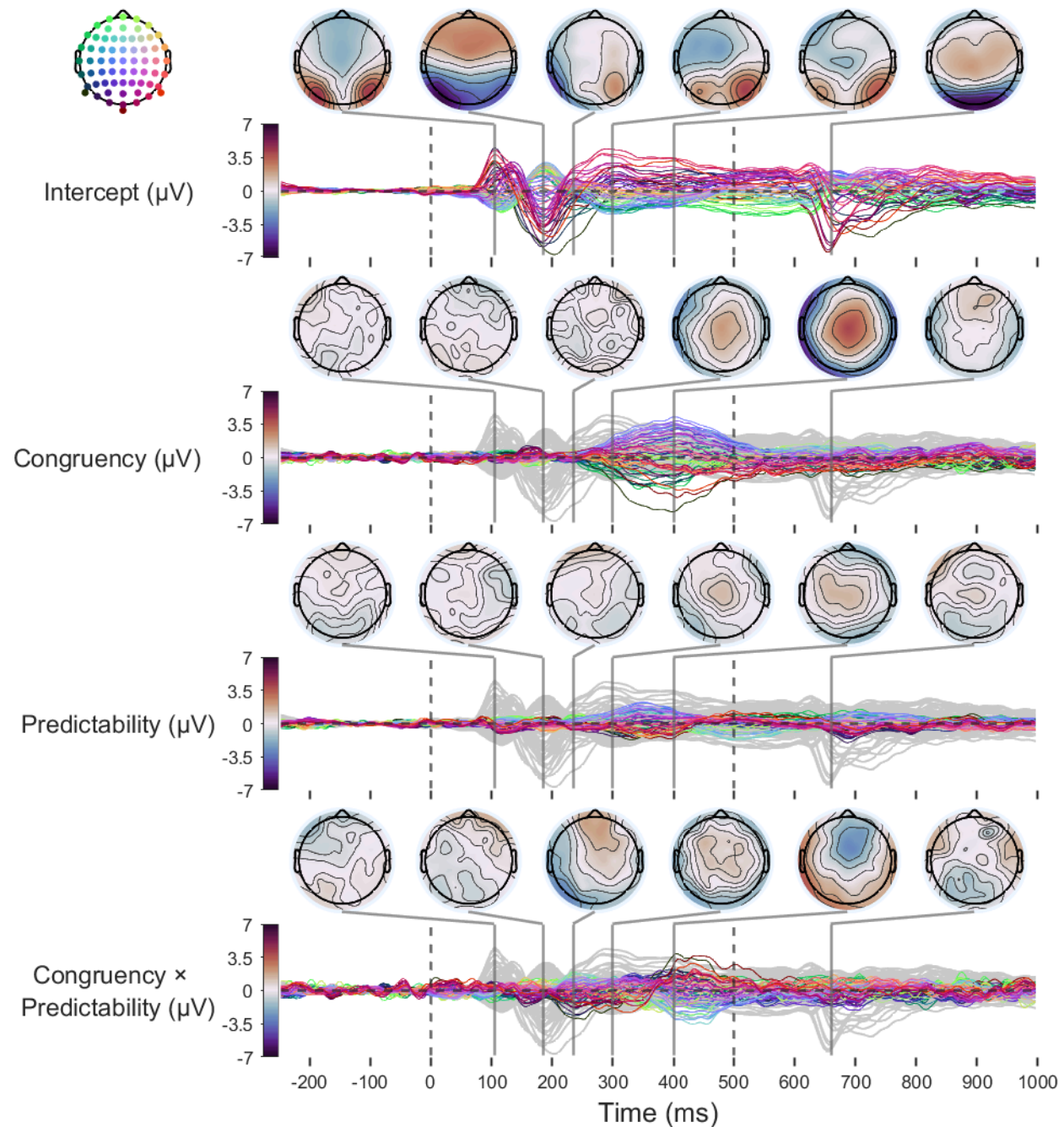
714

715 Results confirmed that the Congruency-Predictability interaction at left
716 occipitotemporal sites was the earliest fixed effect to emerge, and that the effect was
717 small relative to that observed at later time points. It additionally showed that the switch
718 in direction of the Predictability-Congruency interaction shown to peak at around 400 ms
719 in **Figure 11** exhibits a frontocentral topography. This effect was sustained until around
720 475 ms. Interestingly, if this effect captures changes in an N400 component, then the
721 direction of the N400 modulation was, as was the case for the effect on the N1, arguably
722 inconsistent with a simple predictive coding account. This is because the direction of
723 effects suggests that prediction-congruent words elicited the most negative-going N400
724 amplitudes at the lowest level of predictability. As predictability increased, N400
725 amplitudes elicited by picture-congruent words became less extreme, increasingly
726 approaching the N400 amplitudes elicited by picture-incongruent words (**Supplementary**
727 **Materials J**). The modulation observed in the opposite direction at occipitotemporal sites
728 in **Figure 10** at around 400 ms likely results from the use of average reference.

729 The scalp-wide analysis also revealed that the main effect of picture-word
730 Congruency shown in **Figure 10** indeed peaks at around 400 ms, with a centroparietal
731 topography. Interpreting this as a modulation of the N400 would mean that, at the lowest

Figure 12

Time-course of scalp-wide fixed-effects estimates.



The first dashed vertical line (0 ms) indicates stimulus (word) onset. The second dashed vertical line (500 ms) indicates the time-point at which the word changed colour to green. Topographic plots of fixed effects are highlighted at key time-points. Model intercepts (reflecting average amplitudes at the lowest level of Predictability) are depicted as grey lines on each panel to provide a reference for timing and magnitude of effects.

732 level of predictability, picture-congruent words elicited more negative-going N400s
733 overall than picture-incongruent words did.

734 Finally, this analysis covered an extended period of time, which revealed a clear
735 negative-going posterior component at around 650 ms. This component peaked around
736 150 ms after the stimulus changed colour to indicate that participants could respond.
737 Given the timing and topography of this component, this is likely to reflect an N1
738 response to the colour change.

739 Discussion

740 In the present study, we tested whether a simple predictive coding account could
741 explain online prediction effects on the amplitude of N1 ERP components elicited by
742 words in biasing contexts. We biased expectations for upcoming words via images of
743 varying predictability. Based on a predictive coding framework, we hypothesised that
744 there would be an interaction between picture-word Predictability and Congruency in
745 which N1 amplitude scales with prediction error. Planned analyses failed to find
746 evidence for this hypothesis, and exploratory analyses revealed, despite strong evidence
747 for prediction effects in the N1, that the direction of the interaction was opposite to that
748 expected under the hypothesis. Specifically, increases in Predictability were associated
749 with greater-amplitude N1s for picture-congruent words, and smaller-amplitude N1s for
750 picture-incongruent words. On this basis, we conclude that a simple predictive coding
751 explanation of the N1 cannot explain predictability effects observed in the picture-word
752 verification task used here.

753 In recent years, predictive coding models have been increasingly applied to
754 explain neural phenomena observed during language processing. This includes
755 predictive coding perspectives on the N1 specifically (e.g., Gagl et al., 2020; Huang
756 et al., 2022; Zhao et al., 2019), or its likely generator, vOT (Price & Devlin, 2011), and
757 other areas of language processing. For example, consider the well-researched N400
758 ERP component, generally recognised since its initial identification as capturing activity

759 related to semantic processes (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980). The
760 N400 shows sensitivity to word- and sentence-level surprise or predictability
761 (Delaney-Busch et al., 2019; Lau et al., 2013; Lindborg et al., 2023; Mantegna et al.,
762 2019; Van Petten & Kutas, 1990), in a manner that may be consistent with predictive
763 coding (Bornkessel-Schlesewsky & Schlewsky, 2019; Eddine et al., 2023; Rabovsky &
764 McRae, 2014). Similar interpretations have been made of other signals, as capturing
765 prediction errors for phonological, semantic, or syntactic representations (Fitz & Chang,
766 2019; Gagnepain et al., 2012; Van Petten & Luka, 2012; Ylinen et al., 2016, 2017).
767 Indeed, emerging evidence is beginning to support the broader contention that
768 naturalistic language comprehension utilises a predictive coding hierarchy spanning the
769 language network (Caucheteux et al., 2023; Schuster et al., 2021; Shain et al., 2020;
770 L. Wang et al., 2023). In this way, evidence for predictive coding in language reflects the
771 growing, although not definitive, empirical evidence for predictive coding models in
772 perception more generally (Clark, 2013; Heilbron & Chait, 2018; Hodson et al., 2024;
773 Walsh et al., 2020).

774 We do not believe our findings refute the existence of predictive coding
775 mechanisms during the N1. This is informed by our review of the literature outlined in the
776 Introduction, in which we found evidence broadly consistent with a predictive coding
777 interpretation of the N1. Instead, we argue that a simple predictive coding account of the
778 N1, in which the component's amplitude straightforwardly indexes prediction error in a
779 manner dependent on prediction certainty, is insufficient to explain the pattern of effects
780 we observed in the picture-word verification task we used here. For a predictive coding
781 model to better account for these data, it would require elaboration. One feature that
782 may be relevant is the nature of the task. We elected to use a picture-word verification
783 task as it encourages explicit prediction of word forms from non-linguistic contexts.
784 However, this task paradigm may alter predictive processing of word forms in two key
785 ways. First, participants will have soon learned that the observed word form only

786 matches its preceding image 50% of the time, which could have interacted with the effect
787 of Predictability (prediction certainty) in unexpected ways. Second, the requirement for
788 explicit verification of prediction congruency may have encouraged artificial processing
789 strategies that are not representative of naturalistic word recognition and reading
790 processes. To better understand whether and how such factors influence any possible
791 predictive coding effects on the N1, we could manipulate prediction error magnitude and
792 precision while the participant's task instructions do not explicitly require processing of
793 the cue. For instance, we could use a picture-word priming design (Sperber et al., 1979;
794 Vanderwart, 1984), presenting picture-word pairs, as in the current study, but ask
795 participants to respond with lexical decisions. Here, prediction error magnitude could be
796 operationalised as the orthographic distance between the string (whether word or
797 non-word), and precision as the predictability of a word given its picture. We believe that
798 such an approach could provide insight into whether, and which, features of the
799 paradigm we used could have resulted in the unexpected pattern of results. Finally, it is
800 possible that dynamics of predictive processing were influenced by the slow presentation
801 rate employed in the present study, relative to more naturalistic reading paradigms.
802 Indeed, previous research has highlighted the importance of presentation rate in
803 prediction effects during reading (e.g., Dambacher et al., 2012), and recent findings have
804 shown that unpredictability in stimulus presentation timing (e.g., with jittered
805 inter-stimulus intervals) may interfere with predictive processes, as indexed by the
806 mismatch negativity component (Tsogli et al., 2022). This explanation of our results could
807 be tested by study designs examining how the congruency-predictability interaction
808 varies over stimulus onset asynchronies of different durations. In sum, while predictive
809 coding mechanisms may ultimately underlie the pattern of effects we observed, the
810 simple account we have tested requires elaboration, informed by insights from other
811 paradigms, for it to explain why our current pattern of effects is opposite to that expected.

812 Our study is not the first to identify patterns in evoked responses that seemingly

813 run counter to a simple predictive coding model (e.g., Bowman et al., 2013; Eisenhauer
814 et al., 2022; Mangun & Hillyard, 1991; Vidal-Gran et al., 2020). One suggested
815 elaboration to a simple predictive coding model that could allow expected stimuli to elicit
816 greater evoked responses than unexpected stimuli supposes that, in such cases,
817 expected stimuli may benefit from greater precision than is allotted to deviant stimuli
818 (Bowman et al., 2023; Heilbron & Chait, 2018; Kok et al., 2012). In our experiment,
819 matched picture-congruent and -incongruent words followed the same pictures, such
820 that predictability, which we used as a measure of top-down prediction precision, should
821 have been identical prior to word presentation. However, if top-down effects can
822 penetrate early stages of visual processing that precede the N1, it is conceivable that
823 processing after word presentation, but prior to the N1, could have up-weighted the
824 precision of information in picture-congruent words' representations, resulting in the
825 observed pattern of effects. In simulations, Bowman et al. (2023) recently demonstrated
826 that precision-modulated predictive coding models can indeed produce "contra-vanilla"
827 patterns in prediction errors if prediction-congruent stimuli benefit from higher precision,
828 but that this should be expected to affect the evoked response non-linearly. Specifically,
829 the latency of the evoked response should be shorter for the prediction-congruent
830 stimulus. Our findings did indeed reveal a latency difference in the N1 offset period,
831 between congruent and incongruent words at the highest level of predictability (**Figure**
832 **10**), but the direction of this difference was opposite to that predicted by Bowman et al.
833 (2023), with a shorter offset period for picture-incongruent words. As such, it is unclear
834 how an elaboration based on differential precision between picture-congruent and
835 -incongruent words may relate to our findings.

836 We acknowledge the possibility that the insufficiency of predictive coding
837 accounts to explain the data we observed may reflect a more fundamental shortcoming.
838 Indeed, an enduring criticism of predictive coding models is that some evidence for them
839 may also be explained by alternative models (de Lange et al., 2018; Hodson et al.,

840 2024). To speculate, predictive coding models may account for activity in the N1 in
841 previously tested paradigms without accurately describing the underlying neural
842 processes. For instance, Luthra et al. (2021) showed that, in spoken word recognition,
843 interactive activation models may provide an alternative account of the ERP amplitude
844 reduction observed in response to prediction violations, without invoking key features of
845 predictive coding models. Indeed, effects indicative of predictive processing may emerge
846 in a system that lacks any representations of, or mechanisms implementing,
847 predictions or prediction errors, instead only implementing "pattern completion"
848 (Falandays et al., 2021). It is tentatively possible that the picture-word verification
849 paradigm we applied here may be a scenario that employs the same neurocognitive
850 processes in the N1 as those employed in other paradigms, but elicits cognitive
851 dynamics whose corresponding neural activity reveals differences from a predictive
852 coding model. It is possible that processing indexed by the N1 can only be explained by
853 a model distinct from the predictive coding framework, even though predictive coding
854 models may correlate with patterns of activity seen in most paradigms. Justifying the
855 development of such a model, distinct from predictive coding, would require much more
856 evidence for the shortcomings of a predictive coding account, and we do not believe our
857 study provides the insights necessary to speculate on the form such a model could take.

858 Such further insights may be provided by an approach that examines patterns in
859 the representational content of neural activity, rather than univariate patterns of overall
860 activity. Such an approach has been exploited previously as a way of comparing
861 prediction error models, in which neural signals represent unexplained content, with
862 sharpening models of language processing, in which neural signals contain sharper
863 representations of predicted content (Desimone, 1996; Grill-Spector et al., 2006). While
864 these models can account for similar patterns in overall neural activity, they predict
865 dissociable patterns in corresponding representational content (Blank & Davis, 2016).
866 For instance, Blank and Davis (2016) employed a Congruency (matching, neutral) ×

867 Precision (signal quality: 4 or 12 vocoder channels) design in an fMRI experiment on
868 speech perception. Representational similarity analyses of fMRI activity from the
869 posterior superior temporal sulcus showed a pattern consistent with the representational
870 content expected under a prediction error account, and not a sharpening account.
871 Analyses of EEG signals in a similar paradigm by Sohoglu and Davis (2020) also show
872 evidence for patterns of representational enhancement and suppression that match a
873 prediction error account, from 100 ms after stimulus presentation. Further evidence for
874 prediction error accounts of early speech perception processes is seen in fMRI and MEG
875 analyses of two-syllable words, where precision is quantified as the predictability of
876 syllable two, given syllable one (Sohoglu et al., 2023). In contrast, however, an MEG
877 study on the representation of lexical-semantic information during *visual* word
878 recognition found evidence more consistent with a sharpening account (Eisenhauer
879 et al., 2022). Although their use was motivated by a need to disentangle two
880 explanations of evoked-response patterns that are both consistent with predictive
881 coding, we believe that such analyses, focusing on representational content, may also
882 provide an avenue to further investigate the pattern we observed that was seemingly
883 inconsistent with predictive coding. This could reveal whether the N1's modulation is
884 accompanied by the representation of more or less stimulus-relevant information, and
885 may more clearly point to the underlying mechanisms.

886 Representational content is also of particular importance when testing predictive
887 coding accounts because it determines the depth in the hierarchy to which top-down
888 predictions can be conveyed and effectively implemented. This is because in a
889 hierarchical model of predictive coding, where levels of the hierarchy utilise different
890 representational formats, the interaction between ascending input and descending
891 predictions must involve some mapping of higher-level onto lower-level representations.
892 For instance, if semantic context can influence processing that is closer to sensory input
893 and indexed by early ERP components (e.g., Enge et al., 2023; Getz & Toscano, 2019;

894 Segalowitz & Zheng, 2009), then higher-level semantic information must be translated
895 into predictions of upcoming lower-level sensory signals. In the case of our study's
896 modulation of the N1, if the N1 is implicated in visual-orthographic processing (Bentin
897 et al., 1999; Brem et al., 2018; Ling et al., 2019; Maurer, Brandeis, & McCandliss, 2005),
898 then predictions of upcoming words must be translated into a visual-orthographic code.
899 Such a mapping could be expected to be very computationally lossy; predictions for
900 visual-orthographic features of a single word should be expected to also confer
901 facilitation for words that are orthographically similar, yet picture-incongruent (Kim & Lai,
902 2012). In contrast, a later ERP more directly implicated in semantic processing, like the
903 N400, may be expected to be less limited by such mappings.

904 From one perspective, mapping of predictions to lower-level representations may
905 be considered a requisite for a phenomenon to be considered top-down modulation
906 (Rauss et al., 2011). This relates to a long-standing debate on whether prediction effects
907 at the lexical level of language processing necessitate top-down input informed by
908 higher-level semantic processes, or could instead result from perhaps more
909 parsimonious intralexical effects (Fodor, 1983; Forster, 1979). A similar argument could
910 be made that context effects on the N1 could be interpreted as intra-*orthographic*,
911 resulting from local interactions in a possible *orthographic module*. As an example, the
912 orthographic features of the word form *fish* may preactivate features of the word form
913 *chips* simply through learned co-occurrence rather than top-down modulation, entirely
914 within an orthographic processing module that possesses nothing approaching a
915 semantic representation. Such facilitation could be implemented via an extension to
916 classic interactive activation models (e.g., McClelland & Rumelhart, 1981) in which there
917 are excitatory lateral connections between word-level units whose strength is determined
918 by co-occurrence frequency. We consider this point to highlight an advantage of
919 paradigms such as ours, that use non-linguistic contexts (e.g., task instructions, images,
920 etc.) to cue upcoming words and word forms. Effects of context that map across

921 representations in this way necessitate transfer of information across levels of the
922 processing hierarchy, and may thus be considered stronger evidence for an influence of
923 top-down predictions.

924 An aspect of the predictive coding account that our design did not fully test also
925 relates to this idea of representational mapping. We dichotomised the variable of
926 congruency (prediction error magnitude), with orthographic Levenshtein distance
927 maximised between picture-congruent and -incongruent word forms. However,
928 prediction error magnitude should also be expected to vary continuously, from
929 unpredicted word forms that are less to more orthographically similar to the predicted
930 word form. This is comparable to Gagl et al.'s (2020) use of a pixel distance metric to
931 calculate the continuous distance between a presented word form and a context-neutral
932 prior. Such an approach could be applied to biasing contexts by instead calculating the
933 orthographic distance between a presented word form and a context-informed prior,
934 where the probability of observing certain pixels (or orthographic features) could be
935 up-weighted proportional to prediction certainty. We believe such an approach could
936 provide useful insights in elucidating the pattern of effects we observed.

937 We note that exploratory analyses at the typical latency of the N400 revealed a
938 pattern which also appears to run counter to a simple predictive coding account of
939 predictability effects. This is seemingly inconsistent with interpretations of the N400 as
940 indexing prediction error (Bornkessel-Schlesewsky & Schlewsky, 2019; Eddine et al.,
941 2023; Rabovsky & McRae, 2014). At the lowest level of predictability, we observed
942 greater N400 amplitudes for picture-congruent words, than for picture-incongruent
943 words. As predictability increased, meanwhile, N400 amplitudes became less extreme
944 for picture-congruent words, rather than becoming more extreme for picture-incongruent
945 words. We caution against over-interpreting these results. In addition to these results
946 being entirely exploratory, we used an average EEG reference, rather than the more
947 standard mastoid reference for later centroparietal components like the N400.

948 Furthermore, when elicited by words in context, such as a sentence or picture, evidence
949 suggests that the N400 indexes both prediction and integration processes (Nieuwland
950 et al., 2018). Nevertheless, that our design elicited effects on the N400 that are
951 seemingly inconsistent with existing findings from more traditional experimental designs
952 may point to our specific experimental design playing a key role in the pattern of effects
953 we observed in the N1.

954 In sum, we tested a simple predictive coding account of the word-elicited N1, but
955 failed to find evidence in favour of it. Exploratory analyses suggest that the pattern of
956 effects in the Congruency-Predictability interaction were in the opposite direction to that
957 expected under a simple predictive coding model. We argue that such a model is
958 insufficient to explain the pattern of effects we observed, and we have identified avenues
959 of future research that could better delineate how predictive processes interact with
960 processing during the N1.

References

961

- 962 Allison, T., McCarthy, G., Nobre, A., Puce, A., & Belger, A. (1994). Human extrastriate
963 visual cortex and the perception of faces, words, numbers, and colors. *Cerebral*
964 *Cortex*, 4(5), 544–554. <https://doi.org/10.1093/cercor/4.5.544>
- 965 Appelbaum, L. G., Liotti, M., Perez, R., Fox, S. P., & Woldorff, M. G. (2009). The
966 temporal dynamics of implicit processing of non-letter, letter, and word-forms in
967 the human visual cortex. *Frontiers in Human Neuroscience*, 3, 1–11.
968 <https://doi.org/10.3389/neuro.09.056.2009>
- 969 Barr, D. J., Levy, J., Scheepers, C., & Tily, H. J. (2013). Random effects structure for
970 confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and*
971 *Language*, 68(Supplement), 1–63. <https://doi.org/10.1016/j.jml.2012.11.001>
- 972 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects
973 models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
974 <https://doi.org/10.18637/jss.v067.i01>
- 975 Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999).
976 ERP manifestations of processing printed words at different psycholinguistic
977 levels: Time course and scalp distribution. *Journal of Cognitive Neuroscience*,
978 11(3), 235–260. <https://doi.org/10.1162/089892999563373>
- 979 Blank, H., & Davis, M. H. (2016). Prediction Errors but Not Sharpened Signals Simulate
980 Multivoxel fMRI Patterns during Speech Perception [Publisher: Public Library of
981 Science]. *PLOS Biology*, 14(11), e1002577.
982 <https://doi.org/10.1371/journal.pbio.1002577>
- 983 Bornkessel-Schlesewsky, I., & Schlewsky, M. (2019). Toward a neurobiologically
984 plausible model of language-related, negative event-related potentials. *Frontiers*
985 *in Psychology*, 10(Article 298). <https://doi.org/10.3389/fpsyg.2019.00298>

- 986 Bowman, H., Collins, D. J., Nayak, A. K., & Cruse, D. (2023). Is predictive coding
987 falsifiable? *Neuroscience & Biobehavioral Reviews*, *154*, 105404.
988 <https://doi.org/10.1016/j.neubiorev.2023.105404>
- 989 Bowman, H., Filetti, M., Janssen, D., Su, L., Alsufyani, A., & Wyble, B. (2013). Subliminal
990 Saliency Search Illustrated: EEG Identity and Deception Detection on the Fringe
991 of Awareness [Publisher: Public Library of Science]. *PLOS ONE*, *8*(1), e54258.
992 <https://doi.org/10.1371/journal.pone.0054258>
- 993 Branzi, F. M., Martin, C. D., & Paz-Alonso, P. M. (2022). Task-relevant representations
994 and cognitive control demands modulate functional connectivity from ventral
995 occipito-temporal cortex during object recognition tasks. *Cerebral Cortex*, *32*(14),
996 3068–3080. <https://doi.org/10.1093/cercor/bhab401>
- 997 Brem, S., Halder, P., Bucher, K., Summers, P., Martin, E., & Brandeis, D. (2009). Tuning
998 of the visual word processing system: Distinct developmental ERP and fMRI
999 effects. *Human Brain Mapping*, *30*(6), 1833–1844.
1000 <https://doi.org/10.1002/hbm.20751>
- 1001 Brem, S., Hunkeler, E., Mächler, M., Kronschnabel, J., Karipidis, I. I., Pleisch, G., &
1002 Brandeis, D. (2018). Increasing expertise to a novel script modulates the visual
1003 N1 ERP in healthy adults. *International Journal of Behavioral Development*, *42*(3),
1004 333–341. <https://doi.org/10.1177/0165025417727871>
- 1005 Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS)
1006 phase ii: 930 new normative photos. *PLoS ONE*, *9*(9), 1–10.
1007 <https://doi.org/10.1371/journal.pone.0106953>
- 1008 Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual
1009 support on lexical processing: Prediction takes precedence. *Cognition*, *136*,
1010 135–149. <https://doi.org/10.1016/j.cognition.2014.10.017>
- 1011 Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan.
1012 *Journal of Statistical Software*, *80*(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>

- 1013 Caucheteux, C., Gramfort, A., & King, J.-R. (2023). Evidence of a predictive coding
1014 hierarchy in the human brain listening to speech. *Nature Human Behaviour*, 1–12.
1015 <https://doi.org/10.1038/s41562-022-01516-2>
- 1016 Chang, C. Y., Hsu, S. H., Pion-Tonachini, L., & Jung, T. P. (2020). Evaluation of Artifact
1017 Subspace Reconstruction for automatic artifact components removal in
1018 multi-channel EEG recordings. *IEEE Transactions on Biomedical Engineering*,
1019 67(4), 1114–1121. <https://doi.org/10.1109/TBME.2019.2930186>
- 1020 Cheimariou, S., Farmer, T. A., & Gordon, J. K. (2019). Lexical prediction in the aging
1021 brain: The effects of predictiveness and congruency on the N400 ERP
1022 component. *Aging, Neuropsychology, and Cognition*, 26(5), 781–806.
1023 <https://doi.org/10.1080/13825585.2018.1529733>
- 1024 Chen, Y., Davis, M. H., Pulvermüller, F., & Hauk, O. (2013). Task modulation of brain
1025 responses in visual word recognition as studied using EEG/MEG and fMRI.
1026 *Frontiers in Human Neuroscience*, 7, 1–14.
1027 <https://doi.org/10.3389/fnhum.2013.00376>
- 1028 Chen, Y., Davis, M. H., Pulvermüller, F., & Hauk, O. (2015). Early visual word processing
1029 is flexible: Evidence from spatiotemporal brain dynamics. *Journal of Cognitive
1030 Neuroscience*, 27(9), 1738–1751. https://doi.org/10.1162/jocn_a_00815
- 1031 Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of
1032 cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204.
1033 <https://doi.org/10.1017/S0140525X12000477>
- 1034 Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G.,
1035 Hénaff, M.-A., & Michel, F. (2000). The visual word form area: Spatial and
1036 temporal characterization of an initial stage of reading in normal subjects and
1037 posterior split-brain patients. *Brain*, 123, 291–307.

- 1038 Compton, P. E., Grossenbacher, P., Posner, M. I., & Tucker, D. M. (1991). A
1039 cognitive-anatomical approach to attention in lexical access. *Journal of Cognitive*
1040 *Neuroscience*, 3(4), 304–312. <https://doi.org/10.1162/jocn.1991.3.4.304>
- 1041 Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., &
1042 Halgren, E. (2000). Dynamic statistical parametric mapping: Combining fMRI and
1043 MEG for high-resolution imaging of cortical activity. *Neuron*, 26(1), 55–67.
1044 [https://doi.org/10.1016/S0896-6273\(00\)81138-1](https://doi.org/10.1016/S0896-6273(00)81138-1)
- 1045 Dambacher, M., Dimigen, O., Braun, M., Wille, K., Jacobs, A. M., & Kliegl, R. (2012).
1046 Stimulus onset asynchrony and the timeline of word recognition: Event-related
1047 potentials during sentence reading. *Neuropsychologia*, 50(8), 1852–1870.
1048 <https://doi.org/10.1016/j.neuropsychologia.2012.04.011>
- 1049 Delaney-Busch, N., Morgan, E., Lau, E., & Kuperberg, G. R. (2019). Neural evidence for
1050 Bayesian trial-by-trial adaptation on the N400 during semantic priming. *Cognition*,
1051 187, 10–20. <https://doi.org/10.1016/j.cognition.2019.01.001>
- 1052 de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception?
1053 *Trends in Cognitive Sciences*, 22(9), 764–779.
1054 <https://doi.org/10.1016/j.tics.2018.06.002>
- 1055 Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of
1056 single-trial EEG dynamics including independent component analysis. *Journal of*
1057 *Neuroscience Methods*, 134, 9–21.
1058 <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- 1059 Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention.
1060 *Proceedings of the National Academy of Sciences*, 93(24), 13494–13499.
1061 <https://doi.org/10.1073/pnas.93.24.13494>
- 1062 Dikker, S., & Pylkkänen, L. (2011). Before the N400: Effects of lexical-semantic violations
1063 in visual cortex. *Brain and Language*, 118(1-2), 23–28.
1064 <https://doi.org/10.1016/j.bandl.2011.02.006>

- 1065 Eaton, J. W., Bateman, D., Hauberg, S., & Wehbring, R. (2020). *GNU Octave version*
1066 *6.1.0 manual: A high-level interactive language for numerical computations.*
1067 <https://www.gnu.org/software/octave/doc/v6.1.0/>
- 1068 Eberhard-Moscicka, A. K., Jost, L. B., Fehlbaum, L. V., Pfenninger, S. E., & Maurer, U.
1069 (2016). Temporal dynamics of early visual word processing - Early versus late N1
1070 sensitivity in children and adults. *Neuropsychologia*, *91*, 509–518.
1071 <https://doi.org/10.1016/j.neuropsychologia.2016.09.014>
- 1072 Eddine, S. N., Brothers, T., Wang, L., Spratling, M., & Kuperberg, G. R. (2023, April). A
1073 predictive coding model of the N400. <https://doi.org/10.1101/2023.04.10.536279>
- 1074 Eisenhauer, S., Gagl, B., & Fiebach, C. J. (2022). Predictive pre-activation of
1075 orthographic and lexical-semantic representations facilitates visual word
1076 recognition. *Psychophysiology*, *59*(3), 1–26. <https://doi.org/10.1111/psyp.13970>
- 1077 Enge, A., Süß, F., & Rahman, R. A. (2023). Instant effects of semantic information on
1078 visual perception. *Journal of Neuroscience*.
1079 <https://doi.org/10.1523/JNEUROSCI.2038-22.2023>
- 1080 Falandays, J. B., Nguyen, B., & Spivey, M. J. (2021). Is prediction nothing more than
1081 multi-scale pattern completion of the future? *Brain Research*, *1768*, 147578.
1082 <https://doi.org/10.1016/j.brainres.2021.147578>
- 1083 Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language
1084 comprehension. *Psychophysiology*, *44*(4), 491–505.
1085 <https://doi.org/10.1111/j.1469-8986.2007.00531.x>
- 1086 Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in*
1087 *Human Neuroscience*, *4*. <https://doi.org/10.3389/fnhum.2010.00215>
- 1088 Fitz, H., & Chang, F. (2019). Language ERPs reflect learning through prediction error
1089 propagation. *Cognitive Psychology*, *111*, 15–52.
1090 <https://doi.org/10.1016/j.cogpsych.2019.03.002>

- 1091 Fodor, J. (1983). Input systems as modules. In *The Modularity of Mind* (pp. 47–101). MIT
1092 Press.
- 1093 Forster, K. I. (1979). Levels of processing and the structure of the language processor. In
1094 W. Cooper & E. Walker (Eds.), *Sentence Processing: Psycholinguistic essays*
1095 *presented to Merrill Garrett* (pp. 27–85). Erlbaum.
- 1096 Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews*
1097 *Neuroscience*, 11(2), 127–138. <https://doi.org/10.1038/nrn2787>
- 1098 Gagl, B., Sassenhagen, J., Haan, S., Gregorova, K., Richlan, F., & Fiebach, C. J. (2020).
1099 An orthographic prediction error as the basis for efficient visual word recognition.
1100 *NeuroImage*, 214(August 2019), 116727.
1101 <https://doi.org/10.1016/j.neuroimage.2020.116727>
- 1102 Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for
1103 spoken words in auditory cortex. *Current Biology*, 22(7), 615–621.
1104 <https://doi.org/10.1016/j.cub.2012.02.015>
- 1105 Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch
1106 negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, 120(3),
1107 453–463. <https://doi.org/10.1016/j.clinph.2008.11.029>
- 1108 Getz, L. M., & Toscano, J. C. (2019). Electrophysiological evidence for top-down lexical
1109 influences on early speech perception. *Psychological Science*, 30(6), 830–841.
1110 <https://doi.org/10.1177/0956797619841813>
- 1111 Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural
1112 models of stimulus-specific effects. *Trends in Cognitive Sciences*, 10(1), 14–23.
1113 <https://doi.org/10.1016/j.tics.2005.11.006>
- 1114 Groppe, D. M., Makeig, S., & Kutas, M. (2009). Identifying reliable independent
1115 components via split-half comparisons. *NeuroImage*, 45(4), 1199–1211.
1116 <https://doi.org/10.1016/j.neuroimage.2008.12.038>

- 1117 Hauk, O., Coutout, C., Holden, A., & Chen, Y. (2012). The time-course of single-word
1118 reading: Evidence from fast behavioral and brain responses. *NeuroImage*, *60*(2),
1119 1462–1477. <https://doi.org/10.1016/j.neuroimage.2012.01.061>
- 1120 Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive
1121 Coding in Auditory Cortex? *Neuroscience*, *389*, 54–73.
1122 <https://doi.org/10.1016/j.neuroscience.2017.07.061>
- 1123 Hodson, R., Mehta, M., & Smith, R. (2024). The empirical status of predictive coding and
1124 active inference. *Neuroscience & Biobehavioral Reviews*, *157*, 105473.
1125 <https://doi.org/10.1016/j.neubiorev.2023.105473>
- 1126 Huang, Z., Yang, S., Xue, L., Yang, H., Lv, Y., & Zhao, J. (2022). Level of orthographic
1127 knowledge helps to reveal automatic predictions in visual word processing.
1128 *Frontiers in Neuroscience*, *15*. <https://doi.org/10.3389/fnins.2021.809574>
- 1129 Hyvärinen, A., & Oja, E. (1997). A fast fixed-point algorithm for independent component
1130 analysis. *Neural Computation*, *9*(7), 1483–1492.
1131 <https://doi.org/10.1162/neco.1997.9.7.1483>
- 1132 Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: Predictive
1133 processing, precision and the pulvinar. *Philosophical Transactions of the Royal
1134 Society B: Biological Sciences*, *370*(1668). <https://doi.org/10.1098/rstb.2014.0169>
- 1135 Kherif, F., Josse, G., & Price, C. J. (2011). Automatic top-down processing explains
1136 common left occipito-temporal responses to visual words and objects. *Cerebral
1137 Cortex*, *21*, 103–114. <https://doi.org/10.1093/cercor/bhq063>
- 1138 Kim, A. E., & Gilley, P. M. (2013). Neural mechanisms of rapid sensitivity to syntactic
1139 anomaly. *Frontiers in Psychology*, *4*, 1–15.
1140 <https://doi.org/10.3389/fpsyg.2013.00045>
- 1141 Kim, A. E., & Lai, V. (2012). Rapid interactions between lexical semantic and word form
1142 analysis during word recognition in context: Evidence from ERPs. *Journal of
1143 Cognitive Neuroscience*, *24*(5), 1104–1112. https://doi.org/10.1162/jocn_a_00148

- 1144 Knolle, F., Schröger, E., & Kotz, S. A. (2013). Prediction errors in self- and
1145 externally-generated deviants. *Biological Psychology*, *92*(2), 410–416.
1146 <https://doi.org/10.1016/j.biopsycho.2012.11.017>
- 1147 Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & de Lange, F. P. (2012). Attention
1148 reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex*
1149 (*New York, N.Y.: 1991*), *22*(9), 2197–2206. <https://doi.org/10.1093/cercor/bhr310>
- 1150 Kretzschmar, F., Schlesewsky, M., & Staub, A. (2015). Dissociating word frequency and
1151 predictability effects in reading: Evidence from coregistration of eye movements
1152 and EEG. *Journal of Experimental Psychology : Learning, Memory, and*
1153 *Cognition*, *41*(6), 1648–1662. <https://doi.org/10.1037/xlm0000128>
- 1154 Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language
1155 comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59.
1156 <https://doi.org/10.1080/23273798.2015.1102299>
- 1157 Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the
1158 N400 component of the event related brain potential (ERP). *Annual review of*
1159 *psychology*, *62*, 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- 1160 Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect
1161 semantic incongruity. *Science*, *207*, 203–205.
1162 <https://doi.org/10.1126/science.7350657>
- 1163 Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of
1164 prediction from association in single-word contexts. *Journal of Cognitive*
1165 *Neuroscience*, *25*(3), 484–502. https://doi.org/10.1162/jocn_a_00328
- 1166 Lau, E. F., Namyst, A., Fogel, A., & Delgado, T. (2016). A Direct Comparison of N400
1167 Effects of Predictability and Incongruity in Adjective-Noun Combination. *Collabra*,
1168 *2*(1), 13. <https://doi.org/10.1525/collabra.40>
- 1169 Lin, S. E., Chen, H. C., Zhao, J., Li, S., He, S., & Weng, X. C. (2011). Left-lateralized
1170 N170 response to unpronounceable pseudo but not false Chinese characters -

- 1171 the key role of orthography. *Neuroscience*, 190, 200–206.
1172 <https://doi.org/10.1016/j.neuroscience.2011.05.071>
- 1173 Lindborg, A., Musiolek, L., Ostwald, D., & Rabovsky, M. (2023). Semantic surprise
1174 predicts the N400 brain potential. *Neuroimage: Reports*, 3(1), 100161.
1175 <https://doi.org/10.1016/j.ynirp.2023.100161>
- 1176 Ling, S., Lee, A. C. H., Armstrong, B. C., & Nestor, A. (2019). How are visual words
1177 represented? Insights from EEG-based visual word decoding, feature derivation
1178 and image reconstruction. *Human Brain Mapping*, 40(17), 5056–5068.
1179 <https://doi.org/10.1002/hbm.24757>
- 1180 Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading.
1181 *Cognitive Psychology*, 88, 22–60. <https://doi.org/10.1016/j.cogpsych.2016.06.002>
- 1182 Luthra, S., Li, M. Y. C., You, H., Brodbeck, C., & Magnuson, J. S. (2021). Does signal
1183 reduction imply predictive coding in models of spoken word recognition?
1184 *Psychonomic Bulletin & Review*, 28(4), 1381–1389.
1185 <https://doi.org/10.3758/s13423-021-01924-x>
- 1186 Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials
1187 indicate changes in perceptual processing during visual-spatial priming. *Journal of*
1188 *Experimental Psychology: Human Perception and Performance*, 17(4),
1189 1057–1074. <https://doi.org/10.1037/0096-1523.17.4.1057>
- 1190 Mantegna, F., Hintz, F., Ostarek, M., Alday, P. M., & Huettig, F. (2019). Distinguishing
1191 integration and prediction accounts of ERP N400 modulations in language
1192 processing through experimental design. *Neuropsychologia*, 134, 107199.
1193 <https://doi.org/10.1016/j.neuropsychologia.2019.107199>
- 1194 MATLAB. (2022). *MATLAB Version 9.13.0 (R2022b)*. The MathWorks Inc.
- 1195 Maurer, U., Brandeis, D., & McCandliss, B. D. (2005). Fast, visual specialization for
1196 reading in English revealed by the topography of the N170 ERP response.
1197 *Behavioral and Brain Functions*, 1, 1–12. <https://doi.org/10.1186/1744-9081-1-13>

- 1198 Maurer, U., Brem, S., Bucher, K., & Brandeis, D. (2005). Emerging neurophysiological
1199 specialization for letter strings. *Journal of Cognitive Neuroscience*, *17*(10),
1200 1532–1552. <https://doi.org/10.1162/089892905774597218>
- 1201 Maurer, U., Rossion, B., & McCandliss, B. D. (2008). Category specificity in early
1202 perception: Face and word N170 responses differ in both lateralization and
1203 habituation properties. *Frontiers in Human Neuroscience*, *2*, 1–7.
1204 <https://doi.org/10.3389/neuro.09.018.2008>
- 1205 McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context
1206 effects in letter perception: Part 1. An account of basic findings. *Psychological*
1207 *Review*, *88*(5), 375–407. <https://doi.org/10.1037/0033-295X.88.5.375>
- 1208 Nieuwland, M. S. (2019). Do 'early' brain responses reveal word form prediction during
1209 language comprehension? A critical review. *Neuroscience and Biobehavioral*
1210 *Reviews*, *96*, 367–400. <https://doi.org/10.1016/j.neubiorev.2018.11.019>
- 1211 Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E.,
1212 Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F.,
1213 Husband, E. M., Ito, A., Kazanina, N., Kogan, V., Kohút, Z., Kulakova, E.,
1214 Mézière, D., Politzer-Ahles, S., Rousselet, G. A., ...
1215 Von Grebmer Zu Wolfsturn, S. (2018). Dissociable effects of prediction and
1216 integration during language comprehension: Evidence from a large-scale study
1217 using brain potentials, 1–23. <https://doi.org/10.1101/267815>
- 1218 Nobre, A. C., Allison, T., & McCarthy, G. (1994). Word recognition in the human inferior
1219 temporal lobe. *372*, 260–263. <https://doi.org/10.1038/372260a0>
- 1220 Pastore, M., & Calcagni, A. (2019). Measuring distribution similarities between samples:
1221 A distribution-free overlapping index. *Frontiers in Psychology*, *10*(1089), 1–8.
1222 <https://doi.org/10.3389/fpsyg.2019.01089>

- 1223 Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of*
1224 *Neuroscience Methods*, 162(1-2), 8–13.
1225 <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- 1226 Penolazzi, B., Hauk, O., & Pulvermüller, F. (2007). Early semantic context integration
1227 and lexical access as revealed by event-related brain potentials. *Biological*
1228 *Psychology*, 74(3), 374–388. <https://doi.org/10.1016/j.biopsycho.2006.09.008>
- 1229 Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A
1230 theory and review. *Psychological Bulletin*, 144(10), 1002–1044.
1231 <https://doi.org/10.1037/bul0000158>
- 1232 Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and
1233 comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347.
1234 <https://doi.org/10.1017/S0140525X12001495>
- 1235 Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated
1236 electroencephalographic independent component classifier, dataset, and website.
1237 *NeuroImage*, 198, 181–197. <https://doi.org/10.1016/j.neuroimage.2019.05.026>
- 1238 Pleisch, G., Karipidis, I. I., Brem, A., Röthlisberger, M., Roth, A., Brandeis, D.,
1239 Walitza, S., & Brem, S. (2019). Simultaneous EEG and fMRI reveals stronger
1240 sensitivity to orthographic strings in the left occipito-temporal cortex of typical
1241 versus poor beginning readers. *Developmental Cognitive Neuroscience*, 40,
1242 1–13. <https://doi.org/10.1016/j.dcn.2019.100717>
- 1243 Powell, M. (2009). The BOBYQA algorithm for bound constrained optimization without
1244 derivatives. *Cambridge NA Report NA2009/06*, University of Cambridge,
1245 Cambridge, 26–46. <https://doi.org/10.1.1.443.7693>
- 1246 Price, C. J., & Devlin, J. T. (2011). The Interactive Account of ventral occipitotemporal
1247 contributions to reading. *Trends in Cognitive Sciences*, 15(6), 246–253.
1248 <https://doi.org/10.1016/j.tics.2011.04.001>

- 1249 R Core Team. (2021). *R: A language and environment for statistical computing*. R
1250 Foundation for Statistical Computing. <https://www.r-project.org/>
- 1251 Rabovsky, M., & McRae, K. (2014). Simulating the N400 ERP component as semantic
1252 network error: Insights from a feature-based connectionist attractor model of word
1253 meaning. *Cognition*, *132*(1), 68–89.
1254 <https://doi.org/10.1016/j.cognition.2014.03.010>
- 1255 Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional
1256 interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*,
1257 *2*(1), 79–87. <https://doi.org/10.1038/4580>
- 1258 Rauschecker, A. M., Bowen, R. F., Parvizi, J., & Wandell, B. A. (2012). Position sensitivity
1259 in the visual word form area. *Proceedings of the National Academy of Sciences of*
1260 *the United States of America*, *109*(24). <https://doi.org/10.1073/pnas.1121304109>
- 1261 Rauss, K., Schwartz, S., & Pourtois, G. (2011). Top-down effects on early visual
1262 processing in humans: A predictive coding framework. *Neuroscience and*
1263 *Biobehavioral Reviews*, *35*(5), 1237–1253.
1264 <https://doi.org/10.1016/j.neubiorev.2010.12.011>
- 1265 Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and
1266 word skipping during reading: Effects of word length and predictability. *Journal of*
1267 *Experimental Psychology: Human Perception and Performance*, *37*, 514–528.
1268 <https://doi.org/10.1037/a0020990>
- 1269 Rodrigues, A. P., Rebola, J., Pereira, M., Van Asselen, M., & Castelo-Branco, M. (2019).
1270 Neural responses of the anterior ventral occipitotemporal cortex in developmental
1271 dyslexia: Beyond the visual word form area. *Investigative Ophthalmology and*
1272 *Visual Science*, *60*(4), 1063–1068. <https://doi.org/10.1167/iovs.18-26325>
- 1273 Rousselet, G. A. (2012). Does filtering preclude us from studying ERP time-courses?
1274 *Frontiers in Psychology*, *3*, 1–9. <https://doi.org/10.3389/fpsyg.2012.00131>

- 1275 Schuster, S., Himmelstoss, N. A., Hutzler, F., Richlan, F., Kronbichler, M., & Hawelka, S.
1276 (2021). Cloze enough? Hemodynamic effects of predictive processing during
1277 natural reading. *NeuroImage*, *228*, 117687.
1278 <https://doi.org/10.1016/j.neuroimage.2020.117687>
- 1279 Segalowitz, S. J., & Zheng, X. (2009). An ERP study of category priming: Evidence of
1280 early lexical semantic access. *Biological Psychology*, *80*(1), 122–129.
1281 <https://doi.org/10.1016/j.biopsycho.2008.04.009>
- 1282 Sereno, S. C., Brewer, C. C., & O'Donnell, P. J. (2003). Context effects in word
1283 recognition: Evidence for early interactive processing. *Psychological Science*,
1284 *14*(4), 328–333. <https://doi.org/10.1111/1467-9280.14471>
- 1285 Sereno, S. C., Hand, C. J., Shahid, A., Mackenzie, I. G., & Leuthold, H. (2019). Early
1286 EEG correlates of word frequency and contextual predictability in reading.
1287 *Language, Cognition and Neuroscience*, *35*(5), 625–640.
1288 <https://doi.org/10.1080/23273798.2019.1580753>
- 1289 Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word
1290 recognition: Evidence from eye movements and event-related potentials.
1291 *NeuroReport*, *9*(10), 2195–2200.
1292 <https://doi.org/10.1097/00001756-199807130-00009>
- 1293 Shain, C., Blank, I. A., van Schijndel, M., Schuler, W., & Fedorenko, E. (2020). fMRI
1294 reveals language-specific predictive coding during naturalistic sentence
1295 comprehension. *Neuropsychologia*, *138*, 107307.
1296 <https://doi.org/10.1016/j.neuropsychologia.2019.107307>
- 1297 Sohoglu, E., Beckers, L., & Davis, M. H. (2023, October). Convergent neural signatures
1298 of speech prediction error are a biological marker for spoken word recognition
1299 [Pages: 2023.10.03.560649 Section: New Results].
1300 <https://doi.org/10.1101/2023.10.03.560649>

- 1301 Sohoglu, E., & Davis, M. H. (2020). Rapid computations of spectrotemporal prediction
1302 error support perception of degraded speech (A. J. King, P. Kok, P. Kok, C. Press,
1303 & E. C. Lalor, Eds.). *eLife*, *9*, e58077. <https://doi.org/10.7554/eLife.58077>
- 1304 Sperber, R. D., McCauley, C., Ragain, R. D., & Weil, C. M. (1979). Semantic priming
1305 effects on picture and word processing. *Memory & Cognition*, *7*(5), 339–345.
1306 <https://doi.org/10.3758/BF03196937>
- 1307 STAN Development Team. (2023). Stan Modeling Language Users Guide and Reference
1308 Manual, 2.32. <https://mc-stan.org>
- 1309 Strijkers, K., Bertrand, D., & Grainger, J. (2015). Seeing the same words differently: The
1310 time course of automaticity and top-down intention in reading. *Journal of*
1311 *Cognitive Neuroscience*, *27*(8), 1542–1551. https://doi.org/10.1162/jocn_a_00797
- 1312 Taha, H., Ibrahim, R., & Khateb, A. (2013). How does arabic orthographic connectivity
1313 modulate brain activity during visual word recognition: An ERP study. *Brain*
1314 *Topography*, *26*(2), 292–302. <https://doi.org/10.1007/s10548-012-0241-2>
- 1315 Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters
1316 can produce artifactual effects and incorrect conclusions in ERP studies of
1317 language and cognition. *Psychophysiology*, *52*(8), 997–1009.
1318 <https://doi.org/10.1111/psyp.12437>
- 1319 Taylor, J. E., Beith, A., & Sereno, S. C. (2020). LexOPS: An R package and user
1320 interface for the controlled generation of word stimuli. *Behavior Research*
1321 *Methods*, *52*, 2372–2382. <https://doi.org/10.3758/s13428-020-01389-1>
- 1322 Thaler, L., Schütz, A. C., Goodale, M. A., & Gegenfurtner, K. R. (2013). What is the best
1323 fixation target? The effect of target shape on stability of fixational eye movements.
1324 *Vision Research*, *76*, 31–42. <https://doi.org/10.1016/j.visres.2012.10.012>
- 1325 Tsogli, V., Jentschke, S., & Koelsch, S. (2022). Unpredictability of the “when” influences
1326 prediction error processing of the “what” and “where”. *PLOS ONE*, *17*(2),
1327 e0263373. <https://doi.org/10.1371/journal.pone.0263373>

- 1328 Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word
1329 frequency in event-related brain potentials. *Memory & Cognition*, *18*(4), 380–393.
1330 <https://doi.org/10.3758/BF03197127>
- 1331 Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension:
1332 Benefits, costs, and ERP components. *International Journal of Psychophysiology*,
1333 *83*(2), 176–190. <https://doi.org/10.1016/j.ijpsycho.2011.09.015>
- 1334 Vanderwart, M. (1984). Priming by pictures in lexical decision. *Journal of Verbal Learning*
1335 *and Verbal Behavior*, *23*(1), 67–83.
1336 [https://doi.org/10.1016/S0022-5371\(84\)90509-7](https://doi.org/10.1016/S0022-5371(84)90509-7)
- 1337 VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of
1338 recognition. *Frontiers in Psychology*, *2*, 1–6.
1339 <https://doi.org/10.3389/fpsyg.2011.00365>
- 1340 Veale, J. F. (2014). Edinburgh Handedness Inventory - Short Form: A revised version
1341 based on confirmatory factor analysis. *Laterality*, *19*(2), 164–177.
1342 <https://doi.org/10.1080/1357650X.2013.783045>
- 1343 Vidal, C., Content, A., & Chetail, F. (2017). BACS: The Brussels Artificial Character Sets
1344 for studies in cognitive psychology and neuroscience. *Behavior Research*
1345 *Methods*, *49*(6), 2093–2112. <https://doi.org/10.3758/s13428-016-0844-8>
- 1346 Vidal-Gran, C., Sokoliuk, R., Bowman, H., & Cruse, D. (2020). Strategic and
1347 Non-Strategic Semantic Expectations Hierarchically Modulate Neural Processing.
1348 *eNeuro*, *7*(5). <https://doi.org/10.1523/ENEURO.0229-20.2020>
- 1349 Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the
1350 neurophysiological evidence for predictive processing as a model of perception.
1351 *Annals of the New York Academy of Sciences*, *1464*(1), 242–268.
1352 <https://doi.org/10.1111/nyas.14321>

- 1353 Wang, F., & Maurer, U. (2017). Top-down modulation of early print-tuned neural activity
1354 in reading. *Neuropsychologia*, *102*, 29–38.
1355 <https://doi.org/10.1016/j.neuropsychologia.2017.05.028>
- 1356 Wang, F., & Maurer, U. (2020). Interaction of top-down category-level expectation and
1357 bottom-up sensory input in early stages of visual-orthographic processing.
1358 *Neuropsychologia*, *137*, 107299.
1359 <https://doi.org/10.1016/j.neuropsychologia.2019.107299>
- 1360 Wang, J., Deng, Y., & Booth, J. R. (2019). Automatic semantic influence on early visual
1361 word recognition in the ventral occipito-temporal cortex. *Neuropsychologia*, *133*,
1362 107188. <https://doi.org/10.1016/j.neuropsychologia.2019.107188>
- 1363 Wang, L., Schoot, L., Brothers, T., Alexander, E., Warnke, L., Kim, M., Khan, S.,
1364 Hämäläinen, M., & Kuperberg, G. R. (2023). Predictive coding across the left
1365 fronto-temporal hierarchy during language comprehension. *Cerebral Cortex*,
1366 *33*(8), 4478–4497. <https://doi.org/10.1093/cercor/bhac356>
- 1367 White, A. L., Palmer, J., Boynton, G. M., & Yeatman, J. D. (2019). Parallel spatial
1368 channels converge at a bottleneck in anterior word-selective cortex. *Proceedings*
1369 *of the National Academy of Sciences of the United States of America*, *116*(20),
1370 10087–10096. <https://doi.org/10.1073/pnas.1822137116>
- 1371 Woolnough, O., Donos, C., Rollo, P. S., Forseth, K. J., Lakretz, Y., Crone, N. E.,
1372 Fischer-Baum, S., Dehaene, S., & Tandon, N. (2021). Spatiotemporal dynamics of
1373 orthographic and lexical processing in the ventral visual pathway. *Nature Human*
1374 *Behaviour*, *5*(3), 389–398. <https://doi.org/10.1038/s41562-020-00982-w>
- 1375 Yeatman, J. D., Rauschecker, A. M., & Wandell, B. A. (2013). Anatomy of the visual word
1376 form area: Adjacent cortical circuits and long-range white matter connections.
1377 *Brain and Language*, *125*(2), 146–155.
1378 <https://doi.org/10.1016/j.bandl.2012.04.010>

- 1379 Ylinen, S., Bosseler, A., Junntila, K., & Huotilainen, M. (2017). Predictive coding
1380 accelerates word recognition and learning in the early stages of language
1381 development. *Developmental Science*, 20(6), e12472.
1382 <https://doi.org/10.1111/desc.12472>
- 1383 Ylinen, S., Huuskonen, M., Mikkola, K., Saure, E., Sinkkonen, T., & Paavilainen, P.
1384 (2016). Predictive coding of phonological rules in auditory cortex: A mismatch
1385 negativity study. *Brain and Language*, 162, 72–80.
1386 <https://doi.org/10.1016/j.bandl.2016.08.007>
- 1387 Zhao, J., Kipp, K., Gaspar, C., Maurer, U., Weng, X., Mecklinger, A., & Li, S. (2014). Fine
1388 neural tuning for orthographic properties of words emerges early in children
1389 reading alphabetic script. *Journal of Cognitive Neuroscience*, 26(11), 2431–2442.
1390 https://doi.org/10.1162/jocn_a_00660
- 1391 Zhao, J., Maurer, U., He, S., & Weng, X. (2019). Development of neural specialization for
1392 print: Evidence for predictive coding in visual word recognition. *PLoS Biology*,
1393 17(10), 1–17. <https://doi.org/10.1371/journal.pbio.3000474>