1   **Decoding Predictions and Violations of Object Position and Category in Time-**

2   **resolved EEG**

3   Christopher J. Whyte[1, 2*], Amanda K. Robinson[1,2], Tijl Grootswagers[1], Hinze

4   Hogendoorn[3], Thomas A. Carlson[1,2].

5

6   Author Affiliations:

7

8   [1] School of Psychology, University of Sydney.

9   [2] Department of Cognitive Science, Macquarie University.

10   [3] Melbourne School of Psychological Sciences, University of Melbourne.

11   * Corresponding author.

12

13

14

15

16

17

18

19

20

21

22   Corresponding author Information:

23   Griffith Taylor Building A19, Manning Rd,

24   Camperdown NSW 2050, Sydney Australia.

25   Email: christopherjackwhyte@gmail.com

26 **Abstract:** Classic models of predictive coding propose that sensory systems use

27 information retained from prior experience to predict current sensory input. Any

28 mismatch between predicted and current input (prediction error) is then fed forward up

29 the hierarchy leading to a revision of the prediction. We tested this hypothesis in the

30 domain of object vision using a combination of multivariate pattern analysis and time-

31 resolved electroencephalography. We presented participants with sequences of

32 images that stepped around fixation in a predictable order. On the majority of

33 presentations, the images conformed to a consistent pattern of position order and

34 object category order, however, on a subset of presentations the last image in the

35 sequence violated the established pattern by either violating the predicted category or

36 position of the object. Contrary to classic predictive coding when decoding position

37 and category we found no differences in decoding accuracy between predictable and

38 violation conditions. However, consistent with recent extensions of predictive coding,

39 exploratory analyses showed that a greater proportion of predictions was made to the

40 forthcoming position in the sequence than to either the previous position or the position

41 behind the previous position suggesting that the visual system actively anticipates

42 future input as opposed to just inferring current input.

43

44

45

46

47

48

49 Key words: Object vision, predictive coding, electroencephalography, multivariate

50 decoding, representation, vision.

Introduction

The human brain processes the position and category of objects within the visual field seemingly without effort. The process of recognising objects, although not apparent via introspection, underpins all of our interactions with the world. Even simple tasks such as making a cup of coffee rely on our ability to rapidly categorise and locate objects within the visual field. The temporal evolution of object recognition has been well characterised experimentally (Carlson, Tovar, Alink & Kriegeskorte, 2013; Cichy, Pantazis, & Oliva, 2014; Grootswagers, Robinson & Carlson, 2019; Grootswagers, Robinson, Shatek & Carlson, 2019; Robinson, Grootswagers & Carlson, 2019), however, the computational architecture underlying this process is a matter of ongoing investigation. A possible clue comes from the highly predictable structure of the visual environment. Objects tend to move along predictable trajectories giving rise to eye movement strategies such as smooth pursuit (Barnes, 2008). And, contextual knowledge of a scene greatly constrains the category of objects that are likely be to present (Bar, 2004). Given the exorbitant metabolic demands of neural processing (Stone, 2018), and the importance of prospective computation for survival (Hopfield, 1994), it would be surprising if the brain did not exploit the inherent redundancy in visual input (resulting from the structured nature of the environment) in the service of perception. In fact, for *any* system responding to an input signal that retains information from the input (i.e., has a non-zero form of memory), the retention of non-predictive information is formally equivalent to energetic inefficiency (Still, Sivak, Bell & Crooks, 2012). The question is, therefore, not whether brains predict but how.

The prospective goal of perception lies at the heart of a family of models in computational neuroscience collectively referred to as predictive processing models

76   (Clark, 2016). Predictive processing models, which includes both predictive coding

77   (Bastos et al, 2012; Friston, 2005; Friston & Kiebel, 2009; Rao & Ballard, 1999) and

78   active inference (Friston et al., 2017; Parr, Da Costa & Friston, 2020), have shown

79   great promise in accounting for a wide range of visual phenomena from extra-classical

80   receptive field effects (Rao & Ballard, 1999) and repetition suppression (Auksztulewicz

81   & Friston, 2016), to selective attention (Feldman & Friston, 2010; Mirza, Adams,

82   Friston & Parr, 2019), and even visual awareness (Parr, Corcoran, Friston & Hohwy,

83   2019; Whyte & Smith, 2020). Of the predictive processing models, by far the most

84   popular is the classic predictive coding model proposed by Rao and Ballard (1999)

85   and later built upon by Friston (2005). Unlike feed-forward neural networks, predictive

86   coding depicts perception as a process of top-down model testing aimed at minimising

87   the difference between an internal model of the world and sensory input. The internal

88   model generates cascades of descending predictions that meet bottom-up signals at

89   each level of the visual hierarchy. The mismatch between the prediction and the

90   bottom-up signal (prediction error) is fed forward to the next level in the hierarchy

91   leading to a revision of the prediction. (Bastos et al, 2012; Friston, 2005, 2010; Clark,

92   2016; Hohwy, 2013, Rao & Ballard, 1999). In line with this view, there is now

93   considerable evidence from functional magnetic resonance imaging (fMRI) suggesting

94   that prediction has a silencing effect on neural responses that is orthogonal to other

95   top-down processes such as attention (Kok et al, 2011; Richter, Ekman & de Lange,

96   2018), and that higher levels in the cortical hierarchy send predictions to subordinate

97   levels of the hierarchy (Summerfield et al, 2006).

98   Also in line with predictive processing models, a body of research conducted in

99   magnetoencephalography and electroencephalography (M/EEG) has shown that

100 expectation also has substantial effects in the temporal domain. When stimuli are

101   expected, stimulus features such as orientation can be decoded even before stimulus

102   onset (Kok, Mostert & De Lange, 2017). Decoding of object position in apparent motion

103   paradigms has a latency advantage when the target stimulus moves along a

104   predictable trajectory (Hogendoorn & Burkitt, 2018), and the violation of the orientation

105   and identity of faces has a dissociable mismatch ERP effect across the dorsal and

106   ventral streams (Robinson et al., 2018).

107   Here we used time-resolved multivariate pattern analysis and EEG (MVPA;

108   Carlson et al, 2013; Cichy et al, 2014; Grootswagers, Wardle & Carlson, 2017) to

109   investigate the temporal effects of prediction and prediction error at different levels of

110   the visual hierarchy. We presented participants with sequences of images that

111   stepped around fixation in a predictable order. On the majority of sequences, the

112   images conformed to the pattern of position and category order, however, on a subset

113   of the sequences the last image in the sequence violated the established pattern by

114   violating either the predicted category (high level) or the predicted position (low level)
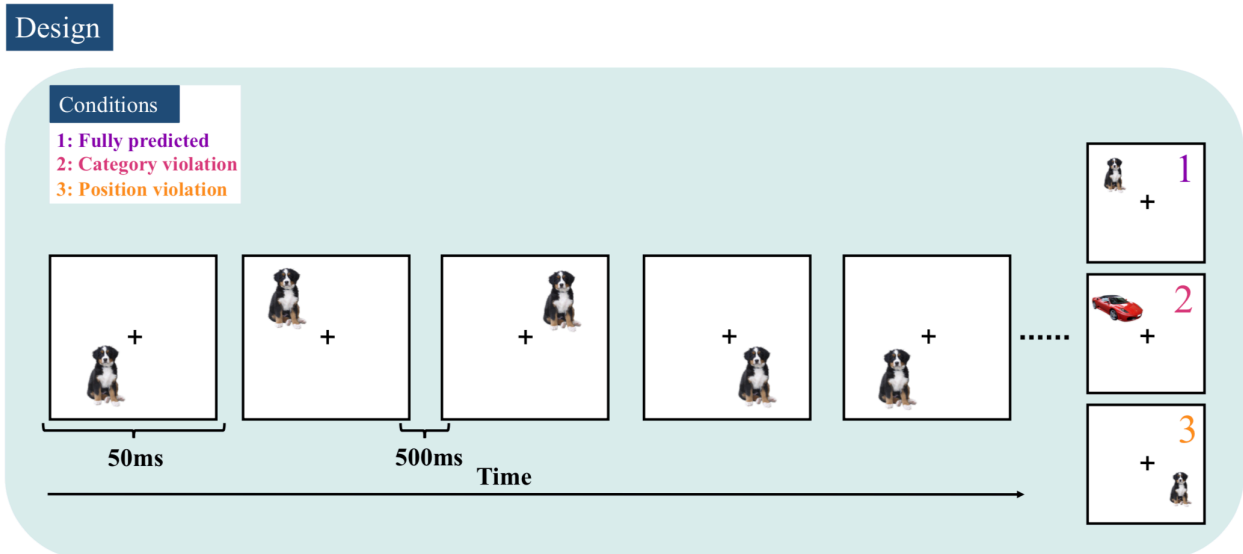
115   of the object.

116

117

Figure 1. Experimental design. Stimuli were presented for 50ms with 500ms ISI in sequence around four possible locations that were equidistant from the fixation cross and either repeated or alternated the category of the stimulus (i.e., dog or car). The last image in the sequence either conformed to the pattern (1: fully predicted - purple), or violated the established pattern by violating either the the predicted category (2: category violation - pink), or the predicted position (3: position violation - yellow). This example shows a clockwise repeating object sequence.

Based upon the classic formulation of predictive coding (Friston, 2005; Rao & Ballard, 1999), and the structure of visual hierarchy (Felleman & Van Essen, 1991), we generated four related hypotheses (see Figure 1). First, we expected that due to prediction error signals, there would be above chance decoding between predicted stimuli, and stimuli that have violated a prediction, for both violations of position and category. Second, since the generation of prediction error is hypothesised to alter the content of representations (c.f., King, Schurger, Naccache & Dehaene, 2014) we expected representations to be less separable when predictions were violated leading to reduced decodability. Third, given the relative independence of the dorsal and ventral streams (Ungerleider & Haxby, 1994) we expected that category violation,

136   which relies on processing within the ventral stream, would not interact with position

137   representation, which relies on processing within the dorsal stream, and vice versa.

138   Several aspects of this study including hypotheses, design, and analysis choices were

139   pre-registered (https://osf.io/hkedz/).
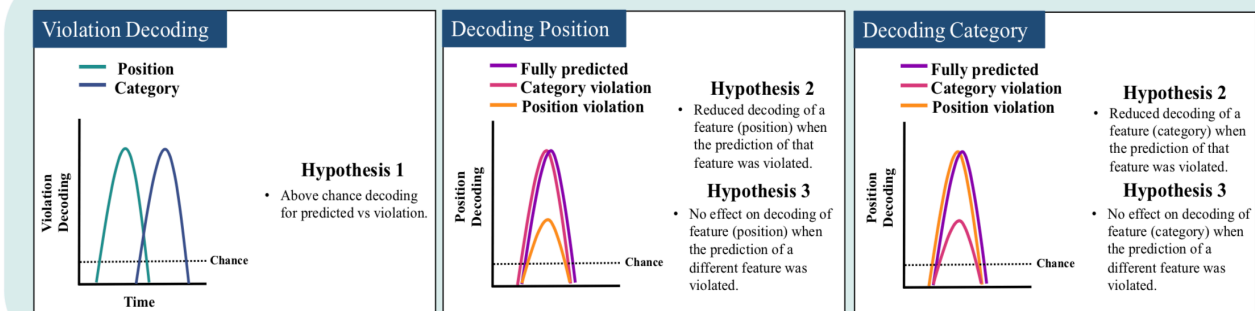
140

141

142   Figure 2. Hypotheses for each condition. Hypothesis 1 – the increase in prediction error on violation

143   trials was expected to lead to above chance decoding between prediction and violation trials for both

144   position and category. Category violation decoding is shifted rightward in the figure because category

145   is extracted at a higher point of the hierarchy than position so the appearance of prediction errors related

146   to category processing is expected to occur later in time. Hypothesis 2 – given that prediction error

147   leads to a revision of the content of representations we expected representations to be less separable

148   when predictions were violated thereby lowering decoding accuracy. Hypothesis 3 – given the relative

149   functional independence of the dorsal and ventral streams we expected that the violation of a feature

150   that is not the target of decoding would have no effect on decoding accuracy.

151

152

153

154

155

156

157 **Methods**

158

159 *Stimuli and procedure*

160

161 We recruited 34 adult participants (21 female) aged between 18 – 27 years old

162 (average 20.15) from the University of Sydney in exchange for course credit or

163 payment. Participants viewed sequences of dog and car images (obtained from the

164 free image site www.pngimage.com) that appeared in four different positions. Stimuli

165 were presented in sequences of 6 to 10 images that stepped around fixation in a

166 predictable order (50% clockwise, 50% counter clockwise). The stimulus subtended

167 3x3 degrees of visual angle and was presented 4 degrees from fixation. Each stimulus

168 was presented for 50ms with a 500ms inter-stimulus interval. There were two types of

169 predictable sequences: repeating object sequences (50%) and alternating object

170 sequences (50%). During the repeating sequences, the same stimulus (dog or car)

171 was presented throughout the sequence. In the alternating sequences the category of

172 stimulus alternated on each successive presentation in the sequence (e.g., dog, car,

173 dog, car…). There were 490 sequences in total. On the majority of sequences (256

174 out of 448 non target sequences), all stimuli conformed to the pattern of position order

175 (clockwise/counter-clockwise) and object order (repeating/alternating). For the

176 remaining sequences, the last image in the sequence violated the established pattern

177 by either violating the predicted category of the object (*category violation*; e.g., dog-

178 dog-dog-**car** or dog-car-dog-**dog**; 96 sequences) or the predicted location (*position*

179 *violation*; 96 sequences). See Figure 1. Importantly, for position violation sequences

180 the position of the last stimulus was a reversal of the established movement (e.g.,

181   positions 4-1-2-3-*2* or 1-4-3-2-*3*. This ensured that for all conditions, the previous item

182   in the sequence could not be a confound in the decoding analysis.

183   Participants were required to monitor the sequence for inverted stimuli which

184   appeared 8.57% of the time (42 sequences). They were instructed to fixate on the

185   cross in the centre of the monitor, and not to move their eyes. The inversion task kept

186   them alert and attentive without making the predictability of the stimulus task relevant.

187   With the exception of the inversion of the target stimulus, target sequences were

188   identical to predictable non-target sequences. Target sequences (i.e., sequences with

189   inverted stimuli) were excluded from analysis.

190   Participants completed 7 blocks each consisting of 70 sequences. Between

191   each block we presented a 'pattern localiser' consisting of a rapid stream of 120 dog

192   and car images yielding a total of 840 additional presentations (i.e., 12 repeats of each

193   dog and car image at each of the 5 locations). Each image was presented with 50ms

194   ISI and 100ms SOA either centrally (at fixation) or at one of the four experimental

195   locations. The location of the stimulus was shuffled such that there was no statistical

196   regularity in the sequence. The pattern localiser served as an independent source of

197   training data for the decoding analysis.

198

199   *EEG Recordings and Pre-processing*

200

201   Continuous EEG data was recorded with a BrainVision ActiChamp system with a

202   digitised sampling rate of 1000Hz. The 64 electrode system was arranged according

203   to the 10-10 placement system all referenced to Cz. Pre-processing was conducted in

204   MATLAB using the EEGLAB toolbox (Delorme & Makeig 2004). The data were filtered

205   with a high pass filter of 0.1Hz and a low pass filter of 100Hz and down-sampled to

206   250Hz. Epochs were created between -200 to 1000ms relative to the onset of the final

207   image in the sequence (448 epochs).

208

209   *Decoding Analysis*

210

211   We employed an MVPA decoding pipeline to all EEG channels following the

212   recommendations of Grootswagers, Wardle and Carlson (2017) using the

213   CoSMoMVPA toolbox (Oosterhof, Connolly & Haxby, 2016). All decoding was

214   performed within subject using a linear discriminant analysis (LDA) classifier.

215   Statistical analysis was performed at the group level averaging across individual

216   decoding accuracies. To explore the emergence of prediction error signals we

217   compared neural responses of violation trials with neural responses of predictable

218   trials. For violation decoding we used a leave - one block - out cross validation

219   scheme. There were two separate analyses: predictable versus object violation and

220   predictable versus position violation. There were far more predictable sequences, so

221   to ensure balanced data in the test set for every unpredictable trial, we selected a

222   predictable trial that was matched for repeating/alternating sequence,

223   clockwise/counter clockwise sequence, category and position. For *position* decoding

224   we used a cross decoding scheme by training the classifier on data from the pattern

225   localiser using the four peripheral positions and testing the classifier on data from the

226   experimental sequences. For category decoding we again used a cross decoding

227   scheme training on data from the pattern localiser and testing on data from the

228   response sequences. However, for category decoding we trained the classifier on

229   stimuli presented at all 5 locations of the pattern localiser (4 peripheral positions and

230   central) to get a better estimate of position-invariant image category information.

231   In total we decoded 8 contrasts; *violation* split by category and position

232   (contrasts 1-2); *position* (i.e., location 1-4) split by fully predicted, position violation,

233   and category violation conditions (contrasts 3-5) and *category* (i.e., dog vs car) fully

234   predicted, position violation, and category violation conditions (contrasts 6-8).

235

236

237

238   *Exploratory Analysis of Classification Errors*

239

240   At the time point of peak decoding for position we found insufficient evidence to

241   determine if there was more position information in the neural signal for predictable

242   compared with position violation trials (hypothesis 2; Figure 2). In order to investigate

243   this hypothesis further, we examined the predictions made by the classifier. The

244   classifier extracted neural patterns of activation specific to each of the four

245   experimental positions and used these to predict the position of the stimulus on each

246   experimental trial. If there was no position information about the stimulus (e.g., prior

247   to its appearance), the classifier would be expected to predict each of the four

248   locations equally often. If there was position information in the neural signal (e.g.,

249   during retinotopic processing of the stimulus), the classifier would be expected to

250   predict the correct position. However, classification is rarely perfect, and investigating

251   the errors made by the classifier can give insight into other information in the neural

252   signal. For example, in an apparent motion paradigm Blom, Feuerriegel, Johnson,

253   Bode and Hogendoorn (2020) trained a classifier to decode between stimuli presented

254   at the two locations on either side of the target stimulus and then tested the classifier

255   on the location of the target stimulus. For the first ~70ms the majority of the predictions

256   made by the classifier were made to the location behind the location of the target but

257   after ~70ms the majority of predictions were made to the location following the target

258   showing that there was anticipatory information in the EEG signal.

259   We examined the average proportion of predictions made by the classifier to

260   each position. We then sorted the predictions made by the classifier to each of the

261   four locations relative to the expected position. Assuming there was only information

262   about the current stimulus in the EEG signal there should have been an equal

263   proportion of predictions made to each of the three incorrect positions. If, however, the

264   EEG signal also contained information about one of the incorrect positions the

265   classifier might predict one of the incorrect positions more often than the others. For

266   example, when the stimulus was presented in an unexpected position, as was the

267   case on violation trials, predictive information in the signal might have increased the

268   proportion of predictions made by the classifier to the predicted position. To statistically

269   evaluate the evidence for differences in the proportion of predictions made by classifier

270   we took the average proportion of predictions made over a 20ms time-window (86-

271   106ms) centred on the point of peak decoding accuracy (96ms) and used Bayes

272   factors (described below) to evaluate the strength of evidence.

273   It is important to note that unlike the analyses listed above, this analysis was

274   not planned a priori and is therefore considered exploratory.

275

276   *Statistical Inference*

277

278   To calculate the evidence for the null and alternative hypotheses we used JZS

279   Bayes Factors (Rouder et al., 2009). To determine the evidence for the alternative

280   hypothesis of above chance decoding we employed a Cauchy prior with the scale

281   factor set to 0.707, while the prior for the null hypothesis was a point at chance, 0.25

282   for position decoding and 0.5 for all other decoding tests (Morey & Rouder, 2011). To

283   determine the evidence for a non-zero difference between decoding accuracies, we

284   used a uniform prior with a point null set to zero. This same procedure was also used

285   in the exploratory analysis described above. Using these distributions, we computed

286   Bayes factors (Dienes, 2011; Jeffreys, 1998; Rouder, Speckman, Sun, Morey, &

287   Iverson, 2009; Wagenmakers, 2007) which, being a ratio of marginal likelihoods,

288   measures the evidence for the alternative hypothesis relative to the null. For the

289   purpose of plotting the results we thresholded the Bayes factors at BF > 1/3 but < 3

290   as inconclusive evidence either way, BF > 6 for modest evidence for the alternative

291   hypothesis, and BF > 10 for strong evidence for the alternative hypotheses. Because

292   point nulls are biased to the alternative hypothesis as sample size becomes larger

293   (Morey & Rouder, 2011), we took BF < 1/3 as strong evidence in favour of the null.

294

295   **Results**

296

297   *Behaviour*

298

299   Participants performed an orthogonal task to detect inverted stimuli. After inspecting

300   the behavioural responses, we excluded one participant who did not respond to any

301   of the targets. After exclusion, mean accuracy was 91.9% (SD = 10.13). We then

302   excluded another two participants from further analysis whose accuracy was lower

303   than 80%, leaving a total of 31 participants whose data was used in the decoding

304   analysis. We used the inclusion criterion of above 80% accuracy because of the

305   extreme simplicity of the task.

306

307   *Violation decoding*

308

309   We expected the neural signal to contain information about whether a stimulus violated

310   the expected pattern for both position and object (Hypothesis 1). In support of

311   hypothesis 1 we observed strong evidence that stimuli were processed differently if

312   they appeared in unexpected positions (position *violation*) relative to expected

313   positions (Figure 3; green line). Position *violation* decoding was above chance (BF >

314   10) 244ms after stimulus onset with two peaks in accuracy at 258ms and 812ms. Each

315   peak coincided with an increase in evidence for above chance decoding (BF > 10). In

316   contrast, partially contrary to hypothesis 1, we did not observe a difference in

317   processing of stimuli when the category violated the established pattern (*category*

318   *violation)* relative to expected category (Figure 3; blue line). In decoding of violation

319   for category we found strong evidence for the null hypothesis across the trial (BF <

320   1/3). Unsurprisingly, when comparing accuracy between the two conditions we found

321   strong evidence (BF > 10) for a difference in decoding accuracy between the two

322   conditions that coincided with peaks in decoding accuracy for *position* violation
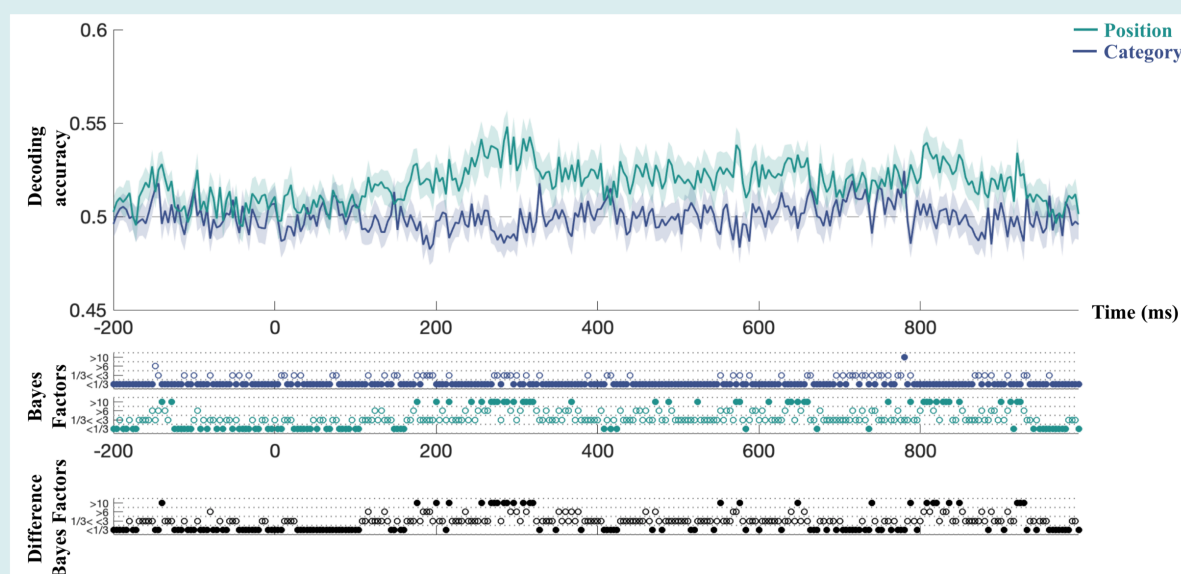
323   decoding*.*

Figure 3. Violation decoding. Mean decoding accuracy for predictable versus position violation (green) and predictable versus category violation (blue). Coloured dotes below the plot indicate the thresholded Bayes factors (BF) for category and position. For BF > 10 and BF < 1/3 which indicate strong evidence for either the alternative or null hypothesis are shown represented by filled in circles. BF > 6 and BF <3 which indicate modest or inconclusive evidence either way are represented by open circles. Black dots indicate the thresholded BF for the difference in decoding accuracy between violation types.

Decoding position

We expected position-related information to be present in the neural signal from an early stage of processing, and for the amount of information to differ for predictable positions compared to position violations (Hypothesis 2). We found strong evidence (BF > 10) for above chance decoding of position (chance = 25%) for *fully predicted*, *category violation*, and *position violation* from 72ms post stimulus onset with a peak at 96ms (Figure 4a). Interestingly, suggestive of hypothesis 2 concerning the disruptive effect of prediction error, peak decoding accuracy for *position violation* was lower than *fully predicted* and *category violation* and was less sustained. However, during the

342   time period where there was a visible difference between decoding accuracies (~85

343   to 120ms) we found inconclusive evidence for the null and alternative hypothesis (1/3

344   > BF < 3). To investigate this in more detail we examined the proportion of predictions

345   made by the classifier for each location (Figure 5). If hypothesis 2 was correct, we

346   would have expected the neural signal to contain information about the predicted

347   position on both violation and predicted trials, indexed by higher numbers of classifier

348   errors for the expected position than other incorrect positions on position violation

349   trials. However, this is not what we found. Intriguingly, at the point of peak decoding

350   accuracy (~96ms), classification error analysis showed a higher number of predictions

351   for the position that followed the expected position (i.e., expected + 1, the next position

352   in the sequence) across all three conditions making up ~25% of classifier output

353   (green line in Figure 5). To evaluate this statistically we examined the differences

354   between the average proportion of classifier predictions made to each position

355   averaged over a 20ms time-window (86-106ms) centred on the point of peak decoding

356   accuracy (96ms). For the *fully predicted* and *category violation* conditions we observed

357   strong evidence BF > 10 that a greater proportion of predictions was made to the next

358   position in the sequence (expected + 1) in comparison to both the previous position

359   (expected – 1) and the position behind the previous position (expected - 2). Similarly,

360   for the *position violation* condition we found strong evidence (BF > 10) that a greater

361   proportion of predictions was made to the next position in the sequence (expected +

362   1) than to the expected position, and modest evidence (BF > 3) that a greater

363   proportion of predictions was made to the next position than to the previous position

364   (expected – 1). We consider the interpretation and significance of this result in the

365   discussion section. Finally, it is worth highlighting that we could not evaluate

366   hypothesis 3 – which predicted that there would be no difference in decoding accuracy

367   when the violated feature was not the target of decoding – as it relied on violation

368   having a disruptive effect as predicted by hypothesis 2.
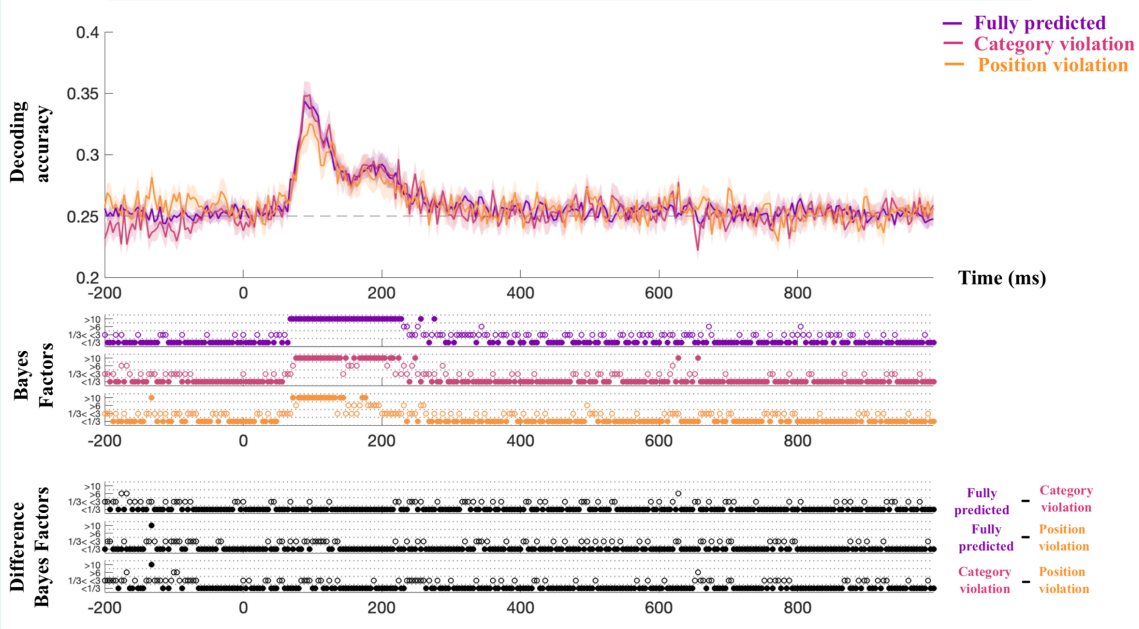
369

370   Figure 4. Position decoding (chance = 25%) for fully predicted, category violation and position violation

371   conditions. Coloured dots below each plot indicate the thresholded Bayes factors for each time point.

372   Black dots indicate the thresholded Bayes factors for the difference in decoding accuracy between
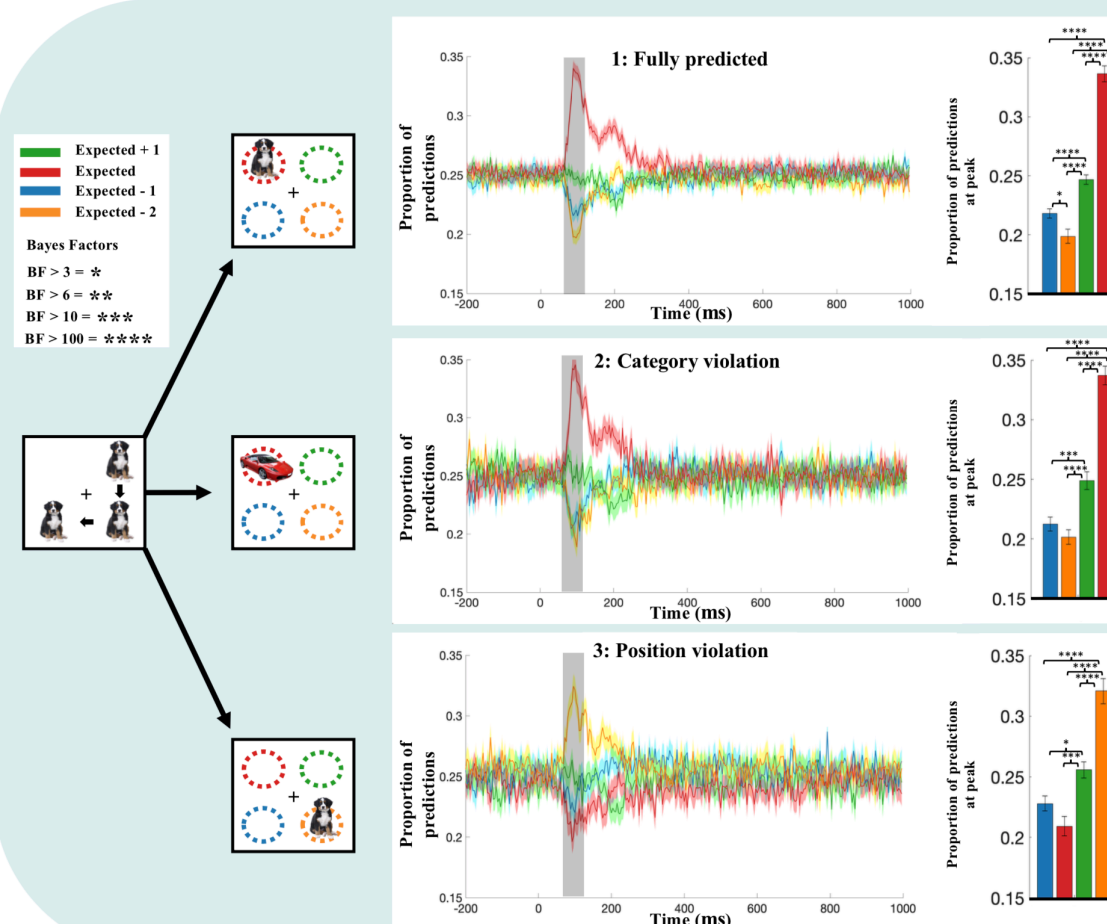
373   conditions.

374

Figure 5. Analysis of classifier output for position decoding across conditions. Left portion of the figure shows an example of the position of the stimulus for each condition in relation to what was expected given the pattern of the preceding sequence. Colours on plots correspond to the stimulus positions. Middle graphs show the proportion of predictions made by the classifier to each position. Across conditions the actual position of the stimulus had the highest proportion of classifier predictions (~32-34%). The right bar plots show the proportion of classifier predictions made to each position at the peak of decoding (shaded portion of graph). Asterisks* above the plots indicate the Bayes factors for the differences in proportion of predictions for each position. Contrary to hypothesis 2, there was a greater proportion of predictions made to the expected + 1 (next) location than either of the other two incorrect locations across conditions. Crucially, in the position violation condition there was exceptionally strong evidence (BF >100) that there was a greater proportion of predictions made to the expected + 1 location than the expected location suggesting the presence of anticipatory information in the EEG signal.

389

Decoding Object Category

391

392 To assess object category information, we decoded car versus dog for the three

393 predictability conditions (see Figure 6). For the *fully predicted* condition there were 6

394 time points between 228-260ms that showed modest (BF > 3) to strong evidence (BF

395 > 10) for above chance decoding of *category*. Similarly, for the *category violation*

396 condition there were 3 time points between 226-276ms that showed modest (BF > 3)

397 to strong evidence (BF > 10) for above chance decoding, and for the *position violation*

398 condition, we found strong evidence for the null hypothesis throughout the trial (BF <

399 1/3) with only a few time-points transiently showing evidence for above chance

400 decoding. In terms of differences in decoding accuracy between conditions, with the

401 exception of a few sparsely distributed and isolated time points, we found strong

402 evidence for the null hypothesis that there was no difference in decoding accuracy

403 from stimulus onset until the end of the epoch (BF < 1/3). These results stand in

404 opposition to hypothesis 2 which forecast that violations of predictions would decrease

405 decoding accuracy if the violation was the target of decoding. Again, we could not

406 evaluate hypothesis 3 as it relied on violations having a disruptive effect on decoding

407 accuracy as predicted by hypothesis 2.
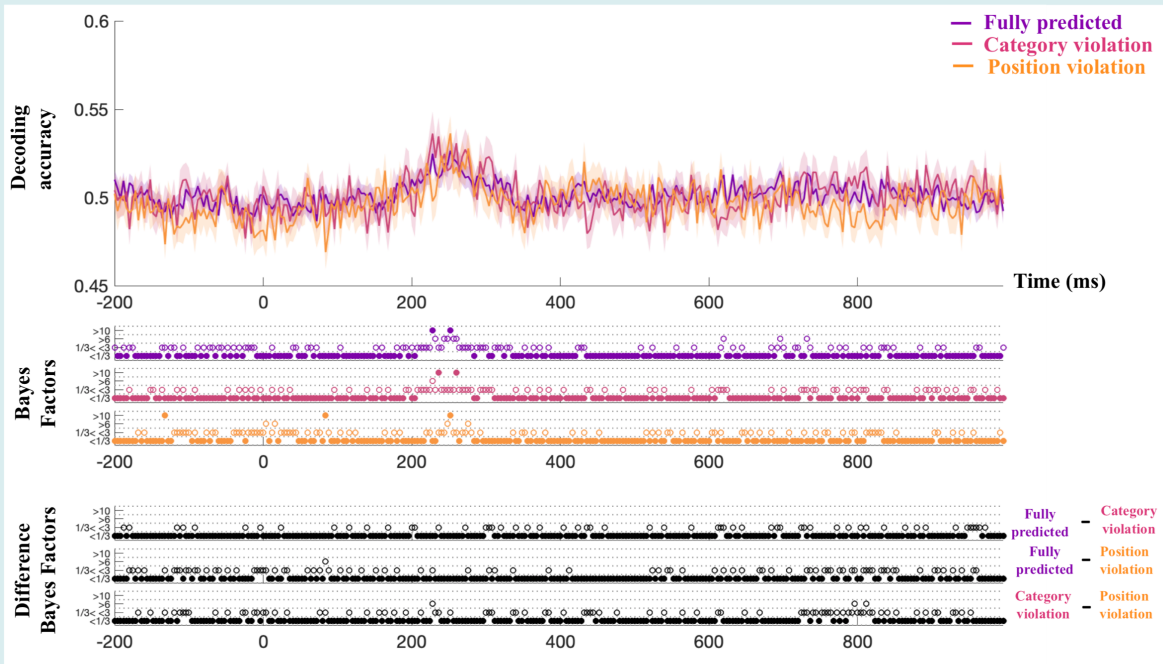
408

Decoding Category

409

410 Figure 6. Category decoding (chance = 50%) for fully predicted, category violation and position violation

411 conditions. Coloured dots below each plot indicate the thresholded Bayes factors for each time point.

412 Black dots indicate the thresholded Bayes factors for the difference in decoding accuracy between

413 conditions.

414

415 **Discussion**

416

417 The aim of this study was to test a set of four hypotheses based on the classic model

418 of predictive coding (Friston, 2005; Rao & Ballard, 1999), and the structure of the

419 visual hierarchy (Felleman & Van Essen, 1991), in the domain of basic object vision

420 using time-resolved EEG. According to the classic model of predictive coding sensory

421 systems use information from prior experience to predict current incoming sensory

422 input, or more concisely, they are trying to predict the present (Clark, 2016). Overall,

423 our results were largely contrary to this idea. The remainder of this paper will discuss

424   the consequence of our results for each hypothesis and propose an explanation of the

425   findings in terms of generalised predictive coding (Friston & Kiebel, 2009; Friston,

426   Stephan, Li & Daunizeau, 2010) and the temporal realignment hypothesis

427   (Hogendoorn & Burkitt, 2019).

428   According to generalised predictive coding (Friston & Kiebel, 2009; Friston,

429   Stephan, Li & Daunizeau, 2010), instead of just aiming to predict current input,

430   predictions are cast in generalised coordinates of motion meaning that predictive

431   signals also represent the velocity, acceleration, and other higher order derivatives of

432   the predicted input allowing sensory systems to extrapolate. For an accessible

433   introduction to the idea of generalised coordinates of motion see Susskind &

434   Hrabovsky, 2014, and for its role in predictive coding see Buckley, Kim, McGregor &

435   Seth, 2017. Similarly, the temporal realignment hypothesis (Hogendoorn & Burkitt,

436   2019) proposes that the brain overcomes the transmission delays inherent to the

437   visual system by having both predictions and prediction errors extrapolate forward in

438   time. The key feature of both of these models is that they posit the existence of

439   temporal predictive signals that carry information about what *will* happen and not just

440   what *is* happening.

441   For *prediction vs violation* of position, we observed strong evidence for above

442   chance decoding which peaked at 258ms. Prima facie, this result seems in line with

443   hypothesis 1 which predicted that the presence of prediction error signals on the

444   violation trials would lead to above chance *violation* decoding. However, decoding was

445   far too late to plausibly reflect an error message which by hypothesis would occur at

446   a similar time-point as peak decoding accuracy for position. In fact, the peak in

447   decoding accuracy for *violation* of position occurred ~150ms later than peak decoding

448   accuracy for position. As such, the time course of the response is more likely due to

449   an orienting of attention (Carlson, Hogendoorn & Verstraten, 2006; Eimer, 2000).) to

450   the unpredicted position. In favour of this interpretation, for *violation* of category we

451   found strong evidence for the null hypothesis of no above chance decoding. If

452   violations of predictions generated error signals that were large enough to be detected

453   at the level of the scalp we would have expected to see above chance decoding of

454   *violation* for category as well as position.

455   Assuming the interpretation put forward above is on track, the lack of a

456   decodable error signal suggests that prediction and prediction error may have a subtler

457   effect than we initially hypothesised. Indeed, considering that the stimulus moved

458   around the screen and did not stay within a consistent set of receptive fields, the short-

459   term changes in synaptic plasticity that are thought to underlie the generation of error

460   related ERPs (Auksztulewicz & Friston, 2016; Garrido, Kilner, Stephan & Friston,

461   2009; Stefanics, Kremláček & Czigler, 2014) would have been reduced, and as such,

462   the changes in voltage that characterise violation effects in ERP would have been less

463   pronounced.

464   With that said, it is still plausible that the presence of predictive signals could

465   have caused the classifier to make more (erroneous) predictions to the expected

466   position on violation trials. Although again, this is not what we found. When decoding

467   position, all three conditions - *fully predicted*, *category violation*, and *position violation*

468   - displayed above chance decoding accuracy with a peak at 96ms. Importantly, peak

469   decoding accuracy for *position violation* seemed lower than *fully predicted* and

470   *category violation* suggesting that prediction error may have had a disruptive effect on

471   position information as proposed in hypothesis 2. However, we found inconclusive

472   evidence differentiating between the null and alternative hypotheses at this time point.

473   To investigate the effect of position violation on position coding in a different way, we

474   inspected the classification output for each of the three conditions. If hypothesis 2 was

475   correct we would have expected to see a greater proportion of predictions to be made

476   to the expected location. Instead, the classifier made a greater proportion of

477   predictions to the next position in the sequence (expected + 1) across all three

478   conditions. Against the classic model of predictive coding this suggests that the visual

479   system actively anticipates future input as opposed to just inferring current input.

480   Crucially, however, this finding is predicted by both generalised predictive coding

481   (Friston & Kiebel, 2009; Friston, Stephan, Li & Daunizeau, 2010), and the temporal

482   realignment hypothesis (Hogendoorn & Burkitt, 2019), which propose that predictions

483   extrapolate forward in time. Further, our results mirror those of Blom, Feuerriegel,

484   Johnson, Bode and Hogendoorn (2020), who found that when a stimulus was a part

485   of a predictable sequence information about of the stimulus' next location was present

486   in the EEG signal 70 - 90ms earlier than would be expected if the evoked response

487   was purely stimulus driven.

488   We modest evidence for above chance classification of category in all three

489   conditions. However, contrary to hypothesis 2, which forecast that violations of

490   predictions would show lower decoding accuracy, we found strong evidence for the

491   null hypothesis that there were no differences between conditions. The lack of effect

492   for category violation has at least two plausible and complementary explanations.

493   First, it may simply be that there was an effect of violation at the neuronal population

494   level but because the cortical representation of objects is weaker in the peripheral

495   parts of the visual field where our stimuli were presented (Levy, Hasson, Avidan,

496   Hendler, & Malach, 2001) the differences could not be seen at the level of the scalp.

497   Indeed, we observed clear category decoding, yet the absolute decoding accuracy

498   was fairly low compared with previous studies using centrally presented objects (e.g.

499   Grootswagers, Robinson & Carlson, 2019; Grootswagers, Robinson, Shatek &

500   Carlson, 2019; Robinson, Grootswagers & Carlson, 2019). Second, like position, it

501   may be that predictions of category are primarily anticipatory in nature and as such we

502   should expect to see a greater proportion of classification errors made to the category

503   of the next stimulus but not the current stimulus. Unfortunately, however, our paradigm

504   did not allow us to interrogate this hypothesis. Since we only had two stimulus

505   categories and our stimulus was presented in an alternating or repeating pattern, the

506   category of the upcoming stimulus was either the same as the current category or the

507   same as the previous category making classifier output uninformative. Still, this

508   hypothesis will be easy to test in future work by simply increasing the number of

509   stimulus categories and presenting the stimuli in predictable sequences at the centre

510   of the screen where there is a stronger cortical representation of object category (as

511   has been done in fMRI; Richter, Ekman & de Lange, 2018). Relatedly, in terms of

512   hypothesis 3, which forecast that violations of the non-target feature would have no

513   effect on decoding, we cannot evaluate its accuracy as it relied on violations having a

514   disruptive effect on decoding (i.e., hypothesis 2). Unfortunately, if we are correct in

515   arguing that the effect of prediction error is reduced the periphery, this hypothesis will

516   be difficult to test using non-invasive techniques.

517   In sum, our results are largely contrary to the classic model of predictive coding

518   (Friston, 2005; Rao & Ballard, 1999) which proposes that sensory systems use prior

519   experience to predict the present (cf. Clark, 2016). Instead, consistent with

520   generalised predictive coding, and the temporal realignment hypothesis, our

521   exploratory analysis suggests that sensory signals are actively anticipate future input,

522   at least for representations of position. This contrary finding, which was predicted by

523   previous theoretical work, represents an important advance in how we should think

524   about prediction in sensory systems. We look forward to future work investigating

525   whether the anticipatory nature of prediction generalises to category representation.

526

527   **Acknowledgments**

530

531   **References**

532   Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual

533   determinants in predictive coding. *cortex*, *80*, 125-140.

534   Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617-

535   629.

536   Barnes, G. R. (2008). Cognitive processes involved in smooth pursuit eye

537   movements. *Brain and cognition*, *68*(3), 309-326.

538   Bell, A. H., Summerfield, C., Morin, E. L., Malecek, N. J., & Ungerleider, L. G. (2016).

539   Encoding of stimulus probability in macaque inferior temporal cortex. *Current*

540   *Biology*, *26*(17), 2280-2290.

541   Blom, T., Feuerriegel, D., Johnson, P., Bode, S., & Hogendoorn, H. (2020).

542   Predictions drive neural representations of visual events ahead of incoming

543   sensory information. *Proceedings of the National Academy of Sciences*.

544   Buckley, C. L., Kim, C. S., McGregor, S., & Seth, A. K. (2017). The free energy

545   principle for action and perception: A mathematical review. *Journal of*

546   *Mathematical Psychology*, *81*, 55-79.

547   Carlson, T. A., Hogendoorn, H., & Verstraten, F. A. (2006). The speed of visual

548   attention: What time is it?. *Journal of Vision*, *6*(12), 6-6.

549   Carlson, T. A., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013).

550   Representational dynamics of object vision: The first 1000 ms. *Journal of*

551   *Vision*, *13*(10), 1. https://doi.org/10.1167/13.10.1

552   Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in

553   space and Time. *Nature Neuroscience*, *17*(3), 455–462.

554   https://doi.org/10.1038/nn.3635

555   Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*.

556   Oxford University Press.

557   Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on?.

558   *Perspectives on Psychological Science*, *6*(3), 274-290.

559   Eimer, M. (2000). The time course of spatial orienting elicited by central and

560   peripheral cues: evidence from event-related brain potentials. *Biological*

561   *Psychology*, *53*(2-3), 253-258.

562   Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in*

563   *human neuroscience*, *4*, 215.

564   Felleman, D. J., & Van Essen, E. (1991). Distributed hierarchical processing in the

565   primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, *1*(1), 1-47.

566   Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the*

567   *Royal Society of London B: Biological Sciences*, *360*(1456), 815-836.

568   Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature Reviews*

569   *Neuroscience*, *11*(2), 127-138.385-397.

570   Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017).

571   Active inference: a process theory. *Neural computation*, *29*(1), 1-49.

572   Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy

573   principle. *Philosophical Transactions of the Royal Society B: Biological*

574   *Sciences*, *364*(1521), 1211-1221.

575   Friston, K., Stephan, K., Li, B., & Daunizeau, J. (2010). Generalised

576   filtering. *Mathematical Problems in Engineering*, *2010*.

577   Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch

578   negativity: a review of underlying mechanisms. *Clinical*

579   *neurophysiology*, *120*(3), 453-463.

580   Grootswagers, T., Robinson, A. K., & Carlson, T. A. (2019). The representational

581   dynamics of visual objects in rapid serial visual processing

582   streams. *NeuroImage*, *188*, 668-679.

583   Grootswagers, T., Robinson, A. K., Shatek, S. M., & Carlson, T. A. (2019). Untangling

584   featural and conceptual object representations. *NeuroImage*, *202*, 116083.

585   Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding dynamic brain

586   patternsfrom evoked responses: A tutorial on multivariate pattern analysis

587   applied to time series neuroimaging data. *Journal of cognitive neuroscience*,

588   *29*(4), 677-697.

589   Hogendoorn, H., & Burkitt, A. N. (2018). Predictive coding of visual object position

590   ahead of moving objects revealed by time-resolved EEG decoding.

591   NeuroImage, 171, 55-61.

592   Hogendoorn, H., & Burkitt, A. N. (2019). Predictive coding with neural transmission

593   delays: a real-time temporal alignment hypothesis. *eNeuro*, *6*(2).

594   Hohwy, J. (2013). The predictive mind. Oxford University Press.

595   Hopfield, J. J. (1994). Physics, computation, and why biology looks so

596   different. *Journal of Theoretical Biology*, *171*(1), 53-60.

Jeffreys, H. (1998). The theory of probability. OUP Oxford.

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. Frontiers in neuroinformatics, 10, 27.

Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. *Frontiers in human neuroscience*, *8*, 666.

Still, S., Sivak, D. A., Bell, A. J., & Crooks, G. E. (2012). Thermodynamics of prediction. *Physical review letters*, *109*(12), 120604.

Stone, J. V. (2016). Principles of Neural Information Theory A Tutorial Introduction.

Susskind, L. and Hrabovsky, G., 2014. *Classical mechanics: the theoretical minimum*. Penguin Books.

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science*, *314*(5803), 1311-1314.

King, J. R., Gramfort, A., Schurger, A., Naccache, L., & Dehaene, S. (2014). Two distinct dynamic modes subtend the detection of unexpected sounds. *PloS one*, *9*(1), e85791.

Kok, P., Mostert, P., & De Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences*, *114*(39), 10473-10478.

Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., & de Lange, F. P. (2011). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral cortex*, bhr310.

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center–periphery organization of human object areas. *Nature neuroscience*, *4*(5), 533.

622 Mirza, M. B., Adams, R. A., Friston, K., & Parr, T. (2019). Introducing a Bayesian

623   model of selective attention based on active inference. *Scientific reports*, *9*(1),

624   1-22.

625 Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval

626   null hypotheses. *Psychological methods*, *16*(4), 406.

627 Parr, T., Corcoran, A. W., Friston, K. J., & Hohwy, J. (2019). Perceptual awareness

628   and active inference. *Neuroscience of consciousness*, *2019*(1), niz012.

629 Parr, T., Da Costa, L., & Friston, K. (2020). Markov blankets, information geometry

630   and stochastic thermodynamics. *Philosophical Transactions of the Royal*

631   *Society A*, *378*(2164), 20190159.

632 Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a

633   functional interpretation of some extra-classical receptive-field effects. *Nature*

634   *neuroscience*, *2*(1), 79.

635 Richter, D., Ekman, M., & de Lange, F. P. (2018). Suppressed sensory response to

636   predictable object stimuli throughout the ventral visual stream. *Journal of*

637   *Neuroscience*, 3421-17.

638 Robinson, A. K., Grootswagers, T., & Carlson, T. A. (2019). The influence of image

639   masking on object representations during rapid serial visual

640   presentation. *NeuroImage*, *197*, 224-231.

641 Robinson, J. E., Woods, W., Leung, S., Kaufman, J., Breakspear, M., Young, A. W.,

642   & Johnston, P. J. (2018). Prediction-error signals to violated expectations

643   about person identity and head orientation are doubly-dissociated across the

644   dorsal and ventral visual streams. *bioRxiv*, 471276.

645   Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009).

646   Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic*

647   *Bulletin & Review*, *16*(2), 225–237.

648   Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human

649   brain. *Current opinion in neurobiology*, *4*(2), 157-165.

650   Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p

651   values. *Psychonomic Bulletin & Review*, *14*(5), 779–804.

652   https://doi.org/10.3758/BF03194105

653   Whyte, C. J., & Smith, R. (2020). The Predictive Global Neuronal Workspace: A

654   Formal Active Inference Model of Visual Consciousness. *bioRxiv*.