
Research Article: Theory/New Concepts | Sensory and Motor Systems

Predictive coding with neural transmission delays: a real-time temporal alignment hypothesis

Hinze Hogendoorn^{1,2} and Anthony N Burkitt³

¹Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, Australia

²Helmholtz Institute, Department of Experimental Psychology, Utrecht University, Utrecht, the Netherlands

³NeuroEngineering Laboratory, Department of Biomedical Engineering, The University of Melbourne, Melbourne, Australia

<https://doi.org/10.1523/ENEURO.0412-18.2019>

Received: 25 October 2018

Revised: 18 March 2019

Accepted: 20 March 2019

Published: 28 March 2019

H.H. and A.B. designed research; H.H. and A.B. performed research; H.H. and A.B. wrote the paper.

Funding: Australian Research Council (ARC)
DP180102268

Conflict of Interest: No. Authors report no conflict of interest.

HH was supported by the Australian Government through the Australian Research Council's Discovery Projects funding scheme (project DP180102268).

Correspondence should be addressed to Hinze Hogendoorn at hhogendoorn@unimelb.edu.au

Cite as: eNeuro 2019; 10.1523/ENEURO.0412-18.2019

Alerts: Sign up at www.eneuro.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Accepted manuscripts are peer-reviewed but have not been through the copyediting, formatting, or proofreading process.

Copyright © 2019 Hogendoorn and Burkitt

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

1 **1. Manuscript Title:**

2 Predictive coding with neural transmission delays: a real-time temporal alignment hypothesis

3

4 **2. Abbreviated Title:**

5 Predictive coding with neural transmission delays

6

7 **3. List all Author Names and Affiliations in order as they would appear in the published article :**

8 Hinze Hogendoorn^{1,2} & Anthony N Burkitt³

9 ¹ Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, Australia

10 ² Helmholtz Institute, Department of Experimental Psychology, Utrecht University, Utrecht, the
11 Netherlands

12 ³ NeuroEngineering Laboratory, Department of Biomedical Engineering, The University of
13 Melbourne, Melbourne, Australia

14

15 **4. Author Contributions: Each author must be identified with at least one of the following:**

16 **Designed research, Performed research, Contributed unpublished reagents/ analytic tools,**
17 **Analyzed data, Wrote the paper.**

18 HH and ANB Designed Research, Performed Research, and Wrote the paper.

19

20 **5. Correspondence should be addressed to (include email address)**

21 Hinze Hogendoorn (hhogendoorn@unimelb.edu.au)

22

23 **6. Number of Figures:**

24 2

25

26 **7. Number of Tables:**

27 0

28

29 **8. Number of Multimedia:**

30 0

31

32 **9. Number of words for Abstract**

33 153

34

35 **10. Number of words for Significance Statement**

36 115

37

38 **11. Number of words for Introduction**

39 439

40

41 **12. Number of words for Discussion:**

42 1309

43

44 **13. Acknowledgements**

45 HH was supported by the Australian Government through the Australian Research Council's
46 Discovery Projects funding scheme (project DP180102268).

47

48 **14. Conflict of Interest**

49 No. Authors report no conflict of interest

50

51 **15. Funding sources**

52 HH was supported by the Australian Government through the Australian Research Council's
53 Discovery Projects funding scheme (project DP180102268).
54

55 **Predictive coding with neural transmission delays: a real-time temporal alignment hypothesis**

56

57 **Abstract:**

58 Hierarchical predictive coding is an influential model of cortical organization, in which sequential
59 hierarchical levels are connected by backward connections carrying predictions, as well as forward
60 connections carrying prediction errors. To date, however, predictive coding models have largely
61 neglected to take into account that neural transmission itself takes time. For a time-varying stimulus,
62 such as a moving object, this means that backward predictions become misaligned with new sensory
63 input. We present an extended model implementing both forward and backward extrapolation
64 mechanisms that realigns backward predictions to minimize prediction error. This realignment has
65 the consequence that neural representations across all hierarchical levels become aligned in real-
66 time. Using visual motion as an example, we show that the model is neurally plausible, that it is
67 consistent with evidence of extrapolation mechanisms throughout the visual hierarchy, that it
68 predicts several known motion-position illusions in human observers, and that it provides a solution
69 to the temporal binding problem.

70

71 **Significance Statement:**

72 Despite the enormous scientific interest in predictive coding as model of cortical processing, most
73 models of predictive coding do not take into account that neural processing itself takes time. We
74 show that when the framework is extended to allow for neural delays, a model emerges that
75 provides a natural, parsimonious explanation for a wide range of experimental findings. It is
76 consistent with neurophysiological data from animals, predicts a range of well-known visual illusions
77 in humans, and provides a principled solution to the temporal binding problem. Altogether, it
78 explains how predictive coding mechanisms cause different brain areas to align in time despite their
79 different delays, and in so doing explains how cortical hierarchies functions in real-time.

80 **Introduction**

81 Predictive coding is a model of neural organisation that originates from a long history of proposals
82 that the brain infers, or predicts, the state of the world on the basis of sensory input (Dayan, Hinton,
83 Neal, & Zemel, 1995; Gregory, 1980; von Helmholtz, 1867). It has been particularly influential in the
84 domain of visual perception (Mumford, 1992; Rao & Ballard, 1999; Spratling, 2008, 2012; Srinivasan,
85 Laughlin, & Dubs, 1982), but has also been extensively applied in audition (see (Bendixen,
86 SanMiguel, & Schröger, 2012; Garrido, Kilner, Stephan, & Friston, 2009) for reviews), the
87 somatosensory system (van Ede, de Lange, Jensen, & Maris, 2011), motor control (Adams, Shipp, &
88 Friston, 2013; Blakemore, Goodbody, & Wolpert, 1998), and decision science (Schultz, 1998;
89 Summerfield & de Lange, 2014), where it accounts for a range of subtle response properties and
90 accords with physiology and neuroanatomy. Although it has been criticized for being insufficiently
91 articulated (Kogo & Trengove, 2015), it has been further developed into a general theory of cortical
92 organization (Bastos et al., 2012; Spratling, 2017) and even been advocated as a realisation of the
93 free energy principle - a principle that might apply to all self-organising biological systems (Friston,
94 2005, 2010, 2018; Friston & Kiebel, 2009).

95

96 An essential principle of predictive coding is a functional organization in which higher organizational
97 units “predict” the activation of lower units. Those lower units then compare their afferent input to
98 this backward prediction, and feed forward the difference: a prediction error (Rao & Ballard, 1999)
99 (where the term “prediction” is used in the strictly hierarchical sense, rather than the everyday
100 temporal sense of predicting the future). This interaction of backward predictions and forward
101 prediction errors characterizes each subsequent level of the processing hierarchy (Figure 1a).

102

103 In this review, we argue that neural transmission delays cause the classical model of hierarchical
104 predictive coding (Rao & Ballard, 1999) to break down when input to the hierarchy changes on a
105 timescale comparable with that of the neural processing. Using visual motion as an example, we

106 present two models that extend the classical model with extrapolation mechanisms to minimize
107 prediction error for time-varying sensory input. One of these models, in which extrapolation
108 mechanisms operate on both forward and back mechanisms, has as a consequence that all levels in
109 the hierarchy become aligned in real-time. We argue that this model not only minimizes prediction
110 error, but also parsimoniously explains a range of spatiotemporal phenomena, including motion-
111 induced position shifts such as the flash-lag effect and related illusions, and provides a natural
112 solution to the question of how asynchronous processing of visual features nevertheless leads to a
113 synchronous conscious experience (the temporal binding problem).

114

115 **Minimizing Prediction Error**

116 The core principle behind the pattern of functional connectivity in hierarchical predictive coding is
117 the minimization of total prediction error. This is considered to be the driving principle on multiple
118 biological time-scales (Friston, 2018):

119

120 At long time-scales, the minimization of prediction error drives evolution and phylogenesis. Neural
121 signaling is metabolically expensive, and there is therefore evolutionary pressure for an organism to
122 evolve a system of neural representation that allows for complex patterns of information to be
123 represented with minimal neural firing. A sparse, higher-order representation that inhibits the
124 costly, redundant activity of lower levels would provide a metabolic, and therefore evolutionary,
125 advantage. Theoretically, the imperative to minimise complexity and metabolic cost is an inherent
126 part of free energy (i.e., prediction error) minimisation. This follows because the free energy
127 provides a proxy for model evidence - and model evidence is accuracy minus complexity (Friston,
128 2010). This means that minimising free energy is effectively the same as self-evidencing (Hohwy,
129 2016) and both entail a minimisation of complexity or computational costs - and their
130 thermodynamic concomitants.

131

132 At the level of individual organisms, at time-scales relevant to decision-making and behaviour, the
133 minimization of prediction error drives learning (den Ouden, Kok, & de Lange, 2012; Friston, 2018).
134 Predictions are made on the basis of an internal model of the world, with the brain essentially using
135 sensory input to predict the underlying cause(s) of that input. The better this internal model fits the
136 world, the better the prediction that can be made, and the lower the prediction error. Minimizing
137 prediction error therefore drives a neural circuit to improve its representation of the world, in other
138 words: learning.

139

140 Finally, at sub-second time-scales relevant to sensory processing, the minimization of prediction
141 error drives the generation of stable perceptual representations. A given pattern of sensory input
142 feeds in to forward and backward mechanisms that iteratively project to each other until a dynamic
143 equilibrium is reached between higher-order predictions and (local) deviations from those
144 predictions. Because this equilibrium is the most efficient representation of the incoming sensory
145 input, the principle of prediction error minimization works to maintain this representation as long as
146 the stimulus is present. This results in a perceptual representation that remains stable over time.
147 Interestingly, because there may be local minima in the function determining total prediction error,
148 a given stimulus might have multiple stable interpretations, as is for example the case for ambiguous
149 stimuli such as the famous Necker cube (Necker, 1832).

150

151 **Hierarchical vs Temporal Prediction**

152 In discussing predictive coding, there is an important distinction to be made regarding the sense in
153 which predictive coding models predict.

154

155 Descriptively, predictive coding models are typically considered (either implicitly or explicitly) to
156 reflect some kind of expectation about the *future*. For example, in perception, pre-activation of the
157 neural representation of an expected sensory event ahead of that event's actual occurrence reflects

158 the nervous system predicting a *future* event (Garrido et al., 2009; Hogendoorn & Burkitt, 2018; Kok,
159 Mostert, & de Lange, 2017). In a decision-making context, a prediction might likewise be a belief
160 about the future consequences of a particular choice (Sterzer, Voss, Schlagenhaut, & Heinz, 2018).
161 These are predictions in the temporal sense of predicting future patterns of neural activity.

162

163 However, conventional models of predictive coding such as the one first proposed by Rao and
164 Ballard are not predictive in this same sense (Rao & Ballard, 1999). Rather than predicting the future,
165 these models are predictive in the *hierarchical* sense of higher areas predicting the activity of lower
166 areas (Bastos et al., 2012; Rao & Ballard, 1999; Spratling, 2012, 2017). These models do not predict
167 what *is going to happen*: rather, by converging on a configuration of neural activity that optimizes
168 the representation of a stable sensory input, they hierarchically predict *what is happening*. The use
169 of the word “prediction” in this context is therefore somewhat unfortunate, as conventional models
170 of predictive coding do not actually present a mechanism that predicts in the temporal way that
171 “prediction” is typically used in ordinary discourse.

172

173 **Predicting the Future**

174 To date, the temporal dimension has been nearly absent from computational work on predictive
175 coding. With the exception of generalized formulations (see below), it is generally implicitly assumed
176 that sensory input is a stationary process, i.e., that it remains unchanged until the system converges
177 on a minimal-prediction-error solution. The available studies of dynamic stimuli in a predictive
178 coding context consider only autocorrelations within a given neuron’s time-series – in other words,
179 the tendency of sensory input to remain the same from moment to moment (Huang & Rao, 2011;
180 Rao, 1999; van Hateren & Ruderman, 1998). Computationally, this prediction is easily implemented
181 as a biphasic temporal impulse response function (Dan, Atick, & Reid, 1996; Dong & Atick, 1995;
182 Huang & Rao, 2011), which is consistent with known properties of neurons in the early visual
183 pathway, including the lateral geniculate nucleus (LGN;(Dan et al., 1996; Dong & Atick, 1995)).

184 However, this is only the most minimal temporal prediction that a neural system might make: the
185 prediction that its input remains unchanged over time.

186

187 A more general framework is provided by generalised (Bayesian) filtering (Friston, Stephan, Li, &
188 Daunizeau, 2010). The key aspect of generalised predictive coding is that instead of just predicting
189 the current state of sensory impressions, there is an explicit representation of velocity, acceleration
190 and other higher orders of motion. This sort of representation has been used to model oculomotor
191 delays during active vision (Perrinet, Adams, & Friston, 2014). However, neither this framework nor
192 predictive coding approaches implicit in Kalman filters (that include a velocity-based prediction)
193 explicitly account for transmission delays.

194

195 As such, what is missing from conventional models of predictive coding is the fact that *neural*
196 *communication itself takes time*. Perhaps because the delays involved in *in silico* simulations are
197 negligible, or perhaps because the model was first articulated for stationary processes, namely static
198 images and stable neural representations, models of neural circuitry have entirely neglected to take
199 into account that neural transmission incurs significant delays. These delays mean that forward and
200 backward signals are misaligned in time. For an event at a given moment, the sensory
201 representation at the first level of the hierarchy needs to be fed forward to the next hierarchical
202 level, where a prediction is formulated, which is then fed back to the original level. In the case of a
203 dynamic stimulus, however, by the time that the prediction arrives back at the first hierarchical level,
204 the stimulus will have changed, and the first level will be representing the new state of that stimulus.
205 As a result, backward predictions will be compared against more recent sensory information than
206 the information on which they were originally based, and which they were sent to suppress. If this
207 temporal misalignment between forward and backward signals would not somehow be
208 compensated, under the classical hierarchical predictive coding model any time-varying stimulus
209 would generate very large prediction errors at each level of representation, which is typically not

210 seen in electrophysiological recordings of responses to stimuli with constant motion (as opposed to
211 unexpected changes in the trajectory, e.g., (Dick & Gray, 2014; McMillan & Gray, 2012; Schwartz,
212 Taylor, Fisher, Harris, & Berry, 2007)).

213

214 Due to neural transmission delays, prediction error is minimized not when a backward signal
215 represents the sensory information that originally generated it, but when it represents the sensory
216 information that *is going to be* available at the lower level by the time the backward signal arrives. In
217 other words, prediction error is minimized when the backward signal *anticipates* the future state of
218 the lower hierarchical level. For stimuli that are changing at a constant rate, estimating that future
219 state requires only rate-of-change information about the relevant feature, and it follows that if such
220 information is available at a given level, it will be recruited to minimize prediction error. When
221 allowing for transmission delays, hierarchical predictions therefore need to become temporal
222 predictions: they need to predict the future.

223

224 **Two Extended Models**

225 A clear example of common, time-varying sensory input is visual motion. Here, we use visual motion
226 to illustrate the limitations of the classical predictive coding model when input is time-varying, and
227 present two extensions to the classical model that would solve these limitations. In this illustration,
228 we consider neural populations at various level of the visual hierarchy that represent position, for
229 instance as a Gaussian population code. Additionally, we will argue that to effectively represent
230 position despite transmission delays, those neural populations will additionally represent velocity.
231 Consequently, each neural population representing a particular position would consist of
232 subpopulations representing the range of stimulus velocities.

233

234 **A. The Classical Predictive Model:** In the classical hierarchical predictive coding model (Rao &
235 Ballard, 1999) (Figure 1a), neural transmission delays mean that backward predictions arrive at a

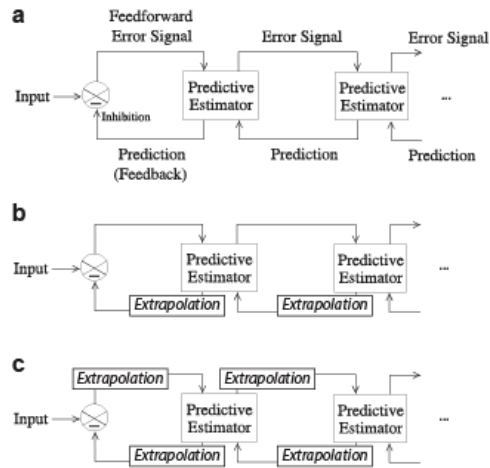
236 level in the processing pathway a significant interval of time after that level fed forward the signal
237 that led to those predictions. Because the input to this level has changed during the elapsed time,
238 this results in prediction error (Figure 2a).

239 **B. The Predictive Model with Extrapolated Feedback:** In this extension to the classical model, each
240 complete forward/backward loop uses any available rate-of-change information to minimize
241 prediction error. In the case of visual motion, this means the circuit will use concurrent velocity
242 information to anticipate the representation at the lower level *by the time the backward signal*
243 *arrives at that level*. In this model, an extrapolation mechanism is implemented only at the backward
244 step of each loop (Figure 1b) to minimize predictive error whilst leaving the forward sweep of
245 information unchanged (Figure 2b).

246

247 **C. The Predictive Model with Real-Time Alignment:** In the classical model, prediction error results
248 from the cumulative delay of both forward and backward transmission. Prediction error is minimized
249 when this cumulative delay is compensated at *any* point(s) in the forward-backward loop. Evidence
250 from both perception (Nijhawan, 1994, 2008) and action (Soechting, Juveli, & Rao, 2009; Zago,
251 McIntyre, Senot, & Lacquaniti, 2009) strongly suggests that at least part of the total delay incurred is
252 compensated by extrapolation at the forward step. Accordingly, in this model, we propose that
253 extrapolation mechanisms work on both forward and backward signals: *any* signal that is sent from
254 one level to another, whether forward or backward, is extrapolated into the future that precisely
255 compensates for the delay incurred by that signal while it is in transit (Figure 1c). In addition to
256 minimizing prediction error, this model has the remarkable consequence of synchronizing
257 representations throughout the hierarchy: under this model, all levels represent a moving object in
258 the same position at the same moment (Figure 2c), independent of where in the hierarchy each level
259 lies.

260



262

263 **Figure 1:** The classical hierarchical predictive coding model and two possible extensions. **[a]** The

264 **Classical Predictive Model.** This model consists of hierarchically organized loops of forward and

265 backward connections. Backward signals carry predictions, and forward signals carry prediction

266 errors. (Rao & Ballard, 1999) **[b]** The **Predictive Model with Extrapolated Feedback.** In order to

267 handle time-varying stimuli such as motion, the classical model can be expanded to include an

268 extrapolation mechanism on the backward step. This would be one way to minimize prediction error

269 for time-varying stimuli. **[c]** The **Predictive Model with Real-Time Alignment.** In this model,

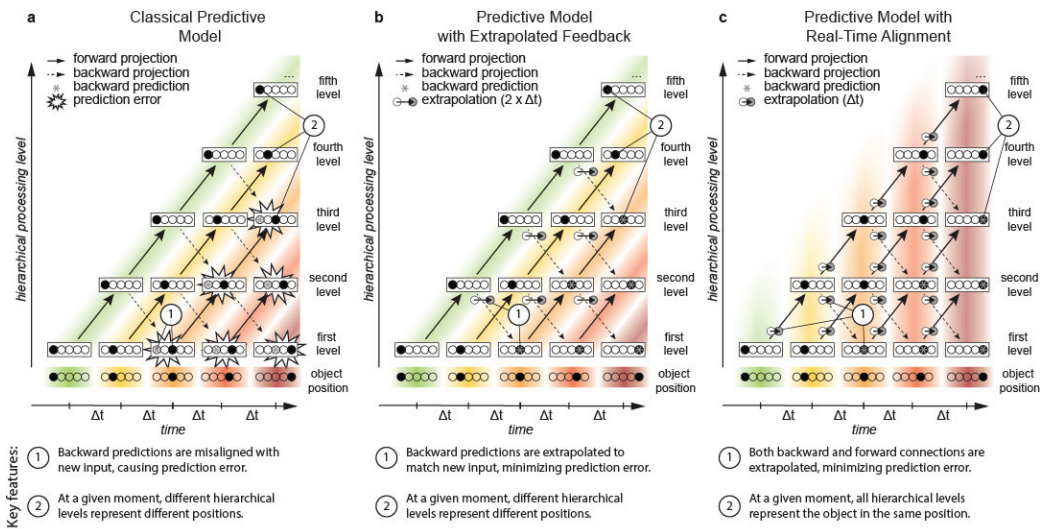
270 extrapolation mechanisms work on both forward and backward steps. Like Model B, this would

271 minimize prediction error, but has the additional consequence that it aligns the content of neural

272 representations across the hierarchy. Diagram labels in (b) and (c) are as in (a) but omitted for

273 clarity.

274



275
276

Figure 2: Three simplified models of the representation of the position of a moving object throughout

277 the visual processing hierarchy under the predictive coding framework. Each rectangle denotes the

278 neural representation of the object's position at a given hierarchical level and at a given time, with

279 the filled circle indicating the object in one of five possible positions. In this simplified representation,

280 all connections are modeled as incurring an equal transmission delay Δt . Coloured bands link

281 corresponding representations and numbered circles highlight core features of each model. **[a].** The

282 **Classical Predictive Model** of predictive coding comprises forward connections from one hierarchical

283 level to the next level (solid lines), and backward connections to the previous level (dashed lines). No

284 allowance is made for neural transmission delays, such that backward connections carry a position-

285 representation (asterisks) that is outdated by the time that signal arrives. The resulting mismatch

286 with more recent sensory input generates large errors, which subsequently propagate through the

287 hierarchy (emphasized with starbursts). **[b].** In the **Predictive Model with Extrapolated Feedback**, an

288 extrapolation mechanism operates on predictive backward projections, anticipating the future

289 position of the object. This mechanism on the backward projections compensates for the total time-

290 cost incurred during both the forward and the backward portion of the loop. This would minimize

291 total prediction error in this simplified model. However, the mechanism would rapidly become more

292 complex when one considers that individual areas tend to send and receive signals to and from

293 *multiple levels of the hierarchy. [c]. In the **Predictive Model with Real-Time Alignment**, extrapolation*
294 *mechanisms compensate for neural delays at both forward and backward steps. This parsimoniously*
295 *minimizes total prediction error, even for more complex connectivity patterns. Additionally, the*
296 *model differs from the first two models in that at any given time, all hierarchical levels represent the*
297 *same position. Conversely, in the first two models, at any given time neural transmission delays mean*
298 *that all hierarchical levels represent a different position. This crucial difference is evident as vertical,*
299 *rather than diagonal coloured bands linking matching representations across the hierarchy. The*
300 *consequence of this hypothesis is therefore that the entire visual hierarchy becomes temporally*
301 *aligned. This provides an automatic and elegant solution to the computational challenge of*
302 *establishing which neural signals belong together in time: the temporal binding problem. It is also*
303 *consistent with demonstrated extrapolation mechanisms in forward pathways and provides a*
304 *parsimonious explanation for a range of motion-induced position shifts.*

305

306 **Evaluating the evidence**

307 We have argued above that the classical model of predictive coding (Bastos et al., 2012; Huang &
308 Rao, 2011; Rao & Ballard, 1999; Spratling, 2017) will consistently produce prediction errors when
309 stimuli are time-varying and neural transmission delays are taken into consideration (Figure 2a). We
310 have proposed two possible extensions to the classical model, each of which would minimize
311 prediction error. Here, we evaluate the evidence for and against each of these two models.

312

313 *Neural Plausibility*

314 Models B and C are both neurally plausible. Firstly, both share key features with existing
315 computational models of motion-position interaction. For example, Kwon and colleagues recently
316 advanced an elegant computational model uniting motion and position perception. An interaction
317 between motion and position signals is a central premise of their model (Kwon, Tadin, & Knill, 2015),
318 such that instantaneous position judgments are biased by concurrent motion stimuli in a

319 qualitatively similar way to the model we propose here. In fact, they demonstrate that this
320 interaction predicts a number of properties of a visual effect known as the curveball illusion, in
321 which an object's internal motion causes its position (and trajectory) to be misperceived. Below, we
322 argue that our extension of hierarchical predictive coding similarly predicts a number of other
323 motion-induced position shifts. One such illusion is the flash-lag effect (Nijhawan, 1994), in which a
324 stationary bar is flashed in alignment with a moving bar, but is perceived to lag behind it.
325 Importantly, Khoei and colleagues recently presented a detailed argument that the flash-lag effect
326 could be explained by a Bayesian integration of motion and position signals, and postulated that
327 such mechanisms might compensate for neural delays (Khoei, Masson, & Perrinet, 2017). This
328 interpretation of the interaction between motion and position signals is entirely consistent with the
329 mechanisms we propose in Models B and C. Finally, we have previously argued that the prediction
330 errors that would necessarily arise under these models when objects unexpectedly change direction,
331 would lead to another perceptual effect known as the flash-grab illusion (Cavanagh & Anstis, 2013;
332 van Heusden, Harris, Garrido, & Hogendoorn, 2019, Figure 1). Consequently, the models we
333 propose are compatible with existing empirical and computational work.

334

335 Secondly, both extensions to the classical model incorporate neural extrapolation mechanisms at
336 each level of the visual hierarchy. This requires firstly that information about rate-of-change be
337 represented and available at each level. In the example of visual motion, it is well-established that
338 velocity is explicitly represented throughout the early visual hierarchy, including the retina (Barlow &
339 Levick, 1965), LGN (Cheong, Tailby, Solomon, & Martin, 2013; Cruz-Martín et al., 2014; Marshel,
340 Kaye, Nauhaus, & Callaway, 2012; Zaltsman, Heimel, & Van Hooser, 2015), and primary visual cortex
341 (Hubel & Wiesel, 1968). As required by both models, velocity information is therefore available
342 throughout the hierarchy. Indeed, it follows from the prediction error minimization principle that if
343 velocity is available at each level, it can and will be used at each level to optimize the accuracy of
344 backward predictions involving that level.

345 Finally, the extrapolation processes posited in both models should cause activity ahead of a moving
346 object's position, preactivating (i.e., priming the neural representation of) the area of space into
347 which the object is expected to move. This activation ahead of a moving object is consistent with
348 reported preactivation of predicted stimulus position in cat (Jancke, Erlhagen, Schöner, & Dinse,
349 2004) and macaque (Subramaniyan et al., 2018) V1, as well as in human EEG (Hogendoorn & Burkitt,
350 2018). Importantly, if the object unexpectedly vanishes, such extrapolation would pre-activate areas
351 of space in which the object ultimately never appeared. This is consistent with psychophysical
352 experiments where an object is perceived to move into areas of the retina where no actual stimulus
353 energy can be detected (Maus & Nijhawan, 2008; Shi & Nijhawan, 2012), as well as with recent fMRI
354 work showing activation in retinotopic areas of the visual field beyond the end of the motion
355 trajectory (Schellekens, van Wezel, Petridou, Ramsey, & Raemaekers, 2016). Although the models
356 presented here are by no means the only models that would be able to explain these results, these
357 results do demonstrate properties predicted by Models B and C.

358

359 *Hierarchical Complexity*

360 Model C is more robust to hierarchical complexity than Model B. Figure 2 shows extremely simplified
361 hierarchies: it omits connections that span multiple hierarchical levels, it represents all time-delays
362 as equal and constant, and it shows only a single projection from each area. In reality, of course, any
363 given area is connected to numerous other areas (Felleman & Van Essen, 1991), each with different
364 receptive field properties. Transmission delays will inevitably differ depending on where forward
365 signals are sent to and where backward signals originate from. Importantly, a given area might
366 receive input about the same moment through different pathways with different lags. A purely
367 backward extrapolation process (Model B) would not naturally compensate for these different lags,
368 because it would not be able to differentiate between the forward signals that led to the prediction.
369 Conversely, a model with extrapolation mechanisms at both forward and backward connections
370 (Model C) *would* be able to compensate for any degree of hierarchical complexity: for a feedback

371 loop with a large transmission delay, prediction error would be minimized by extrapolating further
372 along a trajectory than for a feedback loop with a smaller transmission delay. In this way, each
373 connection would essentially compensate for its own delay.

374

375 *Motion-induced Position Shifts*

376 Model C predicts motion-induced position shifts, namely visual illusions in which motion signals
377 affect the perceived positions of objects. In Model C, the representation in higher hierarchical levels
378 does *not* represent the properties of the stimulus as it was when it was initially detected, but rather
379 the *expected* properties of that stimulus at a certain extent into the future. In the case of visual
380 motion, this means that the forward signal represents the extrapolated, rather than detected
381 position of the moving object. One consequence is that at higher hierarchical levels, a moving object
382 is never represented in the position at which it first appears (Figure 2c: top diagonal). Instead, it is
383 first represented at a certain extent along its (expected) trajectory. Model C therefore neatly
384 predicts a visual phenomenon known as the Fröhlich effect, in which the initial position of an object
385 that suddenly appears in motion is perceived as being shifted in the direction of its motion (Fröhlich,
386 1924; Kirschfeld & Kammer, 1999). Conversely, Model B cannot explain these phenomena, as the
387 feed-forward position representation reflects the veridical location of the stimulus, rather than its
388 extrapolated position (Figure 2b).

389 Model C is also consistent with the much-studied flash-lag phenomenon (Nijhawan, 1994), in which
390 a briefly flashed stimulus that is presented in alignment with a moving stimulus appears to lag
391 behind that stimulus. Although alternative explanations have been proposed (Eagleman &
392 Sejnowski, 2000), a prominent interpretation of this effect is that it reflects motion extrapolation,
393 specifically implemented to compensate for neural delays (Nijhawan, 1994, 2002, 2008). Model C is
394 not only compatible with that interpretation, but it provides a principled argument (prediction error
395 minimization) for why such mechanisms might develop.

396

397 *Neural Computations*

398 There are two key computations that the neural processing hierarchy needs to carry out in order to
399 implement the proposed real-time alignment of Model C. First, it is necessary that the position
400 representation at each successive level incorporates the effect of the neural transmission delay to
401 compensate for the motion, rather than simply reproducing the position representation at the
402 preceding level. Second, this delay-and-motion-dependent shift of the position representation must
403 occur on both the forward and backward connections in the hierarchy. Moreover, these
404 computations must be learnt through a process that is plausible in terms of synaptic plasticity,
405 namely that the change in a synaptic strength is activity-dependent and local. The locality constraint
406 of synaptic plasticity requires that changes in the synaptic strength (i.e., the weight of a connection)
407 can only depend upon the activity of the presynaptic and the postsynaptic neuron. Consequently,
408 the spatial distribution of the synaptic changes in response to a stimulus are confined to the spatial
409 distribution of the position representation of the stimulus, which has important consequences for
410 the structure of the network that emerges as a result of learning.

411

412 By incorporating the velocity of the stimuli into the prediction error minimization principle, it
413 becomes possible for a learning process to selectively potentiate the synapses that lead to the
414 hierarchically organized position representations in both Models B and C. This requires that the
415 appropriate velocity sub-population of each position representation is activated to generate the
416 neural representations illustrated in Figures 2b & c. The prediction error is generated by the extent
417 of misalignment of the actual input and the predicted input position representation on the feedback
418 path, which involves the combined forward and backward paths, as illustrated in Figure 2.
419 Consequently, the prediction error minimization principle results in changes to the weights on both
420 the forward and backward paths. What distinguishes Models B and C is that the probability of
421 potentiation depends upon the extent of the spatial overlap of the position representations at
422 successive times. In Model C this spatial overlap between the forward and backward paths is the

423 same, namely the position representation has changed by the same extent during both the forward
424 and backward transmission delays, assuming that the neural delay time is the same for both paths.
425 However, in Model B the forward and backward paths are quite different: the forward path has
426 complete congruence in the position representation between adjacent levels, whereas the backward
427 path has a position representation that corresponds to the sum of the forward and backward delays,
428 so that the position representations between the two levels on the backward path are much further
429 apart. Since the position representations are local with a distribution that falls off from the center
430 (such as an exponential or power-law fall off), the probability of potentiation of the backward
431 pathway in Model B is correspondingly lower (by an exponential or power-law factor). As a result, a
432 local learning rule that implements the prediction error principle would tend to favor the more equal
433 distribution of delays between forward and backward paths of Model C. Since the same learning
434 principle applies to weights between each successive level of the hierarchical structure, it is capable
435 of providing the basis for the emergence of this structure during development.

436

437 Both Model B and C posit interactions between motion and position signals at multiple levels in the
438 hierarchy. This is compatible with a number of theoretical and computational models of motion and
439 position perception. For example, Eagleman and Sejnowski have argued that for a whole class of
440 motion-induced position shifts, a local motion signal biases perceived position (Eagleman &
441 Sejnowski, 2007) – precisely the local interactions between velocity and position signals that we
442 propose. Furthermore, these instantaneous velocity signals have been shown to affect not only the
443 perceptual localization of concurrently presented targets, but also the planning of saccadic eye
444 movements aimed at those targets (Quinet & Goffart, 2015; van Heusden, Rolfs, Cavanagh, &
445 Hogendoorn, 2018). We argue that, under the hierarchical predictive coding framework, neural
446 processing delays necessarily lead to the evolution of motion-position interactions.

447

448 Furthermore, both Models B and C posit such interactions at multiple levels, including very early
449 levels in the hierarchy. This is consistent with recent work on the flash-grab effect: a motion-induced
450 position shift in which a target briefly flashed on a moving background that reverses direction is
451 perceived shifted away from its true position (Cavanagh & Anstis, 2013). Using EEG, the interaction
452 between motion and position signals that generates the illusion has been shown to occur already
453 within the first 80 ms following stimulus presentation, indicating a very early locus of interaction. A
454 follow-up study using dichoptic presentation revealed that even within this narrow time-frame,
455 extrapolation took place at at least two dissociable processing levels (van Heusden et al., 2019).

456

457 *Temporal Alignment*

458 A defining feature of Model C is that, due to extrapolation at each forward step, all hierarchical areas
459 become aligned in time. Although neural transmission delays mean that it takes successively longer
460 for a new stimulus to be represented at successive levels of the hierarchy (Lamme, Supèr, &
461 Spkreijse, 1998), the fact that the neural signal is extrapolated into the future at each level means
462 that the *representational content* of each consecutive level now runs aligned with the first
463 hierarchical level, and potentially with the world, in real time. In the case of motion, we perceive
464 moving objects where they are, rather than where they were, because the visual system
465 extrapolates the position of those objects to compensate for the delays incurred during processing.
466 Of course, the proposal that the brain compensates neural delays by extrapolation is not new (see
467 (Nijhawan, 2008) for a review). Rather, what is new here is the recognition that this mechanism has
468 not developed for the *purpose* of compensating for delays at the behavioural level, but that it
469 follows necessarily from the fundamental principles of predictive coding.

470

471 The temporal alignment characterizing Model C also provides a natural solution to the problem of
472 temporal binding. Different visual features (such as contours, colour, motion, etc.) are processed in
473 different, specialized brain areas (Felleman & Van Essen, 1991; Livingstone & Hubel, 1988). Due to

474 anatomical and physiological differences, these areas process their information with different
475 latencies, which leads to asynchrony in visual processing. For example, the extraction of colour has
476 been argued to lead the extraction of motion (Arnold, Clifford, & Wenderoth, 2001; Moutoussis &
477 Zeki, 1997b; Zeki & Bartels, 1998). The question therefore arises how these asynchronously
478 processed features are nevertheless experienced as a coherent stream of visual awareness, with
479 components of that experience processed in different places at different times (Zeki & Bartels,
480 1999). Model C intrinsically solves this problem. Because the representation in each area is
481 extrapolated to compensate for the delay incurred in connecting to that area, representations across
482 areas (and therefore features) become aligned in time.

483

484 *Local Time Perception*

485 Model C eliminates the need for a central timing mechanism. Because temporal alignment in this
486 model is an automatic consequence of prediction error minimization at the level of local circuits, no
487 central timing mechanism (e.g., internal clock (Gibbon, 1977)) is required to carry out this alignment.
488 Indeed, under this model, if a temporal property of the prediction loop changed for a particular part
489 of the visual field, then this would be expected to result in *localized* changes in temporal alignment.
490 Although this is at odds with our intuitive experience of the unified passage of time, local adaptation
491 to flicker has been found to distort time perception in a specific location in the visual field
492 (Hogendoorn, Verstraten, & Johnston, 2010; Johnston, Arnold, & Nishida, 2006). Indeed, these
493 spatially localized distortions have been argued to result from adaptive changes to the temporal
494 impulse response function in the lateral geniculate nucleus LGN (Johnston, 2010, 2014), which would
495 disrupt the calibration of any neuron extrapolating a given amount of time into the future. By
496 pointing to adaptation in specific LGN populations, this model has emphasized the importance of
497 local circuitry in explaining the spatially localized nature of this temporal illusion. The
498 counterintuitive empirical finding that the perceived timing of events in specific positions in the

499 visual field can be distorted by local adaptation is therefore consistent with the local compensation
500 mechanisms that form part of Model C.

501

502 *Violated Predictions*

503 Under the assumption that the observer's percept corresponds to the current stable representation
504 of the stimulus at a given level of representation, models B and C produce stable percepts when
505 objects move on predictable trajectories. In Model B, the represented location at any given moment
506 will lag behind the outside world by an amount dependent on the accumulated delay in the
507 hierarchy, and in Model C the represented location will be synchronous with the outside world. In
508 both cases, however, the representation at the appropriate level, and hence the conscious percept,
509 will continue to stably evolve. In contrast, when events do not unfold predictably, such as when
510 objects unexpectedly appear, disappear, or change their trajectory, this introduces new forward
511 information which is, of course, at odds with backward predictions at each level. This gives rise to
512 prediction errors at each level of the hierarchy, with progressively larger prediction errors further up
513 the hierarchy (as it takes longer for sensory information signaling a violation of the status quo to
514 arrive at higher levels). During the intervening time, those areas will continue to (erroneously)
515 extrapolate into a future that never comes to pass.

516

517 The breakdown of prediction error minimization (and therefore accurate perception) in situations
518 where events unfold unpredictably fits with empirical studies showing visual illusions in these
519 situations. As noted above, the flash-lag effect is one such visual illusion. Under Model C, this effect
520 occurs because the position of the moving object can be extrapolated due to its predictability,
521 whereas the flashed object, which is unpredictable, cannot (Nijhawan, 1994). As the sensory
522 information pertaining to the position of the moving object ascends the visual hierarchy, it is
523 extrapolated at each level. Conversely, the representation of the flash is passed on 'as-is', such that a
524 mismatch accumulates between the two. This mismatch between the predictable motion and the

525 unpredictable flash has a parallel in the flash-grab effect, a visual illusion whereby a flash is
526 presented on a moving background that reverses direction (Cavanagh & Anstis, 2013). The result is
527 that the perceived position of the flash is shifted away from the first motion sequence (Blom,
528 Hogendoorn, Liang, & Hogendoorn, 2019) and in the direction of the second motion sequence.
529 Importantly, a recent study parametrically varied the predictability of the flash, and showed that the
530 strength of the resulting illusion decreased as the flash became more predictable (Adamian &
531 Cavanagh, 2016). Likewise, in studies of temporal binding, the asynchrony in the processing of colour
532 and motion is evident only when rapidly moving stimuli abruptly change direction (Moutoussis &
533 Zeki, 1997a). These illusions reveal that accurate perception breaks down when prediction is not
534 possible, consistent with Model C, but not with the classical predictive Model A.

535

536 **Conclusions and Future Directions**

537 Altogether, multiple lines of evidence converge in support of Model C: an extension of the
538 hierarchical predictive coding framework in which extrapolation mechanisms work on both forward
539 and backward connections to minimize prediction error. In this model, minimal prediction error is
540 achieved by local extrapolation mechanisms compensating the specific delays incurred by individual
541 connections at each level of the processing hierarchy. As a result, neural representations across the
542 hierarchy become aligned in real-time. This model provides an extension to classical predictive
543 coding models that is necessary to account for neural transmission delays. In addition, the model
544 predicts and explains a wide range of spatiotemporal phenomena, including motion-induced
545 position shifts, the temporal binding problem, and localized distortions of perceived time.

546

547 *Neural Implementation*

548 We have taken error minimization as the organizing principle of predictive coding, but now extended
549 to incorporate the local velocity of the stimuli. This is in keeping with the approach of previous
550 authors, who have successfully modeled hierarchical prediction error minimization (Bastos et al.,

551 2012; Friston, 2010; Huang & Rao, 2011; Rao & Ballard, 1997, 1999; Spratling, 2017) and motion-
552 position interactions (Kwon et al., 2015) as a Kalman filter (Kalman, 1960). The extension proposed
553 here can, in principle, be implemented straight forwardly by incorporating the local velocity as an
554 additional state variable in a manner analogous to that proposed for motion-position interactions
555 (Kwon et al., 2015). Such an approach, in which the prediction error minimization incorporates the
556 expected changes that occur due to both the motion of the stimulus and the propagation of the
557 neural signals, is possible when the local velocity is one of the state variables that each level of the
558 hierarchy has access to. This could be seen as a special case of generalised predictive coding, where
559 there is an explicit leveraging of (sometimes higher-order) motion representations (Friston et al.,
560 2010). This framework even goes beyond the first order velocity implicit in a Kalman filter to include
561 higher orders of motion. Mathematically, this makes it easy to extrapolate forwards and backwards
562 in time by simply taking linear mixtures of different higher order motion signals using Taylor
563 expansions (Perrinet et al., 2014). This would be one neurally plausible way in which the
564 extrapolation in Figure 2 could be implemented.

565

566 Furthermore, it is important to note that by minimizing prediction error, the system will
567 automatically self-organise as if it “knows” what neural transmission delays are incurred at each
568 step. However, Models B and C do not require these delays to be explicitly represented at all.
569 Rather, if position and velocity are co-represented as state variables, and the system is exposed to
570 time-variant input (such as a moving object), then selective Hebbian potentiation would suffice to
571 strengthen those combinations of connections that cause the backward projection to accurately
572 intercept the new sensory input. This would be one possible implementation of how extrapolation
573 mechanisms calibrate to rate-of-change signals.

574

575 It is important to note that the proposed model extends, rather than replaces, the conventional
576 formulation of predictive coding as first posited by Rao & Ballard (1999). We have not discussed

577 how this model would function in the spatial domain to develop different receptive field properties
578 at each level, as this has been discussed at length by other authors (Huang & Rao, 2011; Jehee,
579 Rothkopf, Beck, & Ballard, 2006; Rao, 1999; Rao & Ballard, 1999) and we intend our model to inherit
580 these characteristics. Our model only extends the conventional model by providing for the situation
581 when input is time-variant. When it is not (i.e., for static stimuli), our model reduces to the
582 conventional model.

583

584 Altogether, the details of how cortical circuits implement extrapolation processing, how those
585 circuits interact with receptive field properties at different levels, and what synaptic plasticity
586 mechanisms underlie the formation of these circuits still remain to be elaborated, and these are key
587 areas for future research.

588

589 *Prediction in the Retina*

590 The proposed model follows from the principle of error minimization within predictive feedback
591 loops. These feedback loops are ubiquitous throughout the visual pathway, including backward
592 connections from V1 to LGN. Although there are no backward connections back to the retina,
593 extrapolation mechanisms have nevertheless been reported in the retina (Berry, Brivanlou, Jordan,
594 & Meister, 1999), and these mechanisms have even been found to produce a specific response to
595 reversals of motion direction, much akin to a prediction error (Chen, Chou, Park, Schwartz, & Berry,
596 2014; Holy, 2007; Schwartz et al., 2007). Indeed, the retina has been argued to implement its own,
597 essentially self-contained predictive coding mechanisms that adaptively adjust spatiotemporal
598 receptive fields (Hosoya, Baccus, & Meister, 2005). In the absence of backward connections to the
599 retina, our model does not directly predict these mechanisms, instead predicting only extrapolation
600 mechanisms in the rest of the visual pathway where backwards connections are ubiquitous. On the
601 long time-scale, compensation for neural delays in the retina provides a behavioural, and therefore

602 evolutionary, advantage, but more research will be necessary to address whether any short-term
603 learning mechanisms play a role in the development of these circuits.

604

605 *The Ubiquity of Velocity Signals*

606 We have emphasized hierarchical mechanisms at early levels of visual processing, consistent with
607 extrapolation in monocular channels(van Heusden et al., 2019), early EEG correlates of
608 extrapolation(Hogendoorn, Verstraten, & Cavanagh, 2015), and evidence that extrapolation
609 mechanisms are shared for both perceptual localization and saccadic targeting (van Heusden et al.,
610 2018). However, an influential body of literature has proposed that the human visual system is
611 organized into two (partly) dissociable pathways: a ventral “what” pathway for object recognition
612 and identification, and a dorsal “where” pathway for localization and motion perception (Aglioti,
613 DeSouza, & Goodale, 1995; Goodale & Milner, 1992; Goodale, Milner, Jakobson, & Carey, 1991;
614 Mishkin & Ungerleider, 1982). Two decades later, the distinction is more nuanced (Gilaie-Dotan,
615 2016; Milner & Goodale, 2008), but the question remains whether velocity signals are truly
616 ubiquitous throughout the visual hierarchy. However, given that the identity of a moving object
617 typically changes much more slowly than its position, it may be that the neuronal machinery that
618 underwrites (generalised) predictive coding in the ventral stream does not have to accommodate
619 transmission delays. This may provide an interesting interpretation of the characteristic physiological
620 time constants associated with the magnocellular stream (directed towards the dorsal stream), as
621 compared to the parvocellular stream implicated in object recognition (Zeki & Shipp, 1988).
622 Consequently, whether real-time temporal alignment is restricted to the “where” pathway or is a
623 general feature of cortical processing remains to be elucidated.

624

625 *The Functional Role of Prediction Error*

626 Throughout this Perspective, we have considered prediction error as something that an effective
627 predictive coding model should minimize. From the perspective of the free energy principle, the

628 imperative to minimise prediction error can be argued from first principles. The sum of squared
629 prediction error is effectively variational free energy, and any system that minimises free energy will
630 therefore self-organise to its preferred physiological and perceptual states. However, we have not
631 addressed the functional role of the prediction error signal itself. As noted by several authors, this
632 signal might serve an alerting or surprise function (see (den Ouden et al., 2012) for a review). In the
633 model proposed here, the signal might have the additional function of correcting a faulty
634 extrapolation (Nijhawan, 2002, 2008; Shi & Nijhawan, 2012). In this role, prediction error signals
635 would work to eliminate lingering neural traces of predictions that were unsubstantiated by sensory
636 input: expected events that did not end up happening. This corrective function has been modelled as
637 a “catch-up” effect for trajectory reversals in the retina (Holy, 2007; Schwartz et al., 2007), and as an
638 increase in position uncertainty (and therefore an increase in relative reliance on sensory
639 information) when objects change trajectory in a recently proposed Bayesian model of visual motion
640 and position perception (Kwon et al., 2015). Further identifying the functional role of these signals is
641 an exciting avenue for future research.

642 **References**

- 643 Adamian, N., & Cavanagh, P. (2016). Localization of flash grab targets is improved with sustained
644 spatial attention. *Journal of Vision*, 16(12), 1266. <https://doi.org/10.1167/16.12.1266>
- 645 Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the
646 motor system. *Brain Structure and Function*. <https://doi.org/10.1007/s00429-012-0475-5>
- 647 Aglioti, S., DeSouza, J. F., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the
648 hand. *Current Biology : CB*, 5(6), 679–685. Retrieved from
649 <http://www.ncbi.nlm.nih.gov/pubmed/7552179>
- 650 Arnold, D. H., Clifford, C. W. ., & Wenderoth, P. (2001). Asynchronous processing in vision: Color
651 leads motion. *Current Biology*, 11(8), 596–600. [https://doi.org/10.1016/S0960-9822\(01\)00156-](https://doi.org/10.1016/S0960-9822(01)00156-7)
652 7
- 653 Barlow, H. B., & Levick, W. R. (1965). The mechanism of directionally selective units in rabbit's retina.
654 *The Journal of Physiology*, 178(3), 477–504. Retrieved from
655 <http://www.ncbi.nlm.nih.gov/pubmed/5827909>
- 656 Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012, November
657 21). Canonical Microcircuits for Predictive Coding. *Neuron*. NIH Public Access.
658 <https://doi.org/10.1016/j.neuron.2012.10.038>
- 659 Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive
660 processing in audition: A review. *International Journal of Psychophysiology*, 83(2), 120–131.
661 <https://doi.org/10.1016/j.ijpsycho.2011.08.003>
- 662 Berry, M. J., Brivanlou, I. H., Jordan, T. A., & Meister, M. (1999). Anticipation of moving stimuli by the
663 retina. *Nature*, 398(6725), 334–338. <https://doi.org/10.1038/18678>
- 664 Blakemore, S.-J., Goodbody, S. J., & Wolpert, D. M. (1998). Predicting the consequences of our own
665 actions: the role of sensorimotor context estimation. *The Journal of Neuroscience*, 18(18),
666 7511–7518. <https://doi.org/Not available>
- 667 Blom, T., Hogendoorn, H., Liang, Q., & Hogendoorn, H. (2019). When predictions fail: correction for

- 668 extrapolation in the flash-grab effect. *Journal of Vision*, *19*, 1–11.
- 669 Cavanagh, P., & Anstis, S. M. (2013). The flash grab effect. *Vision Research*, *91*, 8–20.
670 <https://doi.org/10.1016/j.visres.2013.07.007>
- 671 Chen, E. Y., Chou, J., Park, J., Schwartz, G., & Berry, M. J. (2014). The Neural Circuit Mechanisms
672 Underlying the Retinal Response to Motion Reversal. *Journal of Neuroscience*.
673 <https://doi.org/10.1523/JNEUROSCI.1460-13.2014>
- 674 Cheong, S. K., Tailby, C., Solomon, S. G., & Martin, P. R. (2013). Cortical-Like Receptive Fields in the
675 Lateral Geniculate Nucleus of Marmoset Monkeys. *Journal of Neuroscience*, *33*(16), 6864–6876.
676 <https://doi.org/10.1523/JNEUROSCI.5208-12.2013>
- 677 Cruz-Martín, A., El-Danaf, R. N., Osakada, F., Sriram, B., Dhande, O. S., Nguyen, P. L., ... Huberman, A.
678 D. (2014). A dedicated circuit linking direction selective retinal ganglion cells to primary visual
679 cortex. *Nature*, *507*(7492), 358. <https://doi.org/10.1038/NATURE12989>
- 680 Dan, Y., Atick, J. J., & Reid, R. C. (1996). Efficient Coding of Natural Scenes in the Lateral Geniculate
681 Nucleus: Experimental Test of a Computational Theory. *The Journal of Neuroscience*, *16*(10),
682 3351–3362. Retrieved from <http://www.jneurosci.org/content/jneuro/16/10/3351.full.pdf>
- 683 Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural*
684 *Computation*, *7*(5), 889–904. <https://doi.org/10.1162/neco.1995.7.5.889>
- 685 den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How Prediction Errors Shape Perception,
686 Attention, and Motivation. *Frontiers in Psychology*, *3*(December), 1–12.
687 <https://doi.org/10.3389/fpsyg.2012.00548>
- 688 Dick, P. C., & Gray, J. R. (2014). Spatiotemporal stimulus properties modulate responses to trajectory
689 changes in a locust looming-sensitive pathway. *Journal of Neurophysiology*, *111*(9), 1736–1745.
690 <https://doi.org/10.1152/jn.00499.2013>
- 691 Dong, D. W., & Atick, J. J. (1995). Temporal decorrelation: A theory of lagged and nonlagged
692 responses in the lateral geniculate nucleus. *Network: Computation in Neural Systems*, *6*(2),
693 159–178. https://doi.org/10.1088/0954-898X_6_2_003

- 694 Eagleman, D. M., & Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness.
695 *Science (New York, N.Y.)*, 287(5460), 2036–2038. Retrieved from
696 <http://www.ncbi.nlm.nih.gov/pubmed/10720334>
- 697 Eagleman, D. M., & Sejnowski, T. J. (2007). Motion signals bias localization judgments: A unified
698 explanation for the flash-lag, flash-drag, flash-jump, and Frohlich illusions. *Journal of Vision*,
699 7(4), 3–3. <https://doi.org/10.1167/7.4.3>
- 700 Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral
701 cortex. *Cerebral Cortex*, 1(1), 1–47. <https://doi.org/10.1093/cercor/1.1.1>
- 702 Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B:*
703 *Biological Sciences*, 360(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- 704 Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*,
705 11(2), 127–138. <https://doi.org/10.1038/nrn2787>
- 706 Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience*, 21(8), 1019–1021.
707 <https://doi.org/10.1038/s41593-018-0200-7>
- 708 Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical*
709 *Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1211–1221.
710 <https://doi.org/10.1098/rstb.2008.0300>
- 711 Friston, K., Stephan, K., Li, B., & Daunizeau, J. (2010). Generalised Filtering. *Mathematical Problems*
712 *in Engineering*, 2010. <https://doi.org/10.1155/2010/621670>
- 713 Fröhlich, F. W. (1924). Über die Messung der Empfindungszeit. *Pflügers Archiv Für Die Gesamte*
714 *Physiologie Des Menschen Und Der Tiere*, 202(1), 566–572.
715 <https://doi.org/10.1007/BF01723521>
- 716 Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review
717 of underlying mechanisms. *Clinical Neurophysiology : Official Journal of the International*
718 *Federation of Clinical Neurophysiology*, 120(3), 453–463.
719 <https://doi.org/10.1016/j.clinph.2008.11.029>

- 720 Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*.
721 <https://doi.org/10.1037/0033-295X.84.3.279>
- 722 Gilaie-Dotan, S. (2016). Visual motion serves but is not under the purview of the dorsal pathway.
723 *Neuropsychologia*, *89*, 378–392. <https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2016.07.018>
- 724 Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in*
725 *Neurosciences*, *15*(1), 20–25. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1374953>
- 726 Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation
727 between perceiving objects and grasping them. *Nature*, *349*(6305), 154–156.
728 <https://doi.org/10.1038/349154a0>
- 729 Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of*
730 *London. Series B, Biological Sciences*, *290*(1038), 181–197. Retrieved from
731 <http://www.ncbi.nlm.nih.gov/pubmed/6106237>
- 732 Hogendoorn, H., & Burkitt, A. N. (2018). Predictive coding of visual object position ahead of moving
733 objects revealed by time-resolved EEG decoding. *NeuroImage*, *171*.
734 <https://doi.org/10.1016/j.neuroimage.2017.12.063>
- 735 Hogendoorn, H., Verstraten, F. A. J., & Cavanagh, P. (2015). Strikingly rapid neural basis of motion-
736 induced position shifts revealed by high temporal-resolution EEG pattern classification. *Vision*
737 *Research*, *113*(PA). <https://doi.org/10.1016/j.visres.2015.05.005>
- 738 Hogendoorn, H., Verstraten, F. A. J., & Johnston, A. (2010). Spatially localized time shifts of the
739 perceptual stream. *Frontiers in Psychology*, *1*(NOV), 1–8.
740 <https://doi.org/10.3389/fpsyg.2010.00181>
- 741 Hohwy, J. (2016). The Self-Evidencing Brain. *Noûs*, *50*(2), 259–285.
742 <https://doi.org/10.1111/nous.12062>
- 743 Holy, T. E. (2007). A public confession: the retina trumpets its failed predictions. *Neuron*, *55*(6), 831–
744 832. <https://doi.org/10.1016/j.neuron.2007.09.002>
- 745 Hosoya, T., Baccus, S. A., & Meister, M. (2005). Dynamic predictive coding by the retina. *Nature*,

- 746 436(7047), 71–77. <https://doi.org/10.1038/nature03689>
- 747 Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive*
748 *Science*, 2(5), 580–593. <https://doi.org/10.1002/wcs.142>
- 749 Hubel, D. H., & Wiesel, T. (1968). Receptive Fields and Functional Architecture of monkey striate
750 cortex. *Journal of Physiology*, 195(1), 215–243. [https://doi.org/papers://47831562-1F78-4B52-](https://doi.org/papers://47831562-1F78-4B52-B52E-78BF7F97A700/Paper/p352)
751 [B52E-78BF7F97A700/Paper/p352](https://doi.org/papers://47831562-1F78-4B52-B52E-78BF7F97A700/Paper/p352)
- 752 Jancke, D., Erlhagen, W., Schöner, G., & Dinse, H. R. (2004). Shorter latencies for motion trajectories
753 than for flashes in population responses of cat primary visual cortex. *Journal of Physiology*,
754 556(3), 971–982. <https://doi.org/10.1113/jphysiol.2003.058941>
- 755 Jehee, J. F. M., Rothkopf, C., Beck, J. M., & Ballard, D. H. (2006). Learning receptive fields using
756 predictive feedback. *Journal of Physiology Paris*, 100(1–3), 125–132.
757 <https://doi.org/10.1016/j.jphysparis.2006.09.011>
- 758 Johnston, A. (2010). Modulation of time perception by visual adaptation. In *Attention and Time* (pp.
759 187–200). Oxford University Press.
760 <https://doi.org/10.1093/acprof:oso/9780199563456.003.0014>
- 761 Johnston, A. (2014). Visual Time Perception. In J. S. Werner & L. M. Chalupa (Eds.), *The new visual*
762 *neurosciences* (pp. 749–762). Cambridge, MA: MIT Press.
- 763 Johnston, A., Arnold, D. H., & Nishida, S. (2006). Spatially Localized Distortions of Event Time. *Current*
764 *Biology*, 16(5), 472–479. <https://doi.org/10.1016/j.cub.2006.01.032>
- 765 Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic*
766 *Engineering*, 82(1), 35. <https://doi.org/10.1115/1.3662552>
- 767 Khoei, M. A., Masson, G. S., & Perrinet, L. U. (2017). The Flash-Lag Effect as a Motion-Based
768 Predictive Shift. *PLOS Computational Biology*, 13(1), e1005068.
769 <https://doi.org/10.1371/journal.pcbi.1005068>
- 770 Kirschfeld, K., & Kammer, T. (1999). The Fröhlich effect: a consequence of the interaction of visual
771 focal attention and metacontrast. *Vision Research*, 39(22), 3702–3709. Retrieved from

- 772 <http://www.ncbi.nlm.nih.gov/pubmed/10746140>
- 773 Kogo, N., & Trengove, C. (2015). Is predictive coding theory articulated enough to be testable?
- 774 *Frontiers in Computational Neuroscience*, 9, 111. <https://doi.org/10.3389/fncom.2015.00111>
- 775 Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory
- 776 templates. *Proceedings of the National Academy of Sciences of the United States of America*,
- 777 114(39), 10473–10478. <https://doi.org/10.1073/pnas.1705652114>
- 778 Kwon, O.-S., Tadin, D., & Knill, D. C. (2015). Unifying account of visual motion and position
- 779 perception. *Proceedings of the National Academy of Sciences of the United States of America*,
- 780 112(26), 8142–8147. <https://doi.org/10.1073/pnas.1500361112>
- 781 Lamme, V. A. F., Supèr, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing
- 782 in the visual cortex. *Current Opinion in Neurobiology*, 8(4), 529–535.
- 783 [https://doi.org/10.1016/S0959-4388\(98\)80042-1](https://doi.org/10.1016/S0959-4388(98)80042-1)
- 784 Livingstone, M., & Hubel, D. H. (1988). Segregation of form, color, movement, and depth: Anatomy,
- 785 physiology, and perception. *Science*, 240(4853), 740–749.
- 786 <https://doi.org/10.1126/science.3283936>
- 787 Marshel, J. H., Kaye, A. P., Nauhaus, I., & Callaway, E. M. (2012). Anterior-posterior direction
- 788 opponency in the superficial mouse lateral geniculate nucleus. *Neuron*, 76(4), 713–720.
- 789 <https://doi.org/10.1016/j.neuron.2012.09.021>
- 790 Maus, G. W., & Nijhawan, R. (2008). Motion Extrapolation Into the Blind Spot. *Psychological Science*,
- 791 19(11), 1087–1091. <https://doi.org/10.1111/j.1467-9280.2008.02205.x>
- 792 McMillan, G. A., & Gray, J. R. (2012). A looming-sensitive pathway responds to changes in the
- 793 trajectory of object motion. *Journal of Neurophysiology*, 108(4), 1052–1068.
- 794 <https://doi.org/10.1152/jn.00847.2011>
- 795 Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, 46(3), 774–
- 796 785. <https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2007.10.005>
- 797 Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions

- 798 of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research*, 6(1), 57–77. Retrieved
799 from <http://www.ncbi.nlm.nih.gov/pubmed/7126325>
- 800 Moutoussis, K., & Zeki, S. (1997a). A direct demonstration of perceptual asynchrony in vision.
801 *Proceedings. Biological Sciences*, 264(1380), 393–399. <https://doi.org/10.1098/rspb.1997.0056>
- 802 Moutoussis, K., & Zeki, S. (1997b). Functional segregation and temporal hierarchy of the visual
803 perceptive systems. *Proceedings. Biological Sciences*, 264(1387), 1407–1414.
804 <https://doi.org/10.1098/rspb.1997.0196>
- 805 Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*,
806 66(3), 241–251. <https://doi.org/10.1007/BF00198477>
- 807 Necker, L. A. (1832). LXI. *Observations on some remarkable optical phænomena seen in Switzerland;*
808 *and on an optical phænomenon which occurs on viewing a figure of a crystal or geometrical*
809 *solid. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1(5),
810 329–337. <https://doi.org/10.1080/14786443208647909>
- 811 Nijhawan, R. (1994, July 28). Motion extrapolation in catching. *Nature*. Nature Publishing Group.
812 <https://doi.org/10.1038/370256b0>
- 813 Nijhawan, R. (2002). Neural delays, visual motion and the flash-lag effect. *Trends in Cognitive*
814 *Sciences*, 6(9), 387. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12200181>
- 815 Nijhawan, R. (2008). Visual prediction: Psychophysics and neurophysiology of compensation for time
816 delays. *Behavioral and Brain Sciences*, 31(2), 179–239.
817 <https://doi.org/10.1017/S0140525X08003804>
- 818 Perrinet, L. U., Adams, R. A., & Friston, K. J. (2014). Active inference, eye movements and
819 oculomotor delays. *Biological Cybernetics*, 108(6), 777–801. [https://doi.org/10.1007/s00422-](https://doi.org/10.1007/s00422-014-0620-8)
820 014-0620-8
- 821 Quinet, J., & Goffart, L. (2015). Does the Brain Extrapolate the Position of a Transient Moving Target?
822 *Journal of Neuroscience*, 35(34), 11780–11790. [https://doi.org/10.1523/JNEUROSCI.1212-](https://doi.org/10.1523/JNEUROSCI.1212-15.2015)
823 15.2015

- 824 Rao, R. P. N. (1999). An optimal estimation approach to visual perception and learning. *Vision*
825 *Research*, 39, 1963–1989. Retrieved from <https://homes.cs.washington.edu/~rao/vr99.pdf>
- 826 Rao, R. P. N., & Ballard, D. H. (1997). Dynamic Model of Visual Recognition Predicts Neural Response
827 Properties in the Visual Cortex. *Neural Computation*, 9(4), 721–763.
828 <https://doi.org/10.1162/neco.1997.9.4.721>
- 829 Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional
830 interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87.
831 <https://doi.org/10.1038/4580>
- 832 Schellekens, W., van Wezel, R. J. A., Petridou, N., Ramsey, N. F., & Raemaekers, M. (2016). Predictive
833 coding for motion stimuli in human early visual cortex. *Brain Structure and Function*, 221(2),
834 879–890. <https://doi.org/10.1007/s00429-014-0942-2>
- 835 Schultz, W. (1998). Predictive Reward Signal of Dopamine Neurons. *Journal of Neurophysiology*,
836 80(1), 1–27. <https://doi.org/10.1152/jn.1998.80.1.1>
- 837 Schwartz, G., Taylor, S., Fisher, C., Harris, R., & Berry, M. J. (2007). Synchronized Firing among Retinal
838 Ganglion Cells Signals Motion Reversal. *Neuron*, 55(6), 958–969.
839 <https://doi.org/10.1016/j.neuron.2007.07.042>
- 840 Shi, Z., & Nijhawan, R. (2012). Motion extrapolation in the central fovea. *PLoS ONE*, 7(3), 33651.
841 <https://doi.org/10.1371/journal.pone.0033651>
- 842 Soechting, J. F., Juveli, J. Z., & Rao, H. M. (2009). Models for the extrapolation of target motion for
843 manual interception. *Journal of Neurophysiology*, 102(3), 1491–1502.
844 <https://doi.org/10.1152/jn.00398.2009>
- 845 Spratling, M. W. (2008). Reconciling Predictive Coding and Biased Competition Models of Cortical
846 Function. *Frontiers in Computational Neuroscience*, 2, 4.
847 <https://doi.org/10.3389/neuro.10.004.2008>
- 848 Spratling, M. W. (2012). Predictive coding accounts for V1 response properties recorded using
849 reverse correlation. *Biological Cybernetics*, 106(1), 37–49. <https://doi.org/10.1007/s00422-012->

- 850 0477-7
- 851 Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, *112*, 92–97.
852 <https://doi.org/10.1016/j.bandc.2015.11.003>
- 853 Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the
854 retina. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *216*(1205),
855 427–459. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/6129637>
- 856 Sterzer, P., Voss, M., Schlagenhaut, F., & Heinz, A. (2018). Decision-making in schizophrenia: A
857 predictive-coding perspective. *NeuroImage*. <https://doi.org/10.1016/j.neuroimage.2018.05.074>
- 858 Subramaniyan, M., Ecker, A. S., Patel, S. S., Cotton, R. J., Bethge, M., Pitkow, X., ... Tolias, A. S. (2018).
859 Faster processing of moving compared with flashed bars in awake macaque V1 provides a
860 neural correlate of the flash lag illusion. *Journal of Neurophysiology*, *120*(5), 2430–2452.
861 <https://doi.org/10.1152/jn.00792.2017>
- 862 Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: neural and
863 computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756.
864 <https://doi.org/10.1038/nrn3838>
- 865 van Ede, F., de Lange, F., Jensen, O., & Maris, E. (2011). Orienting Attention to an Upcoming Tactile
866 Event Involves a Spatially and Temporally Specific Modulation of Sensorimotor Alpha- and
867 Beta-Band Oscillations. *Journal of Neuroscience*, *31*(6), 2016–2024.
868 <https://doi.org/10.1523/jneurosci.5630-10.2011>
- 869 van Hateren, J. H., & Ruderman, D. L. (1998). Independent component analysis of natural image
870 sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proc. R.*
871 *Soc. Lond. B*, *265*(1412), 2315–2320. <https://doi.org/10.1098/rspb.1998.0577>
- 872 van Heusden, E., Harris, A. M., Garrido, M. I., & Hogendoorn, H. (2019). Predictive coding of visual
873 motion in both monocular and binocular human visual processing. *Journal of Vision*, *19*(1), 3.
874 <https://doi.org/10.1167/19.1.3>
- 875 van Heusden, E., Rolfs, M., Cavanagh, P., & Hogendoorn, H. (2018). Motion Extrapolation for Eye

- 876 Movements Predicts Perceived Motion-Induced Position Shifts. *The Journal of Neuroscience*,
877 38(38), 8243–8250. <https://doi.org/10.1523/JNEUROSCI.0736-18.2018>
- 878 von Helmholtz, H. (1867). *Handbuch der physiologischen Optik*. Leipzig: Voss.
- 879 Zago, M., McIntyre, J., Senot, P., & Lacquaniti, F. (2009). Visuo-motor coordination and internal
880 models for object interception. *Experimental Brain Research*, 192(4), 571–604.
881 <https://doi.org/10.1007/s00221-008-1691-3>
- 882 Zaltsman, J. B., Heimel, J. A., & Van Hooser, S. D. (2015). Weak orientation and direction selectivity in
883 lateral geniculate nucleus representing central vision in the gray squirrel *Sciurus carolinensis*.
884 *Journal of Neurophysiology*, 113(7), 2987–2997. <https://doi.org/10.1152/jn.00516.2014>
- 885 Zeki, S., & Bartels, A. (1998). The asynchrony of consciousness. *Proceedings. Biological Sciences*,
886 265(1405), 1583–1585. <https://doi.org/10.1098/rspb.1998.0475>
- 887 Zeki, S., & Bartels, A. (1999). Toward a Theory of Visual Consciousness. *Consciousness and Cognition*,
888 8(2), 225–259. <https://doi.org/10.1006/ccog.1999.0390>
- 889 Zeki, S., & Shipp, S. (1988). The functional logic of cortical connections. *Nature*, 335(6188), 311–317.
890 <https://doi.org/10.1038/335311a0>
- 891