Brain Research, doi: 10.1016/j.brainres.2019.146313

Fitting Predictive Coding to the Neurophysiological Data

M. W. Spratling
King's College London, Department of Informatics, London. UK. michael.spratling@kcl.ac.uk

Abstract

Recent neurophysiological data showing the effects of locomotion on neural activity in mouse primary visual cortex has been interpreted as providing strong support for the predictive coding account of cortical function. Specifically, this work has been interpreted as providing direct evidence that prediction-error, a distinguishing property of predictive coding, is encoded in cortex. This article evaluates these claims and highlights some of the discrepancies between the proposed predictive coding model and the neuro-biology. Furthermore, it is shown that the model can be modified so as to fit the empirical data more successfully.

Keywords: predictive processing; primary visual cortex; mismatch neurons; locomotion; sensory integration

1 Introduction

The idea that the brain is engaged in perceptual inference, combining bottom-up evidence with prior knowledge to infer the most likely causes of sensory stimuli, has a long history dating back at least as far as the work of Helmholtz in the 1860s (Clark, 2013). This general idea has inspired more specific theories about how such inference is performed in the brain, including the Bayesian brain hypothesis, analysis-by-synthesis, predictive coding, and the free-energy principle (Barlow, 1994; Chater et al., 2006; Clark, 2013; Friston, 2009, 2005; Griffiths et al., 2008; Griffiths and Tenenbaum, 2006; Kersten et al., 2004; Knill and Richards, 1996; Lee and Mumford, 2003; Ma, 2012; Ma and Jazayeri, 2014; Mumford, 1992; Pouget et al., 2013; Rao and Ballard, 1999; Spratling, 2017b; Vilares and Kording, 2011; Yuille and Kersten, 2006). Largely inspired by the computational model proposed by Rao and Ballard (1999, Figure 1a), which was itself informed by the computational interpretation of cortical structure proposed by Mumford (1992), predictive coding has emerged as a highly influential, neurally-plausible, framework for understanding perceptual inference and brain function in general (Bubic et al., 2010; Clark, 2013; Huang and Rao, 2011; Keller and Mrsic-Flogel, 2018; Kok and de Lange, 2015).

Many neurophysiological experiments designed to test the predictive coding model have found evidence consistent with it (*e.g.*, Alink et al., 2010; Edwards et al., 2017; Egner et al., 2010; Fang et al., 2008; Heilbron and Chait, 2018; Kok et al., 2013; Kok and de Lange, 2015; Kok et al., 2017, 2012; Kok and Turk-Browne, 2018; Murray et al., 2002, 2004; Schellekens et al., 2016; Summerfield and Egner, 2009; Summerfield and Koechlin, 2008; Summerfield et al., 2008; Wacongne et al., 2011). However, possibly the most compelling evidence that predictive coding is performed in cortical circuits comes from two-photon imaging of mouse primary visual cortex (V1) neurons labelled with a genetically encoded calcium indicator. These experiments (reviewed in Fridman and Petreanu, 2017; Keller and Mrsic-Flogel, 2018; Khan and Hofer, 2018; Pakan et al., 2018), found a sub-population of neurons (primarily located in the superficial layers of V1) that were driven by connections from motor regions (M24b and M2) (Leinweber et al., 2017) that conveyed a prediction of optic flow speed based on the animal's running speed (Keller et al., 2012; Zmarz and Keller, 2016). The activity of these V1 neurons was suppressed by optic flow observed visually. Hence, these neurons signalled the mismatch between the predicted and the actual optic flow (Keller et al., 2012; Zmarz and Keller, 2016). Furthermore, these mismatch neurons had localised receptive fields (RFs), so that each signalled the discrepancy in optic flow for a specific region of visual space (Zmarz and Keller, 2016). These localised receptive fields were retinotopically organised such that the mismatch RFs were aligned with the RFs of visually-driven neurons, suggesting that mismatch was calculated in a local circuit (Keller and Mrsic-Flogel, 2018; Zmarz and Keller, 2016).

These experimental results have been interpreted as evidence for predictive coding (Attinger et al., 2017; Busse, 2018; Fridman and Petreanu, 2017; Keller et al., 2012; Keller and Mrsic-Flogel, 2018; Khan and Hofer, 2018; Leinweber et al., 2017; Pakan et al., 2018; Zmarz and Keller, 2016), and particularly, as being consistent with the predictive coding model proposed by Rao and Ballard (1999). Specifically, it is proposed that the mismatch neurons are calculating the error between a prediction of optic flow based on the animal's running speed and the actual optic flow (see Figure 2a). The calculation of prediction-error is a key process in predictive coding, and one which most clearly distinguishes it from other theories of cortical function. Hence, direct evidence for neurons encoding such errors is a significant step in validating the theory of predictive coding.

This article evaluates the proposed interpretation of the experimental results in terms of predictive coding, and identifies several shortcomings in this explanation (Section 2.1). In addition it describes how the predictive coding model can be modified to overcome these issues and to provide a better fit between the neurophysiology and the model (Section 2.2).

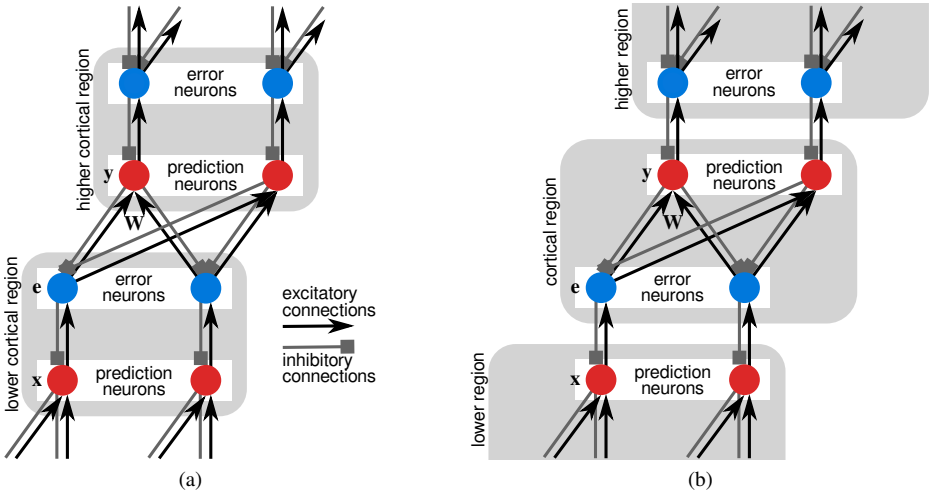

Figure 1: Hierarchical predictive coding models. (a) The model proposed by Rao and Ballard (1999). Each cortical region contains two distinct populations of neurons. To avoid over-crowding the figure only two neurons are shown in each population. The activations of the prediction neurons represent beliefs about the causes underlying the activity of prediction neurons in the preceding cortical region (and hence, indirectly about the causes of the sensory-driven input to cortex). Via the inter-regional feedback connections, the prediction neurons reconstruct the activity pattern of the preceding population of prediction neurons that would be expected given the current estimate of the causes. The activations of the error neurons represent discrepancies between the predicted and actual prediction neuron responses in the earlier cortical region. These errors are sent, via the inter-regional feedforward connections, to update the activations of the subsequent population of prediction neurons, so as to improve the estimate of the causes, and hence, to reduce the errors at the next iteration. Connections from error to prediction neurons within each cortical region allow the discrepancies in the higher cortical region's predictions to influence the predictions made by the lower cortical region. Mathematical equations defining the behaviour of this model are given in Section 2.1.4 and the labels $\mathbf{x}$, $\mathbf{e}$, $\mathbf{y}$, and $\mathbf{W}$ refer to variables in those equations. (b) The same hierarchical predictive coding model, but with a shifted allocation of neural populations to cortical regions (Spratling, 2008b).

2 Results

2.1 Evaluation of the fit between the neurophysiological data and the model

2.1.1 Non-negative firing rates

The behaviour of the V1 mismatch neurons appears to be the opposite of what would be expected for the error neurons in the predictive coding model. Error neurons are portrayed as calculating the difference between the sensory-driven, in this case visually-driven, signal (V) and the top-down prediction (P), *i.e.*, as calculating V-P. Whereas the response properties of the mismatch neurons are more consistent with the calculation of P-V. However, prediction-error (when calculated using subtraction) is a signed quantity: it can take negative as well as positive values (Keller and Mrsic-Flogel, 2018; Rao and Ballard, 1999). For this signed quantity to be encoded in biological neurons it is proposed that there are two types of mismatch neuron: ones that signal positive prediction-errors (also called type I), and those that encode negative prediction-errors (type II) (Attinger et al., 2017; Keller and Mrsic-Flogel, 2018; Rao and Ballard, 1999). It is the latter type of error neuron that corresponds to the mismatch neurons that have been characterised by the neurophysiological results. The proposal that there are both positive and negative error neurons addresses an underlying biological-implausibility with the original model implemented by Rao and Ballard (1999): that error neurons in the model be able to produce both positive and negative firing rates[a].

[a] An alternative solution to this problem would be to use a single population of error neurons to encode the exponential (*i.e.*, the inverse log transform) of the signed prediction errors, or similarly, to calculate the error using division rather than subtraction (Spratling, 2008a; Spratling et al., 2009). When prediction errors are encoded as signed quantities, finding (via gradient descent) the prediction neuron activations that minimise the (sum of squared) prediction errors requires the prediction neurons to be influenced in an additive way by prediction errors (Achler, 2014; Harpur, 1997; Rao and Ballard, 1999). In contrast, when prediction error neurons encode the exponential of the subtractive error, or encode the divisive error, then finding the prediction neuron activations that minimise the error requires the effect of prediction errors be multiplicative (Achler, 2014; Spratling et al., 2009).

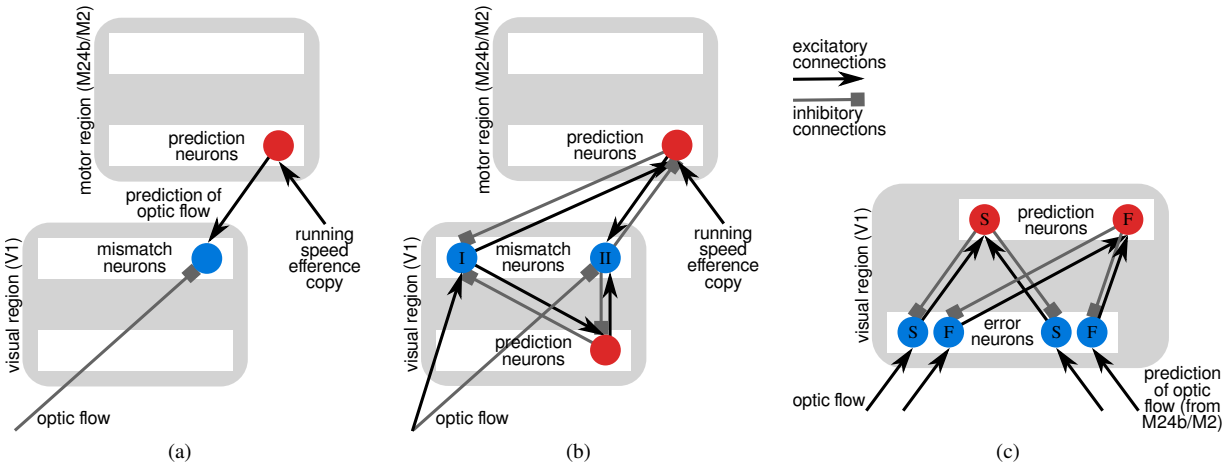

Figure 2: Interpretations of the neurophysiological data in terms of predictive coding. Note that to avoid overcrowding inhibitory inter-neurons have been left out of the figures. (a) The minimal neural circuitry previously proposed to underlie the experimentally observed results (Zmarz and Keller, 2016). Motor regions generate a prediction of the optic flow expected due to locomotion. This prediction is compared to the actual optic flow to determine a prediction-error which is encoded by the mismatch neurons in V1. (b) An extended model to improve the fit between the neurophysiological data and predictive coding, as proposed by Keller and Mrsic-Flogel (2018). The proposed neural circuitry is expanded to include two distinct sub-types of mismatch neurons (types I and II) and to include intra- and inter-regional connections from the mismatch neurons to prediction neurons in both the local region and the motor region (see Section 2.1.1). (c) The neural circuitry proposed in this article. Only one cortical region is shown, as this is sufficient to account for the neurophysiological findings, however, this model could be extended into a hierarchical one by adding connection to and from the prediction neurons to neurons in other regions. The labels S and F indicate neurons tuned to slow and fast optic flow (see Section 2.2.1).

As well as addressing the issue of negative firing rates, Keller and Mrsic-Flogel (2018) also propose that there should be inter-regional connections from the mismatch neurons in V1 to the prediction neurons in the motor regions, and intra-regional connections from the mismatch neurons to prediction neurons in V1. These connections should be excitatory for type I, and inhibitory for type II, mismatch neurons. Such connections are necessary as in a predictive coding model errors are calculated in order to update the predictions, and hence, improve the match between the predicted causes and the true causes of the sensory input. Consequently, the circuitry of the predictive coding model proposed by Keller and Mrsic-Flogel (2018) to explain the neurophysiological data is as shown in Figure 2b.

2.1.2 Fit to mismatch neuron data

Plotting the response (measured using the change in fluorescence, $\frac{\Delta F}{F}$) of the V1 mismatch neurons to different combinations of running speed and visual flow produced the results shown in Figure 3a (Zmarz and Keller, 2016). However, if the mismatch neurons are calculating the non-negative (*i.e.*, half wave rectified) difference between the prediction (P) and the visual signal (V), then the expected firing rate would look more like that shown in Figure 3b (if cortex measures optic flow and running speed on a linear scale), or Figure 3c (if speed is measured on a logarithmic scale). In either case, in order for these neurons to signal negative prediction-errors (calculated using subtraction, see Section 2.1.1) the response should be zero when the visually-driven estimate of optic flow equals or exceeds the prediction based on running speed. Clearly this is not the case for the V1 mismatch neurons, which have a non-zero response when the two estimates of speed are similar.

Calcium-sensitive fluorescence signals provide only an indirect measure of neural spiking activity, and there are thus many potential explanations for this discrepancy. For example, temporal smearing of the optical signal caused, for instance by the timescale of calcium decay, could account for the non-zero values measured when the visual optic flow matches the running speed (Georg Keller, personal communication, 2019). Zmarz and Keller (2016) found that by applying a sigmoid function to the difference between running speed and visual flow speed they could fit the firing rates expected by the predictive coding model to the change in fluorescence measured in cortex. However, it remains possible that there is a more linear relationship between $\frac{\Delta F}{F}$ and neural firing rate. If that were the case, then the equations of the predictive coding model would need to be altered to cope with

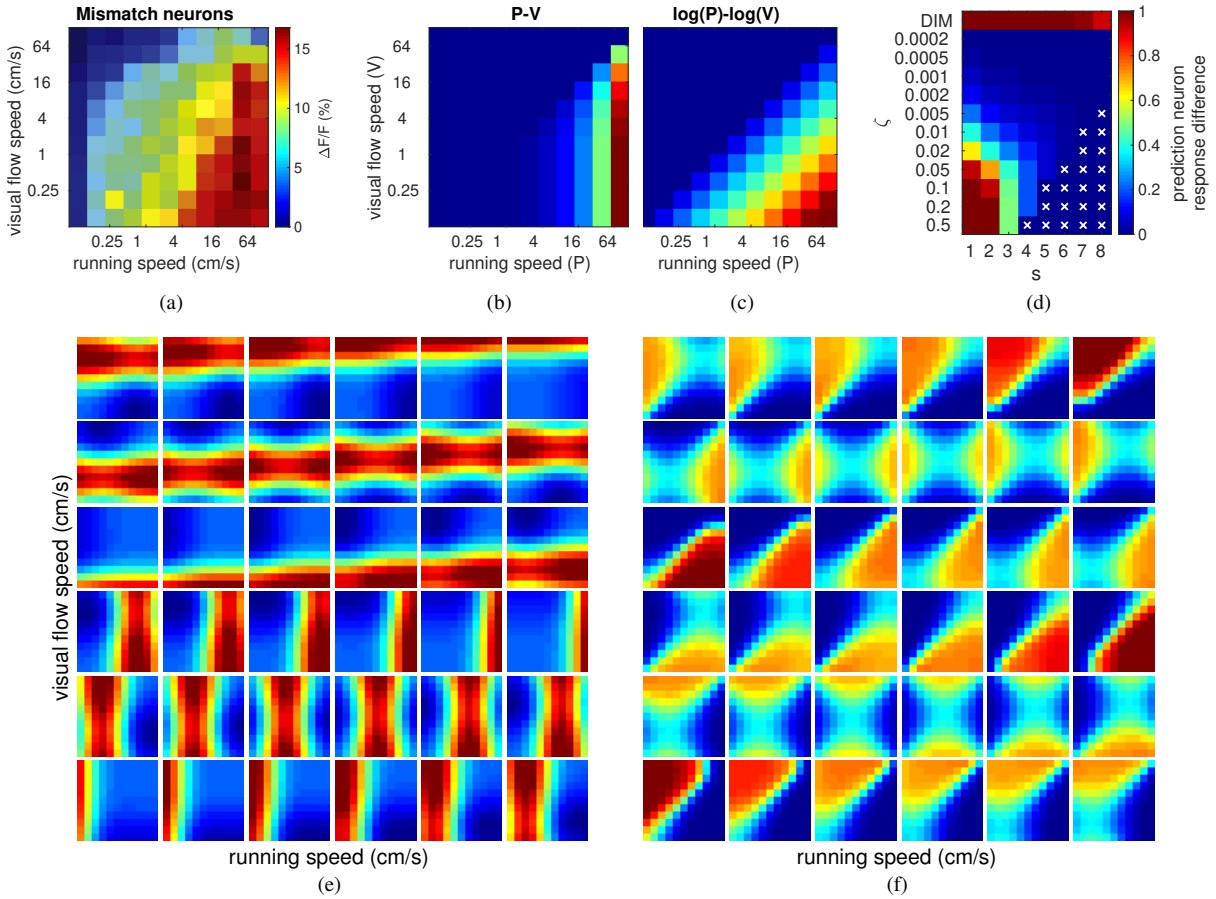

Figure 3: Neurophysiological and simulation results. (a) Population average of mismatch neuron responses recorded from mouse V1 (adapted from Zmarz and Keller, 2016, Fig. 4c). (b & c) The response expected if the mismatch neurons calculate (b) the difference between the prediction (P) and the visual signal (V), or (c) the difference between the logarithms of these values. (d) The difference between the response of the prediction neuron representing the cause of the input to a predictive coding network and the maximum response of all other prediction neurons that represent alternative causes. Results are shown for the Rao and Ballard (1999) implementation of predictive coding (Equation 1) with different values of the parameter, ζ, and for the PC/BC-DIM implementation of predictive coding as described in Section 2.2. In both cases, results are shown for different scales, s, of a synthetic task. White crosses mark conditions which caused the network to be unstable. (e & f) The responses of error neurons in the model proposed in this article (see Figure 2c) to different combinations of running speed and visual flow when the model is implemented using the equations proposed by (e) Rao and Ballard (1999), (f) the PC/BC-DIM algorithm.

mismatch neurons not signalling negative (and positive) prediction-errors but a function of these errors.

2.1.3 Fit to multisensory-integration data

Using a very similar experimental set-up, Saleem et al. (2013) measured the firing-rates of single cells in mouse V1 using multisite electrodes. These experiments found a sub-population of neurons (primarily located in the deep layers of V1) that were driven by both a prediction of optic flow based on the animal's running speed and the visually observed optic flow. The majority of these cells seemed to be integrating locomotor and visual information, to produce a multi-modal estimate of the animal's running speed (Saleem et al., 2013). While the model proposed to account for mismatch neurons (Keller and Mrsic-Flogel, 2018, Figure 2b) does not exclude the possibility of there being another population of neurons that integrate proprioceptive and visual information into a joint estimate of speed, it also does not explain the existence of such neurons.

2.1.4 Lateral inhibition within regions

The original Rao and Ballard (1999) model (see Figure 1a) suggests that the inter-regional connections are many-to-many while the intra-regional connections are one-to-one (or one-to-two if there are separate error neuron populations encoding positive and negative prediction-errors). As a consequence, this model propose that there are many more connections between cortical regions than within cortical regions. However, inter-regional connections have a high cost in terms of physical space (which is limited in the skull) and energy consumption (due to the metabolic cost of sending signals along long-range connections). Much evidence points to evolution having optimised the connectome to minimize these costs by minimising the overall axon wiring length (Bullmore and Sporns, 2012; Cherniak et al., 2004; Ercsey-Ravasz et al., 2013). If the cortex does perform predictive coding, then it is a challenge to explain why it would have been implemented (as envisaged by the Rao and Ballard (1999) model) in such a way as to be maximally inefficient in terms wiring length.

A potential solution is to shift the proposed assignment of neural populations to cortical regions (see Figure 1b), so that the bottom-up input to each cortical region arrives at the error neurons (rather than the prediction neurons) and these neurons connect in a many-to-many pattern to a subsequent population of prediction neurons in the same cortical region (Spratling, 2008b). Such a change in the allocation of neural populations to cortical regions has no effect on the functioning of the model, just its neuro-anatomical interpretation.

This shift in the assignment of neural populations to cortical regions also makes more sense in terms of the functionality of the dense connections between error neurons and the subsequent population of prediction neurons. If the activity of a population of prediction neurons is represented by a vector, $\mathbf{y}$, the feedforward weights to these neurons are represented by a matrix, $\mathbf{W}$, and the output of the prediction neurons in the preceding cortical region is represented by a vector, $\mathbf{x}$ (see Figure 1a), the change in activity of the higher-level prediction neurons is described as follows (see equation 8 of Rao and Ballard, 1999):

$$\frac{\delta \mathbf{y}}{\delta t} = -\vartheta g'(\mathbf{y}) + \eta \left(\mathbf{y}^{td} - \mathbf{y}\right) + \zeta \mathbf{W}\mathbf{x} - \zeta \left(\mathbf{W}\mathbf{W}^T\right)\mathbf{y} \tag{1}$$

Where ϑ, η, and ζ are scalar parameters. The four terms on the right-hand side of this equation are: (1) an activation decay term (where $g'(\mathbf{y}) = \mathbf{y}$ for a Gaussian prior distribution on the responses); (2) a term describing the effect of top-down predictions ($\mathbf{y}^{td}$) from a subsequent cortical region[b]; (3) the feedforward drive; and (4) a term that describes lateral inhibition between the prediction neurons within this population. This lateral inhibition is mediated by the preceding population of error neurons, as is made explicit by rearranging the equation:

$$\frac{\delta \mathbf{y}}{\delta t} = -\vartheta g'(\mathbf{y}) + \eta \left(\mathbf{y}^{td} - \mathbf{y}\right) + \zeta \mathbf{W}\left(\mathbf{x} - \mathbf{W}^T\mathbf{y}\right) \tag{2}$$

Where the term in the last set of brackets is the activation, $\mathbf{e}$, of the preceding population of error neurons, *i.e.*, $\mathbf{e} = \mathbf{x} - \mathbf{W}^T\mathbf{y}$. Lateral inhibition is a process that is typically considered to occur within cortical regions, and not to rely on interactions with populations of neurons in different regions.

The model proposed by Keller and Mrsic-Flogel (2018, Figure 2b) to account for mismatch neurons differs from the model proposed by Rao and Ballard (1999, Figure 1a) in suggesting that the bottom-up, visually-driven, information is not transmitted to the error (or mismatch) neurons in V1 via the preceding population of V1 prediction neurons. Instead, this model proposes that the visual inputs are sent directly to the mismatch neurons and the V1 predictions neurons are reciprocally connected, with dense connections, to the mismatch neurons. As a result, the Keller and Mrsic-Flogel (2018) model proposes a similar number of connections within a cortical region as between two cortical regions. The reciprocal connections between the V1 prediction neurons and the mismatch neurons perform lateral inhibition using connections intrinsic to V1. However, the reciprocal connections between the prediction neurons in the motor area and the mismatch neurons perform lateral inhibition using long-range, inter-regional, connections. Hence, this model inherits from the Rao and Ballard (1999) model the unlikely requirement that lateral inhibitory interactions between neurons in one cortical region are generated through reciprocal connections with a neural population in a separate cortical area.

2.1.5 Ability to scale to complex tasks

One criteria that can be used to evaluate a model of brain function is its ability to scale-up to deal with real-world tasks. The brain is capable of performing complex tasks within rich sensory environments. A model that can only

[b]Other, related, theories of inference in cortical circuits (Adams et al., 2013; Friston, 2009; Friston et al., 2011; Friston, 2005; Shipp, 2016) propose that top-down inputs can selectively modify the gain of error neurons (such that prediction-errors are weighted by their reliability or precision). This would correspond to changing the value of ζ in Equations 1 and 2. In such a model, the information conveyed by the motor regions to V1 might be interpreted as an estimate of the precision of the visual signals under different levels of locomotion, rather than a prediction of optic flow speed. In such models, precision-weighting of the prediction errors is important for allowing fast updating of the predictions while maintaining stability. As will be discussed in Section 2.1.5, the Rao and Ballard (1999) model, which does not use precision-weighting, has issues with stability.

Properties	Configuration Equations	Figure 2b Rao and Ballard (1999)	Figure 2c Rao and Ballard (1999)	Figure 2c PC/BC-DIM
Non-negative firing rates (Section 2.1.1)		✓	✗	✓
Fits mismatch neuron data (Section 2.1.2)		✓	✗	✓
Fits multisensory-integration data (Section 2.1.3)		✗	✓	✓
Lateral inhibition within regions (Section 2.1.4)		✗	✓	✓
Scales to complex tasks (Section 2.1.5)		✗	✗	✓

Table 1: A summary of the desirable properties of a model, and an evaluation of the three models discussed in the main text (in Sections 2.1, 2.2.1, and 2.2.2 respectively) in terms of which properties it does (✓) of does not (✗) possess.

simulate simple tasks with impoverished inputs is therefore unlikely to be a good candidate for a model of cortical function. For the predictive coding model, scaling up requires increasing the dimensionality of the input (*i.e.*, the complexity of the sensory environment) and the number of prediction neurons (to represent the larger number of causes that may underlie the sensory input in this richer environment). However, for the predictive coding model proposed by Rao and Ballard (1999) as the network increases in size it is necessary to reduce the rate at which the prediction neuron responses are updated by decreasing the value of ζ in Equation 1 (Harpur, 1997). If ζ is not decreased the network becomes unstable (all the neural responses oscillate between high and low values and these values become infinitely large over time). However, if ζ is decreased then the activations of the prediction neurons change very slowly and the algorithm becomes impracticably slow.

A simple, synthetic, task can be used to illustrate this issue with scaling in the Rao and Ballard (1999) model. Consider a small patch of image in which the pixels take binary values[c]. Every image within this patch consists of s pixels with a value of 1 (ON) and s pixels with a value of 0 (OFF). Each possible combination of s ON and s OFF pixels has a different underlying cause. As s increases both the size of the image patch increases and the number of possible underlying causes increases. Hence, it is possible to use this task, with increasing values of s, to assess the ability of a predictive coding algorithm to scale. To do so, the images are used as input (via the preceding population of error neurons) to a population of prediction neurons. Every underlying cause is represented by a distinct prediction neuron which has weights that are optimal for representing that cause. When an image is presented to the network, one prediction neuron (the one that represents the cause of the current image) should be strongly active, and all other prediction neurons (which represent other possible causes) should be inactive. Figure 3d shows the difference between the activation of the prediction neuron that represents the cause of the input image and the activation of the next most strongly activated prediction neuron for different values of parameter ζ and for different values of s, or scales of task. It can be seen that as s increase no value of ζ enables the network to correctly identify the underlying cause, and large values of ζ result in network instability (results marked with white crosses). These results were produced using 50 iterations to find the prediction neuron responses. Increasing the number of iterations to 250 did not significantly change the results[d]. Note that the results labelled "DIM" are for a different method of implementing predictive coding that will be described in Section 2.2.

2.2 Improving the fit between the neurophysiological data and the predictive coding model

The effects of locomotion on neural activity in mouse primary visual cortex (Keller et al., 2012; Leinweber et al., 2017; Saleem et al., 2013; Zmarz and Keller, 2016) can be partially explained by the predictive coding model proposed by Keller and Mrsic-Flogel (2018, Figure 2b). However, this model is inconsistent with some aspects of the neuro-biology, as summarised in the second column of Table 1. This section will consider how the outstanding issues can be addressed, and hence, how the fit of the predictive coding model with the empirical data can be improved.

The preceding section has referred to two facets of a prediction coding model: (1) the configuration of the error and prediction neurons and how this is interpreted in terms of cortical circuity; and (2) the mathematical equations that are used to simulate the behaviour of the error and prediction neurons. The model advanced by

[c]Binary inputs are used in order to easily define an unambiguous task, not because predictive coding models are restricted to making predictions about simple ON and OFF states. Indeed, the generative model used in predictive coding is capable of reconstructing continuous valued states (like luminance values) and more specialised algorithms (like Boolean factor analysis and belief propagation) might perform better than predictive coding on binary-valued problems (Friston et al., 2017; Snášel et al., 2008).

[d]Using the non-linear model that is also described in Rao and Ballard (1999), avoids the instability, but still results in the prediction neuron that represents the cause of the input having a response that is not distinctively different from the responses of other prediction neurons. This issue remains if a sparse kurtotic prior (Rao and Ballard, 1999) is imposed on the prediction neuron responses.

Keller and Mrsic-Flogel (2018) uses the configuration shown in Figure 2b and also proposes that prediction-errors are calculated using subtraction, which is consistent with the equations used by Rao and Ballard (1999).

The two properties of a model, configuration and equations, can be varied independently. Section 2.2.1 will consider changing the configuration of the model, while keeping the equations defined by Rao and Ballard (1999). This change solves some of the issues of the Keller and Mrsic-Flogel (2018) model, but it does not solve them all. Section 2.2.2 will consider changing the equations of the model, as well as its configuration. The equations used in Section 2.2.2 are those of the PC/BC-DIM model, which is a version of Predictive Coding (PC) reformulated to make it compatible with Biased Competition (BC) theories of cortical function (Spratling, 2008b) and which uses Divisive Input Modulation (DIM; Spratling et al., 2009) as the method for updating error and prediction neuron activations. DIM calculates prediction-error using division rather than subtraction. It is shown that this change produces a predictive coding model that can more successfully account for the neurophysiological data.

2.2.1 Changing configuration

The results of Saleem et al. (2013, see Section 2.1.3) suggest that V1 is performing multisensory integration in order to produce a combined estimate of the animal's running speed. Multisensory integration can be simulated using a predictive coding model configured as shown in Figure 2c. In this model sensory input from the two different sensory modalities (in this case vision and locomotion) both arrive as inputs to a population of error neurons. Both these sets of inputs are represented using a population code (Deneve et al., 2001; Georgopoulos et al., 1986; Pouget et al., 2003, 2000). In each population each neuron is tuned to a preferred value of speed, such that its response is highest when the measured speed matches its preferred speed, and the neuron's response decreases as the difference between the measured speed and the preferred speed increases. Across the population different neurons have different speed preferences, such that the whole range of possible speeds is represented by the population as a whole. In the model that has been implemented neurons were given Gaussian RFs equally spaced along a logarithmic scale representing speed. 18 neurons were used to provide a population code representing speed based on the optic flow, and another 18 neurons were used to provide a population code representing speed based on locomotion. The prediction neurons also formed a population code. Each prediction neuron had weights that allowed it to receive input (via the error neurons) from neurons representing similar values of speed in the two input population codes. Nine prediction neurons were used in the model. To avoid overcrowding Figure 2c shows only pairs of neurons in each population labelled S and F to indicate tuning to slow and fast speeds, respectively. Note that the error neurons and the subsequent population of prediction neurons are proposed to be located in the same cortical region (V1), this represents a shift in the assignment of neural populations to cortical regions compared to the Rao and Ballard (1999) model, which solves the issue of lateral inhibition relying on long-range, inter-regional, connections (as described in Section 2.1.4).

In the proposed model the response of a particular prediction neuron can be driven by visual or locomotion based input of the appropriate speed. This is consistent with neurophysiological evidence that mouse V1 neurons respond to visual input in the absence of locomotion and to evidence that mouse V1 neurons respond to running in the dark (Keller et al., 2012; Saleem et al., 2013). A prediction neuron will respond most strongly when both inputs are active simultaneously, which is consistent with the neurophysiological data of Saleem et al. (2013, see Section 2.1.3).

The proposed model was simulated using the formula (see Equation 2) proposed by Rao and Ballard (1999), and the responses of all 36 error neurons were recorded for different combinations of running speed and visual flow. These error neuron responses are shown in Figure 3e. It can be seen that none of these model neurons have response properties that are consistent with the mismatch neurons reported by Zmarz and Keller (2016, see Section 2.1.2 and Figure 3a). Furthermore, the equations proposed by Rao and Ballard (1999) require neurons to be able to produce negative firing rates (see Sections 2.1.1) and does not scale to more realistic tasks (see Section 2.1.5). Hence, while this version of the predictive coding model does overcome some short-comings of the model proposed by Keller and Mrsic-Flogel (2018), it has its own limitations as summarised in the third column of Table 1.

2.2.2 Changing configuration and equations

The configuration of the model shown in Figure 2c was simulated for a second time using the equations that define the PC/BC-DIM variation of predictive coding (Spratling, 2008a, 2017b). The responses of the 36 error neurons are shown in Figure 3f. It can be seen that a sub-population of the model error neurons have response properties that are consistent with the mismatch neurons reported by Zmarz and Keller (2016, see Figure 3a). Because the PC/BC-DIM model calculates the errors using a divisive (rather than a subtractive) comparison between the actual and predicted inputs, the sub-population of error neurons that have a large response when running speed exceeds visual flow speed (like the V1 mismatch neurons) have a non-zero response when the two estimates of speed are

similar. This is more consistent with the neurophysiological data than the results produced when calculating the error using subtraction (Figure 3c), and avoids the need for a post-hoc transformation of the output of the model in order to fit the neurophysiological data (Section 2.1.2). As discussed in the preceding section, this configuration of the model also successfully accounts for the neurophysiological data of Saleem et al. (2013, see Section 2.1.3) and proposes that lateral inhibition results from interactions between neurons in local circuits rather than relying on interactions with neurons in distinct cortical regions (Section 2.1.4). Furthermore, the PC/BC-DIM equations allow the model to scale: results of the experiment described in Section 2.1.5 are shown on the row labelled "DIM" in Figure 3d; and models containing 10s of thousands to 10s of millions of prediction neurons have been applied to practical problems in computer vision (Spratling, 2013, 2017a). In addition, the PC/BC-DIM algorithm uses only non-negative firing-rates. Hence, by changing the equations that implement predictive coding the configuration shown in Figure 2c can provide a more comprehensive and successful account of the neuro-biology (see the last column of Table 1).

3 Discussion

This article has shown that a range of neurophysiological data describing the effects of locomotion on neural responses in mouse primary visual cortex (Keller et al., 2012; Leinweber et al., 2017; Saleem et al., 2013; Zmarz and Keller, 2016) can be explained using predictive coding. This was achieved by rejecting the Rao and Ballard (1999) implementation of predictive coding which has many biologically-implausible features (such as needing neurons to be able to produce negative firing rates, predicting that cortical inter-regional connections are more abundant than intra-regional connections, proposing that lateral inhibitory interactions between neurons rely on long-range, inter-regional, connections, and an inability to scale to non-trivial tasks). It was also necessary to reject the neural circuitry proposed by Keller and Mrsic-Flogel (2018): this model, being based on the Rao and Ballard (1999) model, inherits some of the issues of that model as well as failing to fit some of the experimental data. By using a different implementation of predictive coding, PC/BC-DIM, and by configuring the circuity of this model to perform multisensory integration, it was possible to successfully simulate the neurophysiological data while avoiding the issues of the previously proposed model.

Previous work has shown that the same configuration of the PC/BC-DIM model as was used here can perform near optimal (*i.e.*, Bayesian) multisensory integration (Spratling, 2016), consistent with psychophysiological experiments showing that human performance in sensory cue integration tasks is near optimal (Ernst and Jäkel, 2003; Ma and Pouget, 2008; Pouget et al., 2013; Seilheimer et al., 2014). The same model could potentially also explain additional neurophysiological findings. For example, recent experiments have shown that the response to a visual landmark in mouse V1 is modulated by an estimate of the animal's location (Fiser et al., 2016; Saleem et al., 2018). These results could also be simulated using the predictive coding circuity shown in Figure 2c by simply replacing the two sensory inputs with one input encoding visual features (rather than visual optic flow) and one encoding estimated location (rather than the speed of locomotion). In such a model the prediction neurons would respond to specific visual features, but this response would be enhanced when the location estimate matched the expected location for those features. Furthermore, the prediction neurons would respond to location information in the absence of the corresponding visual stimulus, consistent with experimental results interpreted as showing that V1 neurons become predictive of upcoming visual stimuli (Fiser et al., 2016). In addition, in such a model location information in the absence of the corresponding visual stimulus would result in large responses from the error neurons receiving the (unexpected) visual input. This is consistent with experimental results that show that the omission of a predicted visual landmark results in a strong response in V1 (Fiser et al., 2016).

Predictive coding models come in several different forms (Spratling, 2017b). At the computational-level of analysis all these models make the same claim: that the brain is attempting to fit an internal model of the world to the sensory data by minimising the discrepancy between the predictions of the model and the sensory evidence. However, at a more mechanistic-level, different models make different claims and predictions. This has implications for hypothesis-driven empirical research evaluating predictive coding in the brain. In turn such neurophysiological data should play a role in refining and improving models of predictive processing.

4 Experimental procedure

Open-source software, written in MATLAB, which performs the simulations described in this article is available for download from: `http://nms.kcl.ac.uk/michael.spratling/Code/pc_biological_fit.zip`.

The predictive coding networks were simulated using a vector of input values, $\mathbf{x}$, that provided input to a population of error neurons (with activations represented by a vector, $\mathbf{e}$). The error neurons provided input to (and

received connections from) a population of prediction neurons (with activations represented by a vector, $\mathbf{y}$). The connections from the error neurons to the prediction neurons had synaptic weight values represented by a matrix, $\mathbf{W}$.

4.1 Implementation of the Rao and Ballard (1999) algorithm

The prediction neuron activations were updated as described in Equation 2 and the error neuron activations were calculated as $\mathbf{e} = \mathbf{x} - \mathbf{W}^T\mathbf{y}$. As only one pair of error and prediction neurons were simulated, the effects of top-down predictions, $\mathbf{y}^{td}$, (that would arrive from a subsequent population of error neurons in a hierarchical model) were ignored by setting $\mathbf{y}^{td}$ equal to $\mathbf{y}$, so that the second term on the right-hand side of Equation 2 disappeared.

Hence, at each iteration, the prediction neuron responses were updated, such that:

$$\mathbf{y} \leftarrow \mathbf{y} - \vartheta g'(\mathbf{y}) + \zeta \mathbf{W}\left(\mathbf{x} - \mathbf{W}^T\mathbf{y}\right) \tag{3}$$

where $g'(\mathbf{y}) = y$ for the version of the model with a Gaussian prior distribution on the responses, and $g'(\mathbf{y}) = \frac{\mathbf{y}}{1+\mathbf{y}^2}$ for the version of the model with a kurtotic prior. For the non-linear version of the model the following update was used:

$$\mathbf{y} \leftarrow \mathbf{y} - \vartheta g'(\mathbf{y}) + \zeta \mathbf{W}\left(1 - tanh(\mathbf{x} - tanh(\mathbf{W}^T\mathbf{y})\right)^2 \tag{4}$$

All elements of $\mathbf{y}$ were initialised to zero prior to the first iteration. Varying the value of parameter ϑ was not found to improve the results, and the results shown in this article have been produced with a value of $\vartheta = 0$. A value of $\zeta = 0.1$ was used unless stated otherwise.

4.2 Implementation of the PC/BC-DIM algorithm

The activation of the error and prediction neurons were determined using the following equations:

$$\mathbf{e} = \mathbf{x} \oslash \left(\epsilon_2 + \hat{\mathbf{W}}^T\mathbf{y}\right) \tag{5}$$

$$\mathbf{y} \leftarrow (\epsilon_1 + \mathbf{y}) \otimes \mathbf{We} \tag{6}$$

Where $\hat{\mathbf{W}}$ is a matrix representing the same synaptic weight values as $\mathbf{W}$ but such that the rows of $\hat{\mathbf{W}}$ are normalised to have a maximum value of one, $\oslash$ and $\otimes$ indicate element-wise division and multiplication respectively, and ϵ_1 and ϵ_2 are non-negative parameters that, respectively, prevent prediction neurons becoming permanently non-responsive and prevent division-by-zero errors. Values of $\epsilon_1 = 1 \times 10^{-6}$ and $\epsilon_2 = 1 \times 10^{-4}$ were used in all experiments, but the results were not found to be sensitive to these parameter values. All elements of $\mathbf{y}$ were initialised to zero. The values of $\mathbf{e}$ and $\mathbf{y}$ were then iteratively updated with the new values of $\mathbf{y}$ calculated by eq. 6 substituted into eqs. 5 and 6 at the next iteration.

4.3 Synthetic task to test ability to scale

For this task $\mathbf{x}$ was a binary-valued vector with $2s$ elements. s elements in $\mathbf{x}$ had a value of 1 and s elements had a value of 0. Each prediction neuron had weights representing the input generated by one possible cause. Hence, each row of $\mathbf{W}$ was set equal to one possible $\mathbf{x}$, normalised so that the weights summed to one. The number of prediction neurons required to represent all possible causes increases exponentially with s. Specifically, for s equal to integer values from 1 to 8, the corresponding number of causes are: 2, 6, 20, 70, 252, 924, 3432, and 12870. For example, when $s = 2$, the possible input patterns are: $\mathbf{x} = (1,1,0,0)^T$, $\mathbf{x} = (1,0,1,0)^T$, $\mathbf{x} = (1,0,0,1)^T$, $\mathbf{x} = (0,1,1,0)^T$, $\mathbf{x} = (0,1,0,1)^T$, and $\mathbf{x} = (0,0,1,1)^T$ and the weights are:

$$\mathbf{W} = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix}.$$

The network was tested by setting $\mathbf{x}$ equal to one of the possible input patterns, and recording the responses of the prediction neurons after 50 iterations. The response of the prediction neuron with weights corresponding to the presented pattern was then compared to the response of the prediction neuron with the next highest response. The difference between these two prediction neuron responses should be large if the network has correctly represented

the stimulus, *i.e.*, if it can accurately distinguish one cause from another. These results are shown for both predictive coding algorithms in Figure 3d. The results show that when the sensory environment contains distinct, but highly overlapping, stimuli then the Rao and Ballard (1999) algorithm, with the correct choice of parameters, can distinguish one stimulus from another when there are up to around 20 stimuli, while the PC/BC-DIM algorithm can successfully distinguish at least 12870 stimuli.

4.4 Model of multi-sensory integration

For this task $\mathbf{x}$ was a 36 element vector, the first 18 elements represented the outputs of a population of neurons encoding the visual optic flow speed, and the second 18 elements represented the outputs of a population of neurons encoding the predicted optic flow based on running speed. Both these input populations encoded speed using a population code defined on a logarithmic scale. Specifically, the input neurons had Gaussian RFs centred at speed values from 0.0156 to 211.6 cm/s and were equally spaced apart on a logarithmic scale. Each element of $\mathbf{x}$ represented the response of the corresponding neuron in the population code to the sensed speed value.

Each row of $\mathbf{W}$, *i.e.*, each prediction neuron's weights, consisted of two Gaussian RFs. One RF connecting the prediction neuron to the first 18 error neurons, and hence, to the inputs representing visual optic flow speed. The second RF connecting the prediction neuron to the second 18 error neurons, and hence, to the inputs representing running speed. Both RFs were centred at the same speed value in each of these two input spaces. The population of nine prediction neurons had preferences for (*i.e.*, RFs centred at) speed values from 0.0156 to 102.5 cm/s and were equally spaced apart in logarithmic space.

The network was tested by presenting different combinations of speed value to each of the two input populations. These input speed values ranged from 0.0625 to 64 cm/s equally spaced apart on a logarithmic scale. For all combinations of optic flow and running speed, the responses for all 36 error neurons were recorded after 25 iterations. These results are shown in Figure 3e when the network was simulated using the Rao and Ballard (1999) algorithm and Figure 3f when using PC/BC-DIM.

References

Achler, T. (2014). Symbolic neural networks for cognitive capacities. *Biologically Inspired Cognitive Architectures*, 9:71–81.

Adams, R., Shipp, S., and Friston, K. (2013). Predictions not commands: active inference in the motor system. *Brain Structure and Function*, 218(3):611–43.

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., and Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *The Journal of Neuroscience*, 30(8):2960–6.

Attinger, A., Wang, B., and Keller, G. B. (2017). Visuomotor coupling shapes the functional development of mouse visual cortex. *Cell*, 169(7):1291 –1302.e14.

Barlow, H. B. (1994). What is the computational goal of the neocortex? In Koch, C. and Davis, J. L., editors, *Large-Scale Neuronal Theories of the Brain*, chapter 1, pages 1–22. MIT Press, Cambridge, MA.

Bubic, A., von Cramon, D. Y., and Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, 4(25):1–15.

Bullmore, E. and Sporns, O. (2012). The economy of brain network organization. *Nature Reviews Neuroscience*, 13:336–49.

Busse, L. (2018). The influence of locomotion on sensory processing and its underlying neuronal circuits. *e-Neuroforum*, 24(1):A41–A51.

Chater, N., Tenenbaum, J. B., and Yuille, A. (2006). Probabilistic models of cognition: conceptual foundations. *Trends in Cognitive Sciences*, 10:287–91.

Cherniak, C., Mokhtarzada, Z., Rodriguez-Esteban, R., and Changizi, K. (2004). Global optimization of cerebral cortex layout. *Proceedings of the National Academy of Sciences USA*, 101(4):1081–6.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03):181–204.

Deneve, S., Latham, P. E., and Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, 4(8):826–31.

Edwards, G., Vetter, P., McGruer, F., Petro, L. S., and Muckli, L. (2017). Predictive feedback to V1 dynamically updates with sensory input. *Scientific Reports*, 7(16538).

Egner, T., Monti, J. M., and Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *The Journal of Neuroscience*, 30(49):16601–8.

Ercsey-Ravasz, M., Markov, N., Lamy, C., VanEssen, D., Knoblauch, K., Toroczkai, Z., and Kennedy, H. (2013). A predictive network model of cerebral cortical connectivity based on a distance rule. *Neuron*, 80(1):184–97.

Ernst, M. and Jäkel, F. (2003). Learning to combine arbitrary signals from vision and touch. In *4th International Multisensory Research Forum*.

Fang, F., Kersten, D., and Murray, S. O. (2008). Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *Journal of Vision*, 8(7):1–9.

Fiser, A., Mahringer, D., Oyibo, H., Petersen, A., Leinweber, M., and Keller, G. (2016). Experience-dependent spatial expectations in mouse visual cortex. *Nature Neuroscience*, 19:1658–64.

Fridman, M. and Petreanu, L. (2017). Cortical processing: How mice predict the visual effects of locomotion. *Current Biology*, 27(23):R1272–4.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, 13(7):293–301.

Friston, K., Mattout, J., and Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, 104(1-2):137–60.

Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360(1456):815–36.

Friston, K. J., Parr, T., and de Vries, B. (2017). The graphical brain: Belief propagation and active inference. *Network Neuroscience*, 1(4):381–414.

Georgopoulos, A. P., Schwartz, A. B., and Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 233:1416–9.

Griffiths, T. L., Kemp, C., and Tenenbaum, J. B. (2008). Bayesian models of cognition. In Sun, R., editor, *Cambridge Handbook of Computational Cognitive Modeling*. Cambridge University Press, Cambridge, UK.

Griffiths, T. L. and Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, 17(9):767–73.

Harpur, G. F. (1997). *Low Entropy Coding with Unsupervised Neural Networks*. PhD thesis, Department of Engineering, University of Cambridge.

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

Huang, Y. and Rao, R. P. N. (2011). Predictive coding. *WIREs Cognitive Science*, 2:580–93.

Keller, G. B., Bonhoeffer, T., and Hbener, M. (2012). Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron*, 74(5):809–15.

Keller, G. B. and Mrsic-Flogel, T. D. (2018). Predictive processing: A canonical cortical computation. *Neuron*, 100(2):424–35.

Kersten, D., Mamassian, P., and Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 55(1):271–304.

Khan, A. G. and Hofer, S. B. (2018). Contextual signals in visual cortex. *Current Opinion in Neurobiology*, 52:131–8.

Knill, D. C. and Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge University Press, Cambridge, UK.

Kok, P., Brouwer, G. J., van Gerven, M. A. J., and de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *The Journal of Neuroscience*, 33(41):16275–84.

Kok, P. and de Lange, P. F. (2015). Predictive coding in sensory cortex. In Forstmann, U. B. and Wagenmakers, E.-J., editors, *An Introduction to Model-Based Cognitive Neuroscience*, pages 221–44. Springer, New York, NY.

Kok, P., Mostert, P., and Lange, F. P. D. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences USA*, 201705652.

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., and de Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex*, 22:2197–206.

Kok, P. and Turk-Browne, N. (2018). Associative prediction of visual shape in the hippocampus. *The Journal of Neuroscience*, 38(31).

Lee, T. S. and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 20:1434–48.

Leinweber, M., Ward, D. R., Sobczak, J. M., Attinger, A., and Keller, G. B. (2017). A sensorimotor circuit in mouse cortex for visual flow predictions. *Neuron*, 95:1420–32.

Ma, W. J. (2012). Organising probabilistic models of perception. *Trends in Cognitive Sciences*, 16(10):511–8.

Ma, W. J. and Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annual Review of Neuroscience*, 37:205–20.

Ma, W. J. and Pouget, A. (2008). Linking neurons to behavior in multisensory perception: A computational review. *Brain Research*, 1242(0):4–12.

Mumford, D. (1992). On the computational architecture of the neocortex II: the role of cortico-cortical loops.

Biological Cybernetics, 66:241–51.

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., and Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences USA*, 99(23):15164–9.

Murray, S. O., Schrater, P., and Kersten, D. (2004). Perceptual grouping and the interactions between visual cortical areas. *Neural Networks*, 17:695–705.

Pakan, J. M., Francioni, V., and Rochefort, N. L. (2018). Action and learning shape the activity of neuronal circuits in the visual cortex. *Current Opinion in Neurobiology*, 52:88–97.

Pouget, A., Beck, J. M., Ma, W. J., and Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nature Neuroscience*, 16:1170–8.

Pouget, A., Dayan, P., and Zemel, R. S. (2003). Inference and computation with population codes. *Annual Review of Neuroscience*, 26:381–410.

Pouget, A., Zemel, R. S., and Dayan, P. (2000). Information processing with population codes. *Nature Reviews Neuroscience*, 2:125–32.

Rao, R. P. N. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87.

Saleem, A. B., Ayaz, A., Jeffery, K. J., Harris, K. D., and Carandini, M. (2013). Integration of visual motion and locomotion in mouse visual cortex. *Nature Neuroscience*, 16:1864–9.

Saleem, A. B., Diamanti, E. M., Fournier, J., Harris, K. D., and Carandini, M. (2018). Coherent encoding of subjective spatial position in visual cortex and hippocampus. *Nature*, 562:124–7.

Schellekens, W., van Wezel, R. J. A., Petridou, N., Ramsey, N. F., and Raemaekers, M. (2016). Predictive coding for motion stimuli in human early visual cortex. *Brain Structure and Function*, 221(879).

Seilheimer, R. L., Rosenberg, A., and Angelaki, D. E. (2014). Models and processes of multisensory cue combination. *Current Opinion in Neurobiology*, 25:38–46.

Shipp, S. (2016). Neural elements for predictive coding. *Frontiers in Psychology*, 7:1792.

Snášel, V., Moravec, P., Húsek, D., Frolov, A., Řezanková, H., and Polyakov, P. (2008). Pattern discovery for high-dimensional binary datasets. In *Proceedings of the International Conference on Neural Information Processing*, volume 4984 of *Lecture Notes in Computer Science*, pages 861–72. Springer.

Spratling, M. W. (2008a). Predictive coding as a model of biased competition in visual selective attention. *Vision Research*, 48(12):1391–408.

Spratling, M. W. (2008b). Reconciling predictive coding and biased competition models of cortical function. *Frontiers in Computational Neuroscience*, 2(4):1–8.

Spratling, M. W. (2013). Image segmentation using a sparse coding model of cortical area V1. *IEEE Transactions on Image Processing*, 22(4):1631–43.

Spratling, M. W. (2016). A neural implementation of Bayesian inference based on predictive coding. *Connection Science*, 28(4):346–83.

Spratling, M. W. (2017a). A hierarchical predictive coding model of object recognition in natural images. *Cognitive Computation*, 9(2):151–67.

Spratling, M. W. (2017b). A review of predictive coding algorithms. *Brain and Cognition*, 112:92–7.

Spratling, M. W., De Meyer, K., and Kompass, R. (2009). Unsupervised learning of overlapping image components using divisive input modulation. *Computational Intelligence and Neuroscience*, 2009(381457):1–19.

Summerfield, C. and Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13(9):403–9.

Summerfield, C. and Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron*, 59(2):336–47.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., and Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, 11(9):1004–6.

Vilares, I. and Kording, K. (2011). Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, 1224(1):22–39.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., and Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences USA*, 108(51):20754–9.

Yuille, A. and Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, 10(7):301–8.

Zmarz, P. and Keller, G. B. (2016). Mismatch receptive fields in mouse visual cortex. *Neuron*, 92:766–72.