

Does predictive coding have a future?

In the 20th century we thought the brain extracted everything it needed to know from its sensations – the standard ‘sandwich’ model of *stimulus–cognition–response*. The 21st century witnessed a Dennett-like ‘strange inversion’, where the brain became an organ of inference, constructing explanations for sensory samples – and actively resolving uncertainty about what’s going on ‘out there’, beyond its sensory epithelia. If there has been a paradigm shift, then one paper played a key role at the tipping point – a paper whose insights may still be prescient today.

Karl Friston¹

¹ The Wellcome Centre for Human Neuroimaging, University College London, UK

Every decade or so, one reads a paper that makes you think “well, that’s quite remarkable”. In 1999 Rao and Ballard¹ offered a treatment of predictive coding in the visual cortex. On their view, backward connections from higher to lower order visual areas try to predict activity in lower order areas; while the counter stream of ascending, forward connections convey prediction errors; namely, what cannot be predicted. These prediction errors drive expectations in higher levels towards better explanations for lower levels. Using simulations they showed this simple (deep or hierarchical) architecture was not only consistent with neuroanatomy and physiology but could account for range of subtle response properties like ‘end-stopping’ and other extra-classical receptive field effects.

This was a significant achievement in its own right; however, the really remarkable thing – at least for me – was the following: in simulating their little piece of synthetic cortex, neuronal dynamics and connectivity optimised the *same* energy or cost function. I remember reading the methods section several times to convince myself that they could explain all of this computational anatomy and detailed neurophysiology with just one energy function. Surely there was something quite profound about this: here was a truly normative scheme that could explain both fast neuronal dynamics – that underwrite perceptual synthesis – and the slow fluctuations in synaptic efficacy – that mediate perceptual learning – with just one imperative; namely, minimise prediction error.

Self-evidencing and predictive processing

In retrospect, it should not have been quite so remarkable (to me). The predictive coding scheme described by Rao and Ballard has a long pedigree that can be traced back to the students of Plato and Kant to Helmholtz², via epistemological automata and analysis-by-synthesis to perception as hypothesis testing³, via machine learning and information theory to specific proposals for computational architectures in the neocortex^{4,5}. The theme that runs through this legacy is inference and learning to the

Predictive coding

best explanation for our sensorium. In other words, the brain is in the game of optimising its connectivity and dynamics to maximise the evidence for its model of the world⁶.

So what form does this evidence take? If you were a statistician, then it is just Bayesian model evidence – the probability of some data given your model of how those data were generated. If you were in machine learning, you would use a variational bound on log-evidence. If you were an engineer, you would use the cost functions associated with Kalman filters. If you were an information theorist you would use efficiency or minimum description length. Finally, if you commit to predictive coding, you would use (precision weighted) prediction error. Crucially, *these are all the same thing* which, in my writing, is variational free energy⁷.

Predictive coding offered a compelling process theory that lent notions like the Bayesian brain⁸ a mechanistic substance. The Bayesian brain captured a growing consensus that one could understand the brain as a statistical organ, engaging in an abductive inference of an ampliative nature. Predictive coding articulated plausible neuronal processes that were exactly consistent with the imperative to optimise Bayesian model evidence. Within a decade, the Bayesian brain hypothesis and predictive coding became a dominant paradigm in cognitive neuroscience – marking a watershed between 20th-century thinking about the brain as a glorious stimulus-response link and more constructivist 21st century perspectives that emphasised an active sampling of the sensory world. There has been a remarkable uptake of these ideas in fields as diverse as philosophy^{6,9}, ethology and psychoanalysis, with dedicated meetings and books emerging with increasing frequency. But what about neuroscience; has predictive coding told us anything we did not know? In what follows, I rehearse some recent examples where the tenets of predictive coding have pre-empted empirical findings.

The importance of being precise

A recent example is a report from Marques *et al*¹⁰, looking at the functional organisation of cortical feedback inputs to primary visual cortex. In brief, their exceptional results "*show that feedback [FB] inputs show tuning-dependent retinotopic specificity. By targeting locations that would be activated by stimuli orthogonal to or opposite to a cell's own tuning, feedback potentially enhance visual representations in time and space.*"¹⁰ p757.

This particular aspect of feedback was predicted several years based on the optimisation of *precision*, when forming the precision-weighted prediction errors that mediate belief updating. In this setting, precision corresponds to the best estimate of the reliability or inverse variance of prediction errors¹¹. Technically, getting the precision right corresponds to optimising the Kalman gain in Bayesian or Kalman filters¹. Computationally, it underlies the optimal mixing of several streams of sensory input that differ in their reliability – as in multimodal sensory integration⁸. Psychologically, this precision-weighting has been associated with sensory attention and attenuation¹². Physiologically, precision-weighting is thought to be mediated by neuromodulatory mechanisms; for example, classical neuromodulators of synchronous

Predictive coding

gain¹³. In short, most of the interesting bits of predictive coding are about getting the precision right – to select newsworthy, uncertainty resolving prediction errors¹¹.

Precision was missing in 20th-century treatments of predictive coding: interestingly, Rao and Ballard ascribe this Kalman gain to constants they labelled k . Precision has played a key role in taking predictive coding to the next level in cognitive neuroscience: it underwrites computational anatomy of expectation and attentional selection at various levels of hierarchical perception¹¹⁻¹⁴. Failures of the neuromodulatory basis of precision-weighting have figured prominently in explanations for false inference and psychopathology¹⁵, while the electrophysiological and neurochemical correlates of precision engineered, cortical gain control (referred to as excitation-inhibition balance) suddenly acquire a clear teleology.

When applied to problems like figure-ground segregation¹⁶, the precision of prediction errors – say in primary visual cortex – is optimised to produce representational sharpening via lateral inhibition¹⁴. This requires the modulatory effects of descending *predictions of precision* to extend beyond the classical receptive field to produce extraclassical receptive field effects. It further requires the suppression representations that *do not* conform to the attended or inferred stimulus attribute; e.g., orientation or direction. In short, predictive coding predicts exactly the neuromodulation of cells reporting prediction errors (e.g., superficial pyramidal cells) in orthogonal perceptual dimensions or opposite preferences¹⁰.

In short, on the predictive coding view, one would expect that context-sensitive extraclassical receptive field effects would be mediated by excitatory feedback projections to pyramidal cells reporting prediction errors on stimulus attributes that were orthogonal or opposite to the currently inferred or attended attribute. See Figure 1 for a more detailed explanation. This representational sharpening contextualises the formation of prediction errors *per se* (that require top-down retinotopic projections to inhibitory interneurons in the classical receptive field).

It could be said that predictive coding – as a hypothesis for message-passing in cortical hierarchies – has yet to be empirically confirmed^{17,18}. An alternative view of the literature speaks to an enormous amount of anatomical and physiological evidence for predictive coding; particularly, in early visual processing (see ¹⁹ for a list of examples). One could take this view further with specific predictions that have subsequently been confirmed. A nice example (number 6 in the list above) is a spectral asymmetry in forward and backward message-passing during perceptual (visual) synthesis: “*Principal cells elaborating predictions (e.g., deep pyramidal cells) may show distinct (low-pass) dynamics, relative to those encoding error (e.g., superficial pyramidal cells)*”¹⁹ p21). This was subsequently confirmed four years later^{20,21} and is now almost a ‘meme’ when characterising laminar-specific neurophysiological responses^{22,23}.

Predictions that have subsequently been confirmed are not restricted to neurophysiology. For example, in neuroanatomy: “*As an example, a neural inference arising from the earliest formulations of predictive coding is that the source populations of forward and backward pathways should be completely separate, given their functional distinction; this aspect of circuitry - that neurons with extrinsically bifurcating axons*

do not project in both directions - has only recently been confirmed.”²⁴ p. 1792.

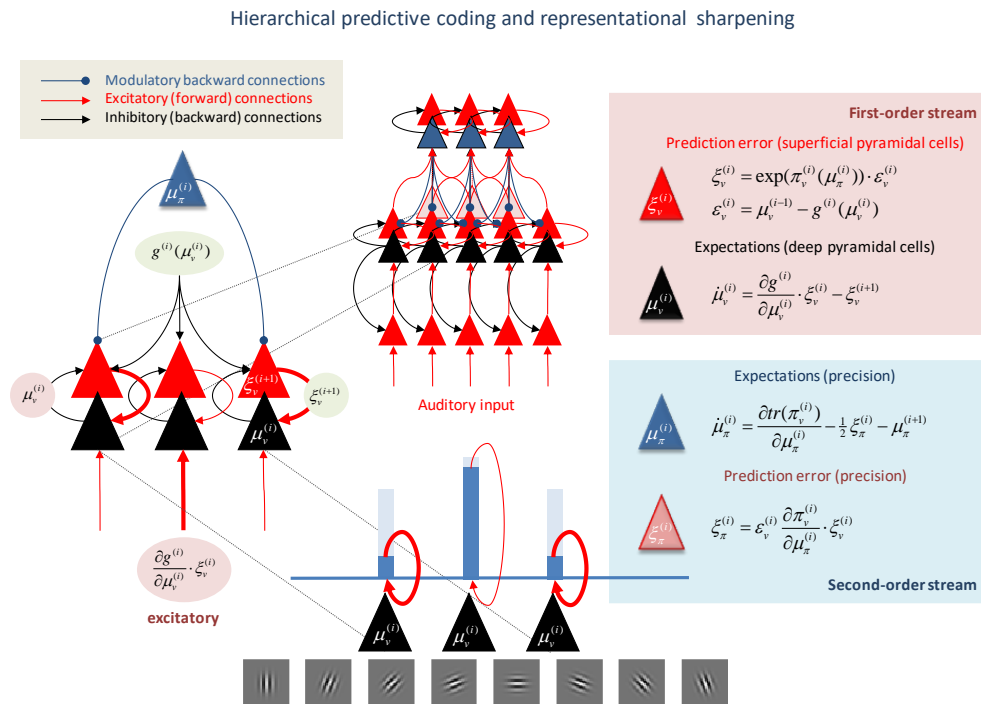


Fig. 1 | Hierarchical predictive coding: schematics that describe the hierarchical message passing implicit in predictive coding based on deep generative models. This message passing is described mathematically on the right as differential equations (i.e., neuronal dynamics). In this scheme, sensory input is conveyed to sensory (e.g., primary visual) cortex via ascending prediction errors (e.g., from the lateral geniculate). Posterior expectations, encoded by the activity of deep pyramidal cells, are driven by ascending prediction errors. At the same time, they are subject to lateral interactions that mediate (empirical) priors. Crucially, prediction errors are modulated by *predictions of their precision* (blue arrows). The predicted precision is based upon the expected precision (based on the sum of squares prediction errors). Heuristically, expectations about precision release posterior expectations from constraints in the vicinity of an inferred attribute or trajectory – and allow them to respond more sensitively to ascending input. This is illustrated in the left panel, which associates the terms in the differential equations with neuronal projections. The key point here is that prediction errors compete for influence over pyramidal cells representing stimulus features (i.e., expectations). If a representation (here, the black triangle in the middle) is released from top-down constraints, it is disinhibited and becomes more sensitive to ascending prediction error. In terms of extra-classical receptive field effects, this corresponds to a representational sharpening (as illustrated in the lower panel). Here, μ denote expectations, while ξ describes precision-weighted prediction errors ϵ . The generative model is denoted by $g(\mu)$ – a function that generates predictions from expectations. For a more detailed description, see¹⁶.

I introduced the target article by noting that perceptual inference (i.e., neurodynamics) and learning (i.e., neuroplasticity) are in the game of optimising the same thing; namely, model evidence or its variational equivalent (i.e., free energy). This remains as prescient today as it was 20 years ago. To see perception, learning and attention (and sensory attenuation) as working hand-in-hand towards the same imperative

Predictive coding

provides an integrative account that may still have an important message. On the one hand, there are still swathes of computational neuroscience that concern themselves almost exclusively with learning (e.g., reinforcement learning) that ignore the inference problem. On the other hand, vanilla predictive processing can often overlook the experience-dependent learning that accompanies evidence accumulation – and the Bayesian model selection (a.k.a., structure learning) of models *per se*. On one view, predictive coding inherits from perceptual psychology, while reinforcement learning is a legacy of behaviourism. This dialectic is also seen in machine learning (a.k.a., classification schemes), with deep learning on the one hand²⁵ and problems of data assimilation and uncertainty quantification on the other. There have been heroic attempts to repair this dialectic (e.g., amortization procedures in machine learning that, effectively, learn how to infer). However, these attempts do not appear to reflect the way that the brain has gracefully integrated perception and learning within the same computational anatomy. This may be important, if we aspire to creating artificial intelligence along neuromimetic lines. In short, perhaps the insight afforded by Rao and Ballard¹ – that learning and perception are two sides of the same coin – still has something important to tell us.

References

- 1 Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci.* **2**, 79-87 (1999).
- 2 Helmholtz, H. in *The Selected Writings of Hermann von Helmholtz* (ed R K Middletown) 384 (Wesleyan University Press, 1878 (1971)).
- 3 Gregory, R. L. Perceptions as hypotheses. *Phil Trans R Soc Lond B.* **290**, 181-197 (1980).
- 4 Dayan, P., Hinton, G. E. & Neal, R. The Helmholtz machine. *Neural Computation* **7**, 889-904 (1995).
- 5 Mumford, D. On the computational architecture of the neocortex. II. *Biol. Cybern.* **66**, 241-251 (1992).
- 6 Hohwy, J. The Self-Evidencing Brain. *Noûs* **50**, 259-285, doi:10.1111/nous.12062 (2016).
- 7 Friston, K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci.* **11**, 127-138 (2010).
- 8 Knill, D. C. & Pouget, A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* **27**, 712-719 (2004).
- 9 Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci.* **36**, 181-204 (2013).
- 10 Marques, T., Nguyen, J., Fioreze, G. & Petreanu, L. The functional organization of cortical feedback inputs to primary visual cortex. *Nature Neuroscience* **21**, 757-764, doi:10.1038/s41593-018-0135-z (2018).
- 11 Clark, A. The many faces of precision. *Front Psychol.* **4**, 270 (2013).
- 12 Feldman, H. & Friston, K. J. Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience* **4**, 215 (2010).
- 13 Parr, T. & Friston, K. J. Working memory, attention, and salience in active inference. *Scientific reports* **7**, 14678, doi:10.1038/s41598-017-15249-0 (2017).
- 14 Kok, P., Jehee, J. F. & de Lange, F. P. Less is more: expectation sharpens representations in the primary visual cortex. *Neuron* **75**, 265-270, doi:10.1016/j.neuron.2012.04.034 (2012).

Predictive coding

- 15 Powers, A. R., Mathys, C. & Corlett, P. R. Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science* **357**, 596-600, doi:10.1126/science.aan3458 (2017).
- 16 Kanai, R., Komura, Y., Shipp, S. & Friston, K. Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**, doi:10.1098/rstb.2014.0169 (2015).
- 17 Heilbron, M. & Chait, M. Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience*, doi:https://doi.org/10.1016/j.neuroscience.2017.07.061 (2017).
- 18 Brodski-Guerniero, A. *et al.* Information-Theoretic Evidence for Predictive Coding in the Face-Processing System. *J Neurosci* **37**, 8273-8283, doi:10.1523/JNEUROSCI.0614-17.2017 (2017).
- 19 Friston, K. Hierarchical models in the brain. *PLoS Comput Biol.* **4**, e1000211 (2008).
- 20 Bastos, A. M. *et al.* Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* **85**, 390-401, doi:10.1016/j.neuron.2014.12.018 (2015).
- 21 Arnal, L. H., Wyart, V. & Giraud, A. L. Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat Neurosci* **14**, 797-801, doi:10.1038/nn.2810 (2011).
- 22 Mejias, J. F., Murray, J. D., Kennedy, H. & Wang, X.-J. Feedforward and feedback frequency-dependent interactions in a large-scale laminar network of the primate cortex. *Science Advances* **2**, doi:10.1126/sciadv.1601335 (2016).
- 23 Michalareas, G. *et al.* Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. *Neuron* **89**, 384-397, doi:10.1016/j.neuron.2015.12.018 (2016).
- 24 Shipp, S. Neural Elements for Predictive Coding. *Front Psychol* **7**, 1792, doi:10.3389/fpsyg.2016.01792 (2016).
- 25 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436-444, doi:10.1038/nature14539 (2015).

Competing interests

The author declares no competing interests.