Corrected: Publisher Correction

HISTORICAL NEWS & VIEWS: NEURAL CODING

Does predictive coding have a future?

In the 20th century we thought the brain extracted knowledge from sensations. The 21st century witnessed a 'strange inversion', in which the brain became an organ of inference, actively constructing explanations for what's going on 'out there', beyond its sensory epithelia. One paper played a key role in this paradigm shift.

Karl Friston

Every decade or so, one reads a paper that makes you think "well, that's quite remarkable". In 1999, Rao and Ballard[1] offered a treatment of visual processing as predictive coding. In 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, i.e., the 'newsworthy' information that cannot be predicted. These prediction errors drive expectations in higher levels toward better explanations for lower levels. Using simulations they showed that this simple (hierarchical) architecture was not only consistent with neuroanatomy and physiology but could also account for a range of subtle response properties such as 'end-stopping' among other extraclassical 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 optimized the same energy or cost function. I remember reading the methods section several times to convince myself that they could explain all of this functional 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: to minimize prediction error.

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 Helmholtz, whose ideas led to epistemological automata, analysis-by-synthesis, and perception as hypothesis testing[2]. Subsequent formalizations within machine learning and information theory then led to specific proposals for computational architectures in the neocortex[3,4]. The theme that runs through this legacy is inference and learning the best explanation for our sensorium. In other words, the brain is in the game of optimizing neuronal dynamics and connectivity to maximize the evidence for its model of the world[5].

So what form does this evidence take? For a statistician, it is just Bayesian model evidence: the probability of observing some data given a model of how those data were generated. In machine learning, the evidence comprises a variational bound on log-evidence. In engineering, it is the cost functions associated with Kalman filters. For an information theorist, it would be the efficiency or minimum description length. Finally, in the realm of predictive coding, the evidence is taken as the (precision weighted) prediction error. Crucially, these are all the same thing, which, in my writing, is variational free energy[6].

Predictive coding offered a compelling process theory that lent notions like the Bayesian brain[7] 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 optimize Bayesian model evidence. Within a decade, the Bayesian brain hypothesis and predictive coding became dominant models 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 emphasized an active sampling of the sensory world. There has been a remarkable uptake of these ideas in fields as diverse as philosophy[5,8], 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.

A recent example is a report from Marques et al.[9], looking at the functional organization 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."[9] (p. 757).

To understand this particular aspect of feedback, we need to consider the role of 'precision-weighted' prediction errors that mediate belief updating. In predictive coding, precision corresponds to the best estimate of the reliability or inverse variance of prediction errors. Heuristically, only precise prediction errors matter for belief updating, where estimating the precision is like estimating the error variance in statistics (i.e., a small standard error corresponds to high precision). Technically, getting the precision right corresponds to optimizing the Kalman gain in Bayesian or Kalman filters[1]. Computationally, it underlies the optimal mixing of sensory streams that differ in their reliability, as in multimodal sensory integration[7]. Psychologically, precision-weighting has been associated with sensory attention and attenuation[10]. Mechanistically, precision-weighting is thought to be mediated by neuromodulatory mechanisms; for example, classical neuromodulators of synchronous gain. In short, most of the interesting bits of predictive coding are about getting the precision right: selecting newsworthy, uncertainty-resolving prediction errors.

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 perception. Failures of the neuromodulatory basis of precision-weighting have figured prominently in explanations for false inference and psychopathology[11], while the electrophysiological and neurochemical

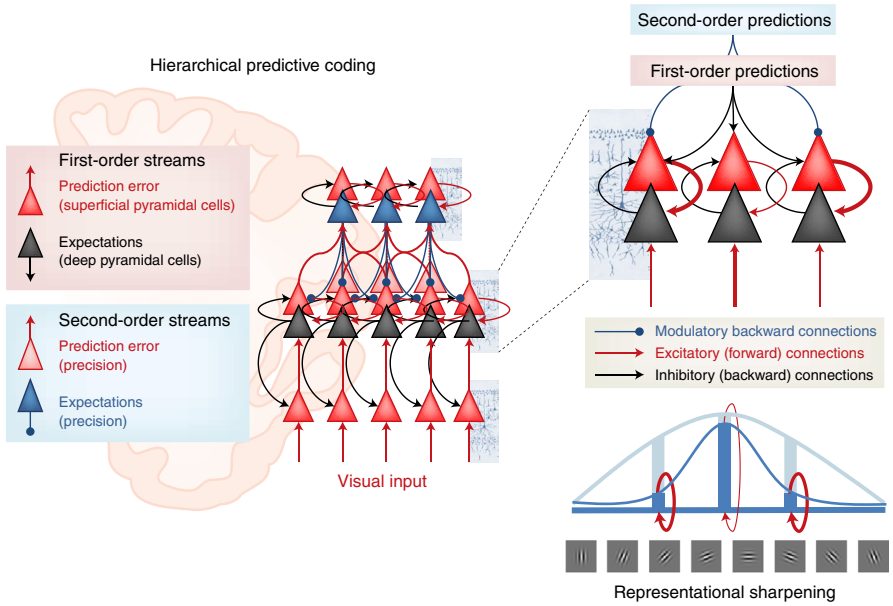

Fig. 1 | Hierarchical predictive coding: schematics that describe the hierarchical message passing implicit in predictive coding based on deep generative models. In this scheme, sensory input is conveyed to sensory (for example, primary visual) cortex via ascending prediction errors (for example, from the lateral geniculate). Posterior expectations, encoded by the activity of deep pyramidal cells, are driven by ascending prediction errors (red arrows). These cells then provide descending predictions (black arrows) that inform prediction errors at the lower level. 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 on the sum of squared prediction errors. This means we have two sets of ascending and descending counter streams: the first dealing with predictions of (first-order) content and the second dealing with (second-order) context; namely, the precision of first-order 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 on the lower right (representational sharpening). The key point here is that prediction errors compete for influence over pyramidal cells representing stimulus features (i.e., expectations). If a representation is released from top-down constraints, it is disinhibited and becomes more sensitive to ascending prediction error. Conversely, if a particular prediction error is afforded greater top-down precision, it effectively pulls the predictive expectation toward its prior mean of zero, as illustrated by the red arrows in the lower right panel. In this example, the activity of the middle deep pyramidal cell (black triangle on the upper right) could encode the expected orientation of a local stimulus (indicated by the Gabor patches on the lower right). In terms of extraclassical receptive field effects, this corresponds to representational sharpening. For a more detailed description of the implicit belief updating and accompanying neuronal dynamics, see ref. [10]. Credit: Katie Vicari/Springer Nature

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 segregation[10], the precision of prediction errors— say in primary visual cortex—is optimized to produce representational sharpening via lateral inhibition. 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 of representations that do not conform to the attended or inferred stimulus attribute. See Fig. 1 for a more detailed explanation. This representational sharpening contextualizes the formation of prediction errors per se and requires top-down retinotopic projections to inhibitory interneurons in the classical receptive field. In short, predictive coding predicts the neuromodulation of cells reporting prediction errors (for example, superficial pyramidal cells) in orthogonal perceptual dimensions or opposite preferences. This is exactly the phenomena reported empirically in Marques et al.[9].

It is sometimes said that predictive coding—as a hypothesis for message-passing in cortical hierarchies—has yet to be empirically confirmed. 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 ref. [12] for a list of examples). One could take this view further with reference to specific predictions that have subsequently been confirmed. A nice example (number 6 in the list presented in ref. [12]) is a spectral asymmetry in forward and backward message-passing during perceptual (visual) synthesis: "[p]rincipal cells elaborating predictions (e.g., deep pyramidal cells) may show distinct (low-pass) dynamics, relative to those encoding error (e.g., superficial pyramidal cells)"[12] (p. 21). This was subsequently confirmed several years later[13,14] and is now almost a 'meme' when characterizing laminar-specific neurophysiological responses.

The predictive validity of predictive coding is not restricted to neurophysiology; it also encompasses neuroanatomy: "[a]s 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."[15] (p. 1792).

I introduced the target article by noting that perceptual inference (i.e., neurodynamics) and learning (i.e., neuroplasticity) are in the game of optimizing 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, attention, and sensory attenuation as working hand-in-hand toward the same imperative provides an integrative account that may still have an important message. There are still swathes of computational neuroscience that concern themselves almost exclusively with learning and ignore the inference problem (for example, reinforcement learning). Conversely, vanilla predictive processing can often overlook the experience-dependent learning that accompanies evidence accumulation, as well as the Bayesian model selection (a.k.a. structure learning) of models per se. This polarization may reflect the differences in conceptual lineage: predictive coding takes its lead from perceptual psychology, while reinforcement learning is a legacy of behaviorism. This dialectic is also seen in machine learning, with deep learning on the one hand and problems of data assimilation and uncertainty quantification on the other. The have been heroic attempts to bridge this

gap (for example, 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 create artificial intelligence along neuromimetic lines. In short, perhaps the insight afforded by Rao and Ballard[1]—that learning and perception are two sides of the same coin—may still have something important to tell us. ❐

Karl Friston
The Wellcome Centre for Human Neuroimaging, University College London, London, UK.
e-mail: k.friston@ucl.ac.uk

Published online: 23 July 2018
https://doi.org/10.1038/s41593-018-0200-7

References

1. Rao, R. P. & Ballard, D. H. *Nat. Neurosci.* **2**, 79–87 (1999).
2. Gregory, R. L. *Philos. Trans. R. Soc. Lond. B* **290**, 181–197 (1980).
3. Dayan, P., Hinton, G. E., Neal, R. M. & Zemel, R. S. *Neural Comput.* **7**, 889–904 (1995).
4. Mumford, D. *Biol. Cybern.* **66**, 241–251 (1992).
5. Hohwy, J. *Noûs* **50**, 259–285 (2016).
6. Friston, K. *Nat. Rev. Neurosci.* **11**, 127–138 (2010).
7. Knill, D. C. & Pouget, A. *Trends Neurosci.* **27**, 712–719 (2004).
8. Clark, A. *Behav. Brain Sci.* **36**, 181–204 (2013).
9. Marques, T., Nguyen, J., Fioreze, G. & Petreanu, L. *Nat. Neurosci.* **21**, 757–764 (2018).
10. Kanai, R., Komura, Y., Shipp, S. & Friston, K. *Phil. Trans. R. Soc. Lond. B* https://doi.org/10.1098/rstb.2014.0169 (2015).
11. Powers, A. R., Mathys, C. & Corlett, P. R. *Science* **357**, 596–600 (2017).
12. Friston, K. *PLoS Comput. Biol.* **4**, e1000211 (2008).
13. Bastos, A. M. et al. *Neuron* **85**, 390–401 (2015).
14. Arnal, L. H., Wyart, V. & Giraud, A. L. *Nat. Neurosci.* **14**, 797–801 (2011).
15. Shipp, S. *Front. Psychol.* **7**, 1792 (2016).

Competing interests

The author declares no competing interests.

SYNAPTIC PLASTICITY

FRETting over postsynaptic PKC signaling

Protein kinases are key regulators of excitatory synapse plasticity. In this issue, using novel optical reporters of protein kinase C (PKC) activity, Colgan et al. identify PKCα as critical for integrating NMDA receptor and neurotrophin signaling to control dendritic spine structural plasticity, synaptic potentiation, and learning and memory.

Mark L. Dell'Acqua and Kevin M. Woolfrey

Scientists have been fascinated for decades by how excitatory synapses on dendritic spines are remodeled by activity—and how this in turn may underlie learning and memory—but the details have been challenging to pin down. There is little doubt that kinase activity is important, but the sheer number of synaptic protein kinases and the intricacies of their functions pose a major challenge to unraveling their specific roles in synaptic plasticity. For example, the PKC family of kinases was first implicated in synaptic plasticity over 30 years ago, yet after a flood of papers in the late 1980s, this field plateaued, not least because of the complexity of PKC signaling. Conventional PKCs (α, β, γ) are signaling hubs that require binding to intracellular Ca^{2+} and phosphatidylserine and diacylglycerol (DAG) at the plasma membrane for activation. They modify a variety of important synaptic substrates. Before the work of Colgan et al.[1], featured in this issue of *Nature Neuroscience*, the identity of the PKC isozyme primarily responsible for long-term potentiation (LTP) and its associated spine enlargement, or structural LTP (sLTP), had remained elusive. By combining powerful imaging techniques in organotypic slices with behavior and electrophysiology across knockout mouse lines, Colgan et al. not only profile the activation of each conventional PKCα,β,γ isozyme in dendritic spines, but also identify key roles for PKCα in LTP, dendritic spine sLTP, and learning and memory[1].

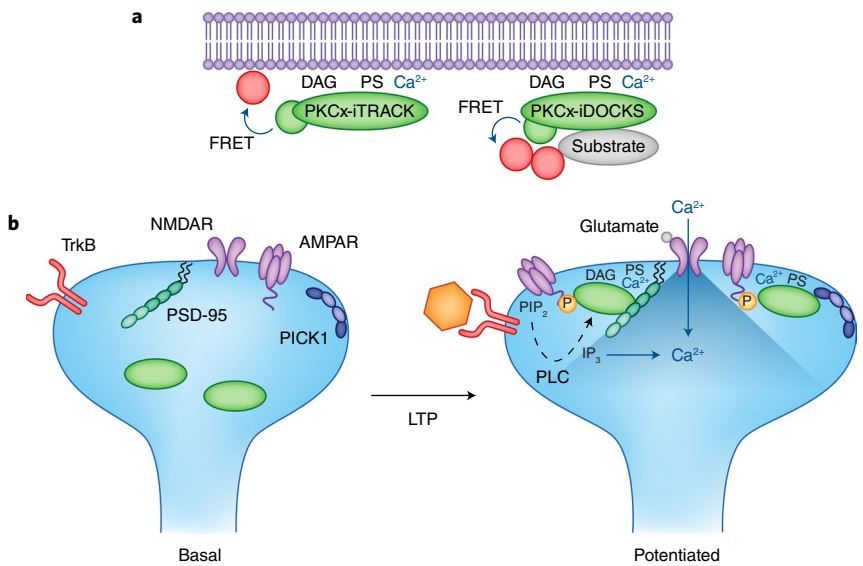

Fig. 1 | New FRET probes reveal that PKCα is a postsynaptic signal integrator that is uniquely responsible for structural long-term potentiation of excitatory synapses on dendritic spines. **a**, FRET-based optical reporters of membrane translocation (iTRACK) and substrate binding (iDOCKS), which reflect kinase activation, for each of the three PKC isozymes, PKCα, β and γ (PKCx). **b**, NMDA receptor (NMDAR) and autocrine BDNF signaling combine to activate PKCα in dendritic spines. Active PKCα is then targeted to postsynaptic microdomains through interactions with PDZ-domain-containing proteins such as PSD-95 and PICK1. AMPAR, AMPA-type glutamate receptor; P, phosphate; PIP_2, phosphatidylinositol-4,5-bisphosphate; PS, phosphatidylserine.

A long-standing question regarding the involvement of conventional PKCs in LTP is whether each isozyme has unique