Perceptual inference: a matter of predictions and errors

Peter Kok

Summary

A recent study finds that separate populations of neurons in inferotemporal cortex code for predictions and prediction errors, providing evidence for predictive coding theories of perception.

Main text

More than a century ago, Helmholtz described perception as a process of unconscious inference — i.e. trying to infer the most likely causes of our sensory inputs given our prior knowledge of the world. In recent decades, interest in this perspective has been rekindled by new insights from computer science and neuroscience, leading to theories on how the brain could accomplish such inference. One highly influential theory is 'predictive coding' [1, 2], according to which cortical regions constantly generate hypotheses (or 'predictions') about the likely causes of their inputs. For instance, when presented with the stimulus in Figure 1A, a region specialised in processing simple geometrical shapes may generate the hypothesis of a white triangle partially occluding three black circles [3]. In addition to representing these predictions, each cortical region also encodes how they differ from current sensory inputs. These 'prediction errors' allow for efficient updating of hypotheses. This coding scheme can account for many properties of how neurons behave in early visual cortex [2, 4], and has received indirect support from neuroimaging [3, 5–8] and electrophysiology [9, 10] in humans, and from electrophysiology in monkeys [11]. However, to date there has been a noticeable lack of evidence for a central tenet of predictive coding theory, and one that distinguishes it from other theories of perceptual inference [12] — that predictions and prediction errors are explicitly and separately represented within a given cortical region (Figure 1B). Consistent with the theory, a new study measuring single unit responses in macaques reports encoding of predictions and prediction errors by separate neural populations in inferotemporal cortex (IT) [13].

In their study, Bell and colleagues [13] presented monkeys with images of faces and fruits, while a latent variable — not revealed to the monkeys — determined the relative probability of each image category. Despite the implicit nature of this manipulation, neural responses in IT (known to be involved in processing complex visual stimuli) were strongly modulated by image predictability. First, averaged over all face-responsive cells, neural firing rates were higher for unexpected vs. than expected faces, consistent with the encoding of prediction errors. Second, multivariate analyses revealed that population activity encoded the probability of a face occurring, even before an image was presented. Thus, IT encodes both predictions about upcoming sensory input, as well as the mismatch between these predictions and the input that was actually received (Figure 1C).

One strong prediction made by predictive coding theories is that these two signals, predictions and prediction errors, are represented in separate populations of neurons. One way to establish this would be to record from neurons in different cortical layers: In the theory, predictions generated by a cortical region are sent back to explain its inputs from lower-level regions, and thus they should reside in feedback-providing deep layers; prediction errors, on the other hand, are sent forward as input to higher-level regions, and thus should reside in superficial layers [1, 14]. Bell and colleagues did not measure the cortical depth of the neurons from which they recorded, but they did address this issue in a different way: by examining the relationship between the signals across neurons. Although the strength of neurons' face (vs. fruit) preference correlated positively with both their encoding of the *a priori* likelihood of a face appearing (i.e., prediction) and their enhanced response for unexpected vs. expected faces (i.e., prediction error), there was strikingly no correlation between prediction and prediction error encoding across neurons. In other words, the population of face-sensitive neurons appeared to consist of two orthogonal subpopulations, one encoding predictions and the other prediction errors.

Future efforts should now be directed toward characterising and localising these subpopulations. One possible explanation for subpopulations of neurons sharing tuning preferences (here, faces) but encoding different variables could be that they reside in different layers of the same columns. That is, neurons within a cortical column are usually tuned for the same visual features, but they differ in their (feedforward, lateral, and feedback) connectivity, and may thus receive and transmit different messages from and to different cortical regions [12, 14]. This leads to the testable prediction that each cortical column contains both prediction and prediction error neurons, in different layers. One exciting prospect is that human fMRI studies are beginning to uncover layer-specific BOLD responses, allowing for the study of these neuronal subpopulations noninvasively in humans [15, 16]. Such future work, both in animals and humans, will shed more light on the exact neural circuitry that implements perceptual inference. This is particularly important because several implementations of predictive coding have been proposed, differing in the nature of feedback connections (inhibitory vs. excitatory) and the localisation of prediction and prediction error neurons [1, 2, 4, 14].

One somewhat surprising feature of the prediction signals revealed by Bell and colleagues is their long-lasting nature. The signal starts before the image is presented, but continues until well after image onset (~500 ms). In other words, the *a priori* likelihood of a face appearing is still being encoded even after an image of a piece of fruit has been presented, whereas one might expect that the prediction error caused by this image would have led to immediate updating of the face prediction. That is, after a fruit has been presented, the likelihood of it being a face is zero. One possibility is that the prediction lingers in anticipation of upcoming trials. More generally, it remains an open question whether this temporal profile reflects the statistical structure of this particular experiment — with face probability being modulated by a slowly changing variable, as opposed to being cued one a trial-by-trial basis — or whether this is a general feature of the cortical encoding of predictions. Prediction signals would indeed be expected to evolve more slowly than prediction error signals, since the former integrate over the latter [14], but perhaps not quite at this timescale.

Another important question is which brain regions are responsible for keeping track of probabilities and regularities in the environment. This is likely to depend strongly on the type of regularity and its timescale, but several candidate regions have been proposed. For instance, the hippocampus can extract statistical regularities from sensory inputs, generate predictions based on these regularities, and send them to visual cortex [17, 18]. Alternatively, regions of prefrontal cortex have also been argued to generate sensory predictions [19, 20].

Predictive coding offers a different perspective than traditional theories of sensory processing on the type of information represented in sensory neurons. It suggests that neurons do not simply encode features of their bottom-up input, but rather hypotheses about what is out there in the world, as well as the mismatch between these hypotheses and current sensory data. The new study by Bell and colleagues brings us one step closer to understanding the neural circuitry underlying this process of perceptual inference, and suggests exciting new avenues for uncovering the neural basis of perception.

References

1. Friston, K. (2005). A theory of cortical responses. Philos. Trans. R. Soc. B Biol. Sci. *360*, 815–836.
2. Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. *2*, 79–87.
3. Kok, P., and de Lange, F. P. (2014). Shape Perception Simultaneously Up- and Downregulates Neural Activity in the Primary Visual Cortex. Curr. Biol. *24*, 1531–1535.
4. Spratling, M. W. (2010). Predictive Coding as a Model of Response Properties in Cortical Area V1. J. Neurosci. *30*, 3531–3543.

5. den Ouden, H. E. M., Friston, K. J., Daw, N. D., McIntosh, A. R., and Stephan, K. E. (2009). A Dual Role for Prediction Error in Associative Learning. Cereb. Cortex *19*, 1175–1185.
6. Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., and Muckli, L. (2010). Stimulus Predictability Reduces Responses in Primary Visual Cortex. J. Neurosci. *30*, 2960–2966.
7. Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., and Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. Nat. Neurosci. *11*, 1004–1006.
8. Kok, P., Jehee, J. F. M., and de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. Neuron *75*, 265–270.
9. Todorovic, A., van Ede, F., Maris, E., and de Lange, F. P. (2011). Prior Expectation Mediates Neural Adaptation to Repeated Sounds in the Auditory Cortex: An MEG Study. J. Neurosci. *31*, 9118–9123.
10. 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. Proc. Natl. Acad. Sci. *108*, 20754–20759.
11. Meyer, T., and Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. Proc. Natl. Acad. Sci. *108*, 19401–19406.
12. Lee, T. S., and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. J. Opt. Soc. Am. A *20*, 1434.
13. Bell, A. H., Summerfield, C., Morin, E. L., Malecek, N. J., and Ungerleider, L. G. (in press). Encoding of stimulus probability in macaque inferior temporal cortex. Curr. Biol.
14. Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. Neuron *76*, 695–711.
15. Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., Goebel, R., and Yacoub, E. (2015). Contextual Feedback to Superficial Layers of V1. Curr. Biol. *25*, 2690–2695.
16. Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., and de Lange, F. P. (2016). Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. Curr. Biol. *26*, 371–376.
17. Schapiro, A. C., Kustner, L. V., and Turk-Browne, N. B. (2012). Shaping of Object Representations in the Human Medial Temporal Lobe Based on Temporal Regularities. Curr. Biol. *22*, 1622–1627.
18. Hindy, N. C., Ng, F. Y., and Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. Nat. Neurosci. *19*, 665–667.
19. Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., and Hirsch, J. (2006). Predictive Codes for Forthcoming Perception in the Frontal Cortex. Science *314*, 1311–1314.
20. Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hämäläinen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., et al. (2006). Top-down facilitation of visual recognition. Proc. Natl. Acad. Sci. U. S. A. *103*, 449–454.

Affiliation
Princeton Neuroscience Institute, Princeton University, 301 Peretsman-Scully Hall, Princeton, NJ 08544, USA

E-mail address
pkok@princeton.edu

Figure

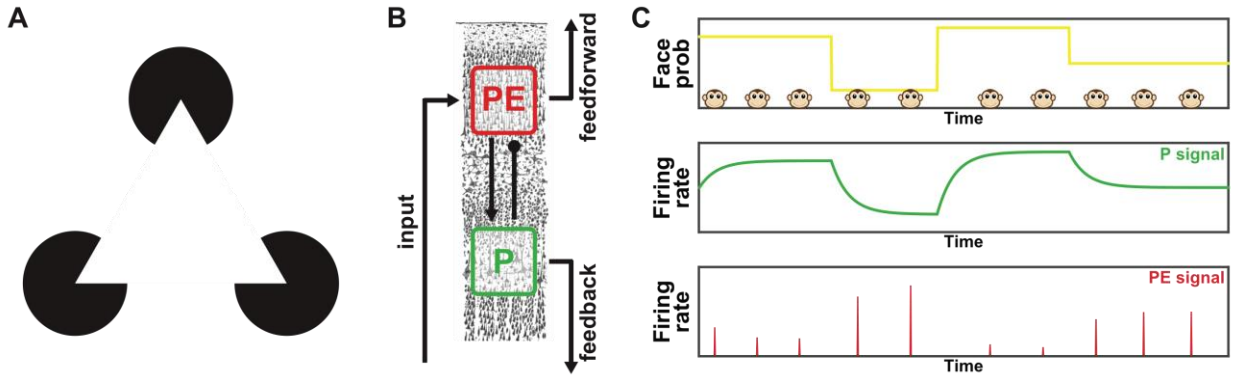

Figure 1. Perceptual inference implemented through predictions and prediction errors.
(A) The Kanizsa illusion, an example of perceptual inference. (B) A simple predictive coding circuit, with separate neurons encoding predictions (P) and prediction errors (PE). (C) Bell and colleagues manipulated the probability of upcoming images being faces (vs. fruits) over time (top panel, yellow trace). Thus, images of faces could be presented either when they were expected or when they were unexpected. Separate subsets of IT neurons encoded the *a priori* probability of a face appearing (P signal, middle panel), and the unexpectedness of face presentations (PE signal, lower panel).