Feature-specific predictive processing: What's in a prediction error?

David Richter[1,2], Cem Uran[3,4], Martin Vinck[3,4], Floris P de Lange[2]

1 Mind, Brain and Behavior Research Center (CIMCYC), University of Granada, Granada, Spain
2 Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, the Netherlands
3 Ernst Strüngmann Institute (ESI) for Neuroscience, Frankfurt am Main, Germany
4 Donders Centre for Neuroscience, Department of Neurophysics, Radboud University Nijmegen, Nijmegen, the Netherlands

Abstract

Despite numerous studies reporting sensory prediction errors - a key component of predictive processing theories - the nature of the surprise represented in these errors remains largely unknown. Here we highlight recent studies, which provide evidence that prediction errors, even in early sensory areas, may reflect high-level surprise, offering new insights into the role of predictive processing in the brain beyond classical accounts of redundancy reduction.

Keywords: convolutional neural networks, deep neural networks, firing rates, fMRI, gamma oscillations, hierarchy, predictability, prediction error, predictive coding, predictive processing, visual cortex

The brain has been described as a "prediction machine" that generates expectations about its environment to facilitate perception and decision-making [1,2]. Studies have demonstrated that neural responses are positively correlated to the amount of surprise that a stimulus elicits [3]. These studies conventionally compared neural responses between two conditions: predicted and surprising stimuli. In an experiment, these two categories can be well defined, e.g. by presenting specific temporal sequences of stimuli that define predicted and surprising stimuli. Indeed, experiments often use simple artificial stimuli (e.g., gratings) and manipulate one specific stimulus feature (e.g., orientation) to create these two conditions [4]. However, natural stimuli are complex constellations of many lower and higher-order features. Consequently, stimulus predictability naturally varies across many dimensions, which we argue is critical to investigate for charting the neurocomputational principles underlying perceptual inference. For example, if we move our eyes towards the location where we hear the honking sound of a car, we can strongly predict some features (e.g., specific rectilinear features and object size) while other features are completely unpredictable (e.g., the color of the car). In other words, a stimulus may be both predicted and surprising, depending on the specific features.

In our opinion this poses a fundamental question: *What* does the visual system predict? Instead of focusing solely on the magnitude of prediction error responses, new research has attempted to expose the contents of prediction errors at different levels of the cortical hierarchy, which can elucidate the brain's internal models and thereby constrain models of predictive processing. We refer to this new paradigm as *feature-specific predictive processing.* Using convolutional neural networks (CNN), which have successfully been used to study the encoding of natural images by the visual ventral stream [5], recent studies have started to investigate this by decomposing sensory surprise into lower- and higher-level visual features.

First, Uran et al. [6] explored how spatial predictability in natural images influences neural responses in macaque V1. Spatial predictability here referred to how well an image part could be predicted by its spatial surround, which was quantified using an inpainting algorithm. Then, they quantified the difference in CNN (Fig 1A) activity patterns between the predicted and actual image patch, at low (early layers of the CNN, tuned to simple features like orientation and contrast) and high (late layers of the CNN, tuned to more complex features, such as objects and complex textures) levels, to determine low- and high-level predictability. They then investigated how low- and high-level predictability modulated distinct aspects of neural activity (Fig 1B). V1 firing rates were weakly associated with low-level predictability but were strongly (and negatively) correlated with high-level predictability.

Second, Heilbron and de Lange [7] used a similar analysis approach, investigating spatial predictability in mice. Results showed that visual cortical responses in mouse primary visual areas were again primarily modulated by high-level feature predictability. Moreover, this modulation by high-level predictability was most pronounced in superficial layers of V1.

Third, Richter et al. used fMRI [8] and EEG [9] in humans to explore how temporal predictability affects the neural response in the visual ventral stream. Participants learned to expect a specific image given a cue. Following learning, participants were occasionally presented with a different, surprising, image, which could be differentially surprising in terms of low-level and high-level features. As expected, the ‘bottom-up’ stimulus feature tuning profile adhered to the classical visual hierarchy: V1 was tuned for simple features, reflected in early layers of the CNN, while higher-order visual regions were tuned to more complex features, reflected in late layers of the CNN (Fig 1C, upper panel). Interestingly however, *all* visual areas from early occipital to higher-order ventral regions, including V1, were upregulated as a function of high-level, but not low-level, surprise (Fig 1C, bottom panel) [8]. Moreover, these modulations by high-level surprise arose relatively early, within 200ms following stimulus onset [9].

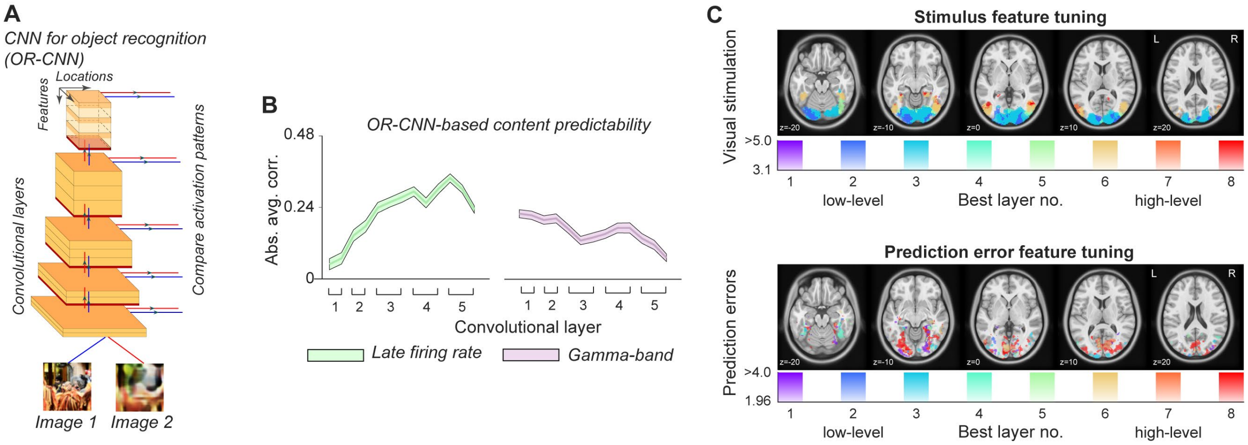

Fig. 1 (A) Object classification CNNs (VGG-16 in Uran et al. [6], and AlexNet in Richter et al. [8,9]) were used to quantify stimulus predictability at different feature levels. **(B)** Correlation of macaque V1 neural activity with CNN predictability. Late multi-unit firing rates showed a significant increase in correlation across layers with image unpredictability, while γ-power showed a significant decrease. Figure adapted from Uran et al. [6]. **(C)** Representational similarity analysis showed a gradient from low-level to high-level visual feature processing along the ventral visual hierarchy during prediction-free contexts (top panel), with early visual cortex (EVC) aligning most with early CNN layers (cold colors) and higher visual areas more with late CNN layers (warm colors). In contrast, prediction error magnitudes increased with high-level visual feature surprise across the visual system, including in EVC (bottom panel). Figure adapted from Richter et al. [8].

We believe that despite differences in the type of predictions (spatial vs. temporal), populations studied (macaques, mice, humans), and recording methods (multi-unit, Neuropixel probes, fMRI, EEG), the above studies converge on a similar motif, namely that high-level visual surprise strongly drives neural activity, even in the earliest cortical visual areas, potentially as a result of feedback. Next, we discuss how these findings align with several models of predictive processing:

(1) Hierarchical predictive coding (HPC). HPC is a major model of cortical function that posits that feedback signals predictions about neural representations in the lower hierarchical level, while feedforward connections convey prediction errors arising from the comparison between top-down predictions and local representations [2]. Arguably, HPC thereby entails that higher areas signal prediction errors about higher-order features, while prediction errors about lower-order features are signaled in lower areas, because each area generates errors relative to predictions from the next hierarchical level. Thus, we argue that the findings of Uran et al. [6], Heilbron and de Lange [7], and Richter et al. [8] that V1 activity increases for high- rather than low-level surprise may be difficult to reconcile with classic HPC implementations that depend on strict hierarchical error representation.

(2) Feedback propagation of error signals. An alternative to classic HPC is that error signals are generated in higher hierarchical levels and transmitted to lower levels via feedback connections, where they lead to increased neural activity. V1 receives feedback from many higher-order sensory areas, allowing a direct modulation by top-down signals from many areas [10]. In this scenario, feedback could serve to increase attention [11], update synaptic weights, aid in learning and novelty detection [12], and recruit additional neural resources when surprising events are encountered. This class of models would predict that V1 activity increases for high-level surprise, which we believe is consistent with the observations of the studies summarized above [6–8]. In this case, error signals may act as a scalar surprise signal that gain-modulates V1 responses, rather than subtract specific predictions [4].

(3) V1 as a comparator circuit for higher-level features. Another possibility, which we consider to be in line with these recent findings, is that low-level areas like V1 may act as a comparator circuit for higher-order features. Local features can provide evidence for objects, and object recognition can be approximated by globally averaging over many local evidence filters [13]. For example, a local patch of yellow in a V1 receptive field provides some evidence for a banana, a sun, or an autumn leave, and if feedback into V1 signals the prediction of a kiwi, a higher-order prediction error could result when a local feature is not compatible with the high-level prediction.

(4) Dendritic HPC. The study of Uran et al. [6] showed that while high-level predictability modulated V1 firing rates, low-level prediction errors primarily affected 30-80 Hz gamma-band synchronization within V1. We consider the dissociation between firing rates and gamma-synchronization to fit well within the recently proposed dendritic HPC model. Here, low-level predictability should reflect an increased cancellation (and efficient encoding) of feedforward excitatory inputs by local inhibitory feedback, leading to a balanced excitation-inhibition state that promotes gamma-synchronization [14]. In dendritic HPC models [15], these feedforward prediction errors would be encoded at the basal dendrites of pyramidal neurons. The increase in V1 firing rates with higher-order predictability, on the other hand, should reflect prediction errors on the apical dendrites, which are the recipient of top-down feedback. Such apical prediction errors may lead to lasting increase in V1 firing rates by generating dendritic plateau potentials.

Notably, V1 receives feedback from many higher-order areas [10], such that feedback concerning high-level features and error signals may reach it via direct projections. The fact that V1 receives feedback from many hierarchical levels, rather than only from the next level, as assumed in some predictive coding models, in our opinion provides a plausible anatomical basis for the finding that V1 activity signals mismatch signals about high-level features.

Finally, does this feature-specific predictive processing strategy make sense from a behavioral standpoint? As the brain may be a “predictive machine” that continuously refines its internal models, a key question is: What are these predictions about? In principle, the brain could predict across all hierarchical levels, including fine-grained details of its sensory input, down to the level of pixels - as classic HPC models may suggest. However, higher-level features are likely more behaviorally relevant and may generally be easier to predict than lower-level features (e.g., rain vs individual drops of rain). Indeed, spatiotemporal prediction tasks like predicting future frames can be learned more effectively at the level of higher-order features rather than at the pixel level [16], and high-level predictions may be particularly effective for self-supervised learning of visual representations [7]. Finally, high-level stimulus predictability shows stronger correlations with human perceptual similarity judgements and salience maps [6]. Thus, learning and behavior may be primarily guided by high-level predictability, and therefore we argue that perhaps the brain’s “prediction game” is focuses on high-level, not low-level features in its sensory input.

Together, these studies demonstrate what kind of predictions the brain makes in richer, more naturalistic, environments. A future challenge will be to better integrate the relevant biological implementation principles (as outlined in e.g. dendritic HPC) with algorithmic and computational goals to generate a unified theory of predictive processing that spans across all of Marr's levels of analysis.

Acknowledgements

Funding

This work was supported by the Marie Skłodowska-Curie Grant "PreVision" (Project 101147241) awarded to DR, and ERC CoG (101000942) "SURPRISE" and NWO Vici VI.C.231.043 awarded to FPdL. MV was supported by an ERC starting grant (850861) SPATEMP, DFG VI Grants (908/5-1 and 908/7-1; 505660261; 520285844; SPP LOOPS), an NWO VIDI Grant, and the Dutch Brain Interface Initiative (DBI2). The funders had no role in the decision to publish or preparation of the manuscript.

Contributions

All authors contributed to writing, reviewing, and editing of the manuscript.

Declaration of Competing Interest

The authors declare the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Friston, K. (2005) A theory of cortical responses. *Phil. Trans. R. Soc. B* 360, 815–836
2. 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. *Nat Neurosci* 2, 79–87
3. de Lange, F.P. *et al.* (2018) How Do Expectations Shape Perception? *Trends in Cognitive Sciences* 22, 764–779
4. Furutachi, S. *et al.* (2024) Cooperative thalamocortical circuit mechanism for sensory prediction errors. *Nature* 633, 398–406
5. Yamins, D.L.K. *et al.* (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8619–8624
6. Uran, C. *et al.* (2022) Predictive coding of natural images by V1 firing rates and rhythmic synchronization. *Neuron* 110, 1240-1257.e8
7. Heilbron, M. and de Lange, F.P. (2025) Higher-level spatial prediction in natural vision across mouse visual cortex. *bioRxiv* DOI: https://doi.org/10.1101/2025.05.15.654212
8. Richter, D. *et al.* (2024) High-level visual prediction errors in early visual cortex. *PLoS Biol* 22, e3002829
9. Richter, D. *et al.* (2025) Rapid Computation of High-Level Visual Surprise. *bioRxiv* DOI: https://doi.org/10.1101/2025.06.20.660166
10. Vezoli, J. *et al.* (2021) Cortical hierarchy, dual counterstream architecture and the importance of top-down generative networks. *NeuroImage* 225, 117479
11. Alink, A. and Blank, H. (2021) Can expectation suppression be explained by reduced attention to predictable stimuli? *NeuroImage* 231, 117824
12. Doron, G. *et al.* (2020) Perirhinal input to neocortical layer 1 controls learning. *Science* 370, eaaz3136
13. Farahat, A. *et al.* (2023) A novel feature-scrambling approach reveals the capacity of convolutional neural networks to learn spatial relations. *Neural Networks* 167, 400–414
14. Vinck, M. *et al.* (2025) Large-scale interactions in predictive processing: oscillatory versus transient dynamics. *Trends in Cognitive Sciences* 29, 133–148
15. Mikulasch, F.A. *et al.* (2023) Where is the error? Hierarchical predictive coding through dendritic error computation. *Trends in Neurosciences* 46, 45–59
16. Luc, P. *et al.* (2017) Predicting Deeper into the Future of Semantic Segmentation. *arXiv* DOI: 10.48550/arXiv.1703.07684