A hierarchy of spatial predictions across human visual cortex during natural vision

Wieger H. Scheurer & Micha Heilbron
wieger.scheurer@donders.ru.nl; micha.heilbron@uva.nl
Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands
Department of Psychology, Brain and Cognition, University of Amsterdam, The Netherlands

Abstract

The predictive processing framework posits that the brain constantly compares incoming sensory signals with self-generated predictions. However, evidence for prediction in sensory cortex mostly comes from artificial paradigms with simple, highly predictable stimuli, leaving it unclear whether the reported sensory prediction effects generalise to perception more broadly. Here, we probe predictions in naturalistic perception, analysing high-resolution 7T functional magnetic resonance imaging (fMRI) responses of human participants viewing tens of thousands natural scenes. We use deep generative models to quantify the inherent spatial predictability of image patches, and relate resulting estimates to brain activity. Our results reveal robust and widespread predictability modulations of BOLD responses across the visual cortex. Higher visual areas were sensitive to more high-level predictability, forming a prediction hierarchy. Effects were stronger in voxels with higher eccentricity receptive fields, aligning with predictive coding and Bayesian theories. These results demonstrate the ubiquity of prediction in vision and inform neurocomputational models of predictive coding and self-supervised learning in the brain.

Keywords: predictive coding; generative models; natural vision; human visual cortex; fMRI; pRF; hierarchical sensory prediction

Introduction

Vision is more intricate than meets the eye, requiring the brain to convert noisy visual input into a coherent percept of the world. Theories of predictive processing postulate that this conversion relies on constant comparison between incoming sensory signals and self-generated predictions (Rao & Ballard, 1999; Friston, 2005), of which the discrepancy serves both perceptual inference and self-supervised learning. While influential (Summerfield & de Lange, 2014), accumulated evidence supporting prediction in sensory cortex has largely relied on artificial, prediction-encouraging experimental paradigms, leaving it unclear whether observed effects generalize to perception more broadly.

Generative AI offers a new way to study prediction (Heilbron et al., 2022; de Lange et al., 2022). Instead of experimentally manipulating stimulus predictability, this involves estimating stimulus predictability using generative models. As these models are stimulus-computable, they allow probing sensory predictions using complex, naturalistic stimuli.

Uran et al. (2022) introduced a way to apply this approach to natural scene perception, using the fidelity with which an image patch could be 'in-painted' as a metric for spatial predictability. They found that neurons in macaque V1 fired more strongly if the image content in their receptive field was less predictable.

Here, we build on their method, investigating the effect of spatial predictability in the Natural Scenes Dataset (NSD (Allen et al., 2022)), analysing BOLD responses throughout the visual system of human participants viewing 73,000 natural images.

Methods

Our analyses involve the stimulus images, fMRI recordings, and population receptive field (pRF) maps from NSD. We use

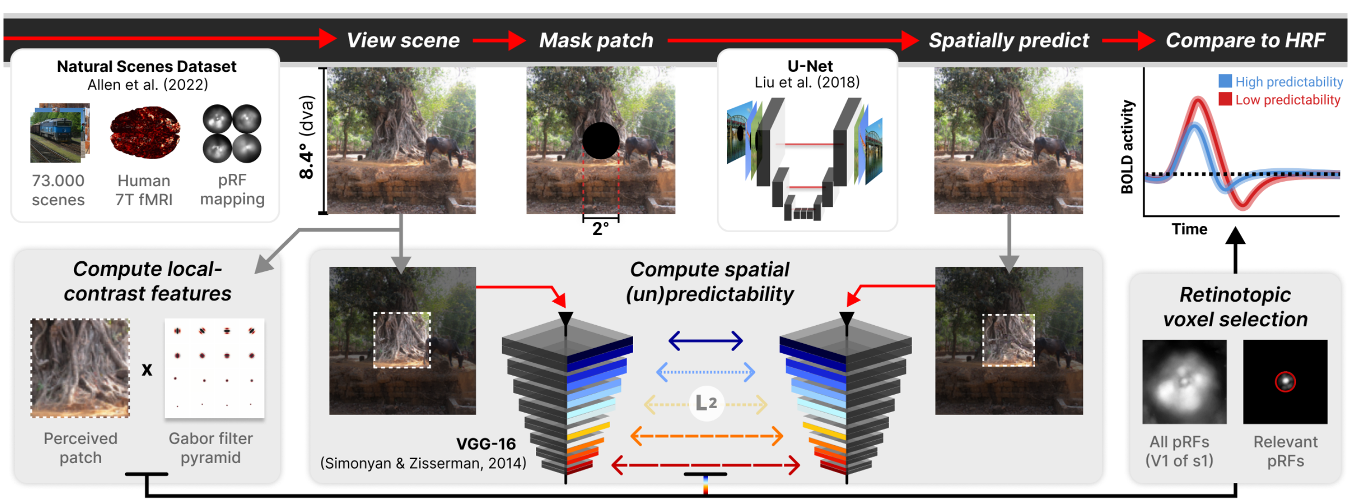

Figure 1: **Graphical outline of the empirical framework**, including data and modelling resources (white blocks) and analytical methods (grey blocks). We quantified the spatial predictability of natural image patches (1 ° visual angle radius) by comparing actual visual content with model-predicted reconstructions from a U-Net autoencoder (Liu et al., 2018). These predictability estimates were regressed on retinotopically (pRF) relevant fMRI BOLD responses from humans viewing 73,000 natural images (NSD (Allen et al., 2022)), controlling for low-level visual features (Gabor pyramid filter outputs).

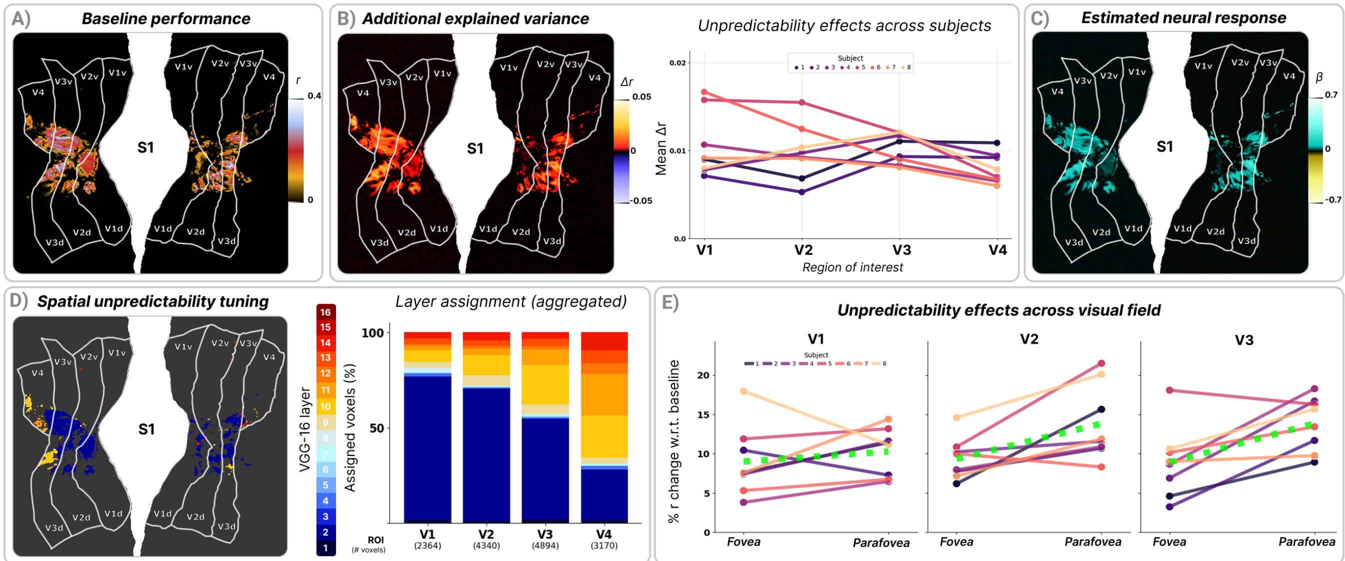

Figure 2: **Results** (n.b.: single-subject surface plots display effects found in full sample) **A)** Baseline model of bottom-up encoded features (gabor pyramid) is substantially predictive of evoked response variance in visual cortex (Pearson correlation (r) between actual HRF amplitude and predictions of cross-validated ridge regression model). **B)** Unpredictability features robustly contribute additional explained neural variance (Δr) over and above baseline. **C)** Direction of neural response modulations by spatial unpredictability is consistently positive, expressed as the mean difference in estimated change of normalised response (β). **D)** Visual cortex is hierarchically tuned for spatial unpredictability, as revealed by voxel assignment of the abstract feature level most predictive of its response variance (argmax(Δr)). **E)** Spatial prediction effects scale proportional to sensory evidence across the visual field, demonstrated by stronger predictive contributions of unpredictability features in more eccentric neural populations.

deep generative models to quantify the inherent spatial predictability of each natural scene and regress these estimates on evoked neural responses to test for prediction effects across human visual cortex, while controlling for low-level visual features using a Gabor pyramid baseline model (Fig 1).

For each image, we select a specific patch of which we quantify its spatial unpredictability and features encoded in the feedforward sweep of visual processing. Specifically, we use a U-Net to inpaint masked patches given their spatial context, and define spatial unpredictability as the discrepancy between model-predicted and ground-truth patches. This is based on the ℓ_2 distance between features of the predicted and observed image patch, at each layer of a deep convolutional neural network (DCNN; VGG-16) (Simonyan & Zisserman, 2014). For bottom-up features we take the first 500 principal components (PCs) of filter outputs from a gabor pyramid convolved with ground-truth patches. We relate these features to the estimated amplitude (single-trial haemodynamic response function (HRF) fit β-coefficients) of scene-evoked responses in the blood oxygenation level-dependent (BOLD) signal using ridge regression with cross-validation. To assess spatial unpredictability effects we examine the contribution of unpredictability features over and above the baseline predictive performance of only bottom-up features.

Results

We first asked whether BOLD responses were modulated by spatial unpredictability, defined as unpredictablity explaining significant additional variance over and above the baseline model. We find this is the case in all participants (Fig 2B). Moreover, when we inspect the coefficients, the direction of modulation is consistently positive: BOLD response is stronger when image content is less predictable, in line with expectation suppression (Summerfield & de Lange, 2014).

We then asked what level of unpredictability visual cortex was most sensitive to. This revealed a hierarchy of predictions where higher-level visual areas become progressively sensitive to more abstract levels of visual predictability. This gradient parallels the established bottom-up feature encoding hierarchy (Güçlü & Van Gerven, 2015). However, it diverges from recent observations in animal models (Heilbron & de Lange, 2023; Schwiedrzik & Freiwald, 2017; Uran et al., 2022) and human fMRI (Richter et al., 2024), which suggested that visual cortex predicts specifically at higher levels of abstraction.

Finally, we asked whether these effects displayed a key hallmark of prediction: sensitivity to sensory reliability. We compared predictability effects in voxels with foveal vs parafoveal RFs, reasoning that voxels with less reliable bottom-up sensory input should rely more on sensory prediction. This is indeed what we find in every brain region where we could perform this analysis: prediction effects are generally stronger when visual input is less reliable (Fig 2D). This is in line with Bayesian inference within predictive processing (Lee & Mumford, 2003), considering sensory precision decreases with eccentricity.

Discussion

These findings provide evidence for the ubiquity of prediction in the human visual system. They confirm predictive processing theories and offer new insights into how the brain generates and uses predictions across different cortical areas and visual field locations—supporting ideas of hierarchical prediction (Heilbron & de Lange, 2023; Friston, 2008; Wacongne et al., 2011) in active Bayesian predictive coding (Rao, 2024).

Together, these results provide a new window on sensory prediction, with implications for computational models of predictive processing and self-supervised learning in visual cortex.

References

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., ... Kay, K. (2022, January). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, *25*(1), 116–126. doi: 10.1038/s41593-021-00962-x

de Lange, F. P., Schmitt, L.-M., & Heilbron, M. (2022, December). Reconstructing the predictive architecture of the mind and brain. *Trends in Cognitive Sciences*, *26*(12), 1018–1019. doi: 10.1016/j.tics.2022.08.007

Friston, K. (2005, April). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. (2008). Hierarchical models in the brain. *PLoS computational biology*, *4*(11), e1000211. doi: 10.1371/journal.pcbi.1000211

Güçlü, U., & Van Gerven, M. A. J. (2015, July). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *Journal of Neuroscience*, *35*(27), 10005–10014. doi: 10.1523/JNEUROSCI.5023-14.2015

Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences*, *119*(32), e2201968119. doi: https://doi.org/10.1073/pnas.2201968119

Heilbron, M., & de Lange, F. (2023). Higher-level spatial prediction during natural scene perception in mouse primary visual cortex. In *2023 Conference on Cognitive Computational Neuroscience.* Oxford, UK: Cognitive Computational Neuroscience. doi: 10.32470/CCN.2023.1364-0

Lee, T. S., & Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. *JOSA a*, *20*(7), 1434–1448. doi: https://doi.org/10.1364/JOSAA.20.001434

Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., & Catanzaro, B. (2018). Image inpainting for irregular holes using partial convolutions. In *Proceedings of the european conference on computer vision (eccv)* (pp. 85–100). doi: https://doi.org/10.1007/978-3-030-01252-6$_6$

Rao, R. P. N. (2024, July). A sensory–motor theory of the neocortex. *Nature Neuroscience*, *27*(7), 1221–1235. doi: 10.1038/s41593-024-01673-9

Rao, R. P. N., & Ballard, D. H. (1999, January). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. doi: 10.1038/4580

Richter, D., Kietzmann, T. C., & de Lange, F. P. (2024). High-level visual prediction errors in early visual cortex. *PLoS Biology*, *22*(11), e3002829. doi: 10.1371/journal.pbio.3002829

Schwiedrzik, C. M., & Freiwald, W. A. (2017). High-level prediction signals in a low-level area of the macaque face-processing hierarchy. *Neuron*, *96*(1), 89–97. doi: https://doi.org/10.1016/j.neuron.2017.09.007

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. doi: https://doi.org/10.48550/arXiv.1409.1556

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. doi: 10.1038/nrn3838

Uran, C., Peter, A., Lazar, A., Barnes, W., Klon-Lipok, J., Shapcott, K. A., ... Vinck, M. (2022, April). Predictive coding of natural images by V1 firing rates and rhythmic synchronization. *Neuron*, *110*(7), 1240–1257.e8. doi: 10.1016/j.neuron.2022.01.002

Wacongne, C., Labyt, E., Van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences*, *108*(51), 20754–20759. doi: 10.1073/pnas.1117807108