# A Hierarchical Structure for Perceptual Awareness in the Human Brain

**Nadine Dijkstra (n.dijkstra@ucl.ac.uk)**
Wellcome Centre for Human Neuroimaging
University College London, United Kingdom

**Oliver Warrington (oliver.warrington.18@ucl.ac.uk)**
Wellcome Centre for Human Neuroimaging
University College London, United Kingdom

**Peter Kok (p.kok@ucl.ac.uk)**
Wellcome Centre for Human Neuroimaging
University College London, United Kingdom

**Stephen M. Fleming (stephen.fleming@ucl.ac.uk)**
Wellcome Centre for Human Neuroimaging
Max Planck Centre for Computational Psychiatry and Aging Research
Department of Experimental Psychology
University College London, United Kingdom

## Abstract

**Accounting for why sensitivity to perceptual input (as assayed by discrimination judgments) is not always accompanied by conscious awareness (as assayed by detection judgments) remains a challenge for theories of perception. Here we test a hypothesis that awareness is supported by higher-order inferences within generative models of perceptual content. In line with model simulations, we show that both detection and discrimination expectations influence reaction times on a categorisation task. By combining a no-report version of our task with functional neuroimaging we reveal a neural dissociation between prediction errors (PEs) on content (discrimination) and awareness of content (detection): content PEs are tracked in posterior sensory cortex while awareness PEs are tracked in prefrontal cortex. Together, our results reveal a hierarchical structure supporting visual detection and discrimination, consistent with a proposal that awareness reflects a higher-order inference within perceptual generative models.**



**Keywords:** awareness; consciousness; predictive processing; perception; signal detection theory


## Introduction

Our perceptual experience is characteristically limited: at any given moment in time, we are aware of only a subset of perceptual inputs. Such failures of awareness do not necessarily reflect failures of sensory processing: there are cases where the content of stimuli is rendered invisible but nevertheless continues to exert influence on behavior (Dehaene et al. (2001); Marcel (1983); Persaud et al. (2011); Weiskrantz, Warrington, Sanders, and Marshall (1974), although, see Peters and Lau (2015)). Within the framework of perceptual decision-making, dissociations between performance and awareness can be modelled as a distinction between discrimination – responding to some aspect of stimulus identity – and detection – responding as to whether a stimulus is perceived or not (Azzopardi & Cowey, 1997; Green & Smets, 1966; Peters & Lau, 2015).

Why we detect some perceptual contents but not others remains a core challenge for computational models of perception (Hohwy & Seth, 2020; Marvan & Havlík, 2021). In this study, we test a core hypothesis that awareness reflects a higher-order inference within a generative model of perceptual content (Lau, 2019, 2007; Fleming, 2020; Morales, 2022). Using a novel experimental paradigm in combination with computational modeling and neuroimaging, we identify distinct signatures of discrimination and detection prediction errors in the human brain.

## Results

### Interrogating prediction errors on content and awareness

To quantify hypotheses regarding the different types of prediction errors, we utilized our recently developed Higher-Order

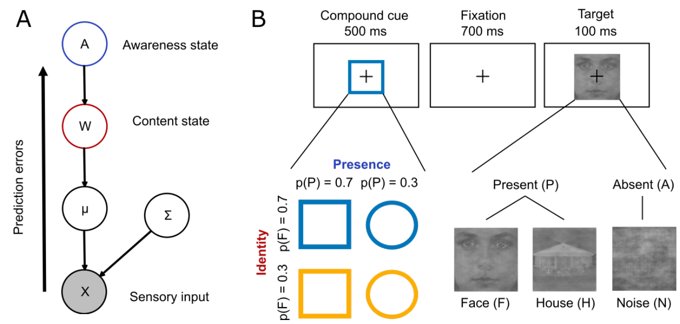


Figure 1: **Model and experimental paradigm.** (A) Graphical representation of the Higher-Order State Space (HOSS) model. Perceptual states W and awareness state A are inferred based on sensory input X. Predictions and prediction errors are generated at all levels of the model (B) Experimental paradigm. Compound cues indicated the probability of stimulus presence (A level) = present vs absent via the shape and stimulus content (W level) = face vs house via the color.

Stace Space (HOSS) model (Fig. 1A). HOSS is a hierarchical Bayesian model in which inference on awareness ($A$) is superordinate with respect to inference on content ($W$). Upon receipt of a sensory sample $X$, the model is inverted to compute the posteriors over $A$ and $W$ by marginalising:

$$P(A|X=x) \propto \sum_{W} P(A)P(W|A)P(X=x|W) \quad (1)$$

$$P(W|X=x) \propto \sum_{A} P(A)P(W|A)P(X=x|W) \quad (2)$$

To be able to independently manipulate predictions about perceptual content and detection/awareness of content, we developed a novel perceptual discrimination task with compound cues (Fig. 1B). The shape of the cue indicated the probability that a stimulus was present (rather than absent) whereas the color of the cue indicated the probability that a stimulus was a face (rather than a house). For example, a yellow circle indicated a high probability that no stimulus would be shown (absence), but if a stimulus was shown, it would likely be a house. In what follows, we refer to predictions about perceptual content that are relevant for discrimination as "content expectations", and to predictions about perceptual presence that are relevant for detection as "presence expectations".

We simulated prediction errors (PEs) on content and presence in each of the 12 conditions (4 compound cues x 3 targets) as the Kullback–Leibler ($KL$) divergence between the prior and posterior distributions within the $A$ and $W$ layers of the model, respectively (Fig. 2A). Content prediction errors ensue when there is a mismatch between the content prediction and the target. For example, a face (F) is predicted but a house (dark blue) is presented, irrespective of whether presence (P) or absence (A) was predicted (Fig. 2A) left). In contrast, presence prediction errors are high when presence (P) was predicted and noise (green) was observed or when

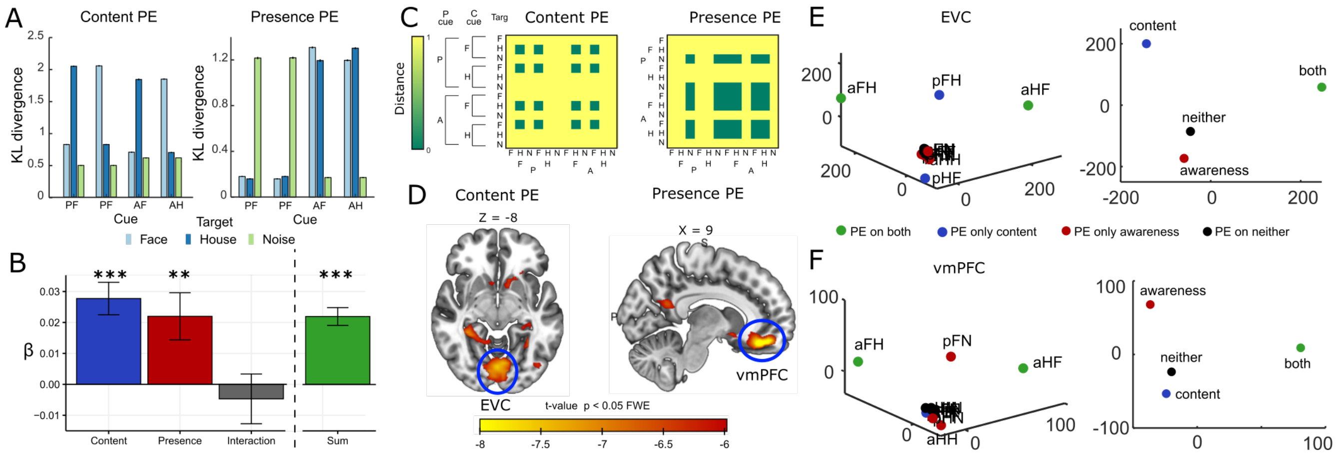


Figure 2: **Model simulations, behaviour and neural results.** (A) Simulated prediction errors (PEs), defined as the KL divergence between prior and posterior, for all experimental conditions. (B) Results from linear mixed effects models explaining reaction times (RTs) from simulated predictions errors. A model containing the summed PEs over both layers (green) best explained the RT data. (C) Representational Dissimilarity Matrices (RDMs) encoding similarities between conditions sharing content prediction errors (left) and awareness prediction errors (right). (D) The content PE RDM correlated negatively with activity in early visual areas (left) whereas the awareness PE RDM correlated negatively with activity in the vmPFC (right). (E) Multi-dimensional scaling (MDS) demonstrates that activity in EVC diverges during conditions with a PE on content (green and blue). (F) In contrast, activity in vmPFC diverges during conditions with a PE on awareness (red and black). Legend: P = present; A = absent; F = face; H = house; N = noise; EVC = early visual cortex; vmPFC = ventromedial prefrontal cortex; **$p<0.005$; ***$p<0.0005$. Error bars reflect standard errors of the mean (SEM)

absence was predicted (A) and presence (face (F) or house (H)) was observed, irrespective of the content predictions (Fig. 2A right).

## Rich representations of content and awareness prediction errors in the brain

We first demonstrated that participants were able to learn and use the compound cues. In a behavioural experiment, thirty-six participants categorized targets following compound-cues. A model-based linear mixed-effects regression analysis predicting reaction times (RT) from the simulated content and presence PEs revealed significant increases in RT for both content PEs ($\beta$ = 0.028, $t(16508.07) = 5.28$, $p = 0.0000013$; (Fig. 2B blue bar) as well as presence PEs ($\beta$ = 0.022, $t(16508.04) = 2.89$, $p = 0.0038$; (Fig. 2B red bar). The interaction between content and presence PEs was not significant (t(16508.04) = -0.58, p = 0.56; Fig. 3B, grey bar), suggesting that the influence of content PE on behaviour was not affected by the degree of presence PE, and vice versa.

Twenty-seven of the participants who had completed the behavioural experiment went on to perform a no-report version of the same task in a 3T MRI scanner while undergoing whole-brain functional neuroimaging. We used representational similarity analysis (RSA) to identify brain regions coding for the two types of prediction errors (Fig. 2C). The content PE RDM correlated negatively with neural acivity in early visual cortex (EVC) whereas the awareness PE correlated with activity in ventromedial prefrontal cortex (vmPFC; Fig. 2D). Multi-dimensional scaling of activity in these areas reveals a rich encoding of PEs: activity patterns encode level-specific PEs (content for EVC: 2E; awareness for vmPFC: 2F) but also diverge according to (a) the exact type of content violation (face, house, present, absent) and (b) whether predictions within the other layer were also violated. This type of PE coding is consistent with a recurrent, hierarchical architecture supporting detection and discrimination.

## Discussion

In this study we set out to test a hypothesis that subjective detection reflects a higher-order inference within a generative model of perception. We developed a novel experimental paradigm that allowed us to independently interrogate prediction errors on stimulus content and presence. In line with a hierarchical architecture for awareness, we found that content PEs were represented in sensory brain areas while presence PEs were represented in frontal brain areas. Our findings are in support of higher-order theories of consciousness (Brown, Lau, & LeDoux, 2019; Lau & Rosenthal, 2011) and suggest, counterintuitively, that inferences about whether something is seen (detection) are higher-order with respect to inferences about what is seen (discrimination)

## Acknowledgments

This work was supported by grants from the Netherlands Organization for Scientific Research, European Research Council, Wellcome and Leverhume Trust.

## References


Azzopardi, P., & Cowey, A. (1997). Is blindsight like normal, near-threshold vision? *Proceedings of the National Academy of Sciences of the United States of America*, *94*(25). doi: 10.1073/pnas.94.25.14190

Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the Higher-Order Approach to Consciousness. *Trends in Cognitive Sciences*, *23*(9), 754–768. doi: 10.1016/j.tics.2019.06.009

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, *4*(7), 752–758. doi: 10.1038/89551

Fleming, S. M. (2020). Awareness as inference in a higher-order state space. *Neuroscience of Consciousness*, *2020*(1), 1–9. doi: 10.1093/nc/niz020

Green, D., & Smets, J. (1966). *Signal detection theory and psychophysics* (Vol. 5). doi: 10.1016/0022-460x(67)90197-6

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, *1*. doi: 10.33735/phimisci.2020.ii.64

Lau, H. (2007). A higher order Bayesian decision theory of consciousness. *Progress in Brain Research*, *168*, 35–48. doi: 10.1016/S0079-6123(07)68004-2

Lau, H. (2019). Consciousness, Metacognition, and Perceptual Reality Monitoring. *PsychArxiv*, 1–17.

Lau, H., & Rosenthal, D. (2011, aug). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373. doi: 10.1016/j.tics.2011.05.009

Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, *15*(2). doi: 10.1016/0010-0285(83)90009-9

Marvan, T., & Havlík, M. (2021). Is predictive processing a theory of perceptual consciousness? *New Ideas in Psychology*, *61*. doi: 10.1016/j.newideapsych.2020.100837

Morales, J. (2022). Introspection Is Signal Detection. *British Journal for the Philosophy of Science*. Retrieved from `https://doi.org/10.1086/715184` doi: https://doi.org/10.1086/715184

Persaud, N., Davidson, M., Maniscalco, B., Mobbs, D., Passingham, R. E., Cowey, A., & Lau, H. (2011). Awareness-related activity in prefrontal and parietal cortices in blindsight reflects more than superior visual performance. *NeuroImage*, *58*(2). doi: 10.1016/j.neuroimage.2011.06.081

Peters, M. A., & Lau, H. (2015). Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *eLife*, *4*. doi: 10.7554/eLife.09651

Weiskrantz, L., Warrington, E. K., Sanders, M. D., & Marshall, J. (1974). Visual capacity in the hemianopic field following a restricted occipital ablation. *Brain*, *97*(1), 709–728. doi: 10.1093/brain/97.1.709