Learning and Inference in the Brain

Wellcome Dept. of Imaging Neuroscience, University College London

Abstract

By formulating the original ideas of Helmholtz on perception, in terms of modern-day theories, one arrives at a model of perceptual inference and learning that can explain a remarkable range of neurobiological facts. Using constructs from statistical physics, machine learning and probability theory the problems of inferring the causes of sensory input and learning the causal structure of their generation can be resolved using exactly the same principles. Furthermore, inference and learning can proceed in a biologically plausible fashion. The ensuing scheme rests on *Empirical Bayes* and hierarchical models of how sensory input is caused. The use of hierarchical models enables the brain to construct prior expectations in a dynamic and context-sensitive fashion. This scheme provides a principled way to understand many aspects of cortical organisation and responses.

In terms of cortical architectures, it predicts that sensory cortices should be arranged hierarchically, that connections should be reciprocal, and that forward and backward connections should show a functional asymmetry (backward connections are both modulatory and driving, whereas forward connections need only be driving). In terms of synaptic physiology it predicts associative plasticity and, for dynamic models, spike–timing-dependent plasticity. In terms of electrophysiology it accounts for classical and extra-classical receptive field effects and long-latency or endogenous components of evoked cortical responses. It predicts the attenuation of responses encoding prediction error with perceptual learning and explains many phenomena like repetition suppression, mismatch negativity (MMN) and the P300 in electroencephalography. In psychophysical terms, it accounts for the behavioural correlates of these physiological phenomena, *e.g.* priming, and global precedence. The final focus of this talk is on perceptual learning as measured with repetition suppression and the implications for empirical studies of coupling among cortical areas.

Overview

- The free energy principle
- Hierarchical dynamic models
- Bayesian inversion and perception
- Neuronal architectures
- Simulating ERPs
- Repetition suppression

The free energy principle

Variational parameters $\lambda = \min_{\lambda} F(y(\lambda_y), \lambda_q)$

Free energy $F = U - H$

Energy $U = -\langle \ln p(y(\lambda_y), u, \theta) \rangle_q$

Entropy $H = -\langle \ln q(u, \theta \mid \lambda_q) \rangle_q$

Action

λ_y

$$F = -\langle \ln p(y(\lambda_y) \mid u, \theta) \rangle_q + D(q \,\|\, p(u, \theta))$$

Perceptual homeostasis $\lambda_y = \min_{\lambda} F(y, \lambda)$

Perception

$\lambda_q = \{\lambda_u, \lambda_\theta\}$

$$F = -\ln p(y) + D(q \,\|\, p(u, \theta \mid y))$$

$$q = q(u \mid \lambda_u) q(\theta \mid \lambda_\theta)$$

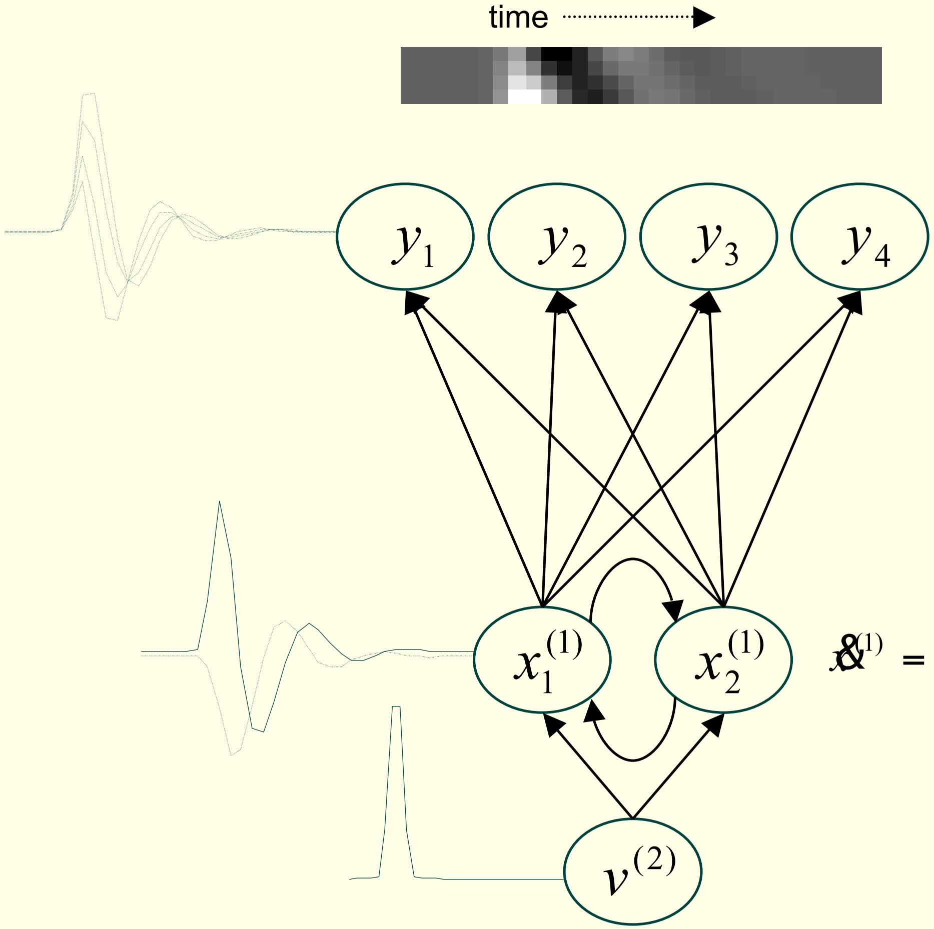
$\lambda_u = \min_{\lambda} F$ Perceptual inference

$\lambda_\theta = \min_{\lambda} F$ Perceptual learning

Hierarchical dynamic models

Hidden and causal states: $u = \{x, v\}$ Conditional estimates: $\lambda_u = \{x, v\}$

$$
\begin{aligned}
y &= \\
v^{(1)} &= g(x^{(1)}, v^{(2)}, \theta^{(1)}) + e^{(1)} \\
\dot{x}^{(1)} &= f(x^{(1)}, v^{(2)}, \theta^{(1)}) \\
\\
v^{(2)} &= g(x^{(2)}, v^{(3)}, \theta^{(2)}) + e^{(2)} \\
\dot{x}^{(2)} &= f(x^{(2)}, v^{(3)}, \theta^{(2)}) \\
&\vdots \\
v^{(n)} &= e^{(n)}
\end{aligned}
$$



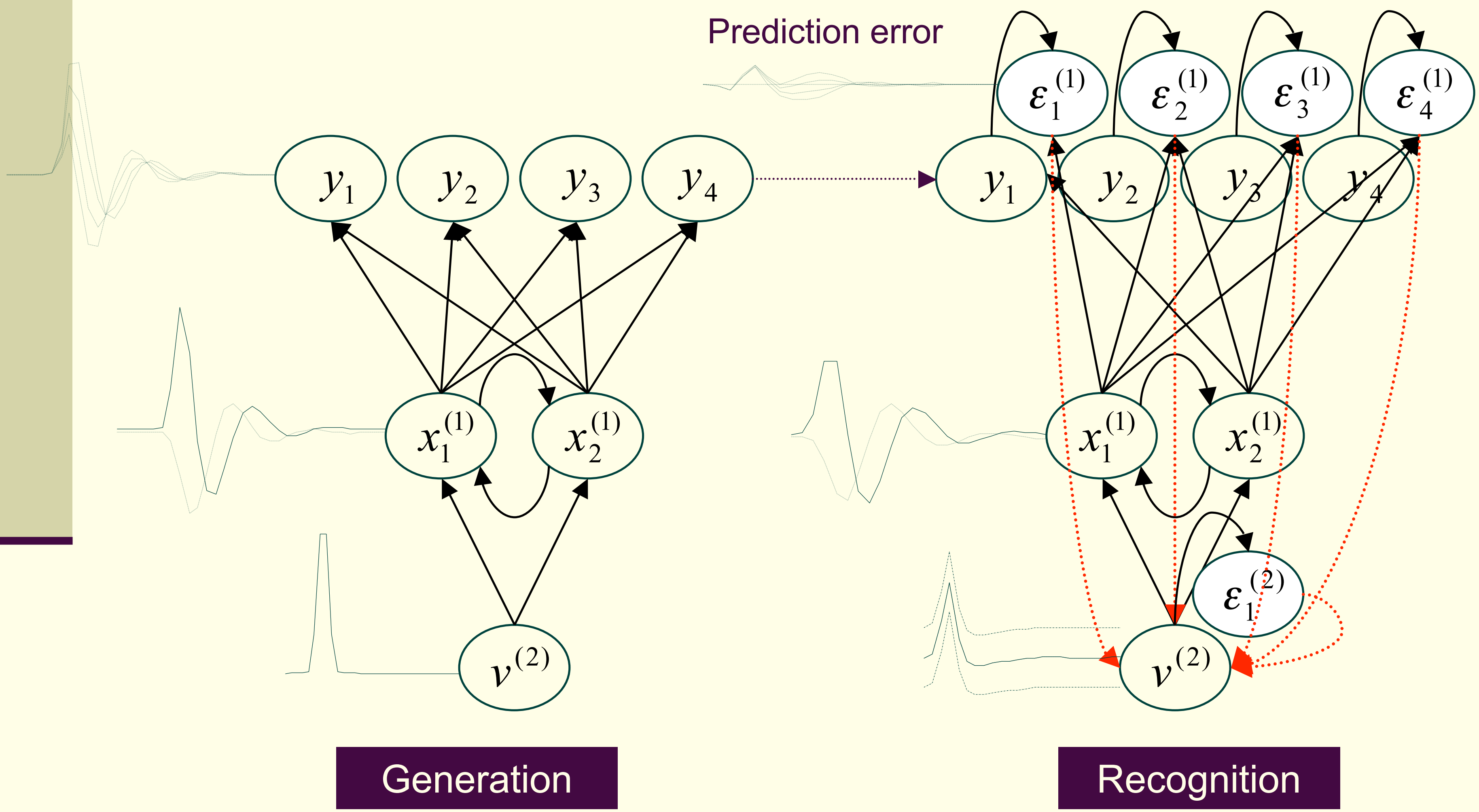
$$y = \theta_3^{(1)} x^{(1)}$$

$$\dot{x}^{(1)} = \theta_1^{(1)} x^{(1)} + \theta_2^{(1)} v^{(2)}$$

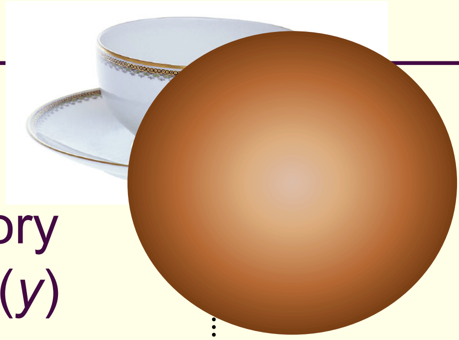
Bayesian inversion and perception

$$v^{(i)} = g(x^{(i)}, v^{(i+1)}) + e^{(i)}$$
$$\dot{x}^{(i)} = f(x^{(i)}, v^{(i+1)})$$

$$\varepsilon^{(i)} = v^{(i)} - g(x^{(i)}, v^{(i+1)})$$
$$\dot{x}^{(i)} = f(x^{(i)}, v^{(i+1)})$$
$$\dot{\mu}^{(i)} = \kappa\, \partial F(t+\tau) \big/ \partial \dot{\mu}^{(i)} = h(\varepsilon^{(i-1)}, \varepsilon^{(i)})$$



Perception and predictive coding



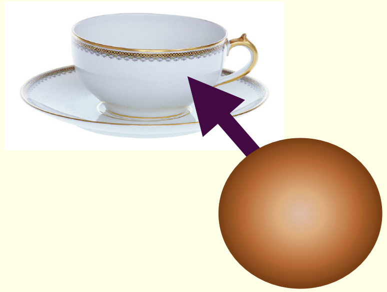
Sensory input (y)

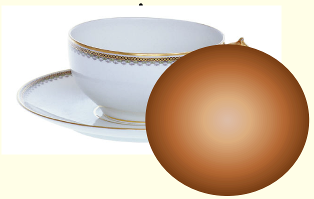
Forward connections provide *Feedback*

$$\varepsilon = y - g(v)$$

causes (v)

Prediction error





Backward connections embody the Generative model

$$g(v)$$

Prediction

Variational Bayes and EM

E-Step

$$\dot{\tilde{\varepsilon}}^{(i)} = h(\varepsilon^{(i-1)}, \varepsilon^{(i)}) = -\kappa\, \overbrace{\frac{\partial \tilde{\varepsilon}^{(i-1)T}}{\partial v^{(i)}} U^{(i)} \tilde{\varepsilon}^{(i-1)}}^{\text{Bottom-up}} \overbrace{- \kappa S^{(i)} \tilde{\varepsilon}^{(i)}}^{\text{Lateral}}$$

$$\dot{x}^{(i)} = f(x^{(i)}, v^{(i+1)})$$

$$\varepsilon^{(i)} = \underbrace{v^{(i)} - g(x^{(i)}, v^{(i+1)})}_{\text{Top-down}}$$

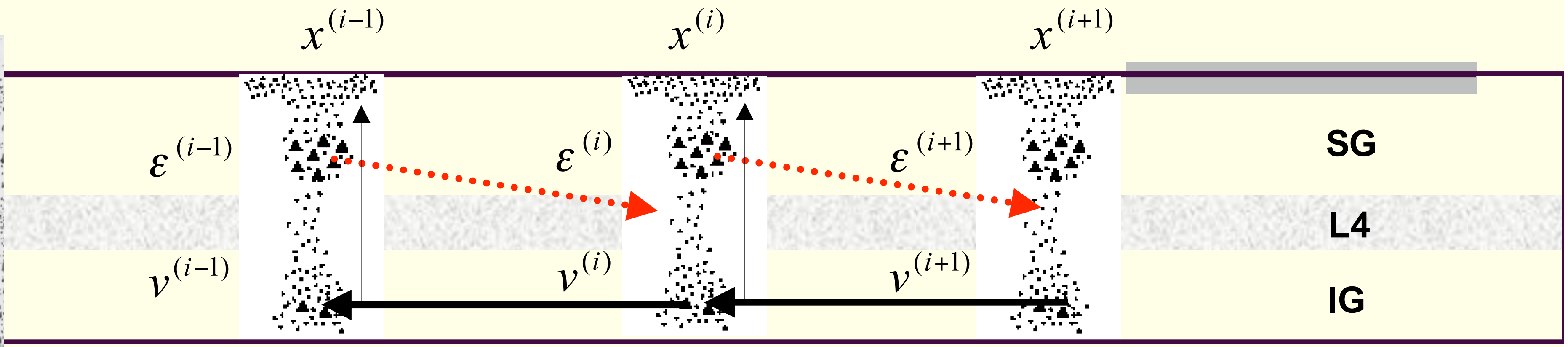
$$\tilde{\varepsilon}(t) = \begin{bmatrix} \varepsilon(t) \\ \dot{\varepsilon}(t) \\ \vdots \end{bmatrix}$$

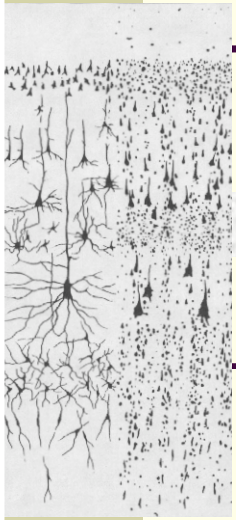
M-Step

$$\Delta\theta = \left\langle -\frac{\partial^2 F}{\partial \theta^2} \right\rangle^{-1} \left\langle \frac{\partial F}{\partial \theta} \right\rangle$$

A dynamic recognition system that minimises embedded prediction error

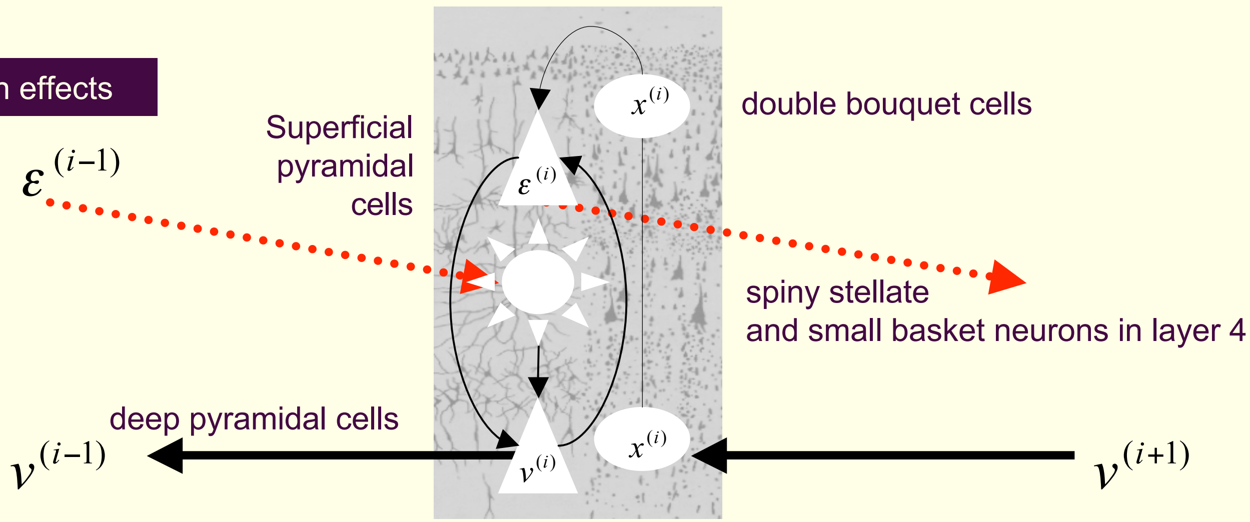
Neuronal architecture





$$\varepsilon^{(i)} = v^{(i)} - g(x^{(i)}, v^{(i+1)})$$

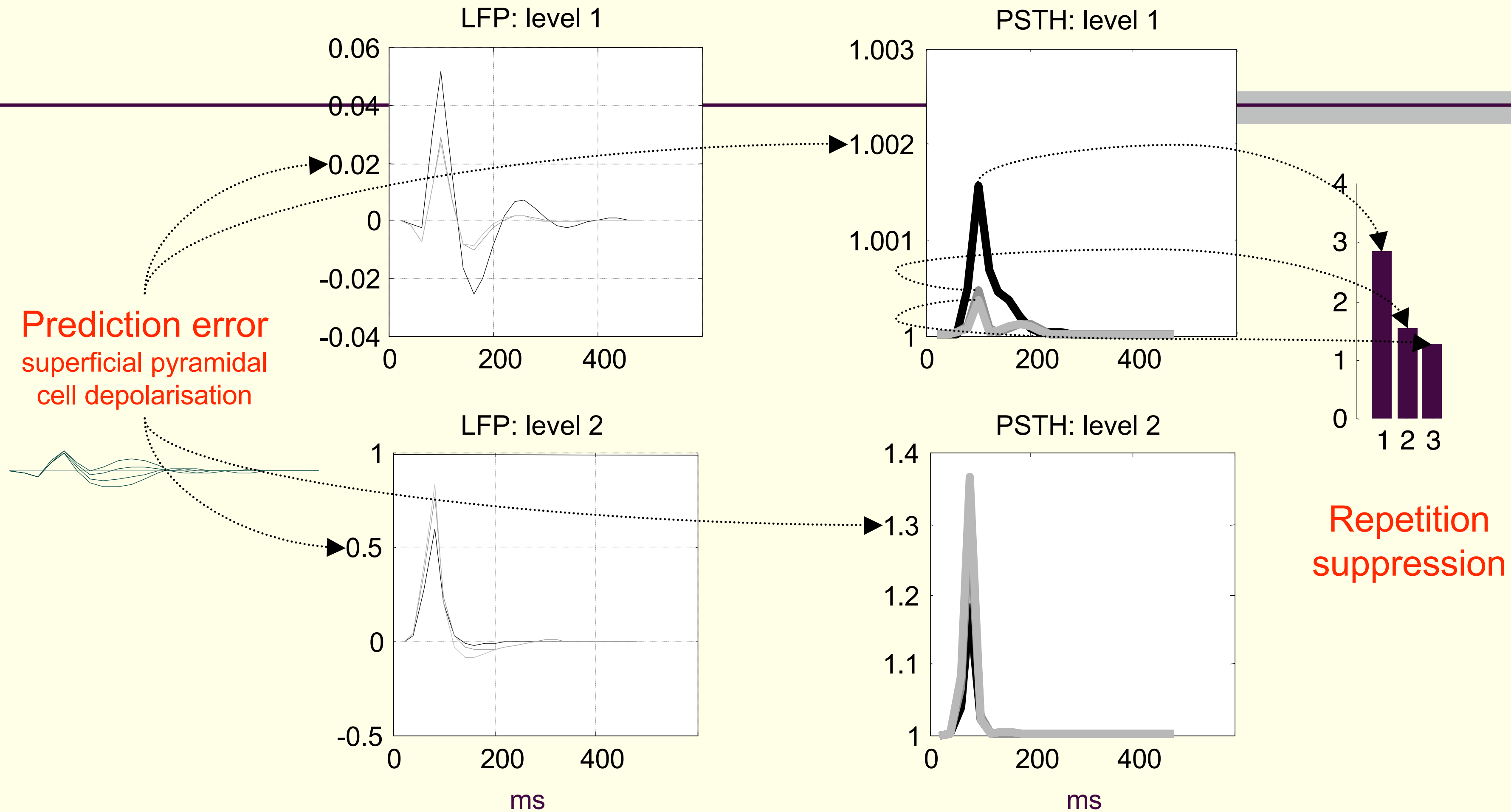
Forward recognition effects



Backward generation effects

$$\dot{\varepsilon}^{(i)} = h(\varepsilon^{(i-1)}, \varepsilon^{(i)}) \qquad \dot{x}^{(i)} = f(x^{(i)}, v^{(i+1)})$$

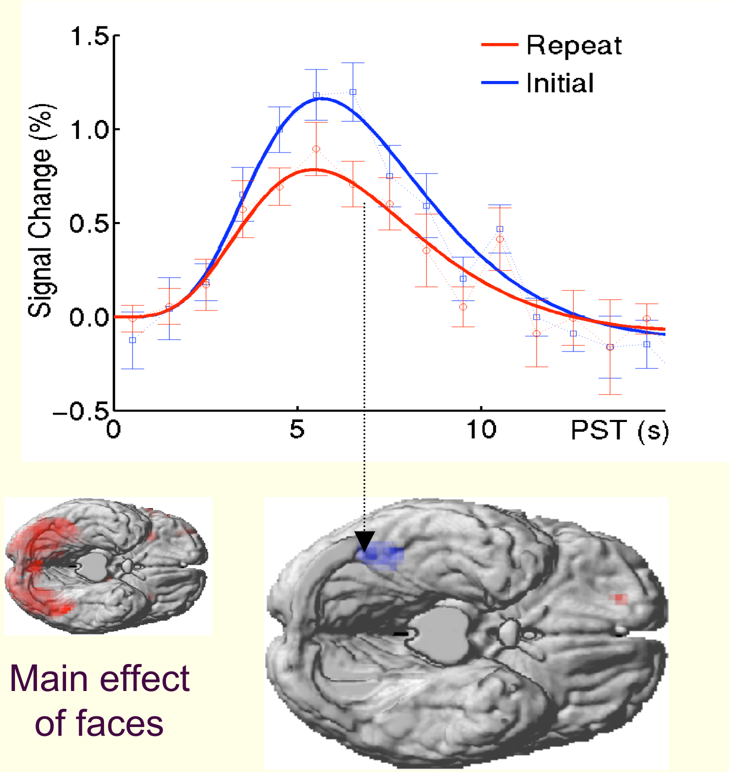
Simulating ERPs



Perceptual inference: error suppression over peristimulus time $\lambda_u = \{v, x\} = \min_{\lambda} F$

Perceptual learning: error suppression over repetitions $\lambda_\theta = \theta = \min_{\lambda} F$

An fMRI example of perceptual learning and suppression of prediction error (free energy)



Main effect of faces

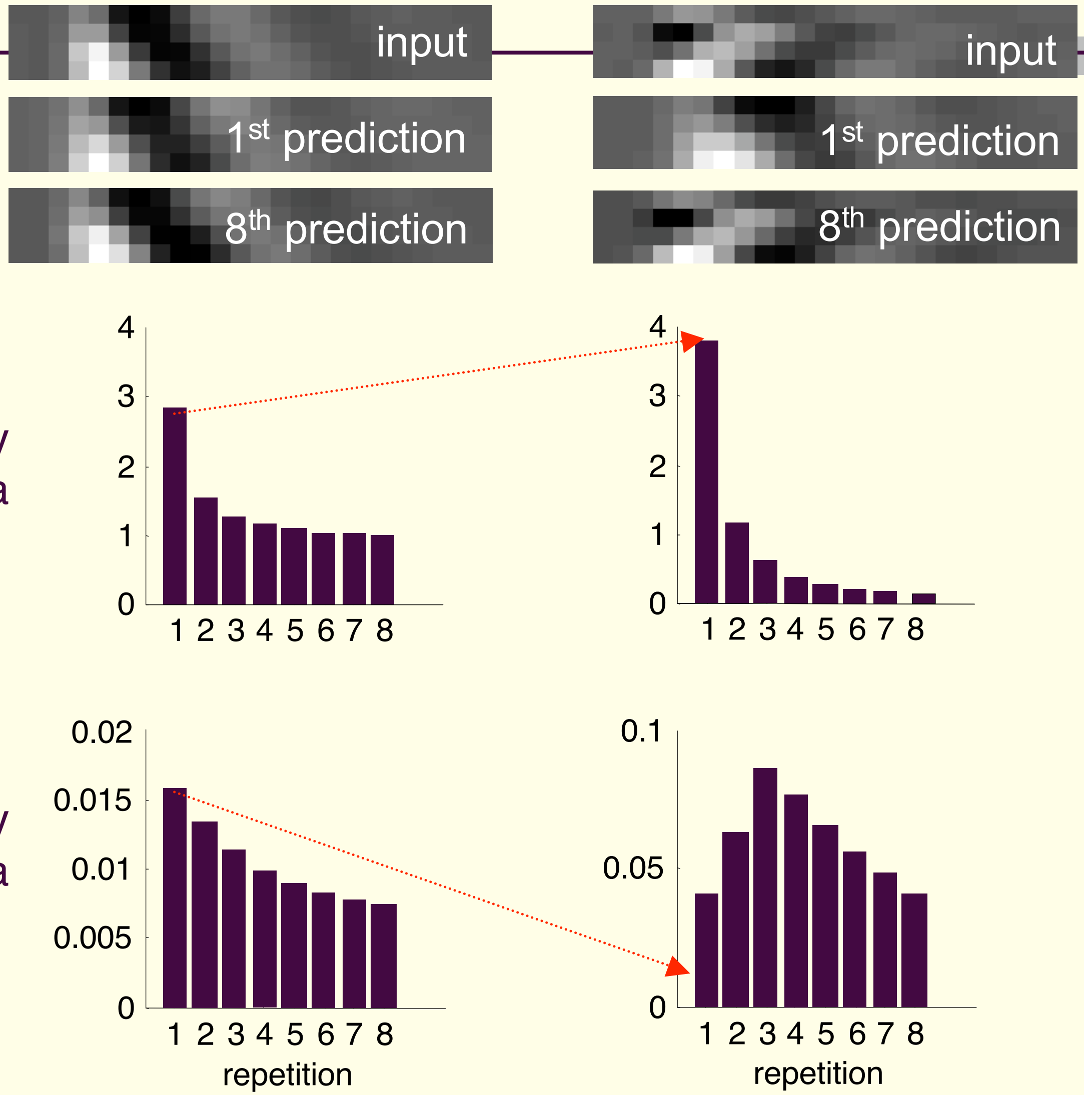
Suppression of inferotemporal responses to repeated faces

Henson *et al* 2000

Repetition suppression

Familiar stimulus

Unfamiliar stimulus



Synaptic activity
In lower area

Synaptic activity
In higher area

Shape perception reduces activity in human primary visual cortex

Scott O. Murray , Daniel Kersten , Bruno A. Olshausen, Paul Schrater, and David L. Woods

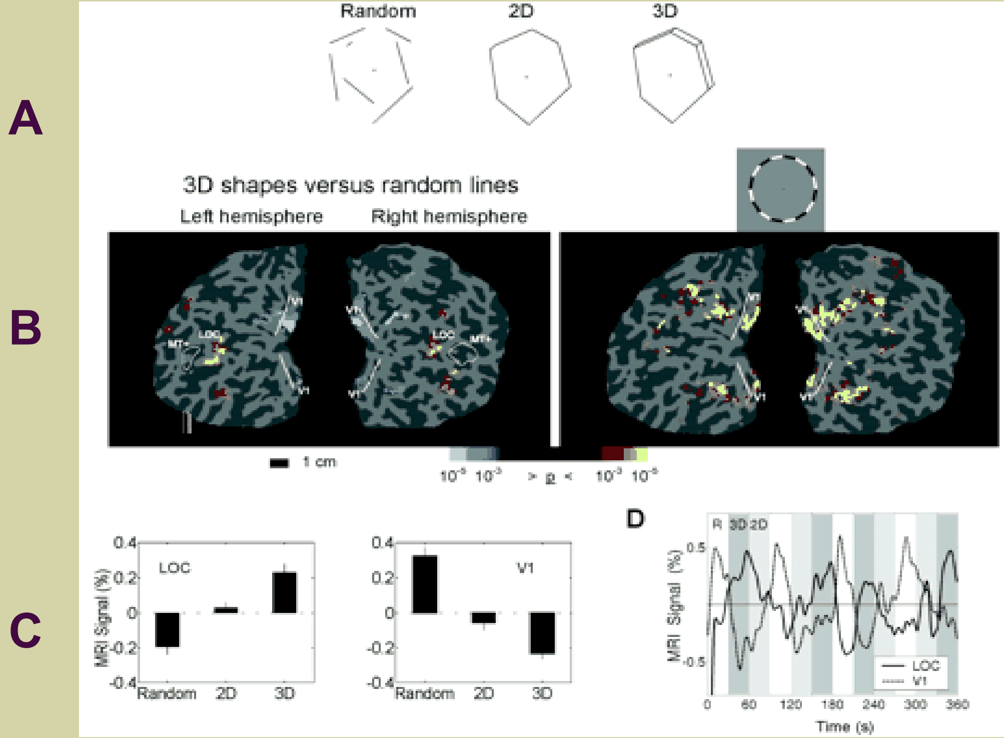


Fig. 1. Experiment 1. (A) Examples of the three different stimulus conditions. (B Left) Areas of increased (red/yellow) and decreased (blue) activity comparing 3D figures to random lines for a representative subject on a flattened representation of occipital cortex. (B Right) A flickering ring stimulus matching the mean eccentricity of the line drawings was used to independently locate the portion of V1 where the line drawing stimuli occurred. The reduced activity for the 3D figures in V1 is restricted to the cortical area representing the stimuli. The solid line indicates the representation of the vertical meridian, marking the boundary of V1. The location of MT+ defined by random dot motion is included as a reference. Fig. 6 shows the relative location of the ROIs and the location of the "cuts" to flatten the cortex. (C) The average percent signal change from the mean for the three conditions averaged over six subjects. All pair-wise comparisons are significant, $P < 0.001$. Error bars are SEM. (D) The average time course of the MRI signal in the LOC (solid line) and V1 (dashed line). Percent signal change is from the mean activation across all three conditions. Periods corresponding to the three conditions, random (R, white), 3D (dark gray), and 2D (light gray), are shown. The dissociation between the LOC and V1 is clearly evident: as activity increases in the LOC, activity in V1 declines.

Activity in higher and lower area

Summary

- A free energy principle can account for some aspects of action and perception
- The architecture of cortical systems speak to hierarchical generative models
- Estimation of hierarchical dynamic models corresponds to a generalised deconvolution of inputs to disclose their causes
- This deconvolution can be implemented in a neuronally plausible fashion by constructing a dynamic system that self-organises when exposed to inputs to suppress its free energy