Neural Coding: Generative Models

Angus Chadwick

School of Informatics, University of Edinburgh, UK

Computational Neuroscience (Lecture 9, 2024/2025)

Outline of Lecture

- Natural image statistics
- Generative models
- Sparse Coding and independent components analysis
- Convolutional Neural Networks
- Predictive Coding

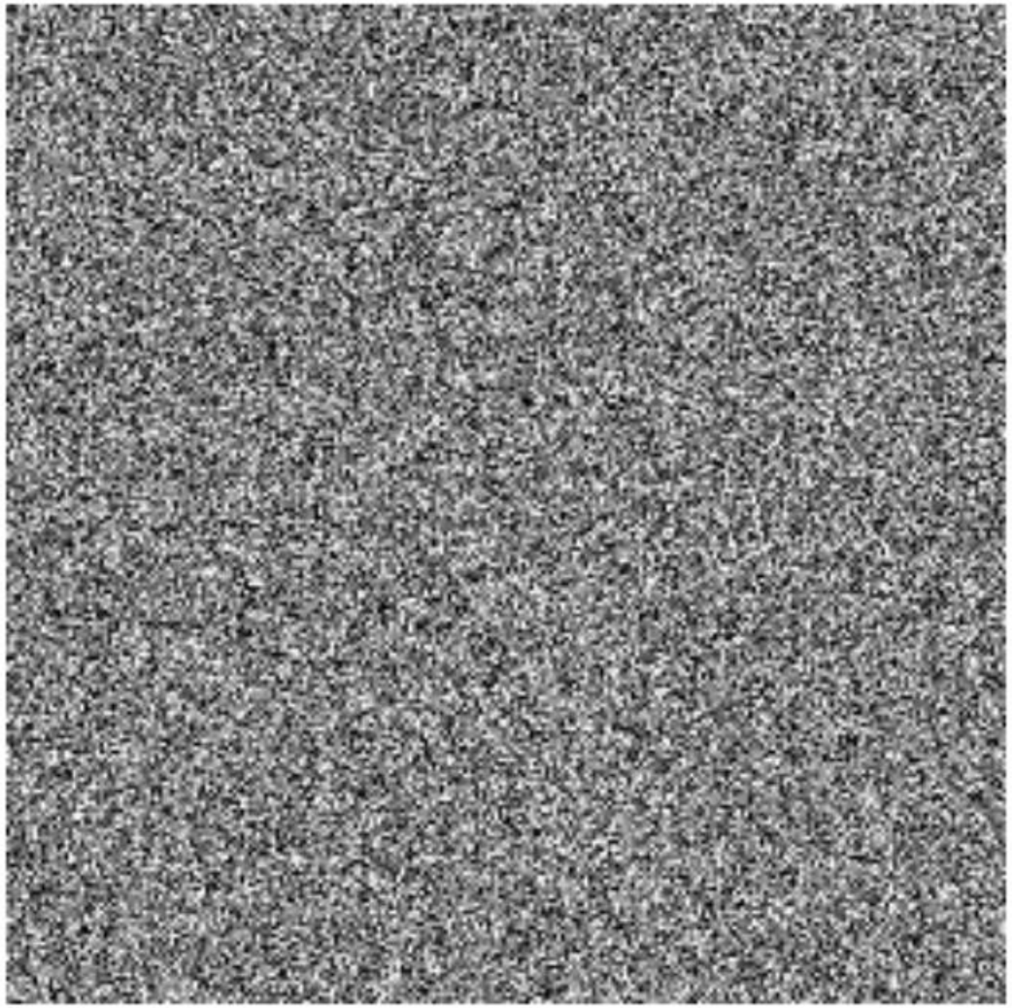
Redundancy Reduction

- Statistics of sensory input are highly redundant
- For example, neighbouring pixels of an image are highly correlated
- Horace Barlow (and earlier Attneave) hypothesised that sensory systems are designed to minimise redundancy (the Efficient Coding Hypothesis)
- One way to do this is to store a generative model for the underlying causes of sensory input (e.g., objects and light sources, although we will consider simpler models)
- Representing sensory input in terms of latent causes should result in an efficient code
- Basic idea goes back to Helmholtz (perception is “unconscious inference”)

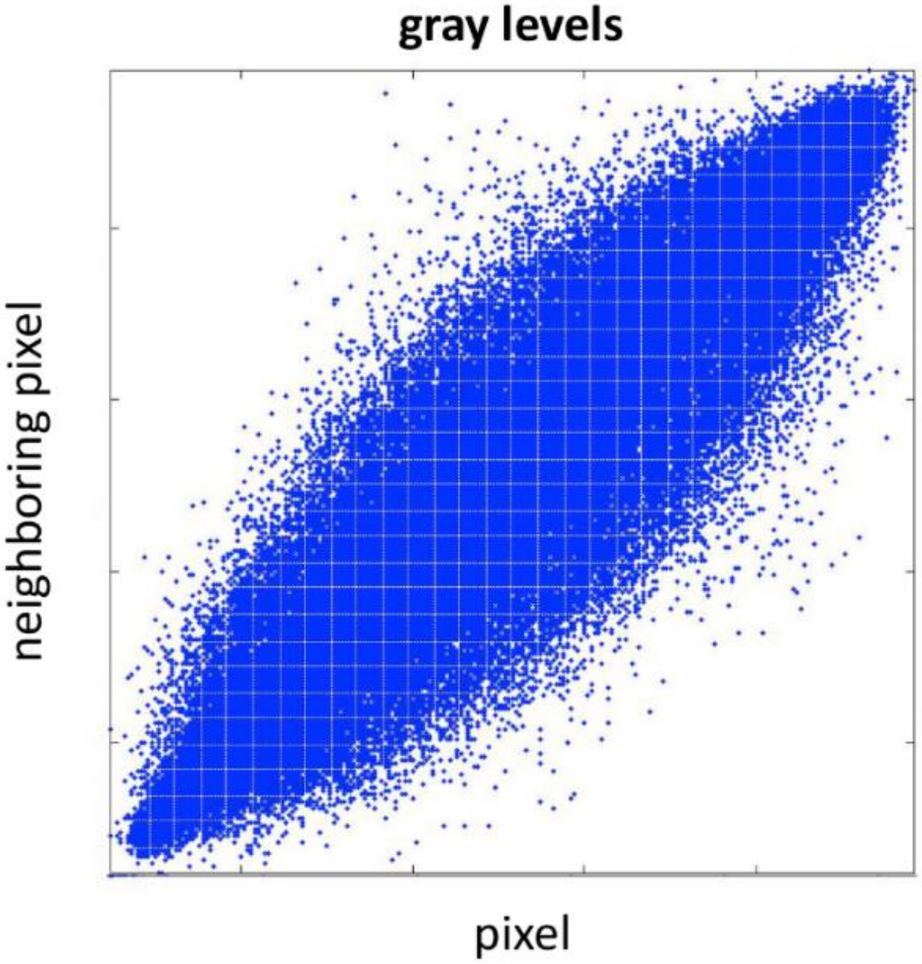
Natural Image Statistics: First Order

- First order statistics: statistics of single pixels

- An image with first order statistics of natural images but higher order statistics removed



Natural Image Statistics: Second Order



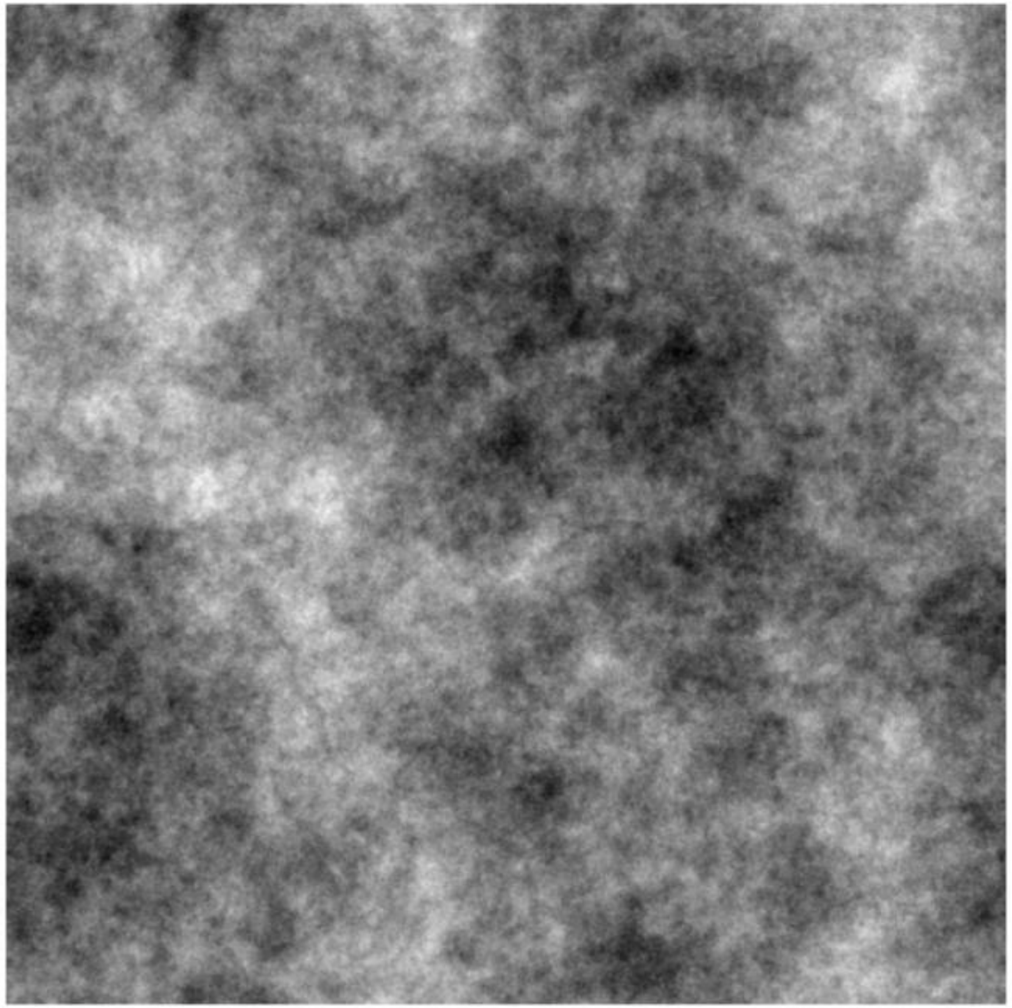
- An example of second-order statistics: correlations between neighbouring pixel values
- These are mostly removed by the retina (an operation called “whitening”)

Natural Image Statistics: Second Order

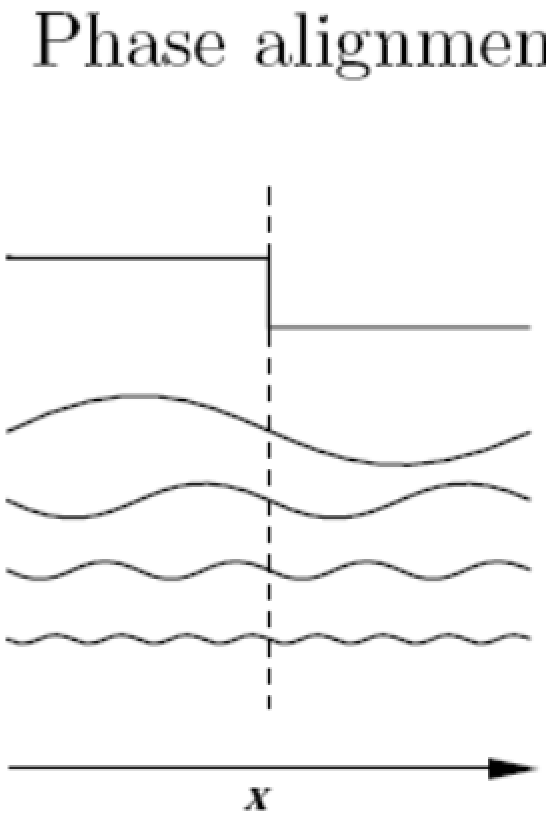
- Second order statistics:

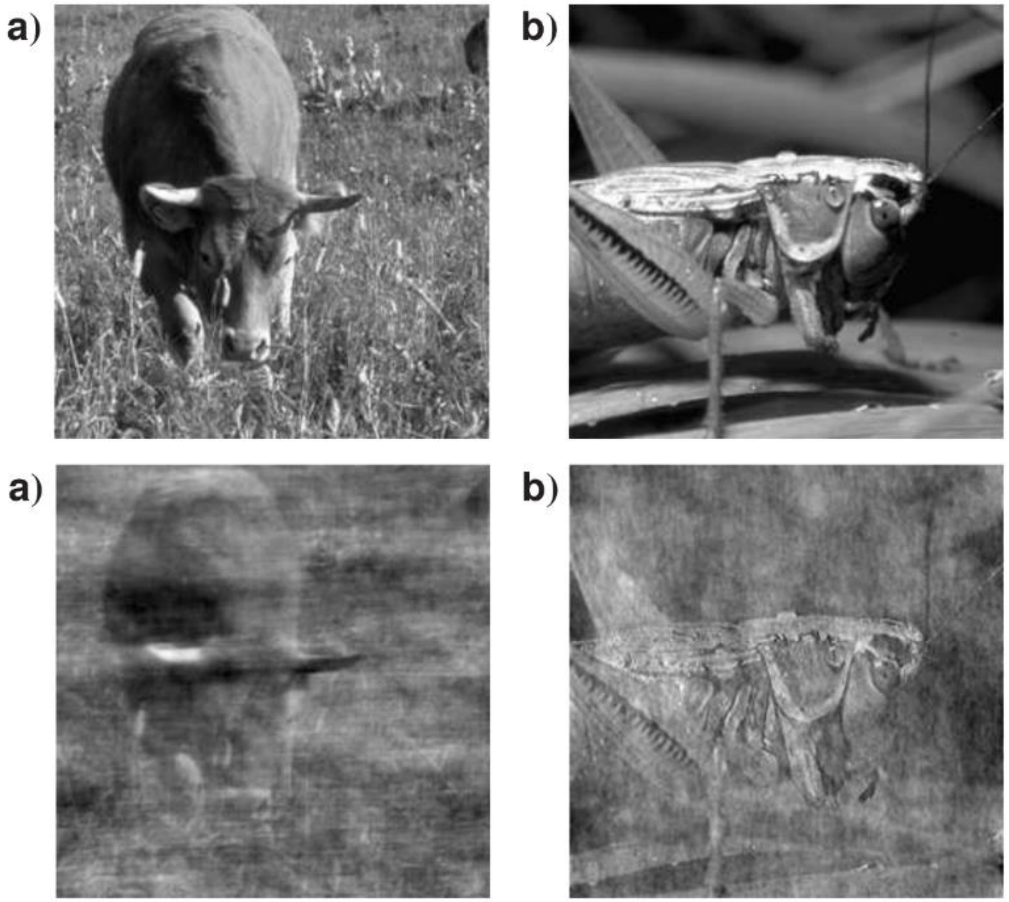
  covariance/correlation/pairwise

- An image with 2nd order statistics of natural images but higher order statistics removed



Natural Image Statistics: Higher Order





- Scrambling phases of different spatial frequencies leads to non-natural images
- Phase alignment is an example of higher-order statistics (i.e. beyond pairwise)

Natural Image Statistics: Efficient Coding

- First-order statistics: histogram equalisation (flatten distribution of pixel intensities)

- Second-order statistics: decorrelation/whitening (flatten power spectrum)

- Higher-order statistics: sparse coding, ICA, etc. can remove some of these (but not all!)

- The efficient coding hypothesis claims that the visual system successively removes first, second, and higher-order statistics of natural images, perhaps using the above algorithms

Representation of Images as Basis Functions

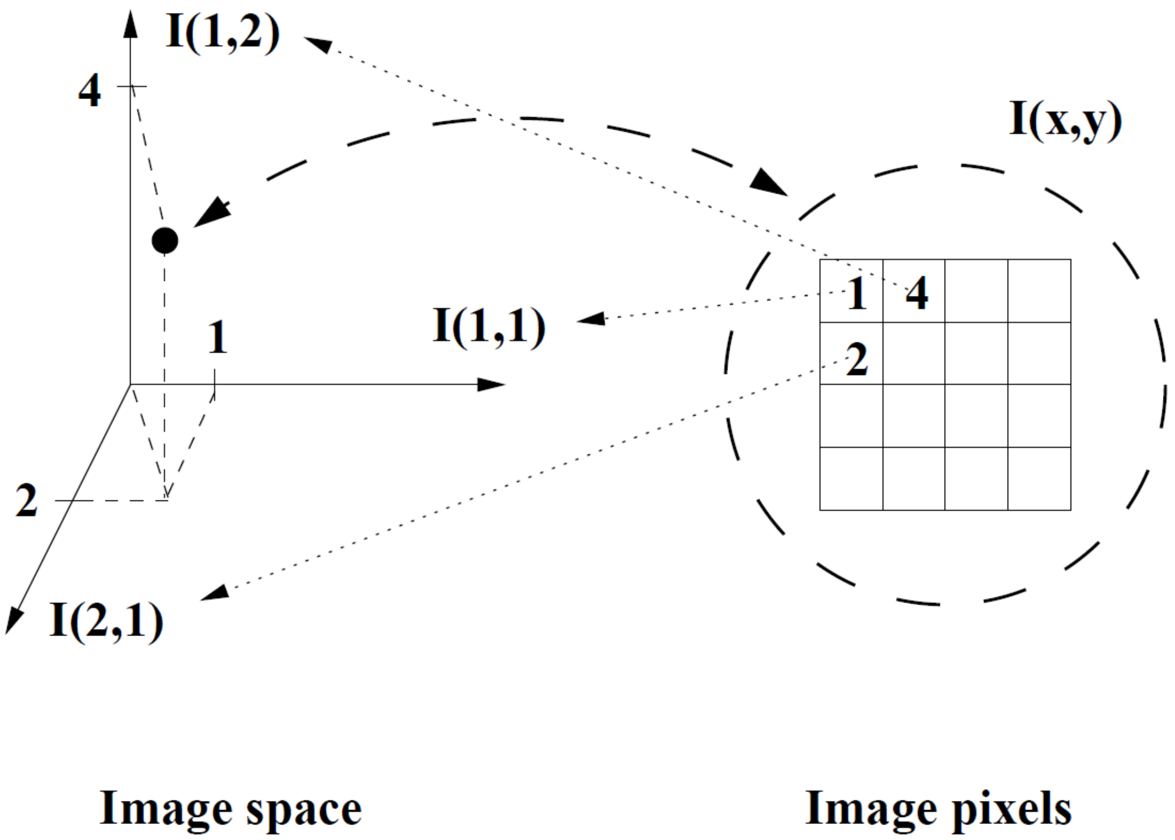
- We can flatten an image into a vector

- The image can be approximated as a weighted sum of basis functions

$$I(x,y) = \sum_{i=1}^{n} A_i(x,y) s_i$$

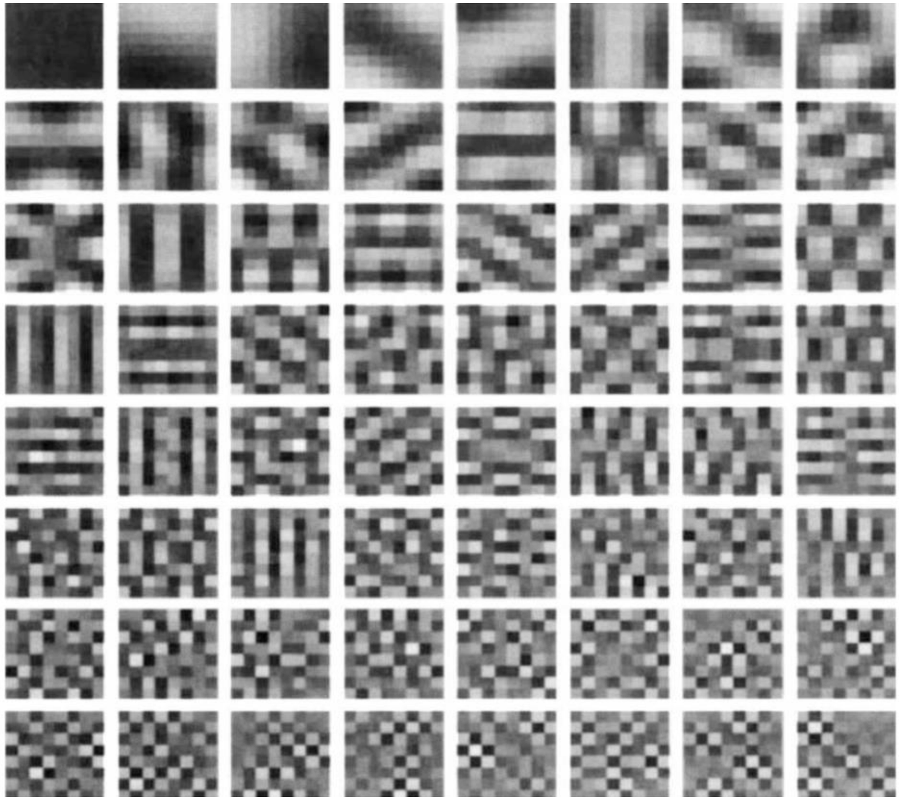
$$\mathbf{I} = A\mathbf{s}$$

- Goal: find a representation where the coefficients s_i are statistically independent of one another (an *efficient code*)

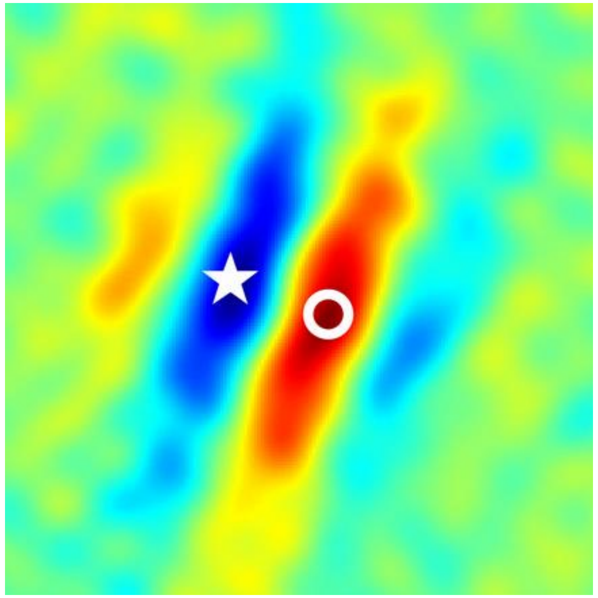


Features Learned by Principal Components Analysis

Principal Components



V1 receptive field



- PCA captures 2nd order statistics of natural images - learned filters are combinations of plane waves at different spatial frequencies.
- Can we find a model that learns Gabor filters, similar to receptive fields in V1?

Generative Models: Learning and Inference

- Assume a model *G* where data ***u*** are generated by a set of latent causes ***h***: $p(\mathbf{u}|\mathbf{h}, G)$

- We can represent the likelihood of an observation ***u*** given the generative model *G* as:

$$p(\mathbf{u}|G) = \sum_{\mathbf{h}} p(\mathbf{u}|\mathbf{h}, G)p(\mathbf{h})$$

Generative Models: Learning and Inference

- Assume a model *G* where data ***u*** are generated by a set of latent causes ***h***: $p(\mathbf{u}|\mathbf{h}, G)$

- We can represent the likelihood of an observation ***u*** given the generative model *G* as:

$$p(\mathbf{u}|G) = \sum_{\mathbf{h}} p(\mathbf{u}|\mathbf{h}, G)p(\mathbf{h})$$

- Given a model *G* and an observation ***u,*** the probability of a set of latent causes ***h*** is:

$$p(\mathbf{h}|\mathbf{u}, G) = \frac{p(\mathbf{u}|\mathbf{h}, G)p(\mathbf{h})}{p(\mathbf{u}|G)}$$

- Learning of the generative model *G* involves maximising the likelihood given a set of observations ***u***, while inference involves finding the latent causes ***h*** under a given model *G* for a single observation ***u***

Generative Models: Examples

- Mixture of Gaussians
- Factor analysis, probabilistic PCA
- Sparse coding
- Independent components analysis

Note: learning and inference are typically hard problems. We often use approximate methods and/or algorithms like expectation-maximisation.

Sparse Coding in V1

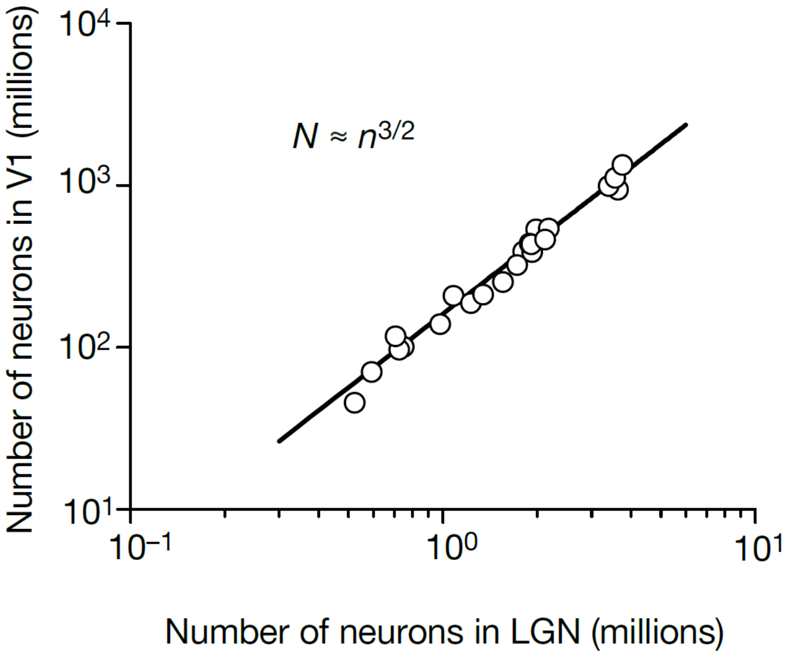
- Two properties of V1 motivate the sparse coding model:

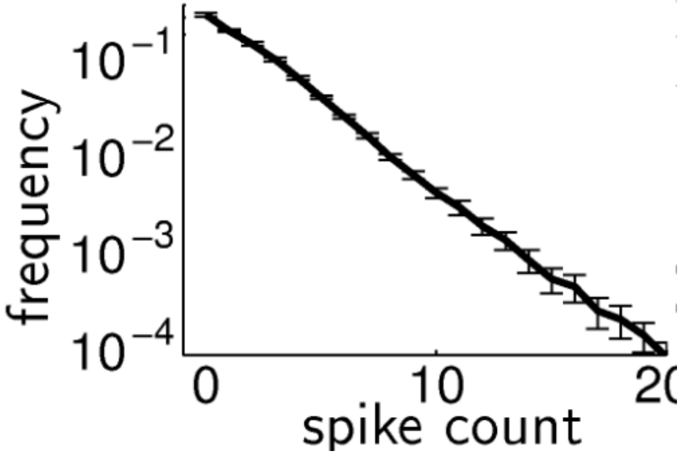
- Overcompleteness: There are many more neurons in V1 than LGN (25:1 in cat). Why?

- Sparseness: Neurons in V1 typically fire sparsely in response to natural images/movies

  (exponential firing rate distribution)

Note: Firing is sparser to natural images than to laboratory stimuli.





Measures of Sparseness

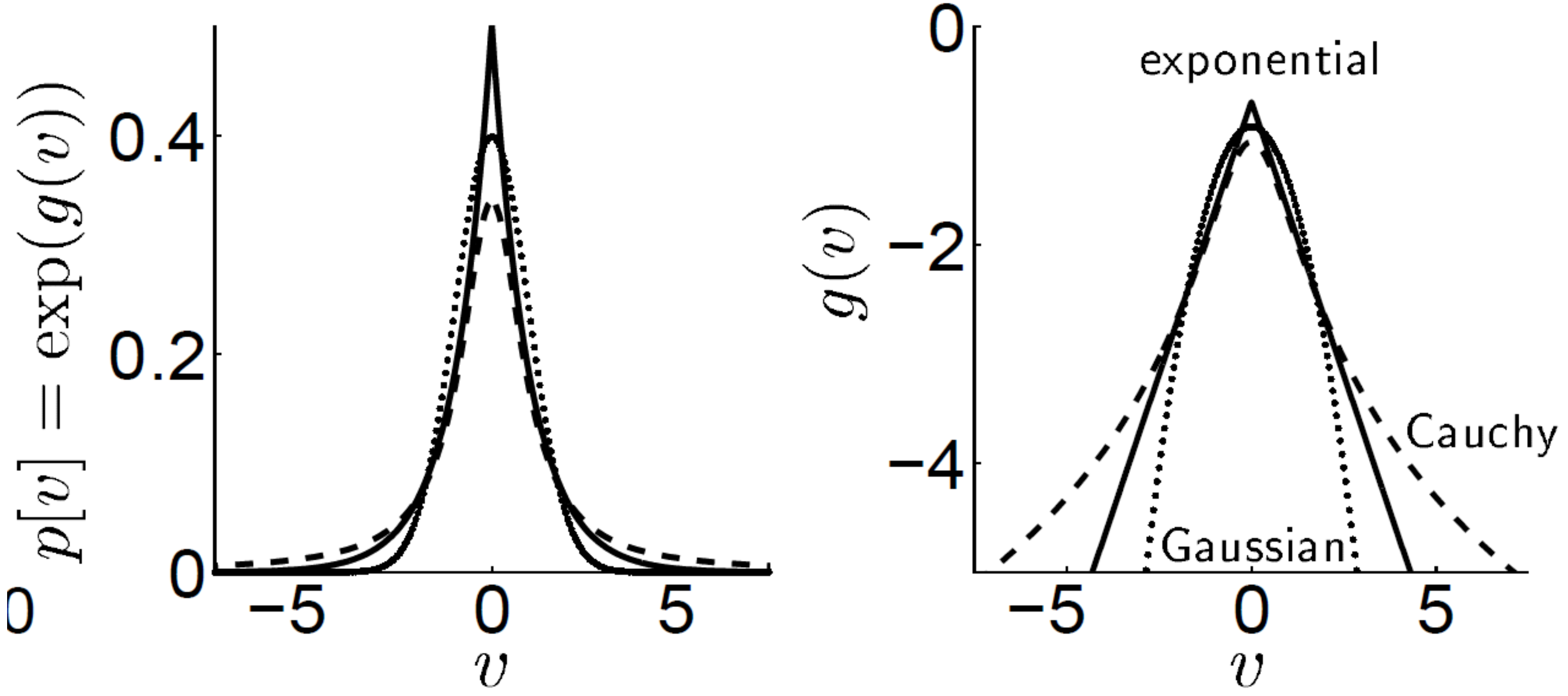
- Distributions that are close to zero most of the time but occasionally far from 0 are called sparse

- Sparse distributions are more likely than Gaussians to generate values near to zero and far from zero (heavy tailed)

$$\text{kurtosis} = \frac{\int p(x)(x-\overline{x})^4 dx}{\left(\left[\int p(x)(x-\overline{x})^2 dx\right]\right)^2} - 3$$

- Gaussian has zero kurtosis, positive means sparser than Gaussian, negative means less sparse

Measures of Sparseness



$p(h) = \exp(g(h))$

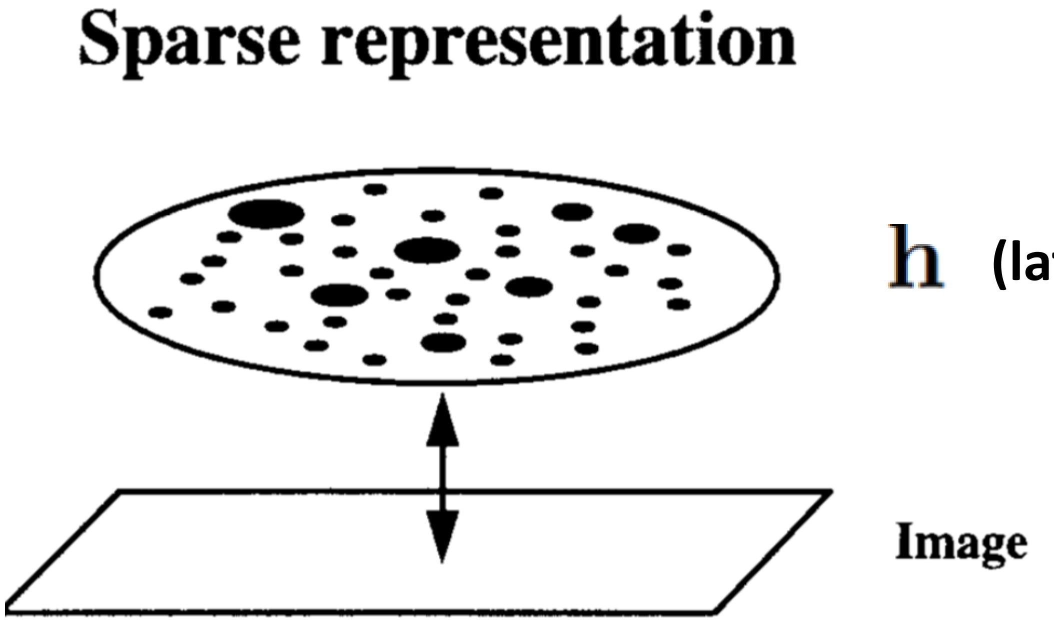
exponential: $g(h) = -|h|$

Cauchy: $g(h) = -\log(1 + h^2)$

Gaussian: $g(h) = -h^2/2$

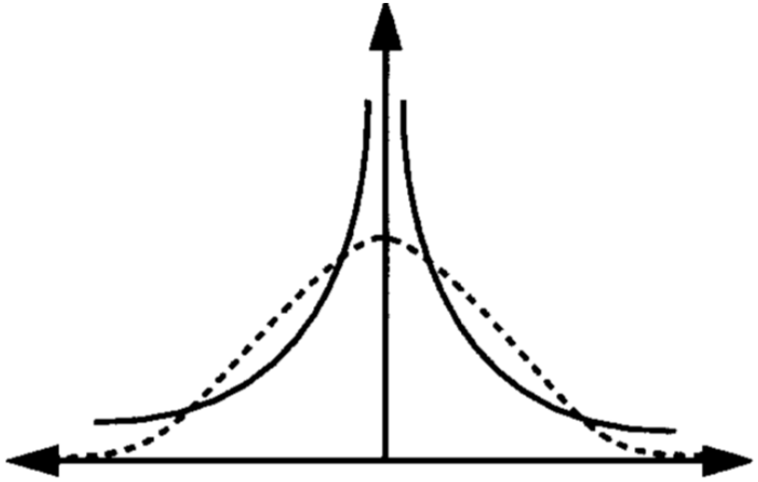
Sparse Coding: Olshausen and Field

- Postulate a generative model for natural images – sparse and independent latent causes
- Assume many more latent causes than observed variables (i.e. pixels)
- Fit model to natural images – does anything that looks like V1 emerge?



$\mathbf{h}$ (latent causes)

$p(h_i)$ (sparse prior)



$$\mathbf{u} = B\mathbf{h} + \mathbf{n}$$

Sparse Coding: Olshausen and Field

Prior: 1) independent latent causes (factorial): $p(\mathbf{h}) = \prod_i p(h_i)$

2) sparse latent causes (exponential or Cauchy)

$$p(h_i) \propto \exp(g(h_i))$$

$$g(h_i) = -|h_i| \qquad\qquad g(h_i) = -\log(1 + h_i^2)$$

Sparse Coding: Olshausen and Field

Prior: 1) independent latent causes (factorial): $p(\mathbf{h}) = \prod_i p(h_i)$

2) sparse latent causes (exponential or Cauchy)

$$p(h_i) \propto \exp(g(h_i))$$

$$g(h_i) = -|h_i| \qquad\qquad g(h_i) = -\log(1 + h_i^2)$$

Likelihood: Linear-Gaussian, from high-D latent causes ***h*** to lower-D image ***u***:

$$p(\mathbf{u}|\mathbf{h}, B) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}||\mathbf{u} - B\mathbf{h}||^2\right]$$

Learning and Inference

Given a set of images ***u***, need to learn a matrix *B* and infer coefficients for ***h*** for each ***u***.

Learning:

$$\langle p(\mathbf{u}|B)\rangle_{\mathbf{u}} = \left\langle \int p(\mathbf{u}|\mathbf{h}, B)p(\mathbf{h})d\mathbf{h} \right\rangle_{\mathbf{u}}$$

Inference:

$$p(\mathbf{h}|\mathbf{u}, B) \propto p(\mathbf{u}|\mathbf{h}, B)p(\mathbf{h})$$

Learning maximises the marginal likelihood of *B*, inference maximises the posterior probability of ***h*** given *B* and a single ***u***

Inference

Given *B,* maximise posterior with respect to ***h***:

$$\log p(\mathbf{h}|\mathbf{u}, B) = -\frac{1}{2\sigma^2}\|\mathbf{u} - B\mathbf{h}\|^2 + \sum_i g(h_i) + \text{const}$$

$$\nabla \log p(\mathbf{h}|\mathbf{u}, B) = \frac{1}{\sigma^2}B^T[\mathbf{u} - B\mathbf{h}] + g'(\mathbf{h})$$

Inference

Given *B,* maximise posterior with respect to ***h***:

$$\log p(\mathbf{h}|\mathbf{u}, B) = -\frac{1}{2\sigma^2}\|\mathbf{u} - B\mathbf{h}\|^2 + \sum_i g(h_i) + \text{const}$$

$$\nabla \log p(\mathbf{h}|\mathbf{u}, B) = \frac{1}{\sigma^2} B^T[\mathbf{u} - B\mathbf{h}] + g'(\mathbf{h})$$

Solve via "coordinate ascent": $\mathbf{h}_{t+1} = \mathbf{h}_t + \epsilon \nabla \log p(\mathbf{h}_t|\mathbf{u}, B)$

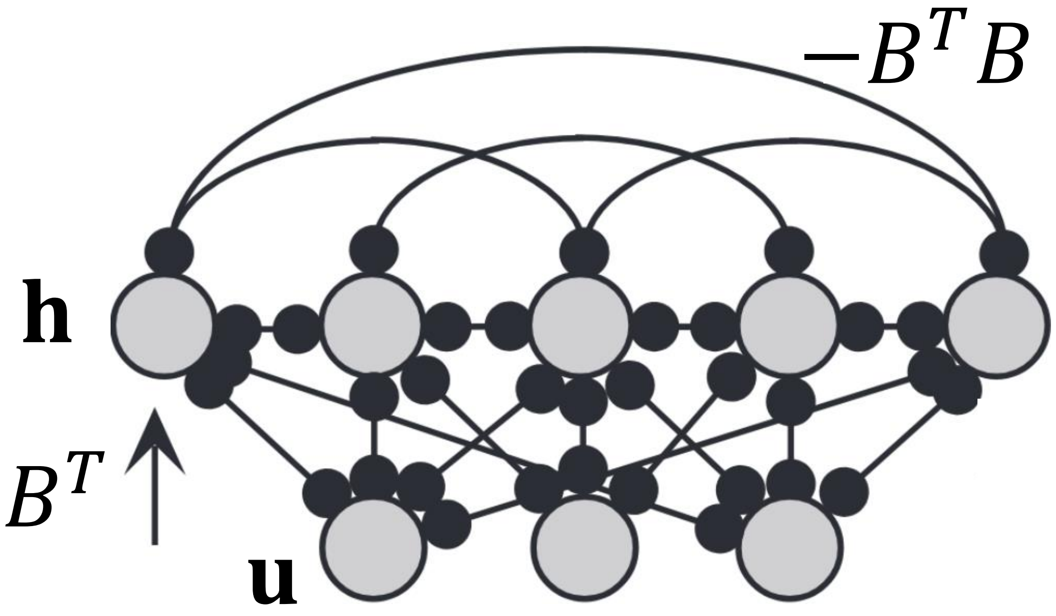
Converges to *local* maximum

Inference: Network Interpretation

Can also be written in terms of dynamics of a network of neurons:

$$\mathbf{h}_{t+1} = \mathbf{h}_t + \epsilon \nabla \log p(\mathbf{h}_t | \mathbf{u}, B)$$

$$\tau \frac{d\mathbf{h}}{dt} = \frac{1}{\sigma} B^T [\mathbf{u} - B\mathbf{h}] + g'(\mathbf{h})$$



Caption

[Figure: Dayan and Abbott]

The network can be interpreted as LGN and V1, with recurrent weights between V1 neurons enforcing the prior and causing convergence to the MAP solution via the network dynamics

Learning

How do we learn the matrix *B* for a given set of images ***u***?

$$\hat{B} = \operatorname{argmax}\left(\langle p(\mathbf{u}|B)\rangle_{\mathbf{u}}\right) = \operatorname{argmax}\left(\left\langle \int d\mathbf{h}\, p(\mathbf{u}|\mathbf{h}, B)p(\mathbf{h}) \right\rangle_{\mathbf{u}}\right)$$

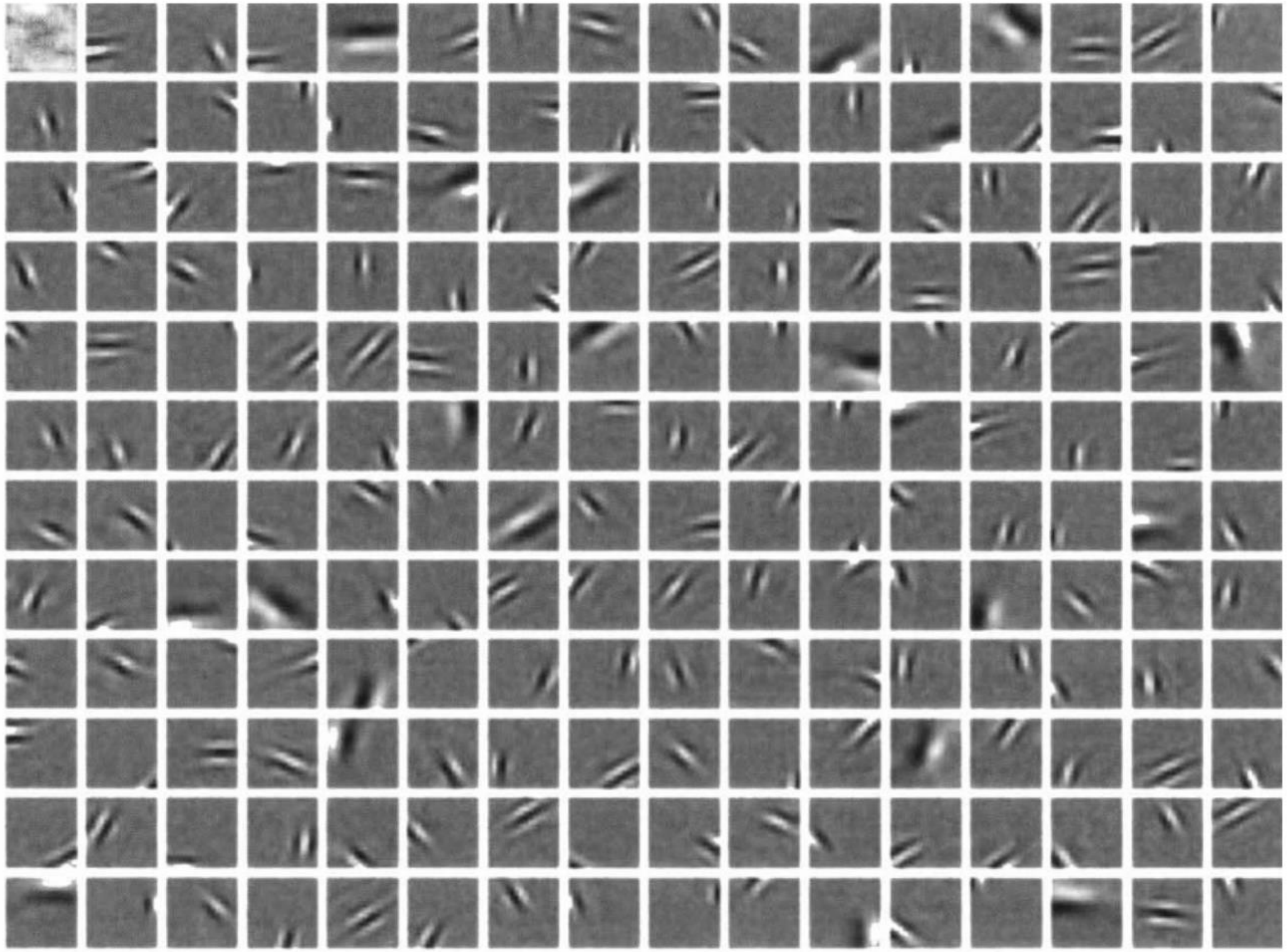
Optimising this is hard due to the integral over the high-dimensional vector ***h***. Learning can be slow to converge.

Olshausen and Field used an approximate posterior to speed up learning (see Dayan and Abbott Ch. 10).

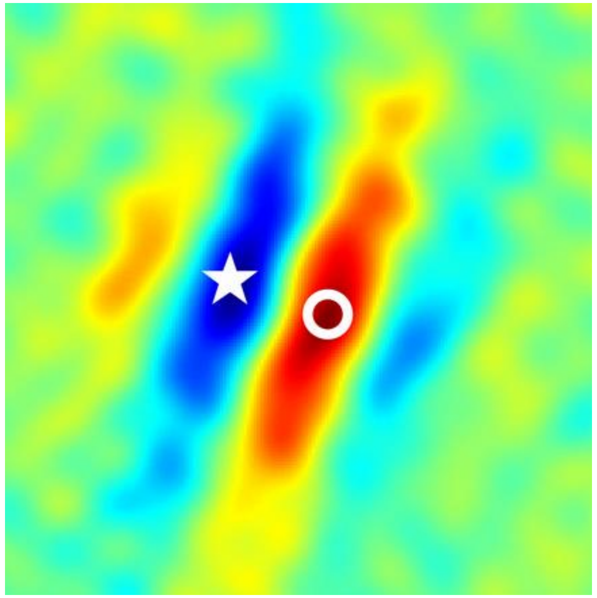
Gabor Receptive Fields via Sparse Coding

- V1-like receptive fields emerge naturally when this algorithm is trained on natural images.

Rows of *B* matrix (reshaped back into a matrix)



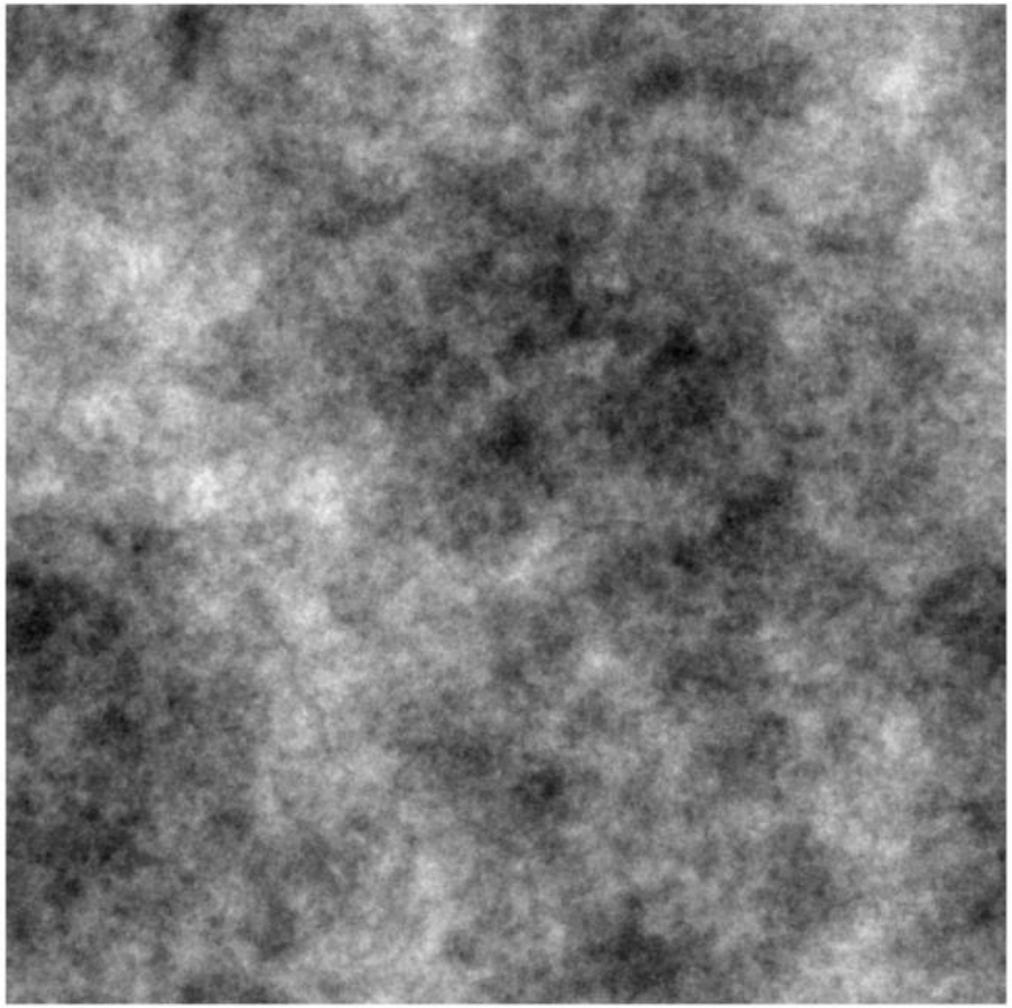
V1 receptive field



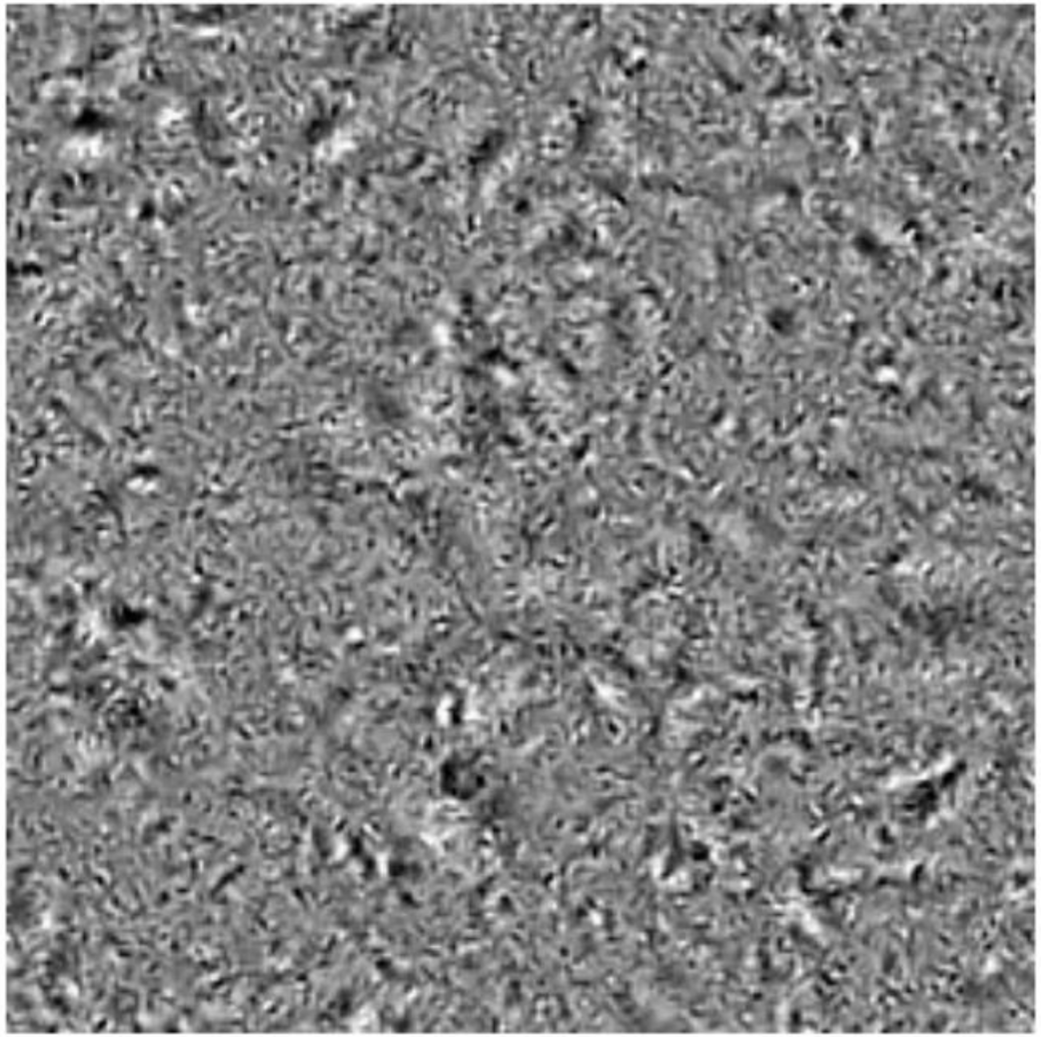
[Figure Olshausen and Field (1996)]

Image Synthesis via Sparse Coding

2nd Order Statistics Model



Sparse Coding Model



Limitations of Sparse Coding

- Model assumptions are ad hoc (why sparse?)
- Learned causes are not independent, violating assumption of model and suggesting not all statistical structure has been captured by model
- Converges to local minimum, sometimes learns multiple copies of same filter
- Choice of number of latent variables, and other hyperparameters is also ad hoc
- Have to whiten images first, otherwise just learns PCA
- Better models involve hierarchies of latent causes (objects, occlusion, luminance, etc.)

Summary: Sparse Coding

- A linear-Gaussian generative model for natural images
- Assumes overcomplete representation (more latents than observed), similar to LGN-V1 cell numbers
- Assumes sparse and independent latent causes
- Gabor filters emerge when trained on natural images
- First model to derive a computational account of V1 receptive fields

Independent Components Analysis

- A similar approach to sparse coding: use a linear generative model (this time with no noise)

$$\mathbf{u} = B\mathbf{h}$$

- Assume that latent variables ***h*** mix to form the observed signal ***u*** – goal is to unmix them

- Various objectives can be used: maximise mutual information between ***u*** and ***h*** (infomax); maximise non-Gaussianity of ***h*** (e.g., Kurtosis), etc.

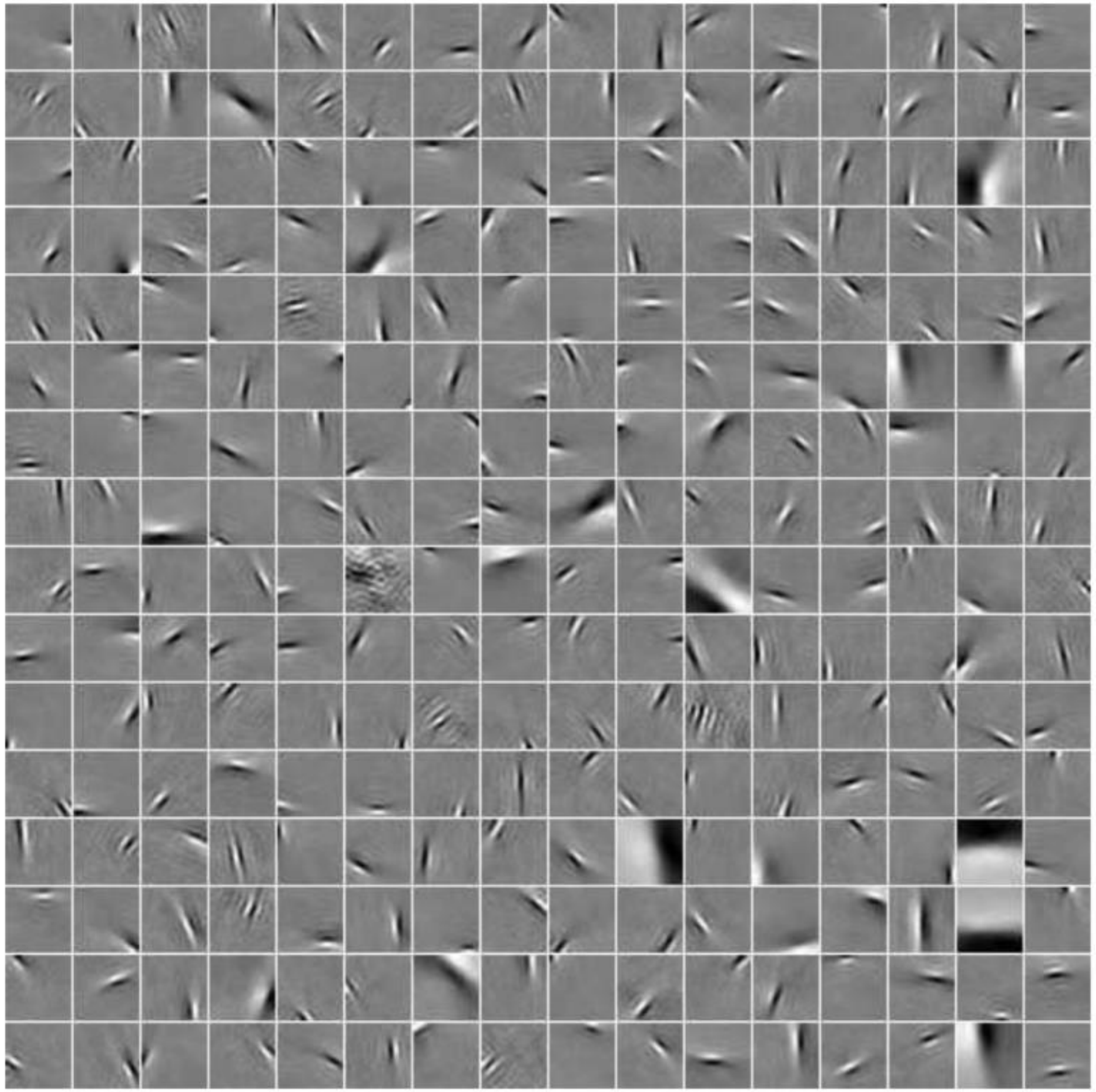
- We will skip the technical details and simply show the behaviour of the model on natural images

Independent Components Analysis

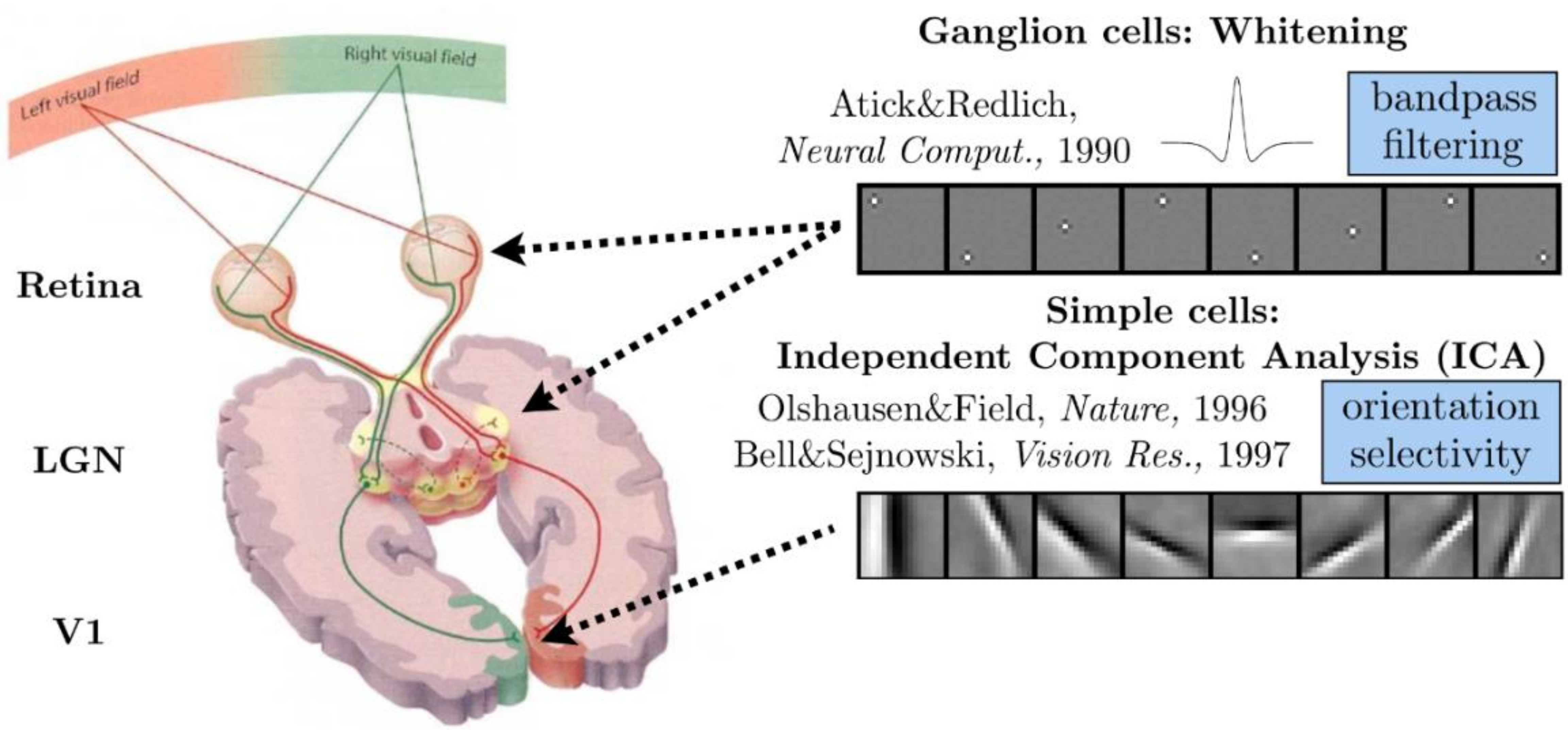
- Middle: observed data, constructed by two mixed independent components
- Left: apply PCA, whitens the data via rotation and scaling (*x* and *y* uncorrelated but still dependent)
- Right: apply ICA, *x* and *y* are now independent (but not Gaussian)

ICA: Learned Receptive Fields

- Basis functions learned by ICA trained on natural images



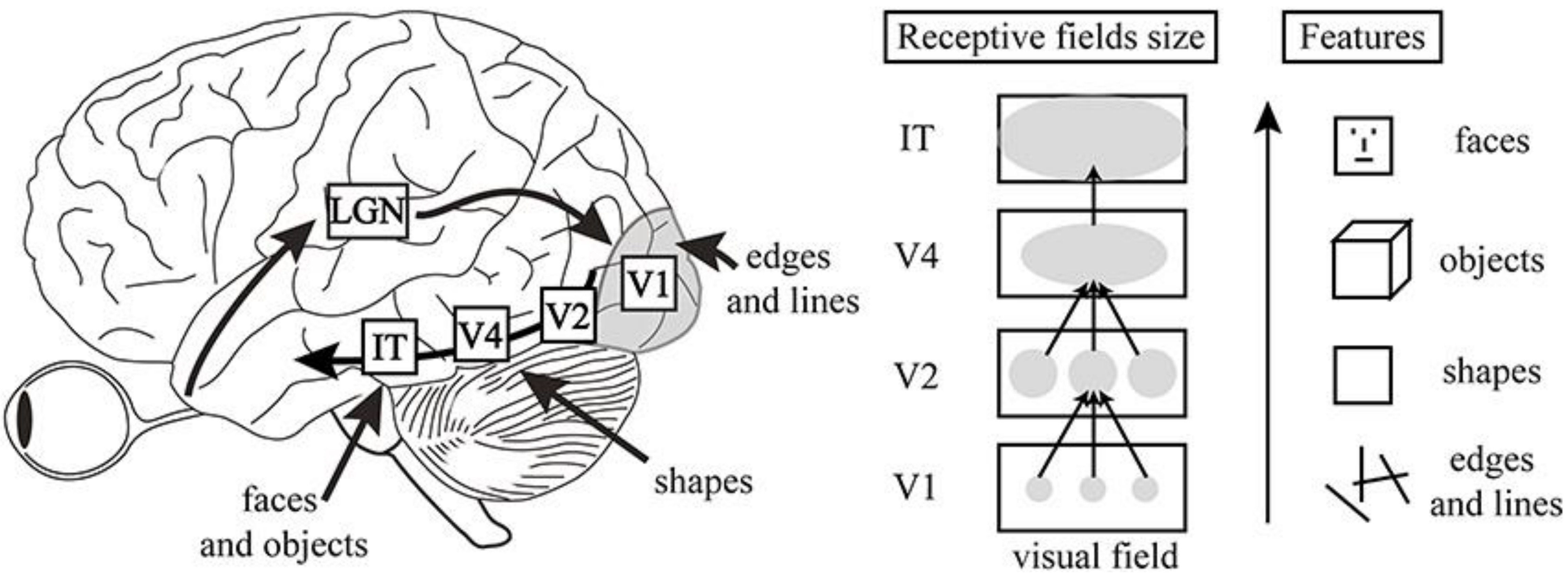
The Early Visual System: Whitening + ICA/Sparse Coding



Sparse Coding vs ICA

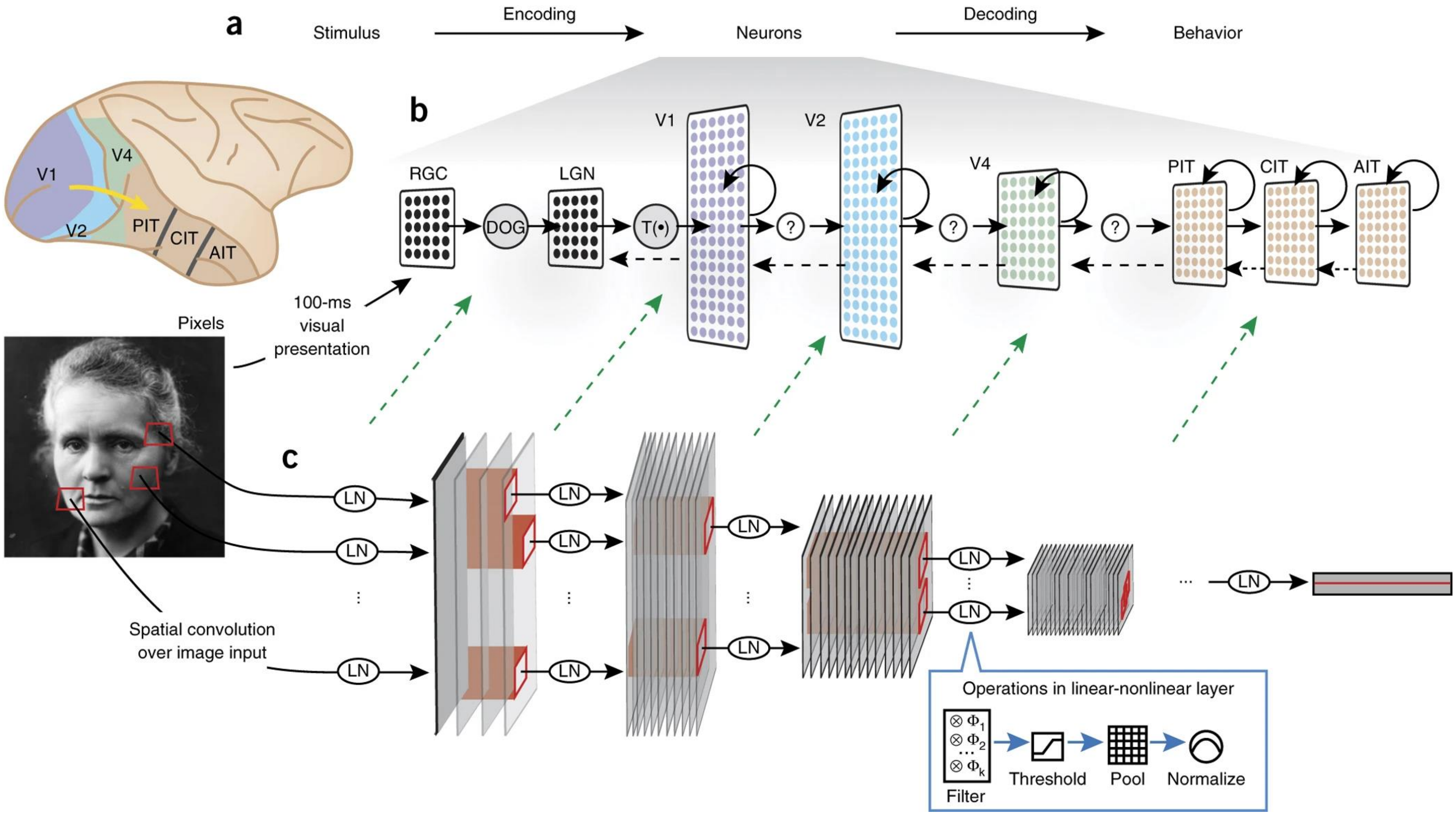
- Both are linear generative models
- Both reproduce V1-like receptive fields (as do various other algorithms)
- Both generate sparse output for natural images
- Different objectives/algorithms (ICA: maximise information/non-Gaussianity, SC: sparse overcomplete representation)
- Both fail to capture representations in higher visual brain areas (but some work has been done to address this…)

Higher Visual Processing



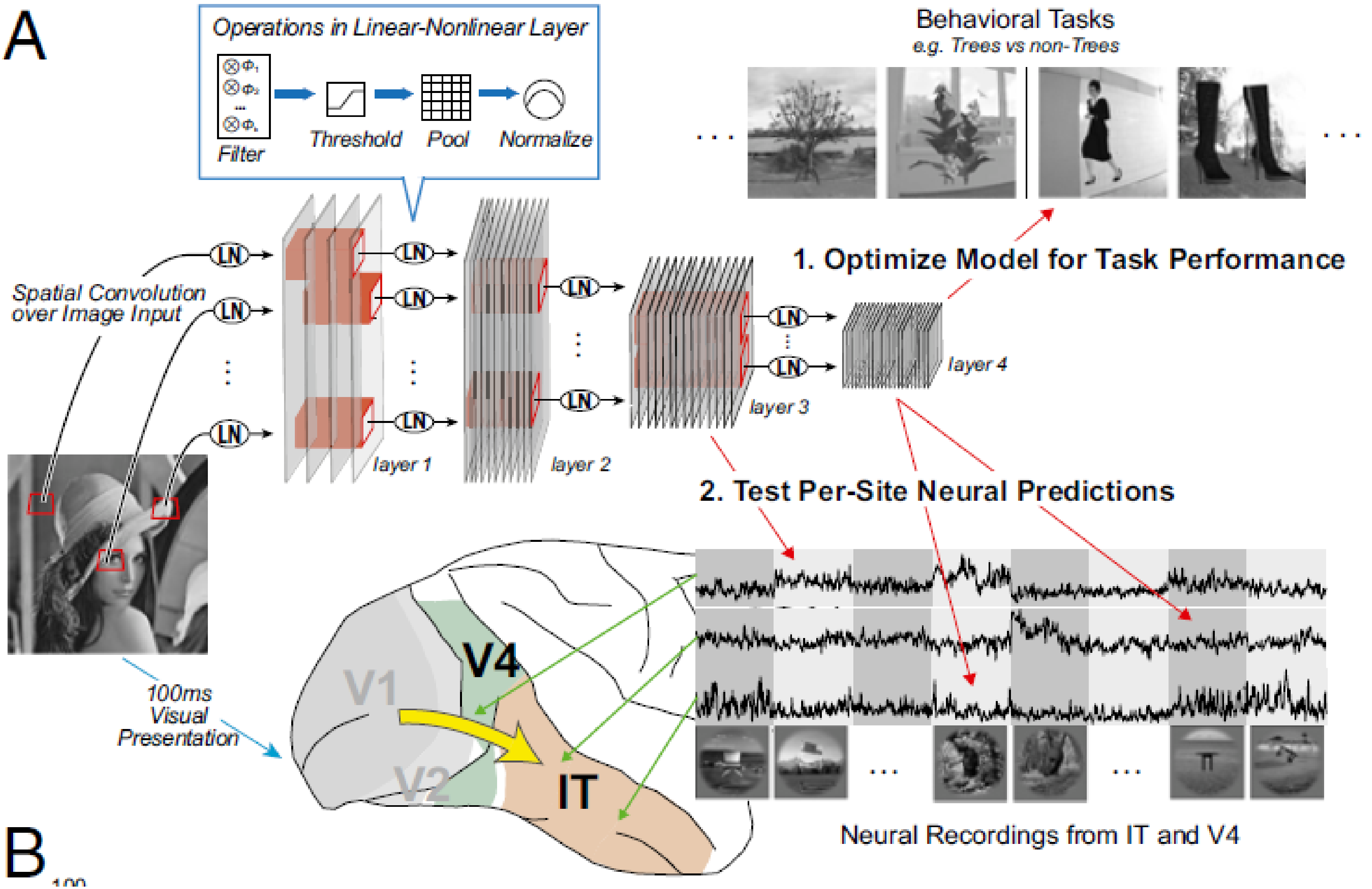
Convolutional Neural Networks

Do CNNs resemble the visual system?



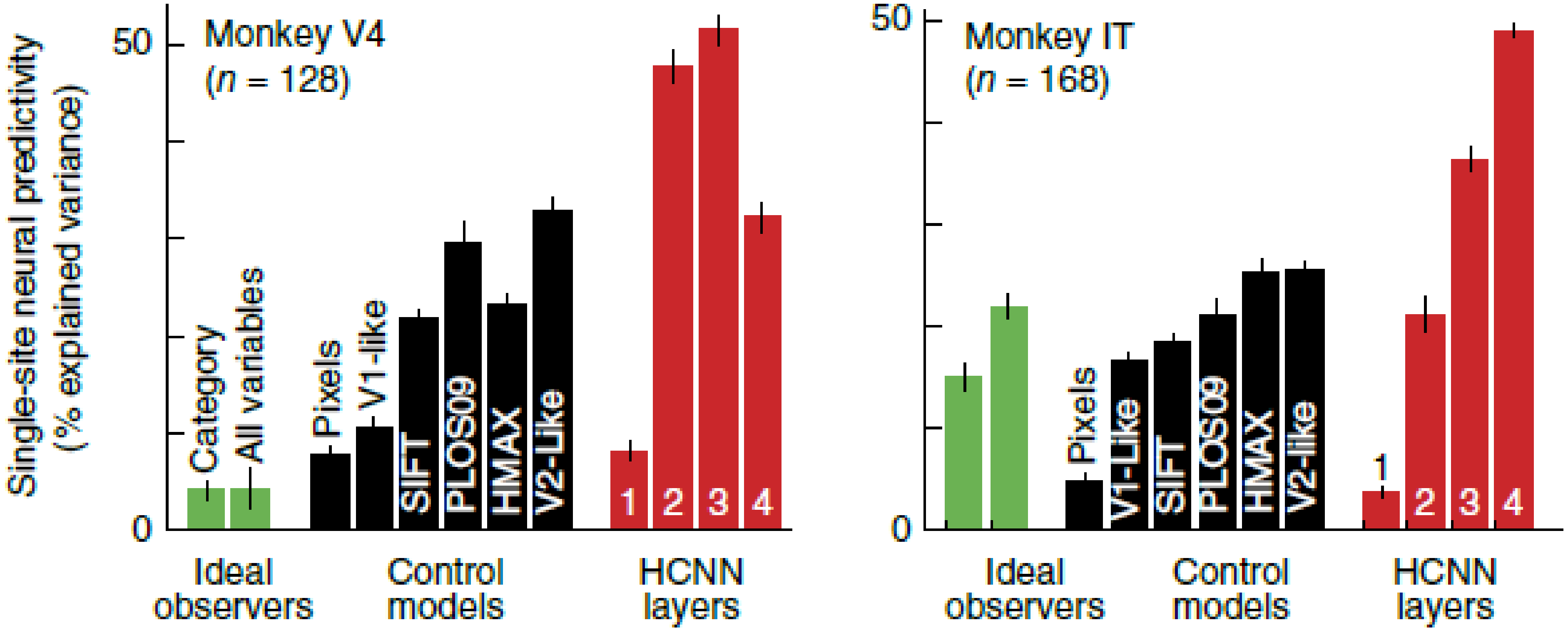
Convolutional Neural Networks

Train CNNs to categorise images, compare learned representations to visual system



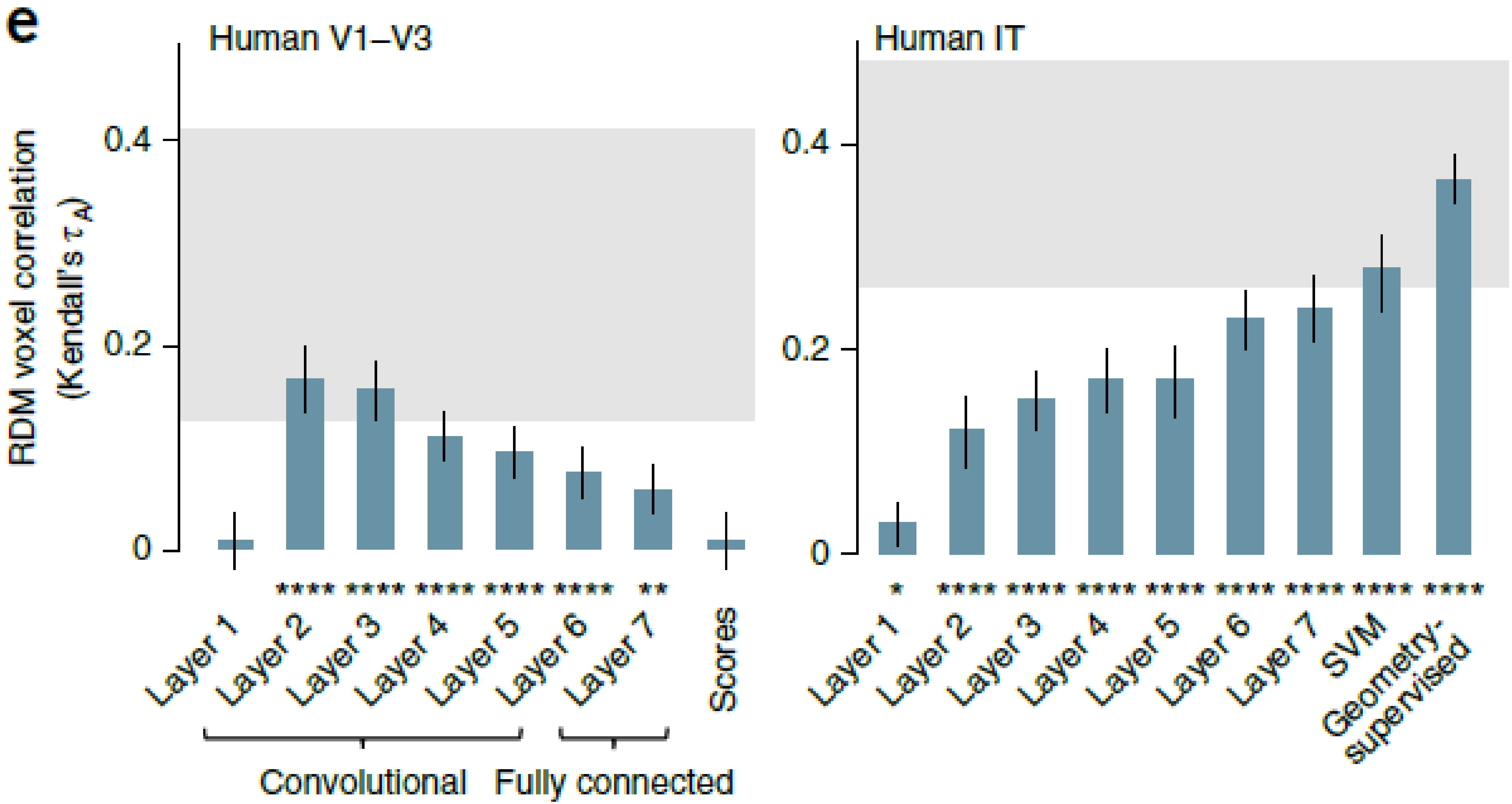
Convolutional Neural Networks

CNN responses better resemble higher visual system layers than other models



Convolutional Neural Networks

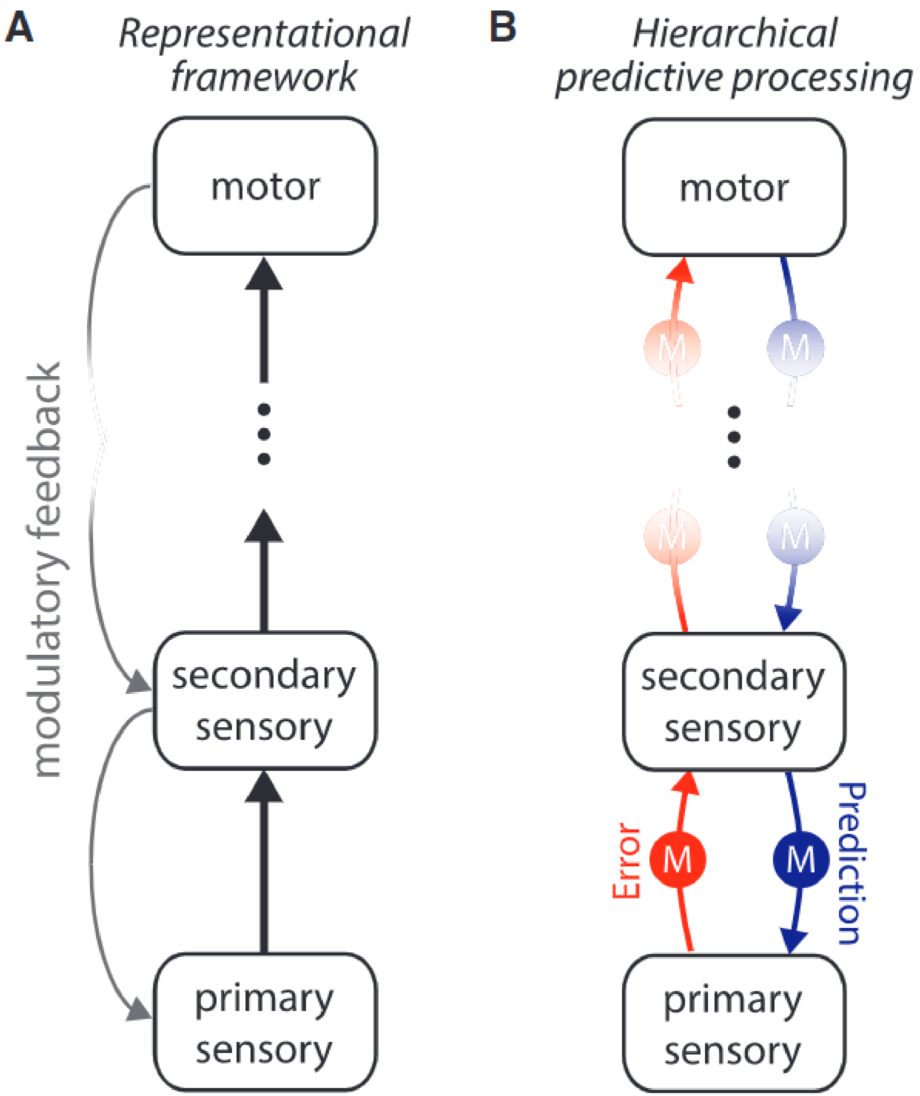
Layers of the CNN roughly correspond to the layers of the visual system

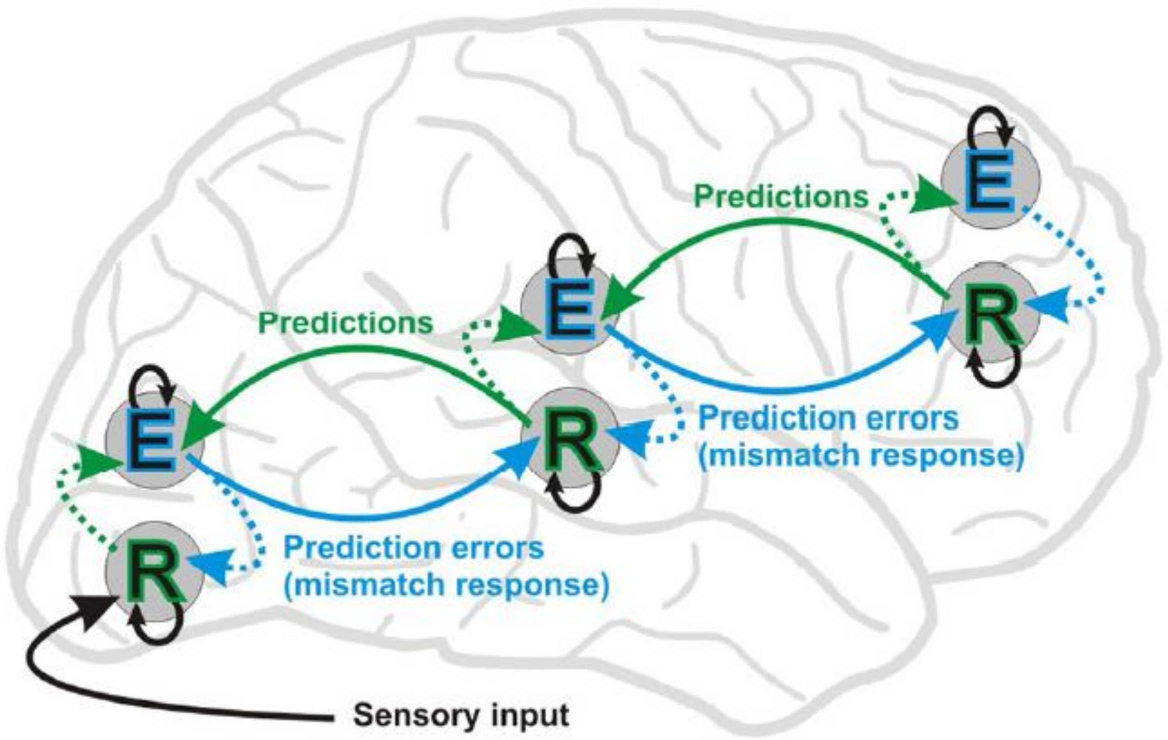


Convolutional Neural Networks: Summary

- CNNs are discriminative, not generative models
- CNNs are supervised, the brain doesn’t get much supervision (mostly reward/punishment)
- CNNs use backprop, the brain can’t use backprop (backprop is non-local)
- Nonetheless, when trained on natural images, CNNs reproduce aspects of the visual system, including deeper layers
- A less principled approach than generative models, and less biologically plausible in terms of learning, but generative models struggle to capture deeper layers of the visual system

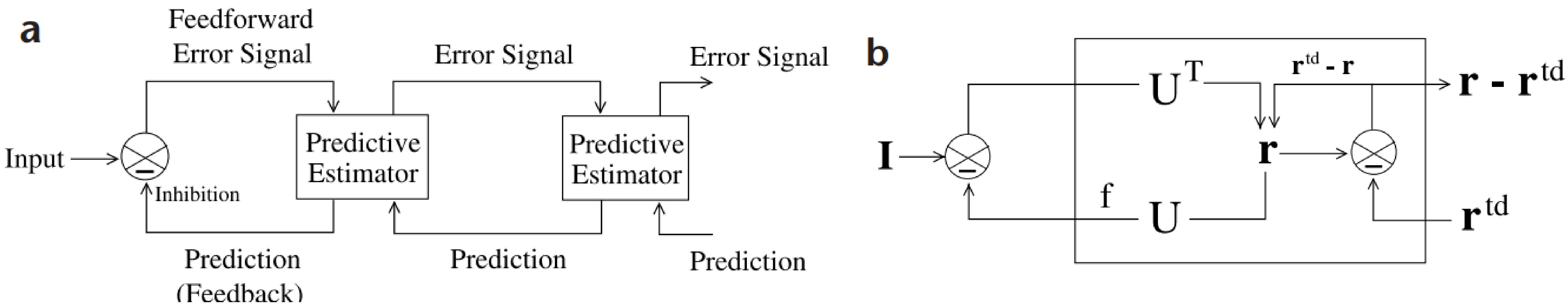
Role of Feedback: Predictive Coding





Mrsic-Flogel and Keller (2018)

Predictive Coding: Generative Model



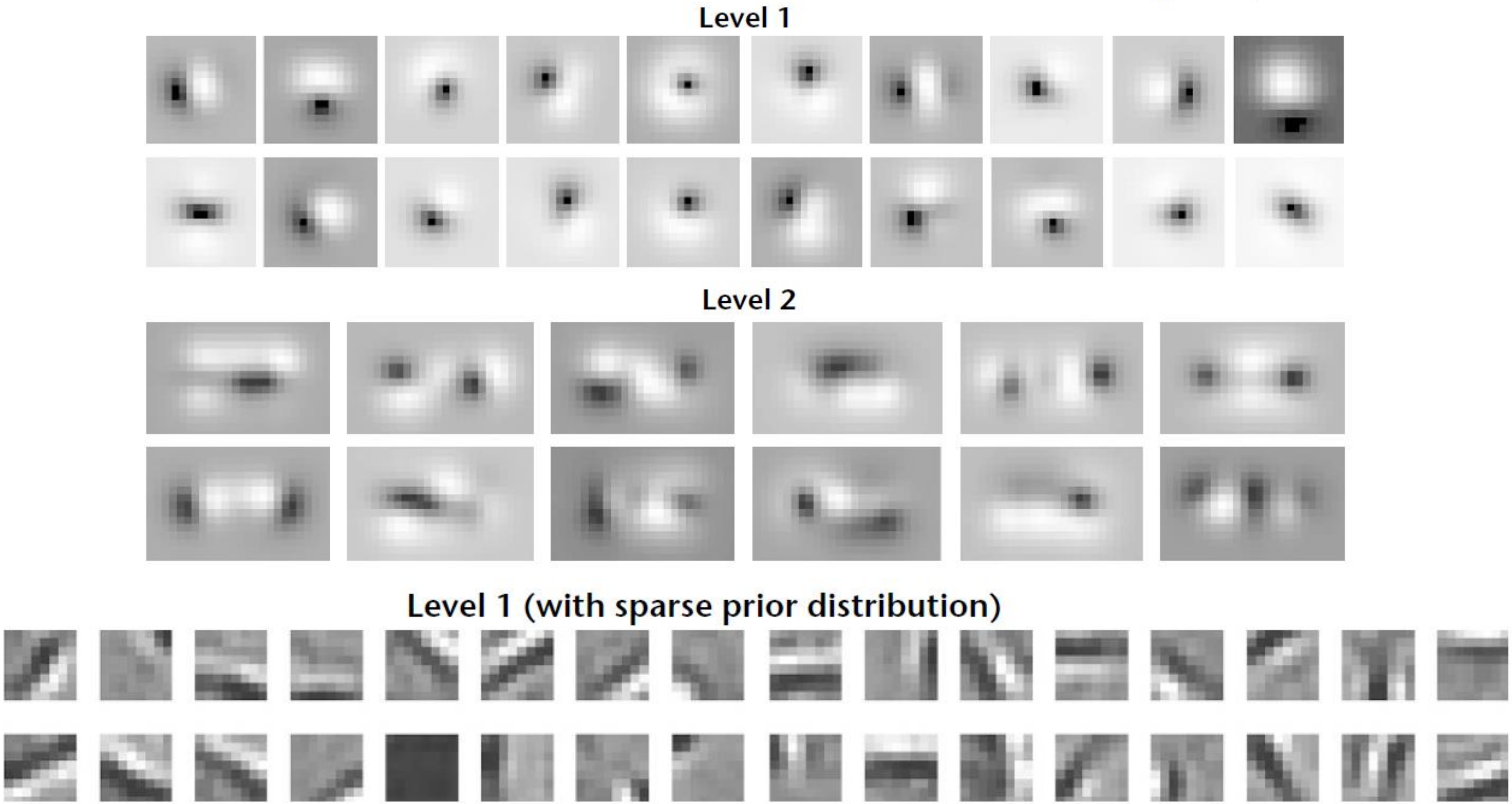
Predictive coding uses the following Hierarchical Generative Model

$$\mathbf{I} = f(\mathrm{U}\mathbf{r}) + \mathbf{n} \qquad \mathbf{r} = \mathbf{r}^{td} + \mathbf{n}^{td} \qquad \mathbf{r}^{td} = f(\mathrm{U}^h\, \mathbf{r}^h)$$

I is the image, *U* is a learned matrix, and ***n*** is Gaussian noise.
This represents a single layer of the hierarchy, which can be stacked to create more layers.

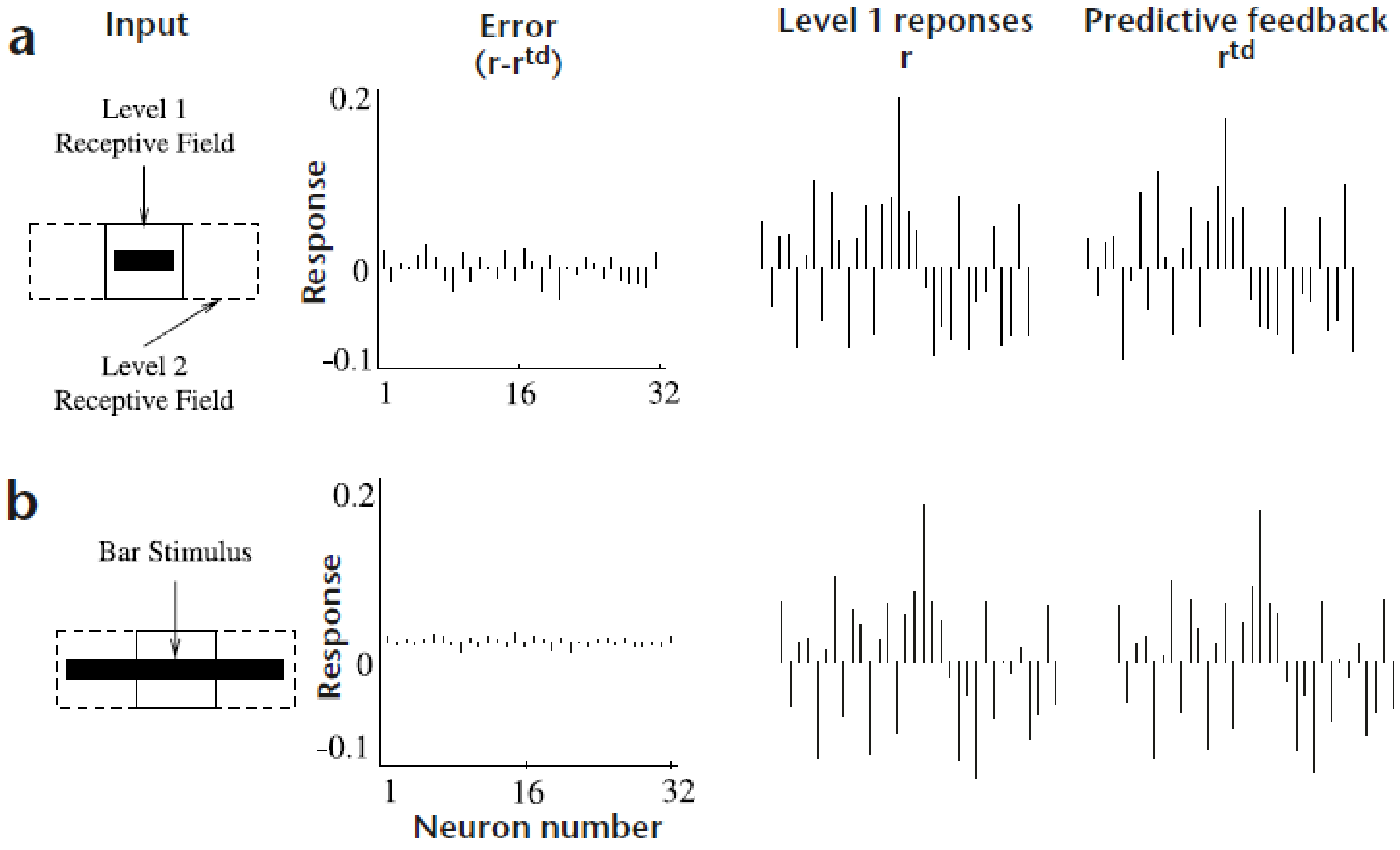
Rao and Ballard (1999)

Predictive Coding: Learned Representations



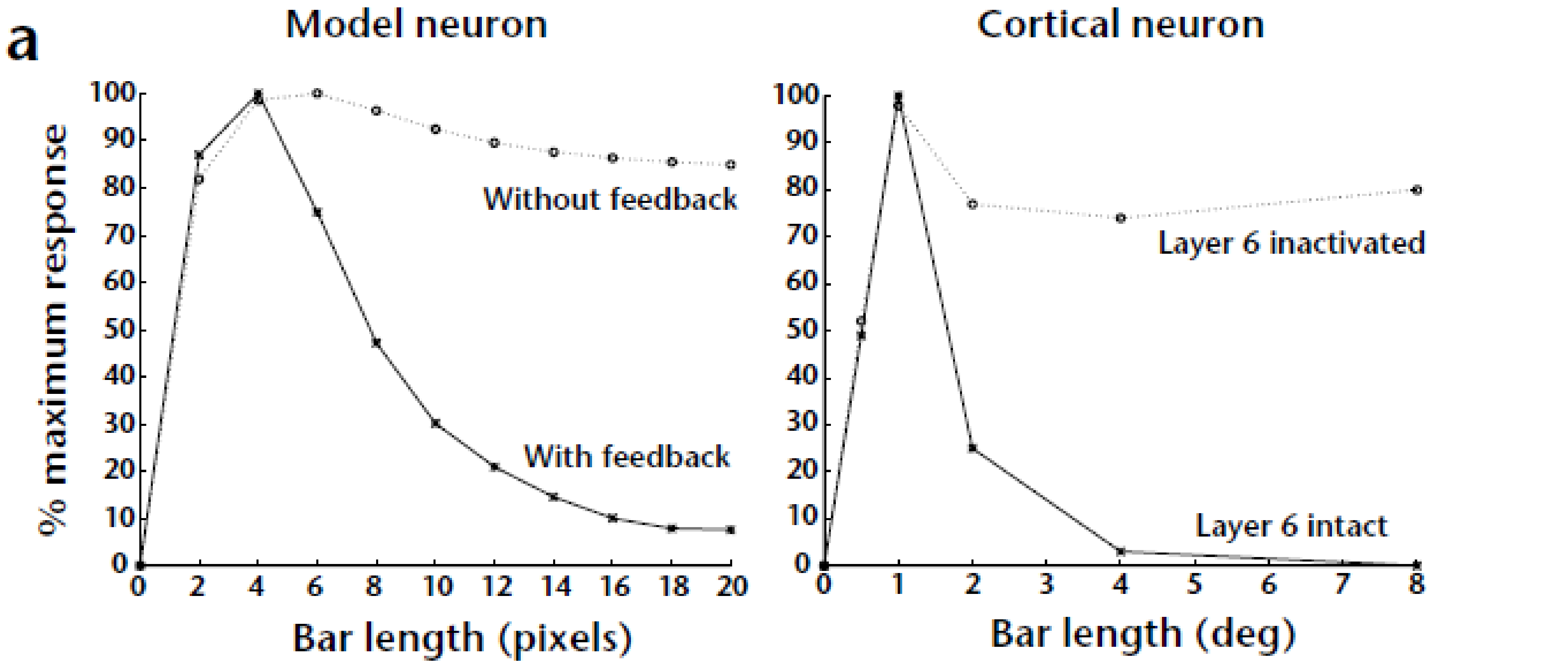
Rao and Ballard (1999)

Predictive Coding: Nonclassical Receptive Fields



Bars which are smaller than a neuron's receptive field generate a prediction error
=> Prediction error neurons should exist that signal this in their firing rate.

Predictive Coding: Nonclassical Receptive Fields



Such neurons are found in V1. Silencing top-down input reveals impact of predictive feedback from higher visual areas. Same effect occurs in model.

Rao and Ballard (1999)

Summary: Predictive Coding

- Predictive coding is a hierarchical generative model for natural images
- Top-down predictions are compared to bottom-up prediction errors
- Reproduces classical V1 receptive fields and non-classical responses related to prediction errors
- Provides a computational account of feedback from higher to lower visual areas
- More recent variants can account for many more properties of visual system (Lotter and Cox, 2020)

Summary: Generative Models

- Generative models attempt to explain data via their underlying causes
- Typically unsupervised: can learn representations from data
- Sensory systems are hypothesised to learn a generative model for the statistics of sensory input, and perform inference to find latent causes
- Sparse coding, ICA, and predictive coding learn V1-like receptive fields
- Convolutional neural networks are not generative models, but can replicate aspects of the visual system
- Generative models address representation but not the goals of sensory systems – sensory systems evolved to guide behaviour, and do not passively represent the world

Bibliography

- Hyvarinen, Hurri and Hoyer (Natural Image Statistics)
- Dayan and Abbott Ch 10
- Olshausen and Field (1996, 1997, 2004)
- Rao and Ballard (1999) (and Mumford 1992)
- Bell and Sejnowski (1995) (and Linsker 1988)