Perception: Inference, Priors and Codes

Maneesh Sahani

Gatsby Computational Neuroscience Unit, UCL

What is Perception *for*

What is Perception *for*

- Control? (After all, the only point of having a brain is to move. . .)

What is Perception *for*

- Control? (After all, the only point of having a brain is to move. . .)
- Forecasting and planning?

What is Perception *for*

- Control? (After all, the only point of having a brain is to move. . .)
- Forecasting and planning?
- Finding prey, mates, forage . . .

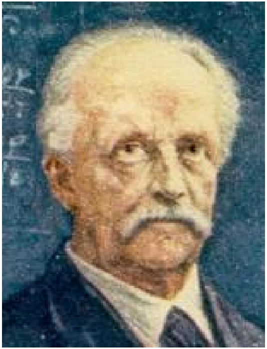
What is Perception *for*

- Control? (After all, the only point of having a brain is to move. . .)
- Forecasting and planning?
- Finding prey, mates, forage . . .

Presumably all of the above, but there is useful intermediate abstraction.

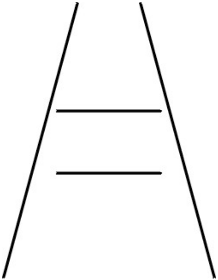
- *work out what's "out there".*

Helmholtz



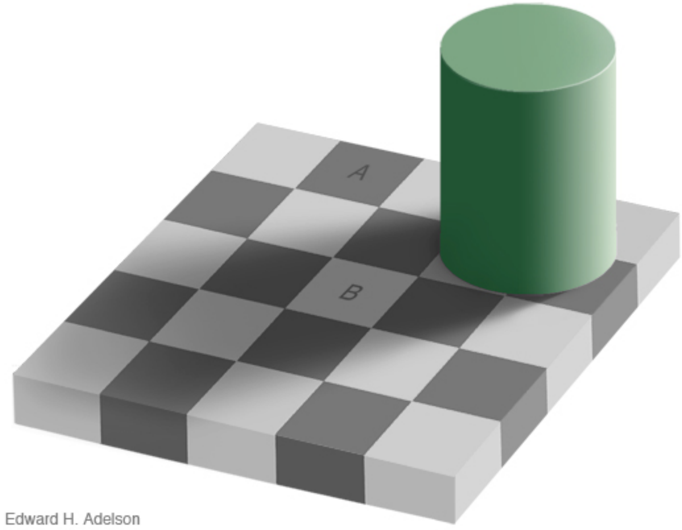
What information, then, can the qualities of such sensations give us about the characteristics of the external causes and influences which produce them? Only this: our sensations are **signs, not images***, of such characteristics.*

Illusions

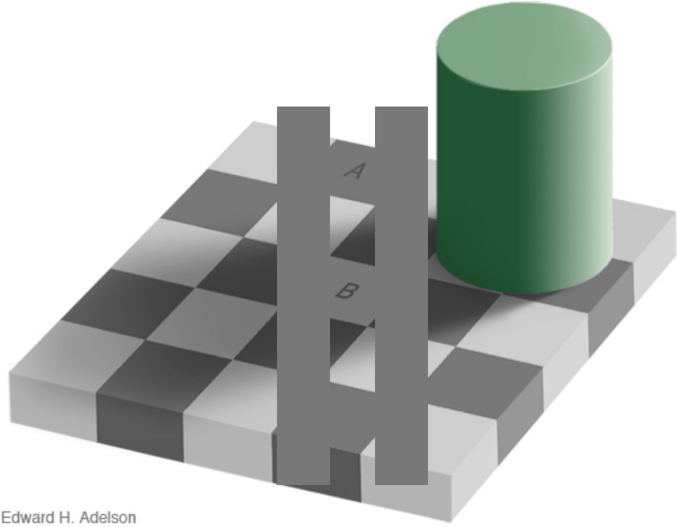


Gregory 1968

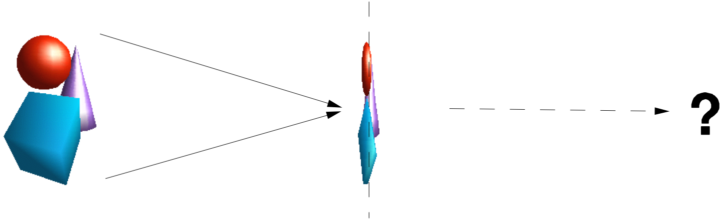
Illusions



Illusions

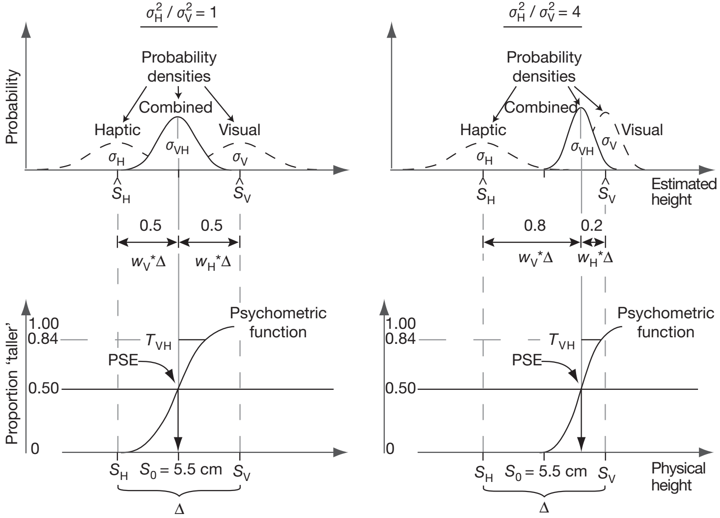


Perception and Generative Models



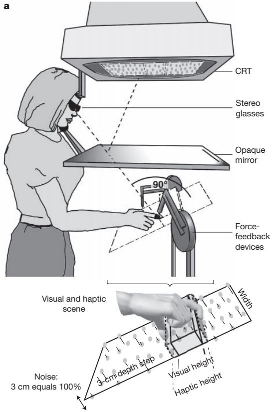
- Sensor activations reflect the state of the world through a (usually non-invertible and noisy) physical transformation.
- The goal of perception is to invert this transformation as best as possible: to infer the state of the world from the sensor signals.
- To do this, we need to know something about the forward (generative) process: both the transformation and the statistics of the world
- . . . and to use every available source of information.

Cue combination



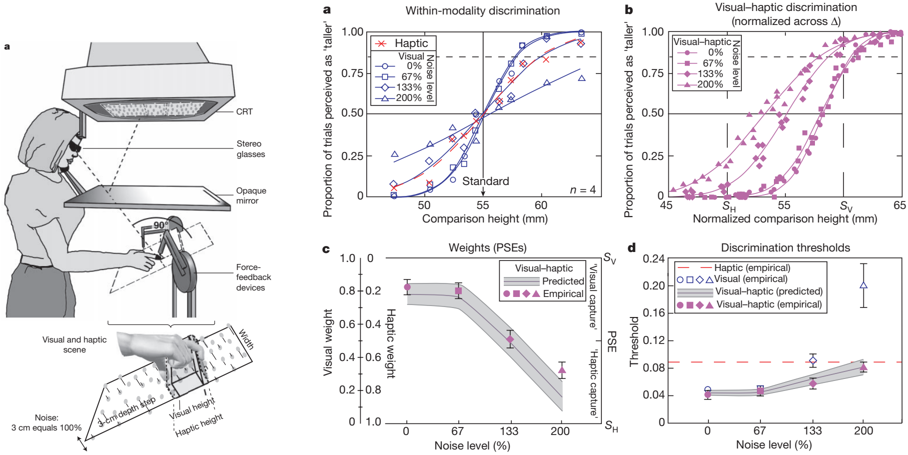
Ernst & Banks 2002

Cue combination



Ernst & Banks 2002

Cue combination

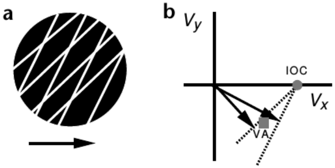


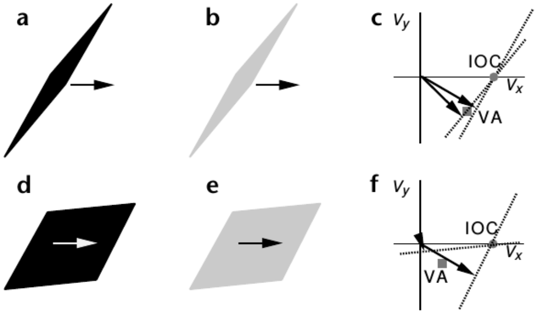
Ernst & Banks 2002

Incorporating priors – long-term priors

https://www.cs.huji.ac.il/~yweiss/Rhombus/rhombus.html

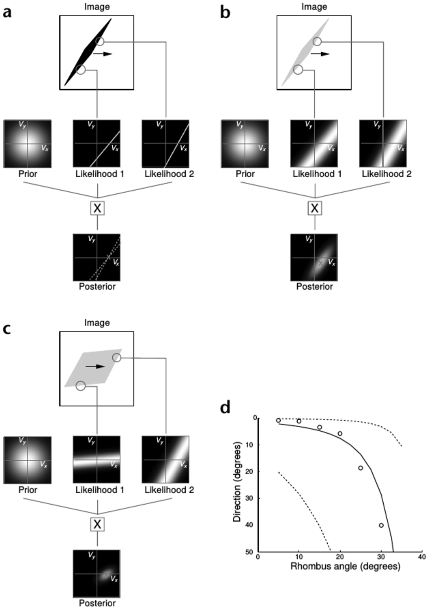
No simple rule





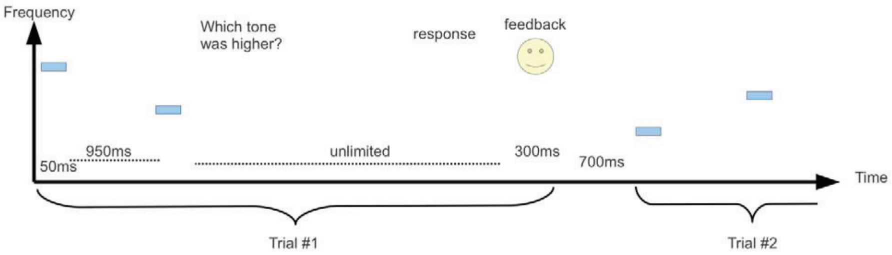
Weiss, Simoncelli, Adelson, 2002

Bayesian inference under a 'slow' prior



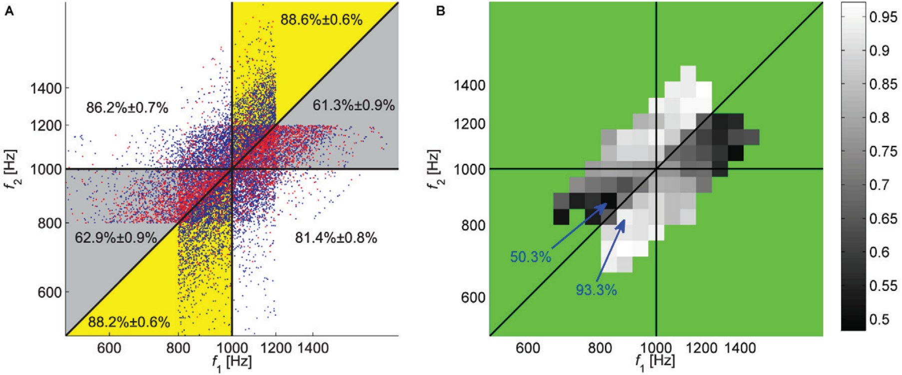
Weiss, Simoncelli, Adelson, 2002

Incorporating priors – short-term adaptation



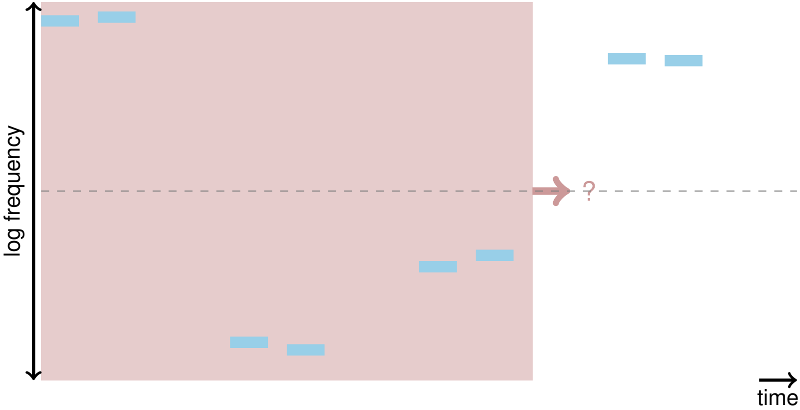
Raviv, Ahissar, Loewenstein, 2012

Frequency discrimination – contraction bias



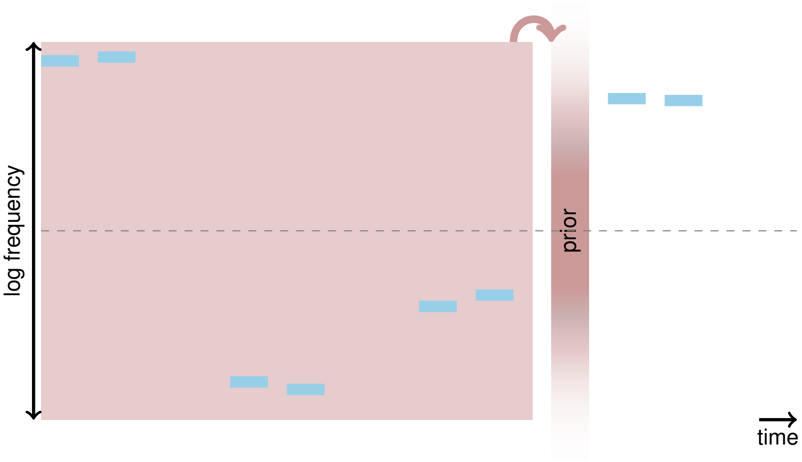
Raviv, Ahissar, Loewenstein, 2012

Prior context



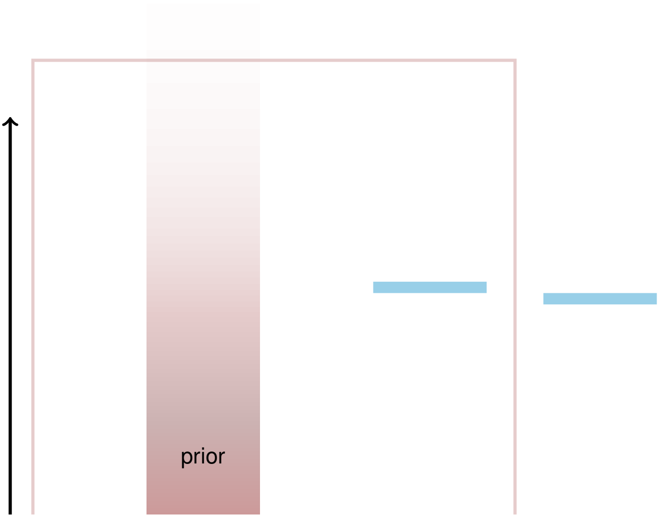
Ashourian, Loewenstein (2011)

Prior context



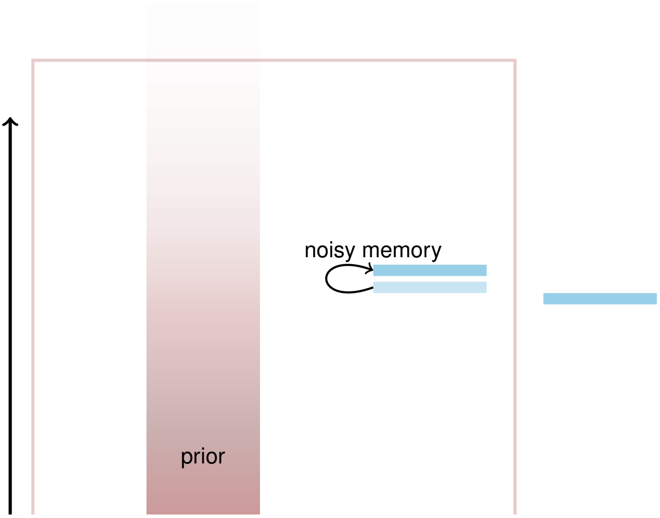
Ashourian, Loewenstein (2011)

Memory



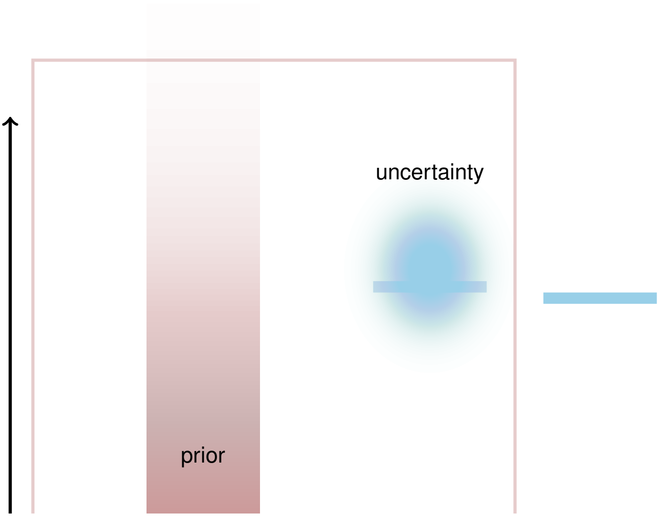
Ashourian, Loewenstein (2011)

Memory



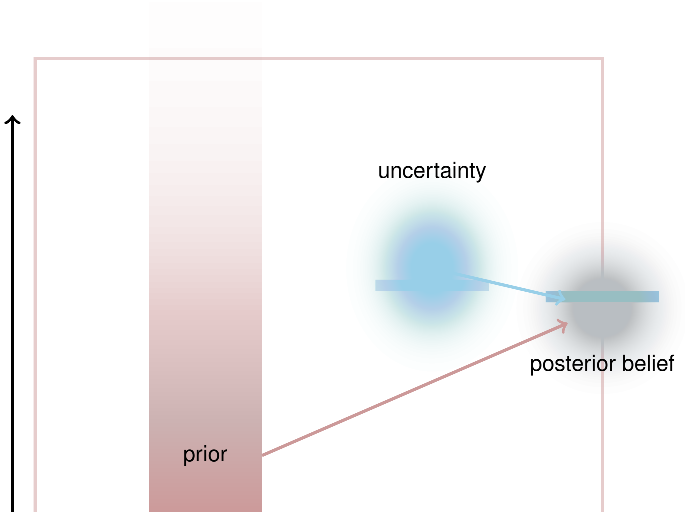
Ashourian, Loewenstein (2011)

Memory



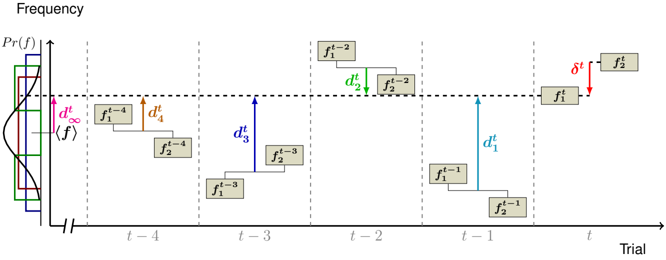
Ashourian, Loewenstein (2011)

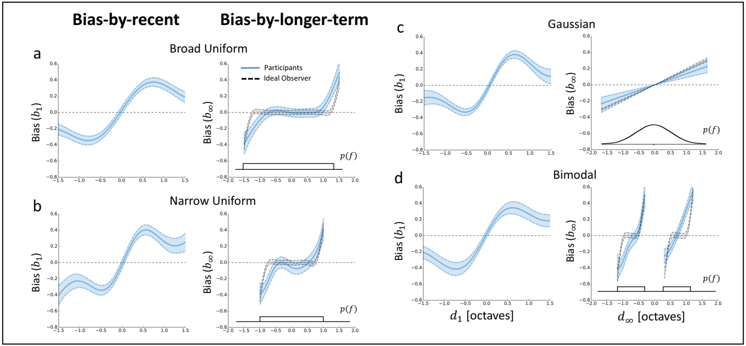
Memory



Ashourian, Loewenstein (2011)

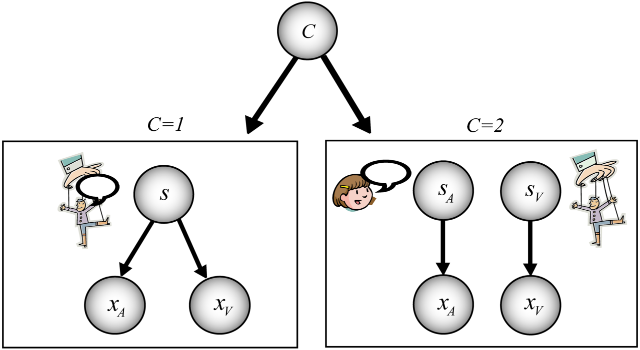
Subjects acquire varied priors





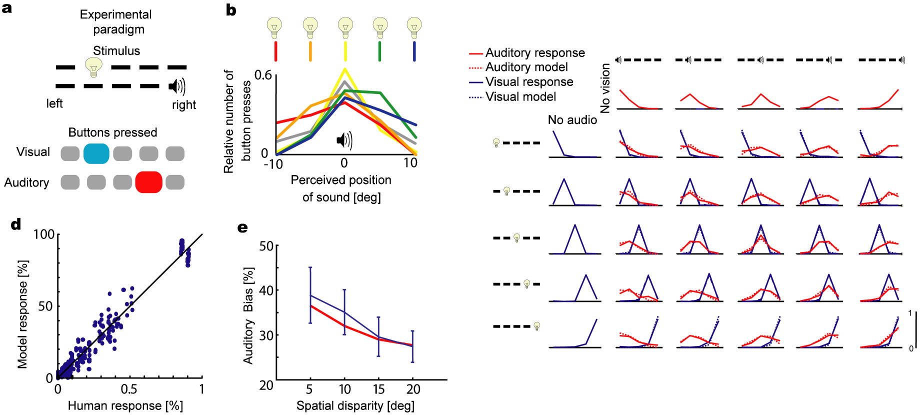
Lieder, Adam et al. (2019)

Structured inference



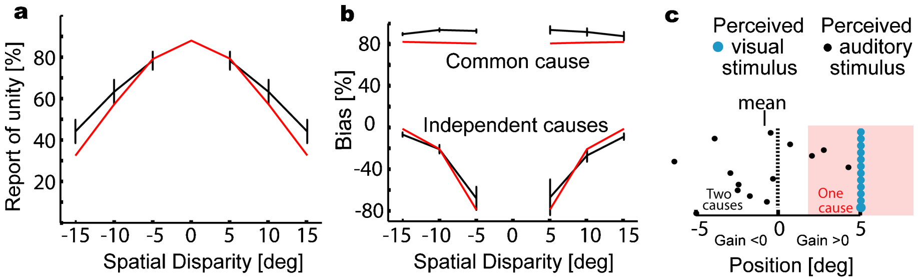
Kördig, Beierholm, et al. 2007

Structured inference



Körding, Beierholm, et al. 2007

Structured inference



Kördig, Beierholm, et al. 2007

Some neural consequences (in theory)

- *Sensory* systems (possibly needed for low-level control) should feed into *Perceptual* systems.
  - See Goodale & Milner on (visual) ventral and dorsal streams.

- Response properties and receptive fields in the perceptual pathway reflect properties of elements within an inferential system.
  - Inference is more than “recognition”: segmentation, grouping, reconstruction, prediction . . . We parse scenes and process and learn about objects we’ve never encountered.
  - Properties of neural systems should align with inferential processing in models of the real world.
  - Representations should have the capacity to represent and manipulate uncertainties, priors and other elements of inference.

Unsupervised Learning

- Sensory data do not come with 'supervision' – no direct information about $\mathbf{z}$.
- Prediction of and/or reinforcement from consequences of actions can help, but input is sparse relative to raw data.
- Most unsupervised learning is based (explicitly or implicitly) on fitting a "generative" world model to data. Power comes from conjunction of principled probabilistic structural priors and likelihood-based objectives for learning.
- Other viewpoints, such as redundancy reduction, are often special cases.

Physical vs. Generic Models

- If the physics is known and simple (or if evolution is lucky), it may be possible to invert the exact physical model. This will give the most accurate results.
    - Often difficult, particularly from an evolutionary standpoint.
    - Not flexible (e.g. if the statistics of the world change).
    - May be difficult to invert.
    - Neocortex appears to be generic.

Physical vs. Generic Models

- If the physics is known and simple (or if evolution is lucky), it may be possible to invert the exact physical model. This will give the most accurate results.
  - Often difficult, particularly from an evolutionary standpoint.
  - Not flexible (e.g. if the statistics of the world change).
  - May be difficult to invert.
  - Neocortex appears to be generic.
- Alternative: evolution has led to a generic generative model, with only some elements of physicality — but that can be successfully adapted by learning to be close enough to the generative process in the world.

Inference and Learning

Latent variable model:

$$\mathrm{P}_\theta(\mathbf{x}_i) = \int d\mathbf{z}\, \mathrm{P}_\theta(\mathbf{x}_i \mid \mathbf{z})\, \mathrm{P}_\theta(\mathbf{z})$$

Inference (find $\mathbf{z}_i$ given $\mathbf{x}_i$ and θ):

$$\mathrm{P}_\theta(\mathbf{z}_i \mid \mathbf{x}_i) = \frac{\mathrm{P}_\theta(\mathbf{x}_i \mid \mathbf{z}_i)\, \mathrm{P}_\theta(\mathbf{z}_i)}{\mathrm{P}_\theta(\mathbf{x}_i)}$$

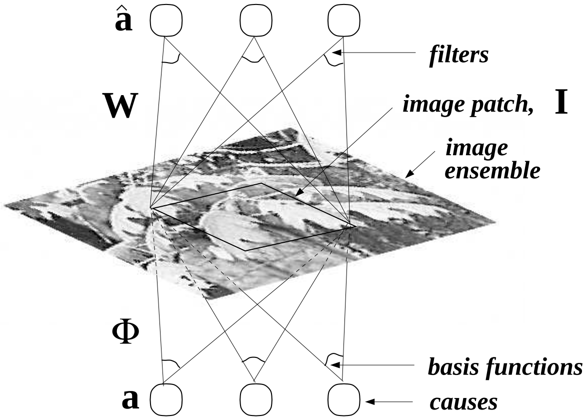
Learning (find θ given $\{\mathbf{x}\}$)

$$\mathrm{P}(\theta \mid \{\mathbf{x}\}) \propto \prod_i \mathrm{P}_\theta(\mathbf{x}_i)\, \mathrm{P}(\theta)$$

often by ML approximation

$$\theta^* = \operatorname*{argmax}_\theta \prod_i \mathrm{P}_\theta(\mathbf{x}_i)$$

Linear Image Codes

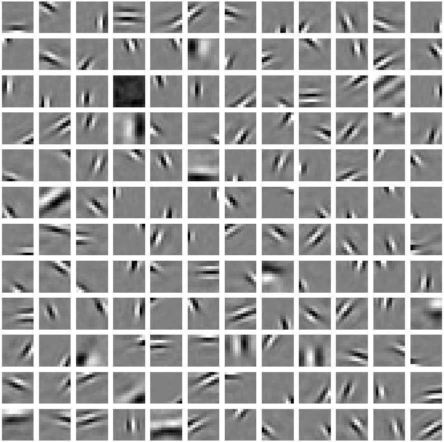


adapted from Bell and Sejnowski (1997)

Sparse Coding

$$E = \min_{\{a_i\}} \underbrace{\sum_{x,y} \left[I(x,y) - \sum_i a_i \phi_i(x,y) \right]^2}_{\log \mathrm{P}\,(\mathbf{I} \mid \mathbf{a})} + \lambda \underbrace{\sum_i S(a_i)}_{\log \mathrm{P}\,(\mathbf{a})}$$

$$S(a) = \log(1 + (a/\sigma)^2)$$

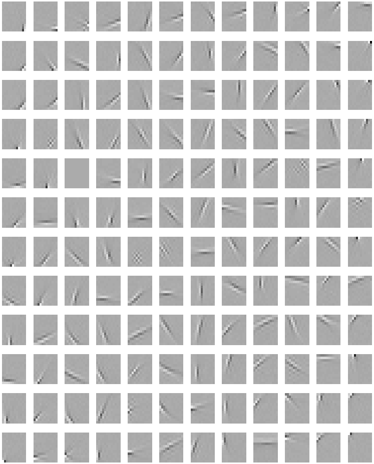


Olshausen & Field (1996)

Infomax

$$E = -H\left[g\left(\sum_{x,y} W_i(x,y)I(x,y)\right)\right]$$

$$g(a) = \frac{1}{1+e^{-a}}$$

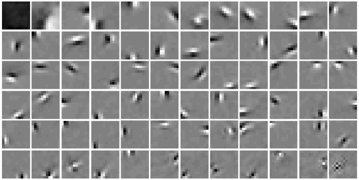


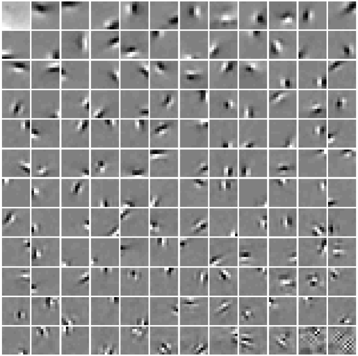
Bell & Sejnowski (1997)

Overcompleteness

$$E = -\int d\mathbf{a}\, \mathrm{P}_{\phi}\left(I \mid \mathbf{a}\right) \mathrm{P}_{S}\left(\mathbf{a}\right)$$

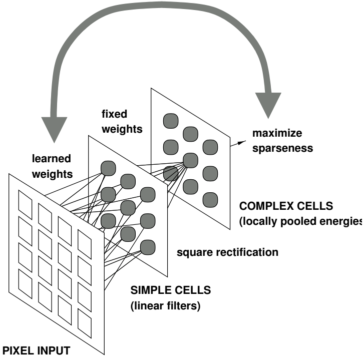
(Integral is approximated by saddle-point method.)

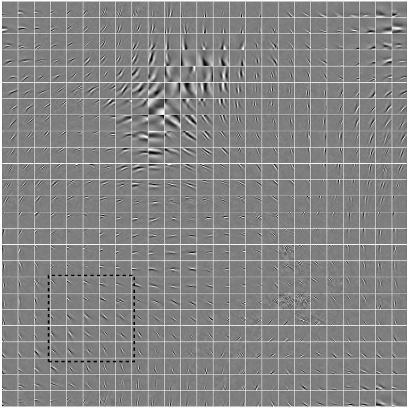




Lewicki & Sejnowski (2000); Lewicki & Olshausen (1999)

Topographic ICA - Hyvärinen & Hoyer

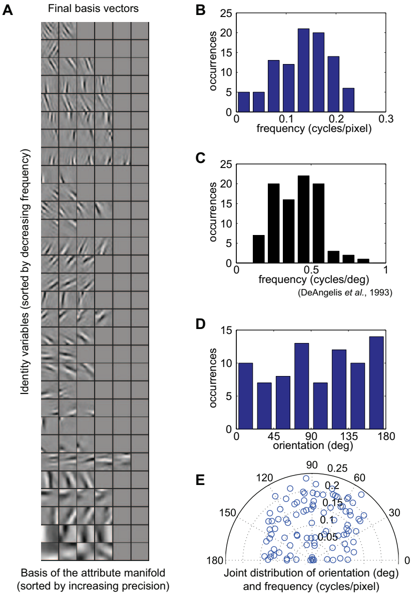




Hyvarinen & Hoyer (2001)

Dynamic constancy

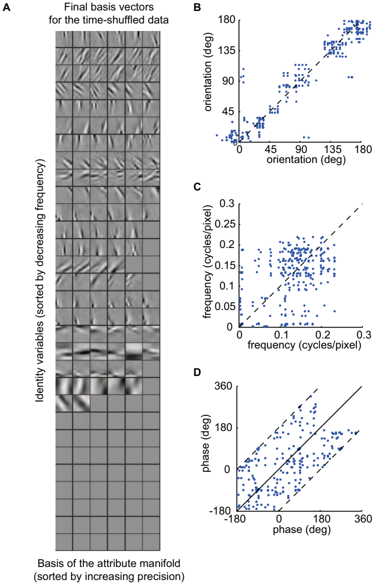
- Dynamic images and latent variables $I(x, y, t) \Rightarrow a_i(t)$.
- Impose prior limiting change in $a_i(t)$.
- With suitably constrained models, results in phase insensitivity (complex cells).



Wiskott & Sejnowski; Körding et al.; Berkes, Turner & Sahani

Dynamic constancy

- Dynamic images and latent variables $I(x, y, t) \Rightarrow a_i(t)$.
- Impose prior limiting change in $a_i(t)$.
- With suitably constrained models, results in phase insensitivity (complex cells).



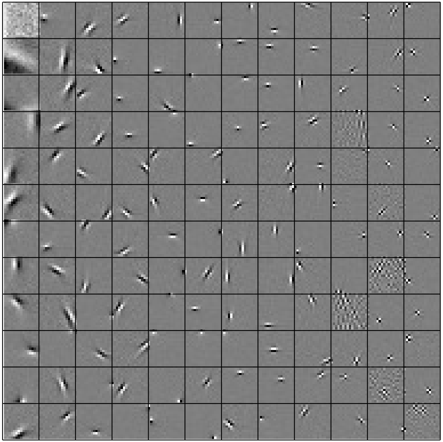
Wiskott & Sejnowski; Körding et al.; Berkes, Turner & Sahani

Recognition models

$$\mathrm{P}\left(I(x,y)\right) = \frac{e^{-E(\hat{\mathbf{a}})}}{\int d\mathbf{b}\ e^{-E(\mathbf{b})}}$$

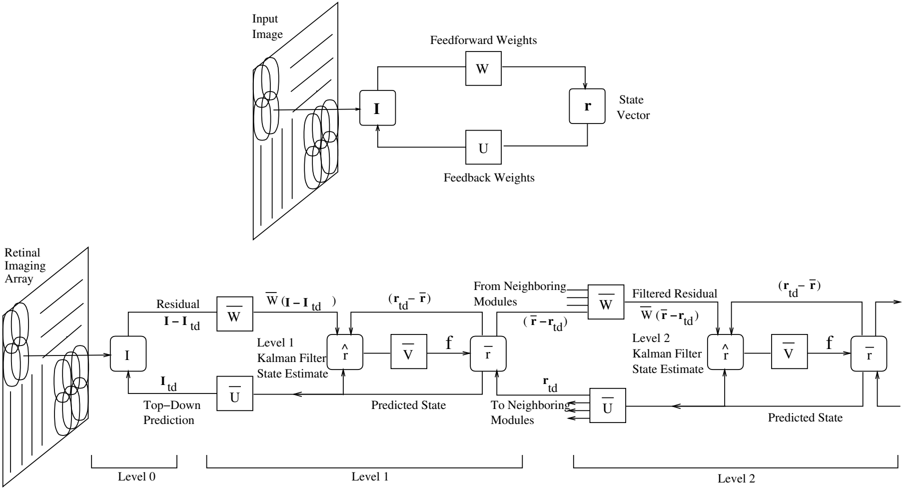
$$E(\hat{\mathbf{a}}) = -\sum_i \log \mathrm{P}_i\left(\hat{a}_i\right)$$

$$\hat{a}_i = \sum_{x,y} W(x,y) I(x,y)$$



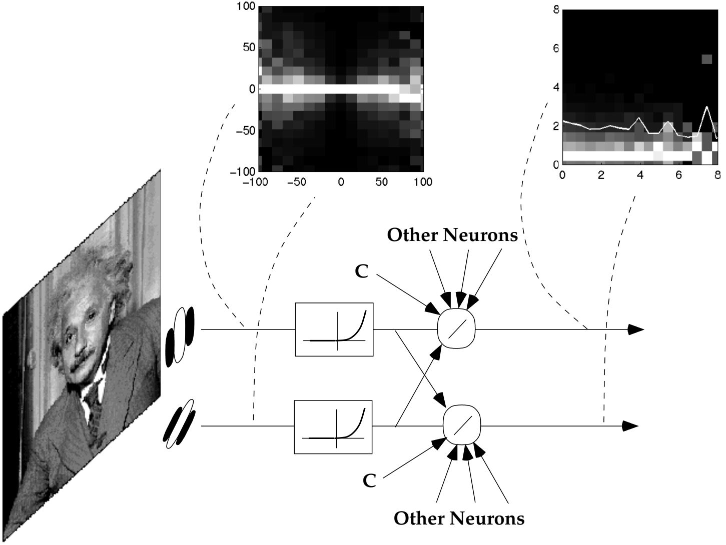
Hinton, Welling, Teh & Osindero (2002)

Feedback cancellation (or predictive coding)



Rao & Ballard (1997) (cf Friston)

Lateral normalization



Wainwright, Schwartz, & Simoncelli 2001

Uncertainty in population codes

- Models relate neural response properties to simple inference, but shouldn't be read literally:
  - Many neurons have very similar responses $\Rightarrow$ population representations of variables.
  - Correspondence usually based on single values, but behaviour (and intermediate stages of computation) seems to require distributional representations.

- Uncertainty is not easy to control, and experimental view on representations is far from settled. Theories can be broadly divided:

- Theory focuses on two broad types of representation:
  - Stochastic (sample-based) representations in time and/or space.
  - Deterministic representations of distributions
    - Linear decoding ("Neural Engineering Framework")
    - Log-linear decoding
    - 'Probabilistic encoding' / Inferential decoding (PPC)
    - Expected-value encodings (DDC)

- Crucial questions around computation, learnability and verifiability.

Probabilistic computation and message passing

- In complex models, beliefs about different variables are interdependent.
- Separate representation of variables $\Rightarrow$ updating based on evolving beliefs and/or priors.
  - Message-passing (generally exact on tree-structured models, but approximate otherwise).
- Many ways to implement message-passing schemes:
  - belief propagation
  - variational messages
  - reparametrisation
  - predictive coding
  - ...
- Hints of many in data, but details are challenging ...