Cognition and Brain Sciences Unit

Neuronal populations, Bayesian inference & learning

Jiaxiang Zhang and Andrea Greve

Bayesian Theory Interest Group 04/03/2014

Questions

MRC | Cognition and Brain Sciences Unit

- 17. How can neuronal populations perform Bayesian inference?
- 18. How can neuronal populations perform Bayesian learning?

- 19. What is the neuronal evidence for representations of uncertainty?
- 20. What is the neuronal evidence for Bayesian inference and learning?

Why do we care?

MRC | Cognition and Brain Sciences Unit

- If neuron(s) can encode probabilities
  - Neural computation ≈ Probabilistic inference
- Human performance sometimes close to Bayes-optimal
  - Perception / motor control
  - Multisensory integration

- Therefore neuron(s) may code both stimulus value and its uncertainty (i.e., probability)

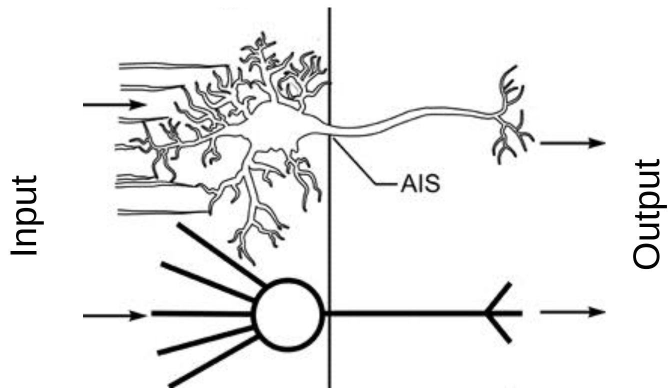
How ?

Neuronal properties

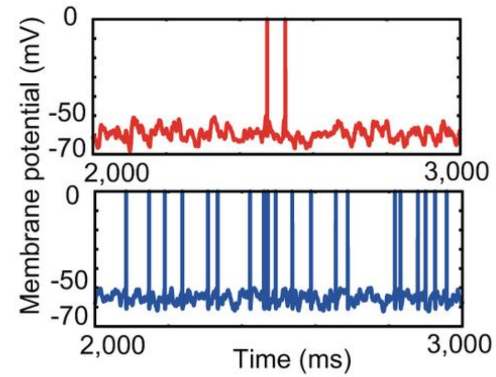
Neurons: Integrate Fire

Biological

Artificial



Spike rate

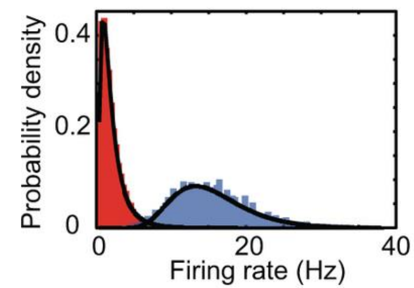


low

high

Integrate-and fire neurons: multiple dendrites and one cell body (soma) receive and integrate synaptic inputs as membrane potentials which are compared to a threshold at the axon initiation segment. If threshold is met, axonal spikes/firings are triggered along a single axon which branches distally to convey outputs.

Firing rate distribution

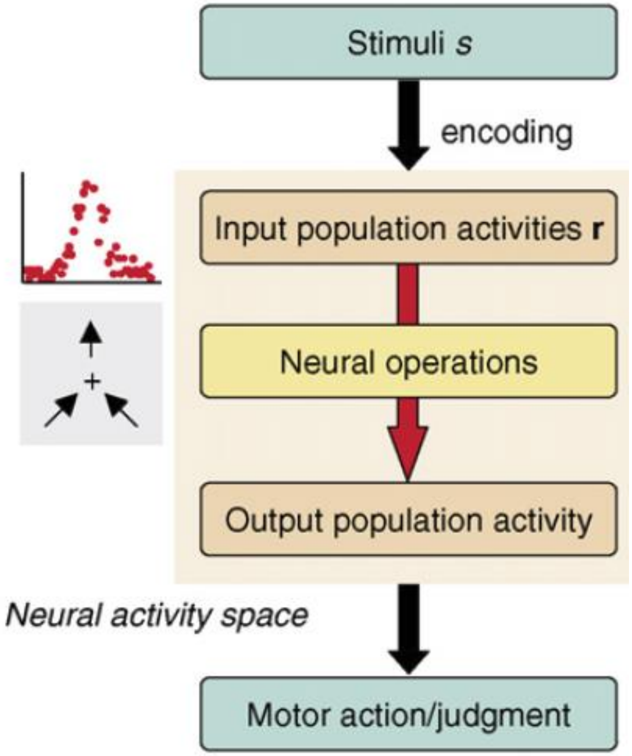


can be fitted by (lognormal) distribution

Mapping neuronal properties

MRC | Cognition and Brain Sciences Unit

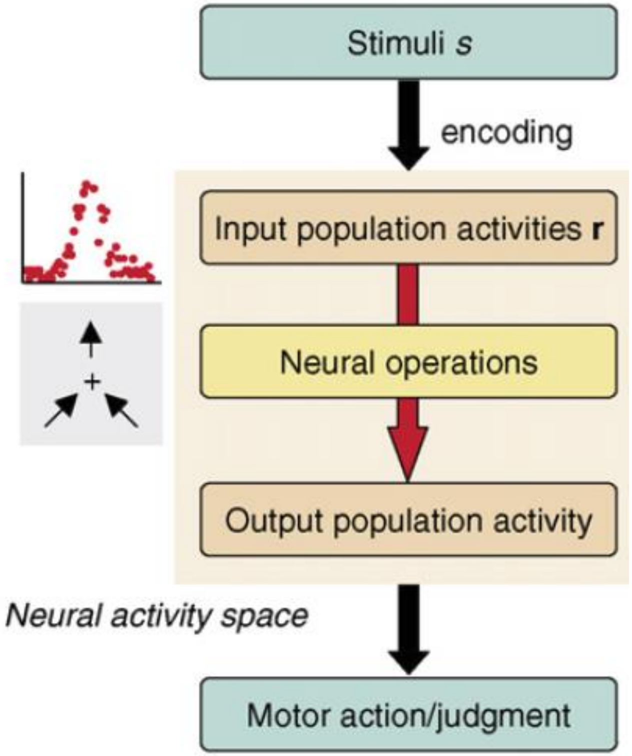
Neural activity space:



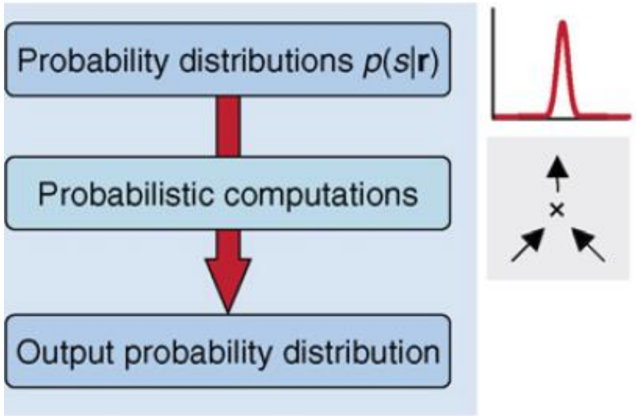
Mapping neuronal properties

MRC | Cognition and Brain Sciences Unit

Neural activity space:



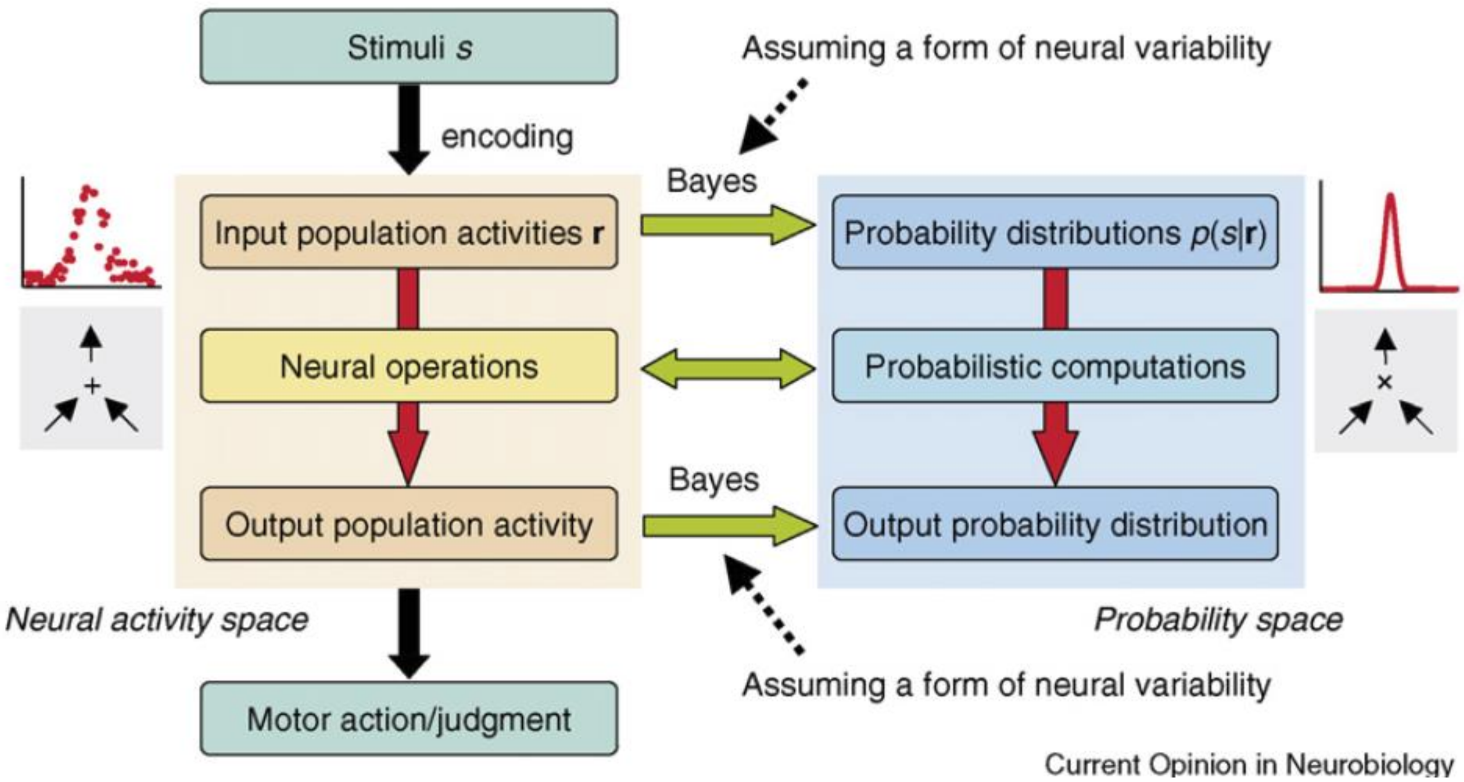
Probability space:



Mapping neuronal properties

MRC | Cognition and Brain Sciences Unit

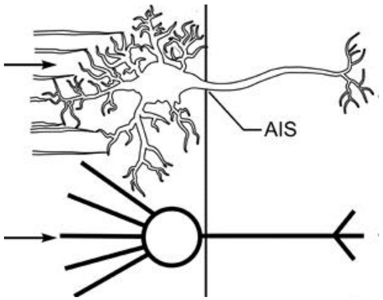
Neural activity space: Probability space:



(Ma et al. 2008)

Mapping neuronal properties

Integrate – Fire Neuron

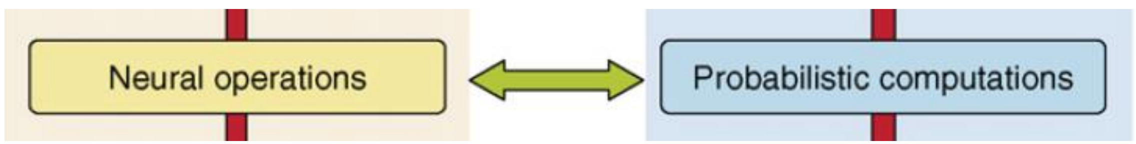


Bayes Theorem

$$P(s|\mathbf{r}) \propto P(\mathbf{r}|s)P(s)$$

?

Summation

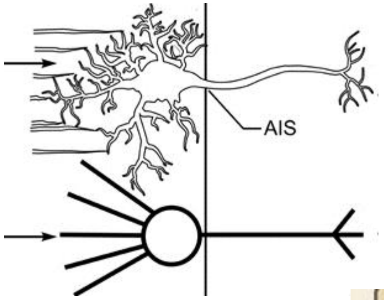


Multiplication

Mapping neuronal properties

MRC Cognition and Brain Sciences Unit

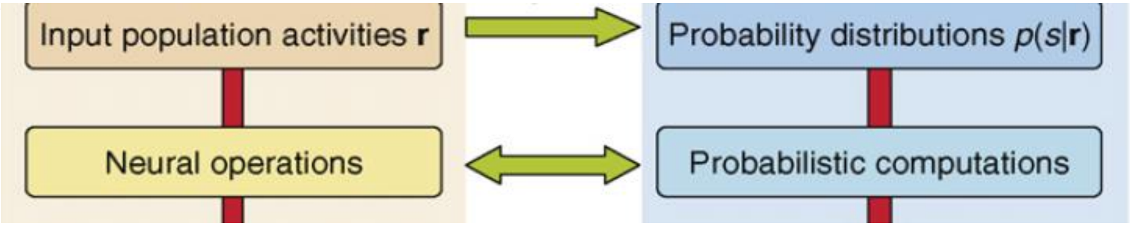
Integrate – Fire Neuron



Bayes Theorem

$$P(s|\mathbf{r}) \propto P(\mathbf{r}|s)P(s)$$

Need population code for probability distribution

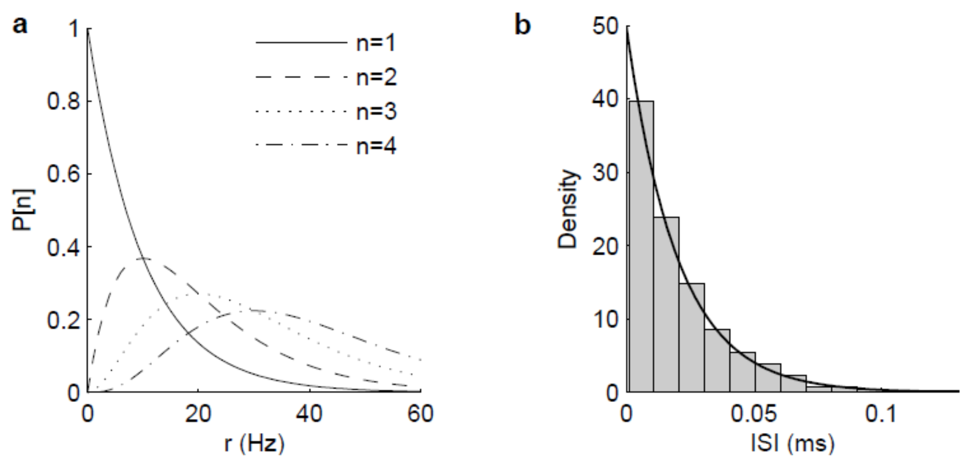


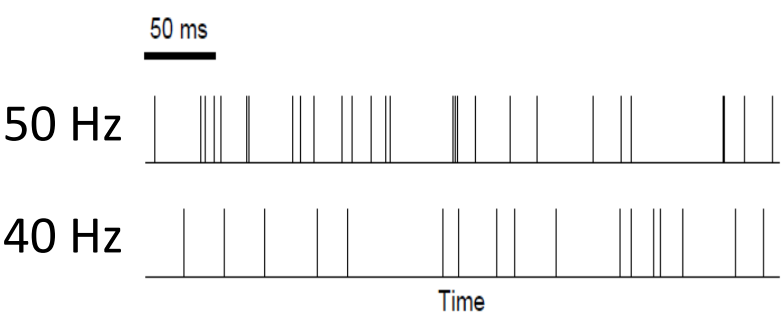
Assumption of neuronal variability

- The spike trains of an individual sensory neuron are from a Poisson process

$$P[n \text{ spikes during } \Delta t] = \frac{e^{-r\Delta t}\,(r\Delta t)^n}{n!}$$

Mean firing rate = *r*



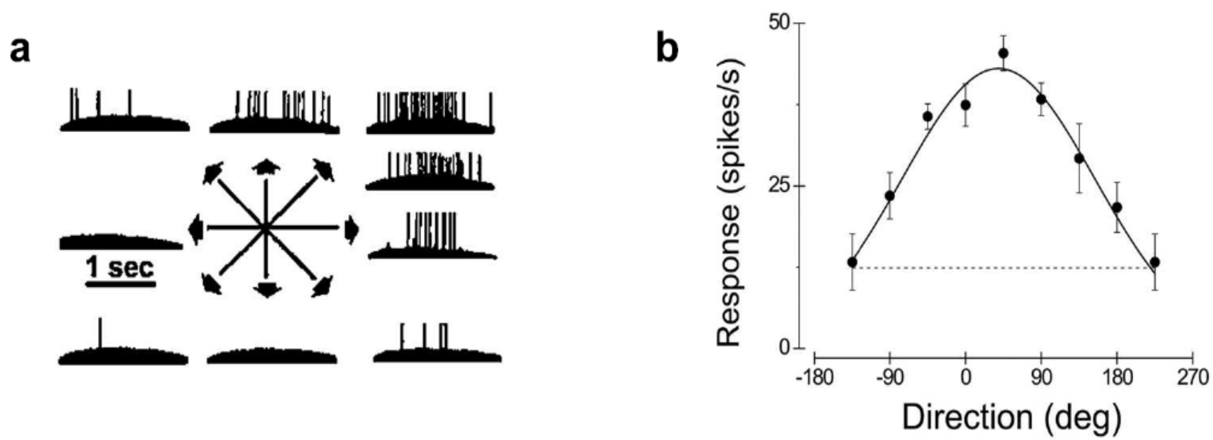


Stimulus-selective firing (tuning)

- Sensory neurons have selective activities to preferred stimulus features (direction/orientation/frequency …)

$$r = f(s)$$

Observed firing rate (r) · Tuning function (f) · Stimulus feature (s)

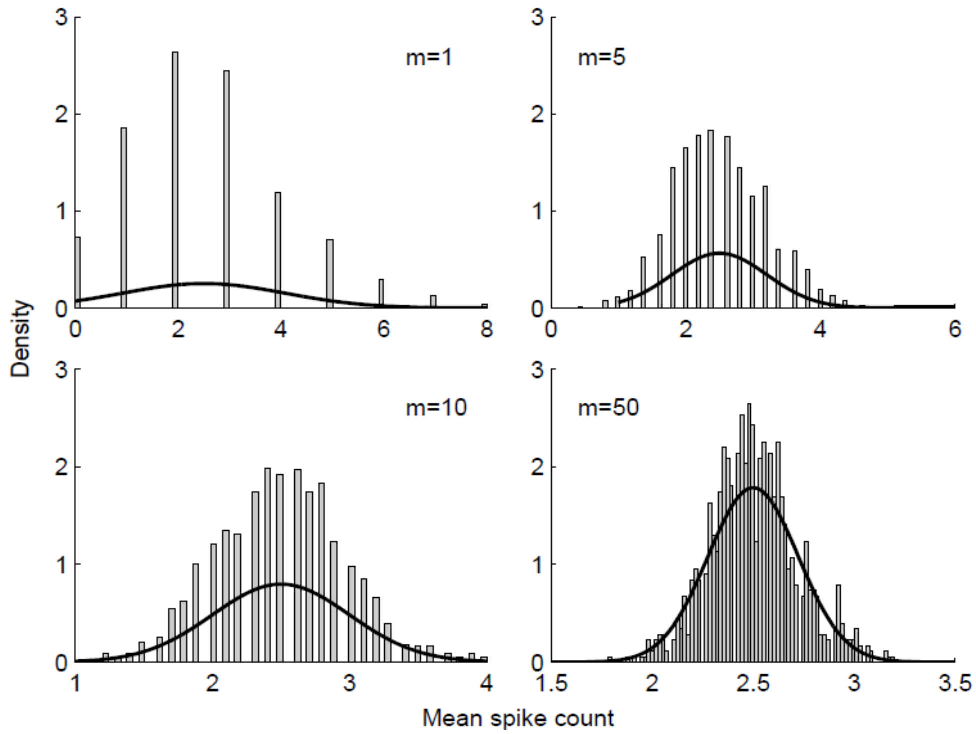


MT neuron in Macaque

DeAngelis and Uka (2003).

Rate coding from population

- Mean firing rate can be reliably encoded in a group of homogenous neurons (central limit theorem)



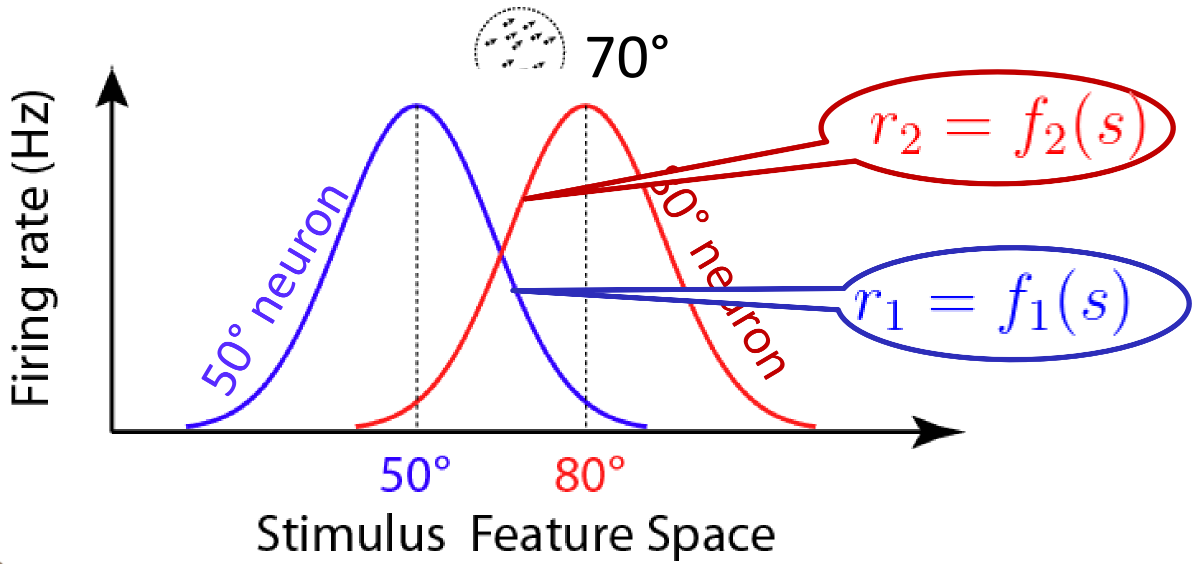
r=50 Hz ~
2.5 spike per 50 ms

Spike count in 50 ms (averaged by population size)

Population with turning curves

MRC | Cognition and Brain Sciences Unit

- Neural activity space (i.e., physical space)
  - Given stimulus feature s
  - N independent Poisson neurons $[1, 2, 3, \ldots, N]$
  - with observed firing rate $\mathbf{r} = [r_1, r_2, \ldots, r_N]$
  - from turning function $\mathbf{F}(s) = [f_1(s), f_2(s), \ldots, f_N(s)]$



Population with turning curves

MRC | Cognition and Brain Sciences Unit

- (Moving to probability space)
- For neuron i, the probability of observing r_i (given s) follows a Poisson distribution with mean (and var) $f_i(s)$.

$$P(r_i|s) = \frac{e^{-f_i(s)} f_i(s)^{r_i}}{r_i!}$$

- Then the probability of observing $\mathbf{r}$ (given s) is:

$$P(\mathbf{r}|s) = \prod_{i=1,2,3,\ldots,N} P(r_i|s) = \prod_{i=1,2,3,\ldots N} \frac{e^{-f_i(s)} f_i(s)^{r_i}}{r_i!}$$

Bayes from neural population

MRC | Cognition and Brain Sciences Unit

$$P(s|\mathbf{r}) \propto P(\mathbf{r}|s)P(s) \propto \prod_i \frac{e^{-f_i(s)} f_i(s)^{r_i}}{r_i!}$$

Probability space

$$= e^{-\sum_i f_i(s)} \prod_i \frac{f_i(s)^{r_i}}{r_i!}$$

Neural space

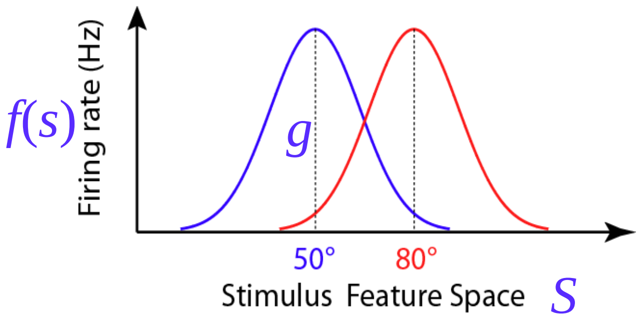
$$\propto \prod_i f_i(s)^{r_i}$$

Further assumption

MRC | Cognition and Brain Sciences Unit

- If neurons have bell shape tuning function:

$$f_i(s) = g * e^{-\frac{1}{2\sigma^2}(s - \hat{s}_i)^2}$$



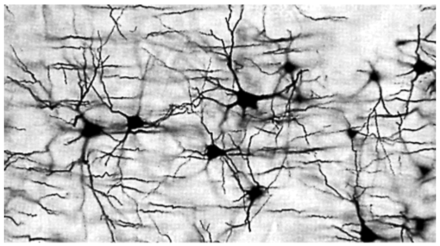
$$P(s|\mathbf{r}) = \mathrm{Norm}(s, \frac{1}{c * g})$$

- Therefore, the population activity **r** automatically (implicitly) encode posterior distribution, assuming the **Bayes** rule and a **certain** form of neuronal variability.

Bayes from neural population

MRC | Cognition and Brain Sciences Unit

Population of Neurons



Bayes Theorem

$$P(s|\mathbf{r}) \propto P(\mathbf{r}|s)P(s)$$

Probability space

Assumption:
Independent Poisson

$$\propto \prod_i \frac{e^{-f_i(s)} f_i(s)^{r_i}}{r_i!}$$

Input population activities **r** → Probability distributions *p*(*s*|**r**)

$$= e^{-\sum_i f_i(s)} \prod_i \frac{f_i(s)^{r_i}}{r_i!} \propto \prod_i f_i(s)^{r_i}$$

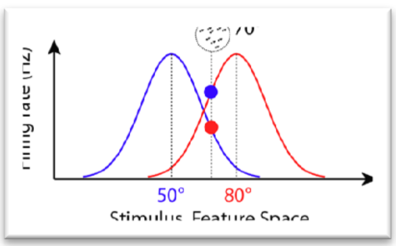
Neural space

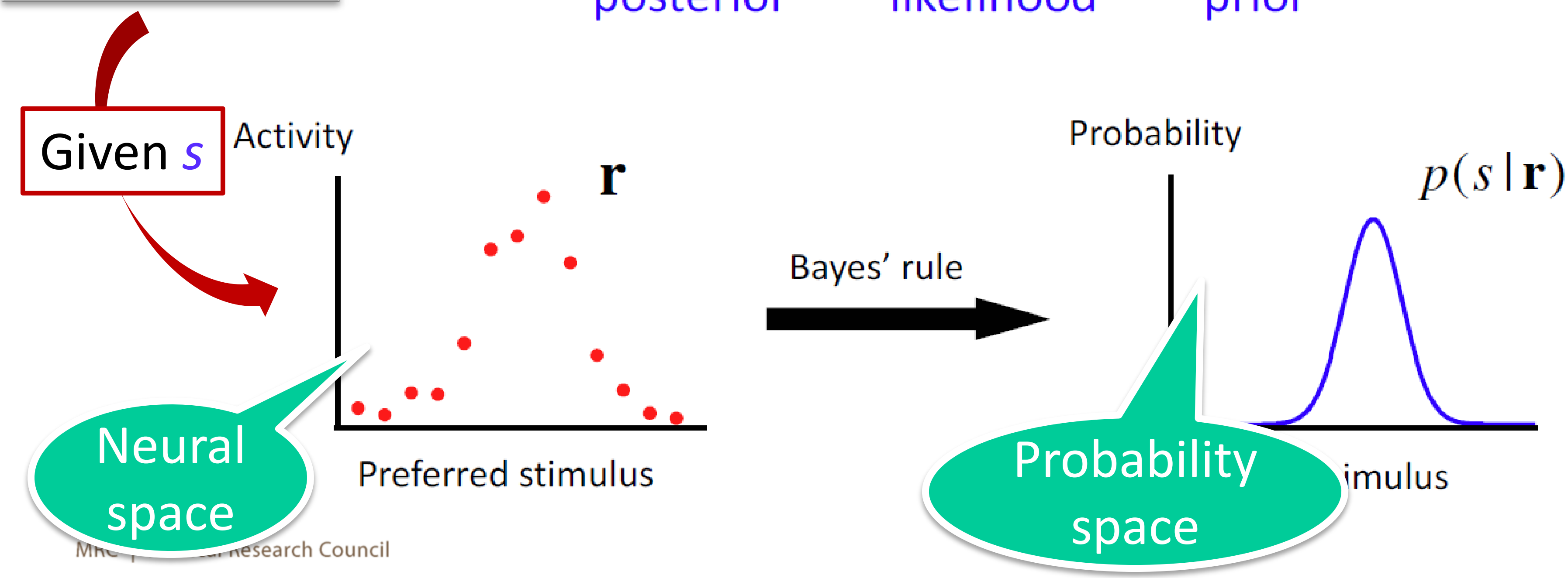
Bayes from neural population

- So Bayes inference can be implemented in neurons:

$$p(s \mid \mathbf{r}) \propto p(\mathbf{r} \mid s)\, p(s)$$

posterior　likelihood　prior



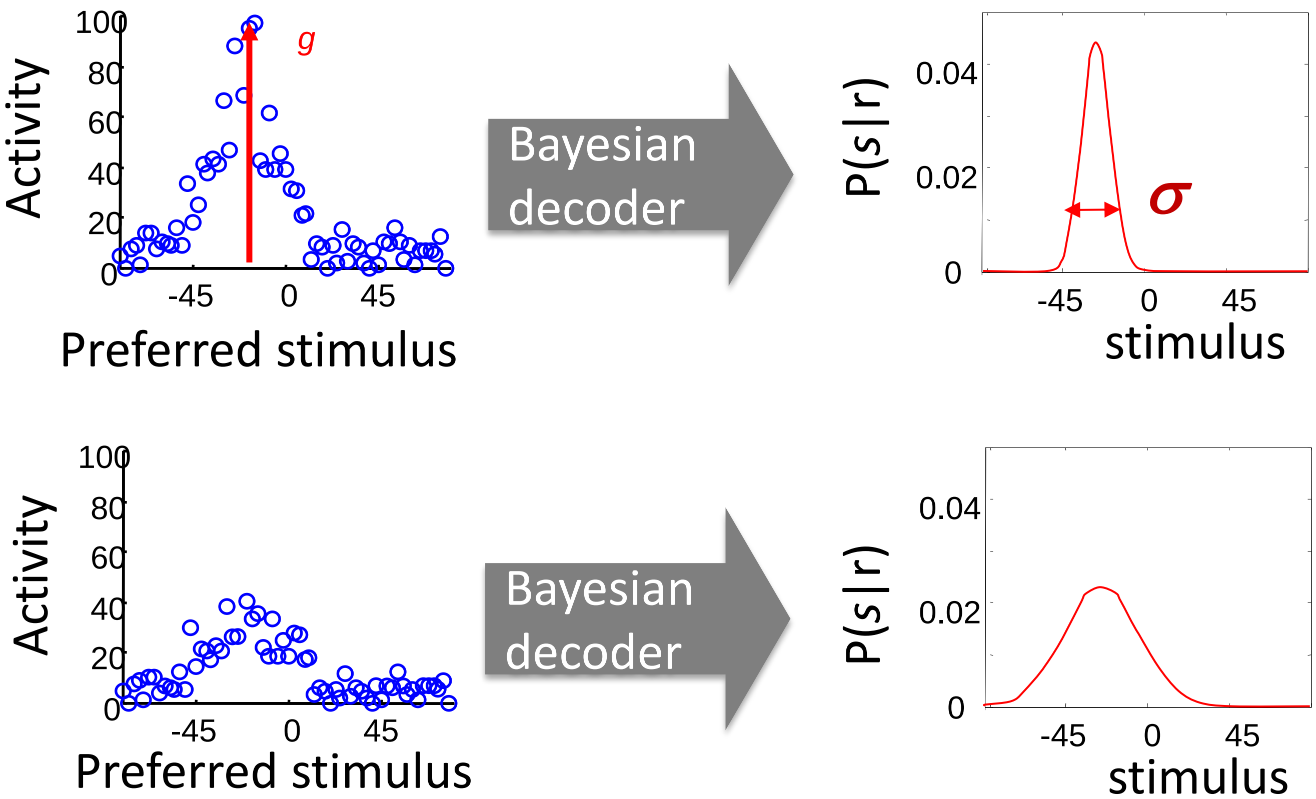


(Un)certainty

MRC | Cognition and Brain Sciences Unit

- The (un) certainty is encoded by g,

$$g \propto 1/\sigma^2$$



High gain,
high certainty

Low gain,
low certainty

All under the assumptions:

MRC | Cognition and Brain Sciences Unit

- Neuronal variability
  - independent Poisson processes
- Tuning function
  - Identically shaped (not necessarily)
  - Gaussian tuning function gives Gaussian posterior
- Those assumptions can take a weaker form

The more general form

MRC | Cognition and Brain Sciences Unit

- Passion variability and identical Gaussian tuning are the special forms of:

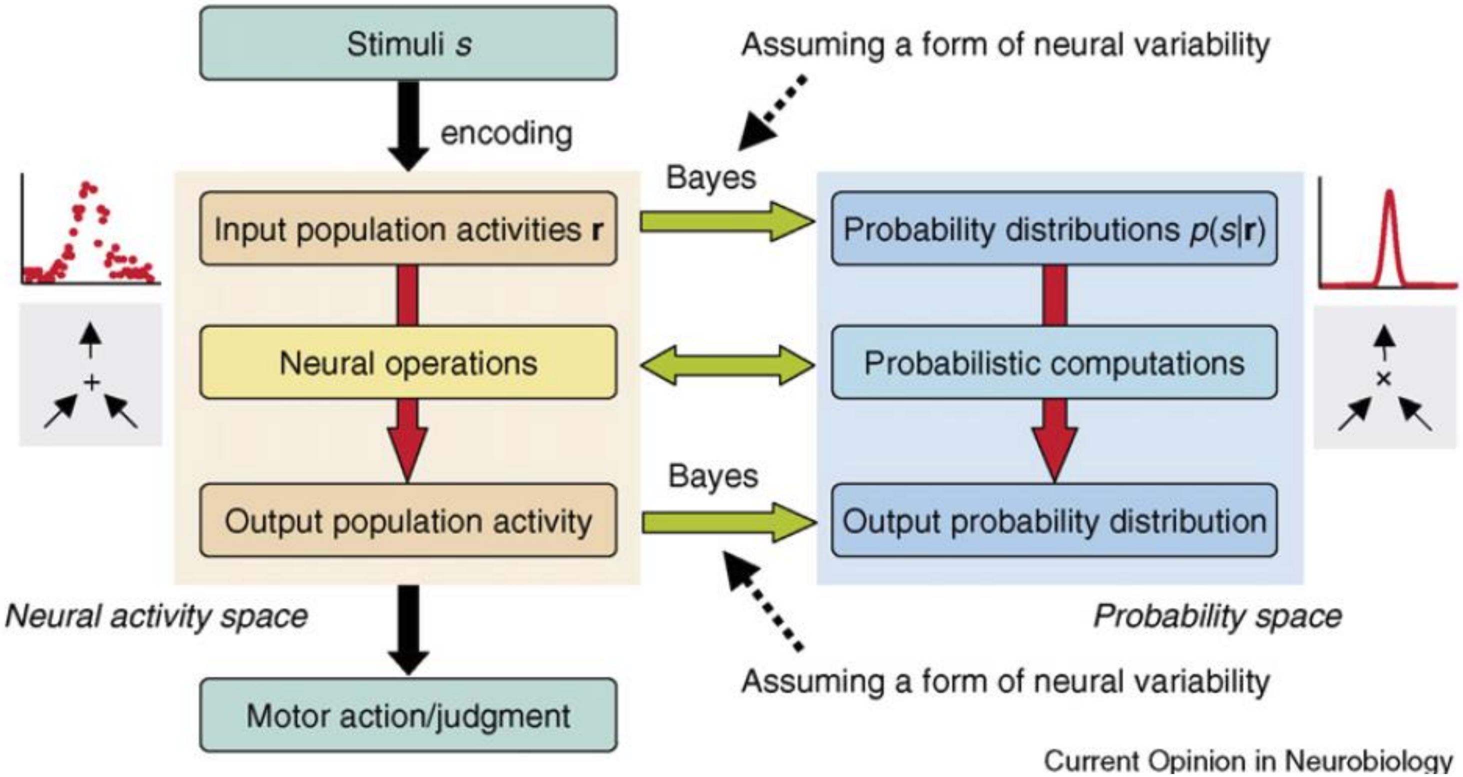
$$P(\mathbf{r}|s) = \phi(\mathbf{r})e^{\mathbf{h}^T(s)\cdot\mathbf{r}}$$

$$\frac{d\mathbf{h}(s)}{ds} = \Sigma_{\mathbf{r}}^{-1}(s)\frac{d\mathbf{f}(s)}{ds}$$

Summary

MRC | Cognition and Brain Sciences Unit

- So Bayes inference can be implemented in neurons:



Questions

MRC | Cognition and Brain Sciences Unit

17. How can neuronal populations perform Bayesian inference?

18. How can neuronal populations perform Bayesian learning?

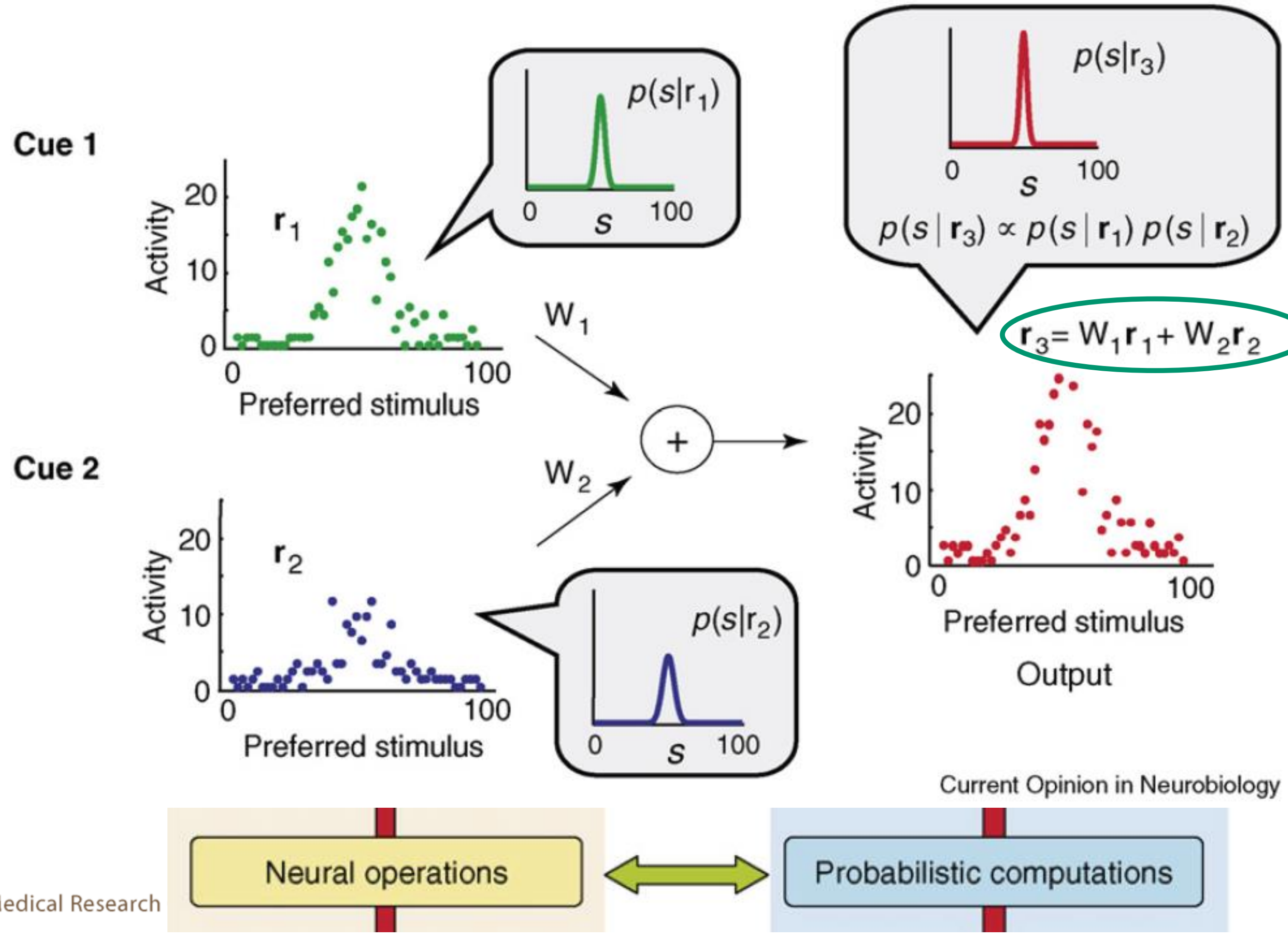
19. What is the neuronal evidence for representations of uncertainty?

20. What is the neuronal evidence for Bayesian inference and learning?

Can we get similar results with realistic networks (integrate and fire neurons)?

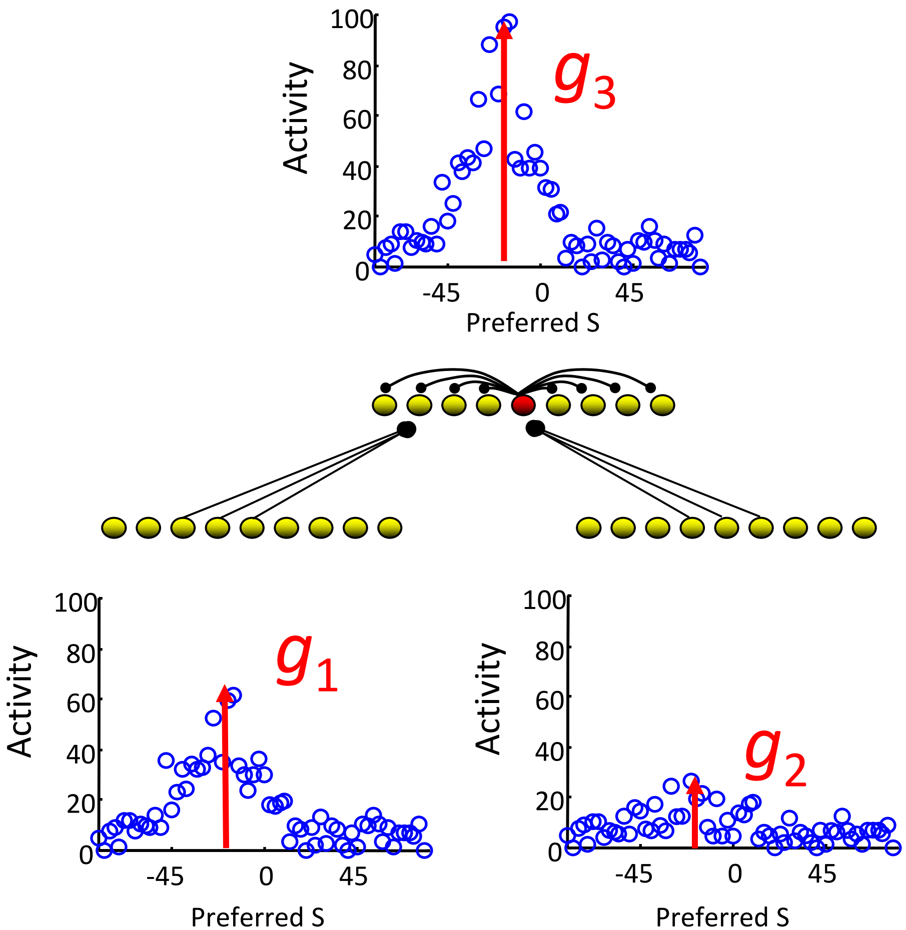
Cue integration

- Optimal cue integration is the linear sum in neural space



(Ma et al. 2008)

Integrate and fire neurons



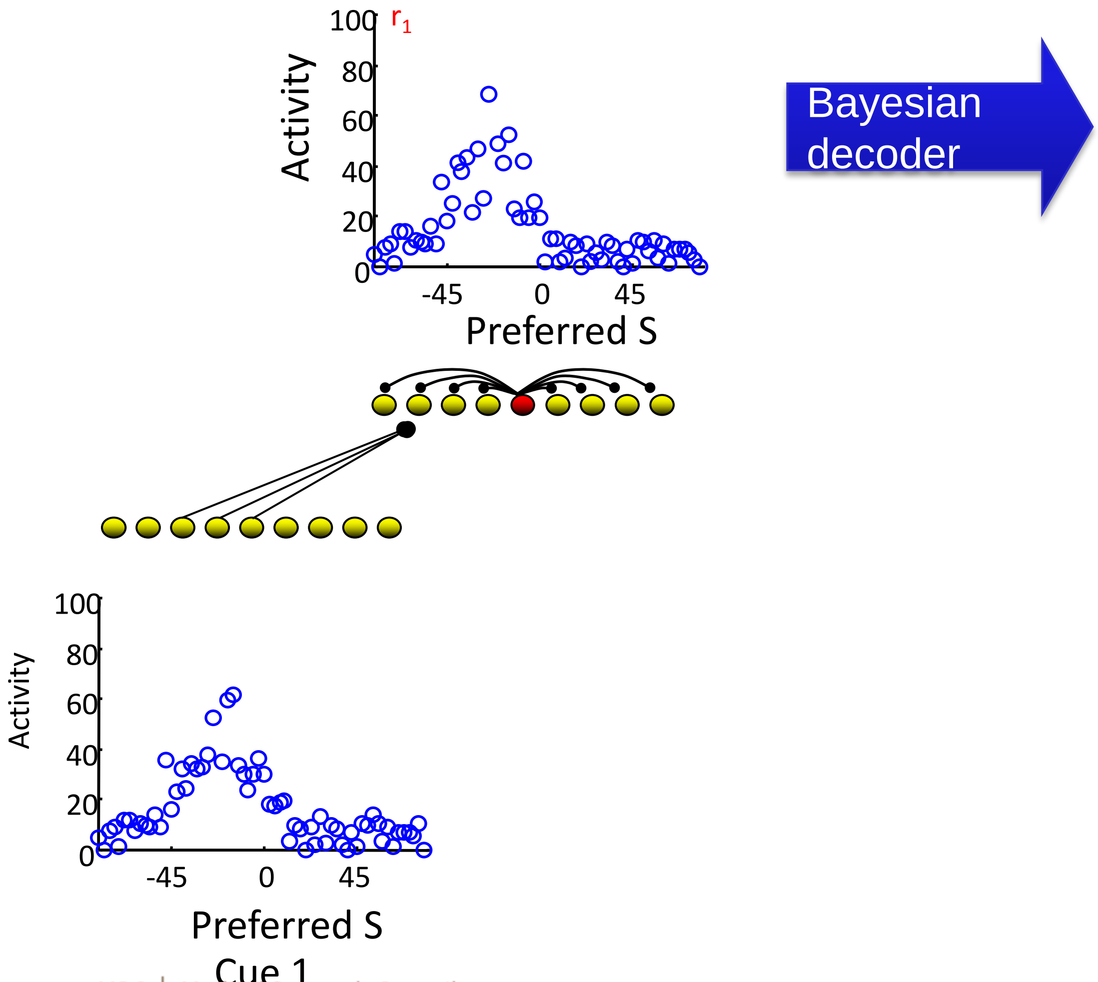
Cue 1 Cue 2

Output layer:

- 1200 conductance-based integrate-and-fire neurons, 1000 excitatory, 200 inhibitory
- Lateral connections
- Fano factors (0.3 to 1)
- Correlated activity

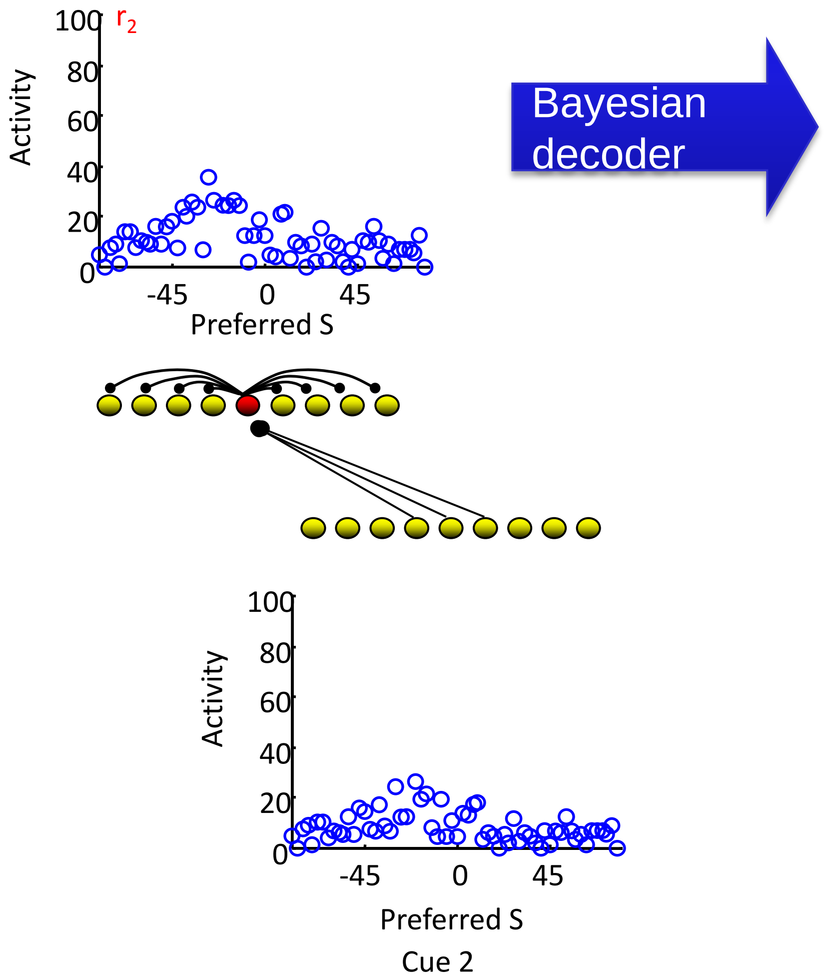
Input: near-Poisson correlated spike trains with different gains and slightly different means

Test cue 1 alone



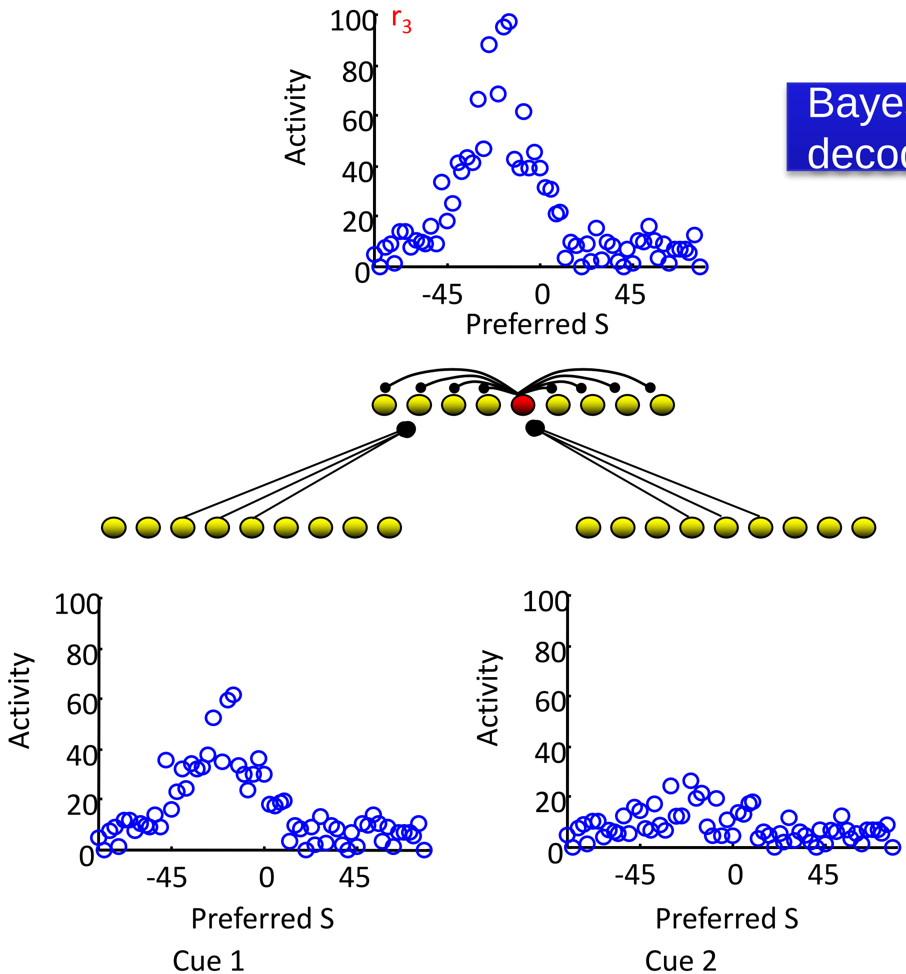
$$p(r_1 | s)$$

Test cue 2 alone



$$p(r_2|s)$$

Test cue1 and cue2 together

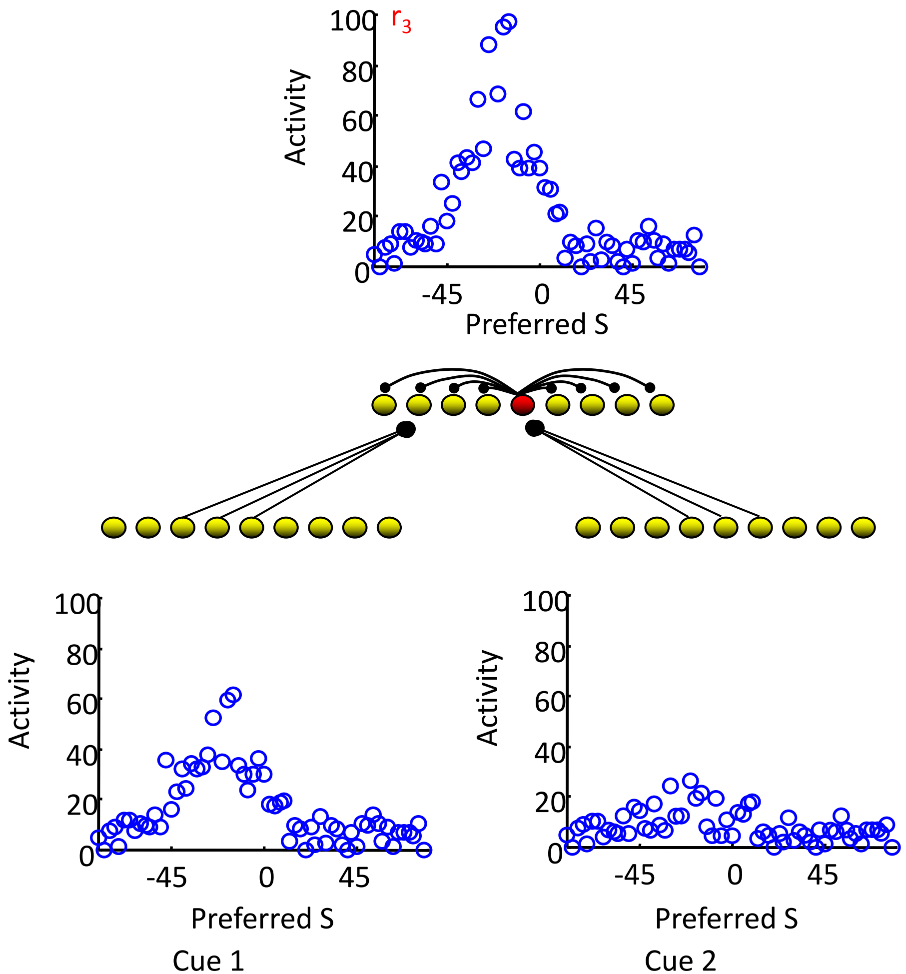


Bayesian decoder

$$p(r_3|s)$$

Compare the distributions

MRC Cognition and Brain Sciences Unit

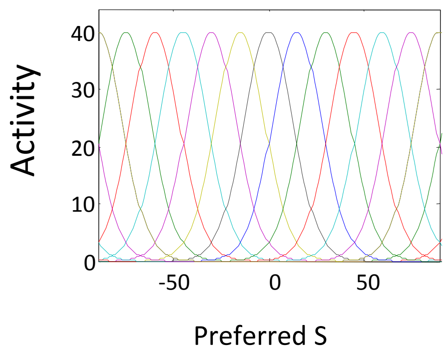


How does $p(r_3|s)$ compare to $p(r_1|s)p(r_2|s)$?

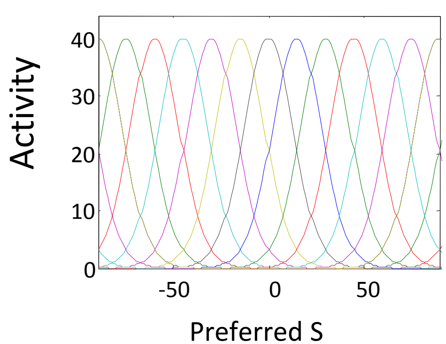
Compare the distributions

MRC | Cognition and Brain Sciences Unit

Cue 1



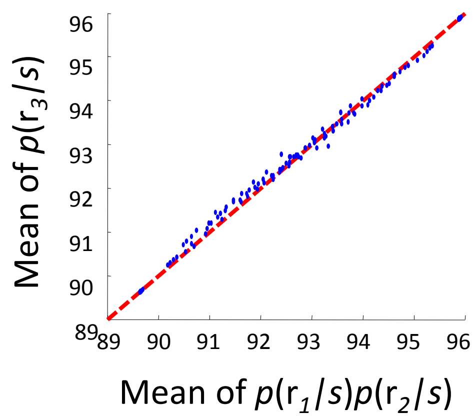
Cue 2



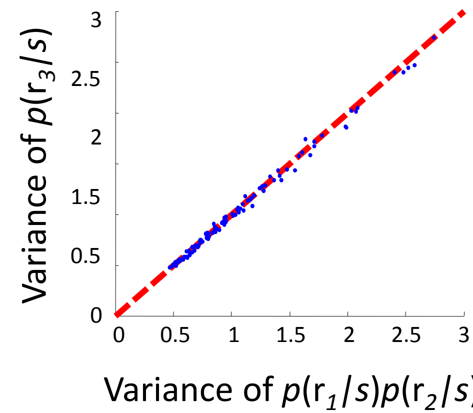
Identical tuning curves

$p(r_3|s)$ versus $p(r_1|s)p(r_2|s)$

Mean

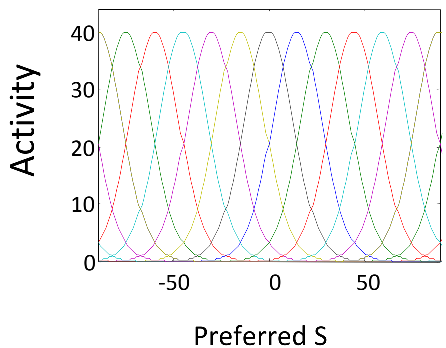


Variance

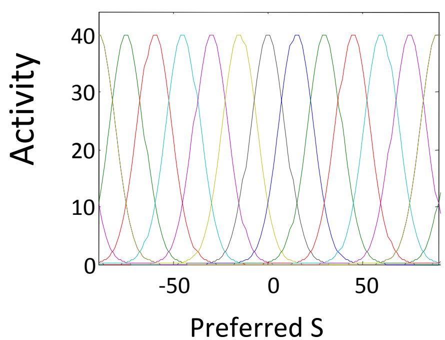


Compare the distributions

Cue 1

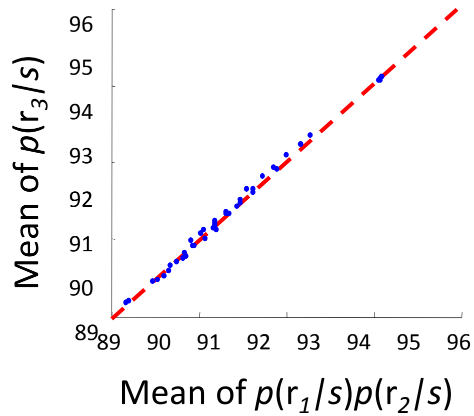


Cue 2

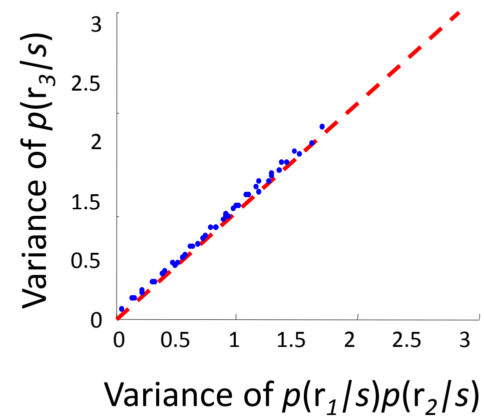


$p(r_3|s)$ versus $p(r_1|s)p(r_2|s)$

Mean



Variance



Different tuning curves and different correlations

Example: Decision-making

MRC | Cognition and Brain Sciences Unit

Accumulating Evidence over time

1. Sensory evidence is accumulated over time
2. Accumulation is stopped at some point
3. Action must be selected

Example: Decision-making

MRC Cognition and Brain Sciences Unit

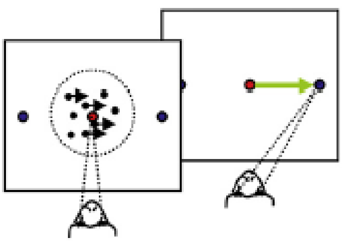
Motion direction task (extensively used in prior studies):

Presentation of random dots, a fraction is moving coherently in one direction

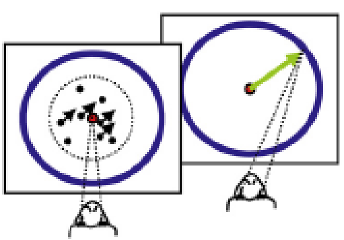
Report direction of movement with a saccadic eye movement to a choice target

Task:

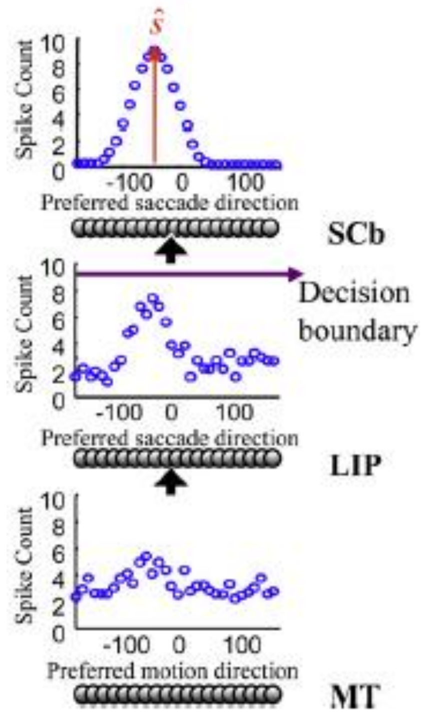
Binary decision making: left or right



Continuous decision making: any direction



Model:



Superior Culliculus(SCb)
Output, decision Layer

Lateral IntraParietal (LIP)
Evidence Accumulation Layer

Middle Temporal (MT)
Input layer

(Beck et al, 2008)

Example: Decision-making

MRC | Cognition and Brain Sciences Unit

Motion direction task (extensively used in prior studies):

Presentation of random dots, a fraction is moving coherently in one direction

Report direction of movement with a saccadic eye movement to a choice target

Model:

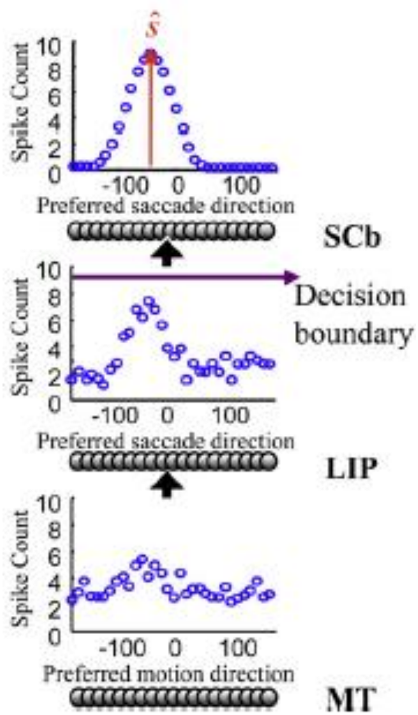
Tuning curves for saccade direction

attractor network

Tuning curves for saccade direction

long time constant (1s), allow to integrate inputs

Tuning curves for direction of motion



$$\mathbf{r}^{\mathrm{LIP}}(t_N) = \sum_{n=1}^{N} \mathbf{r}^{\mathrm{MT}}(t_n)$$

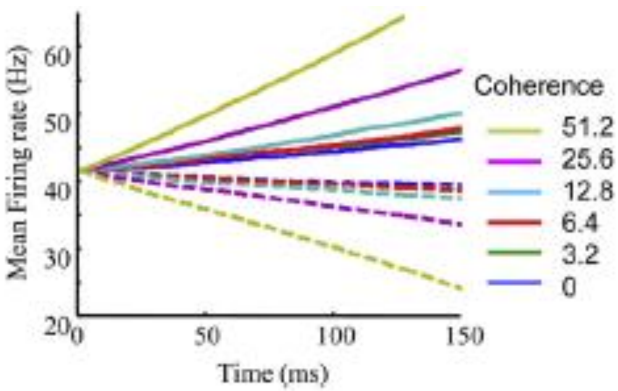
(Beck et al, 2008)

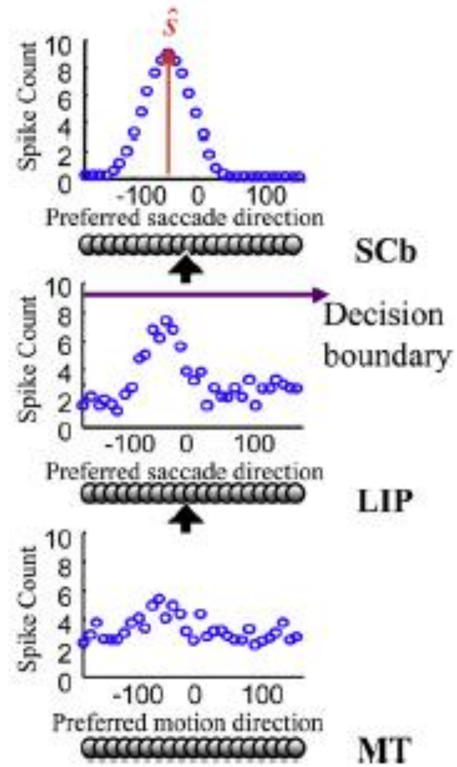
Example: Decision-making

MRC Cognition and Brain Sciences Unit

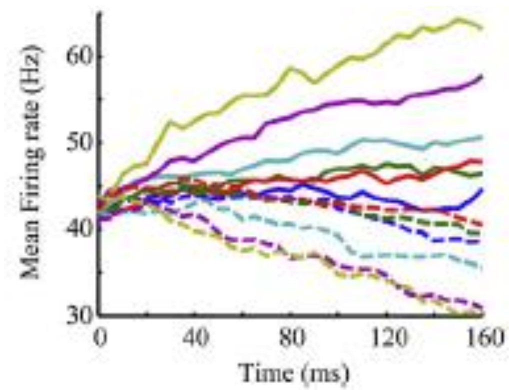
Motion direction task: coherence is reliability of the motion information

Model output:





Empirical evidence



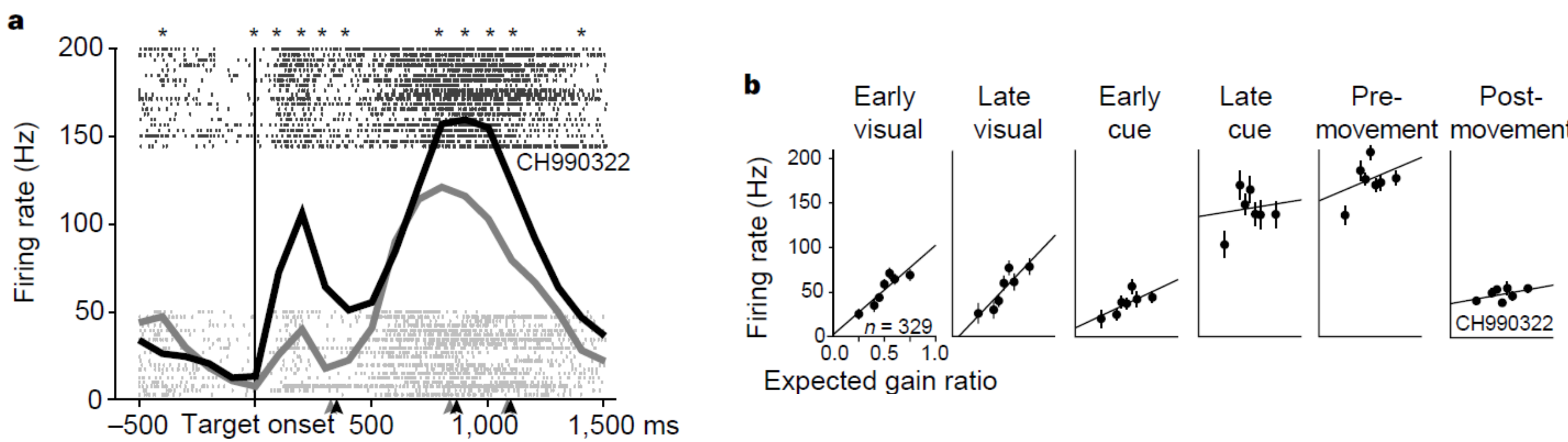
Firing rate over time for two units tuned to 180 (solid line) and 0 (dotted line) for six different levels of coherence.

(Beck et al, 2008)

But what about the prior P(s)

MRC | Cognition and Brain Sciences Unit

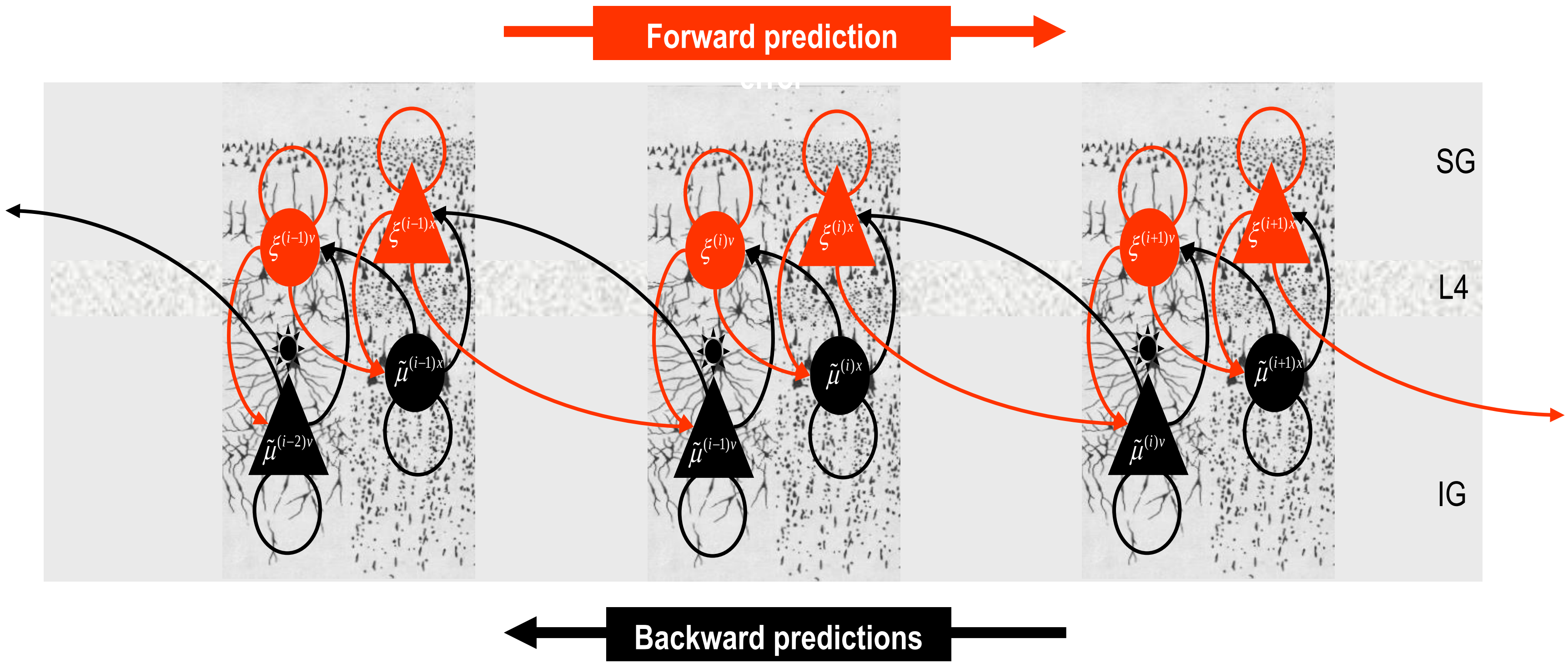
- Local prior ?
  - Prediction: baseline activity in cortex (e.g. before the start of a trial) should encode the prior distribution



(Glimcher and Platt, 1999)

But where does P(s) come from?

- biologically plausible framework of acquiring Priors
  - Optimized (learnt) online in hierarchical generative models (under free-energy principal)

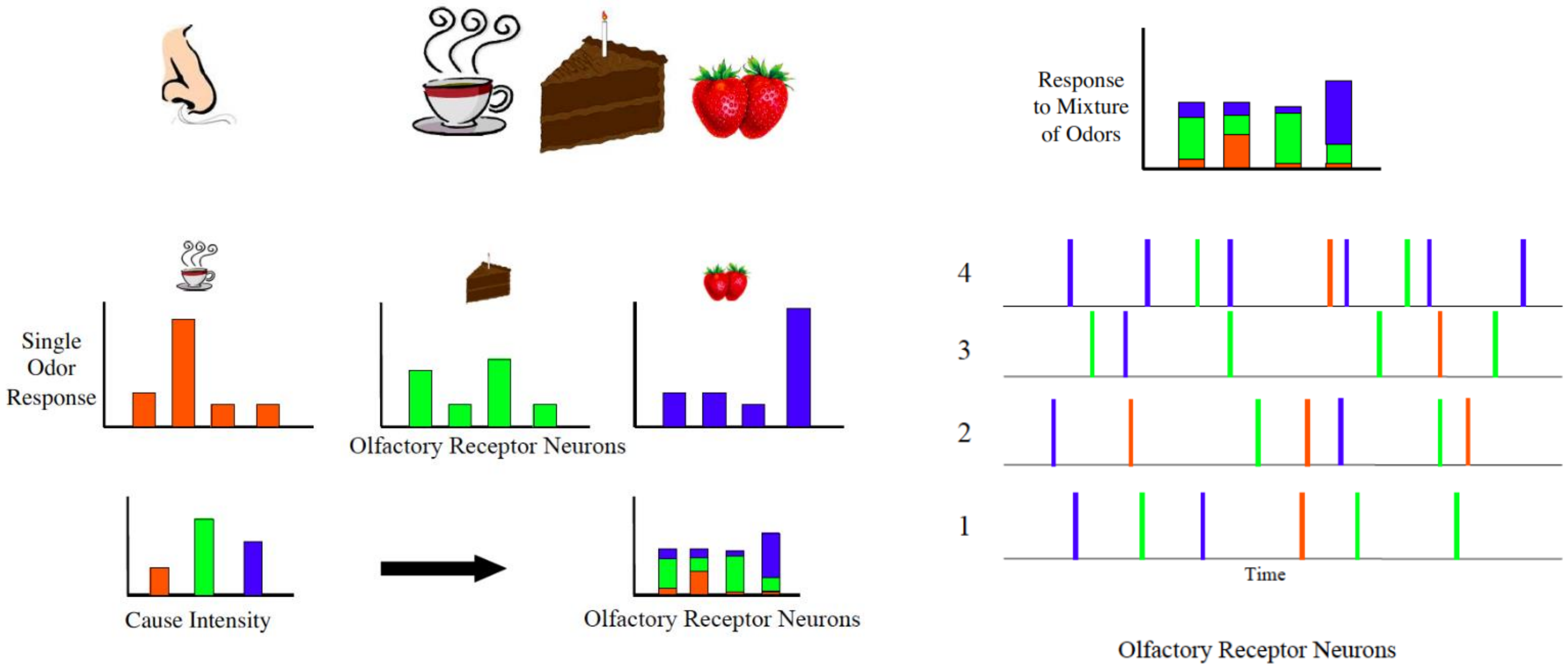


(Friston, 2010)

Bayesian learning

MRC Cognition and Brain Sciences Unit

- So far $p(\mathbf{r}|s)$ is a function only of the stimulus s
- Very often the brain needs to infer the latent causes



Beck et al. (2012) NIPS

Bayesian learning

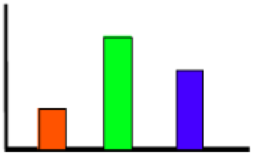
MRC | Cognition and Brain Sciences Unit

- Latent cause $\mathbf{c} = [c_1, c_2, \ldots, c_H]$

- Observed sensory data $\mathbf{r} = [r_1, r_2, \ldots, r_H]$

- Parameter Θ

$$P(\mathbf{r}|\Theta) = \sum_i P(r_i|c_i;\Theta)P(c_i|\Theta)$$



Cause Intensity

- Optimization problem: find maximum likelihood parameters Θ

$$\Theta^* = \mathrm{argmax}_\Theta P(\mathbf{r}|\mathbf{c};\Theta)$$

Expectation-Maximization algorithm

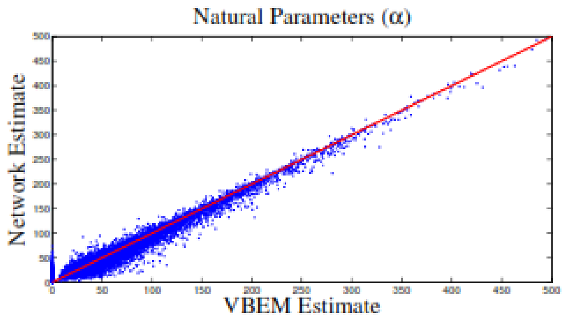
- The lower bound of $\log P(\boldsymbol{r}|\Theta)$ is the free energy

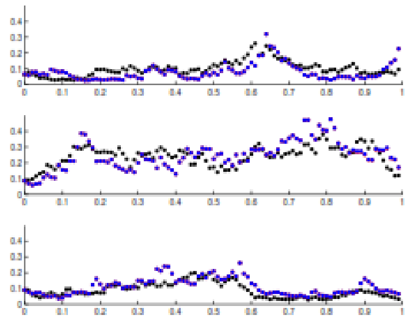
$$\log P(\mathbf{r}|\Theta) \geq \mathbb{F}(\theta, Q) = \sum_i Q(\mathbf{c}|\Theta) \log \frac{P(\mathbf{r}, \mathbf{c}|\Theta)}{P(\mathbf{c}|\Theta)}$$

- *Q* variational distribution
- E-M iteratively optimize ***F***
  - E-step (inference): compute the variational posterior *Q* at current Θ
  - M-step (learning): update Θ based on current *Q*
  - Repeat until converge

Beck et al. (2012) NIPS

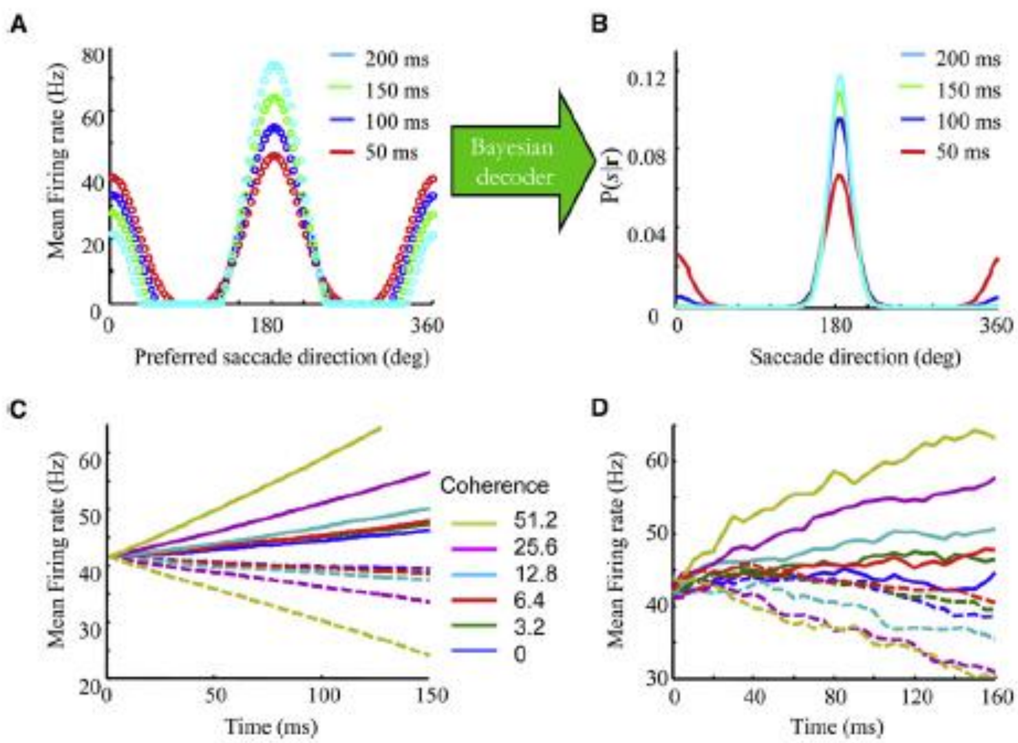
Online learning using EM





Thank you

Example: Decision-making



Accumulating Evidence over time

(Beck et al, 2008)