

# Probabilistic Population Codes in Cortex

## Computational Models of Neural Systems Lecture 7.2

David S. Touretzky  
November, 2025

# Probability: Bayes' Rule

We want to know if a patient has disease  $d$ . Test them.

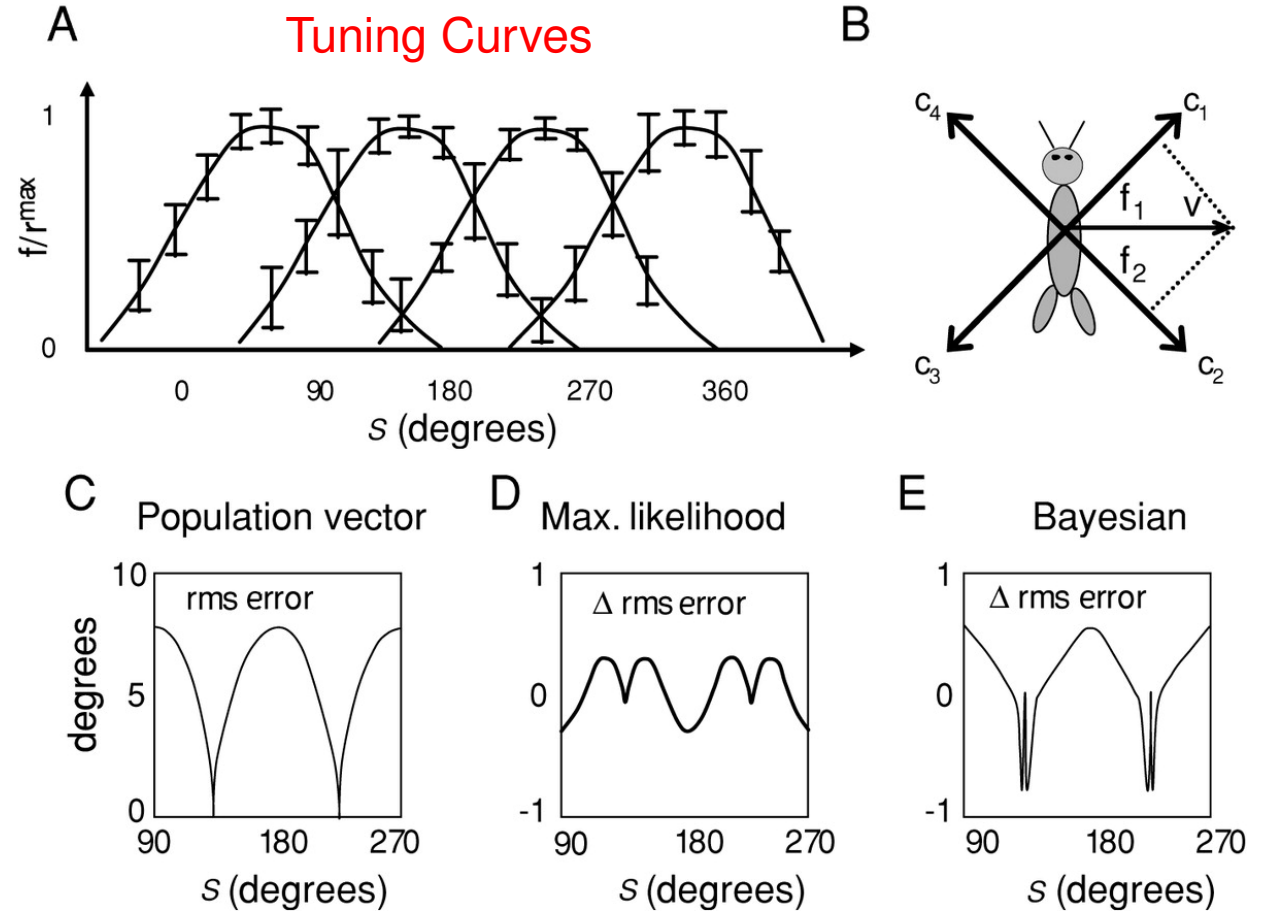
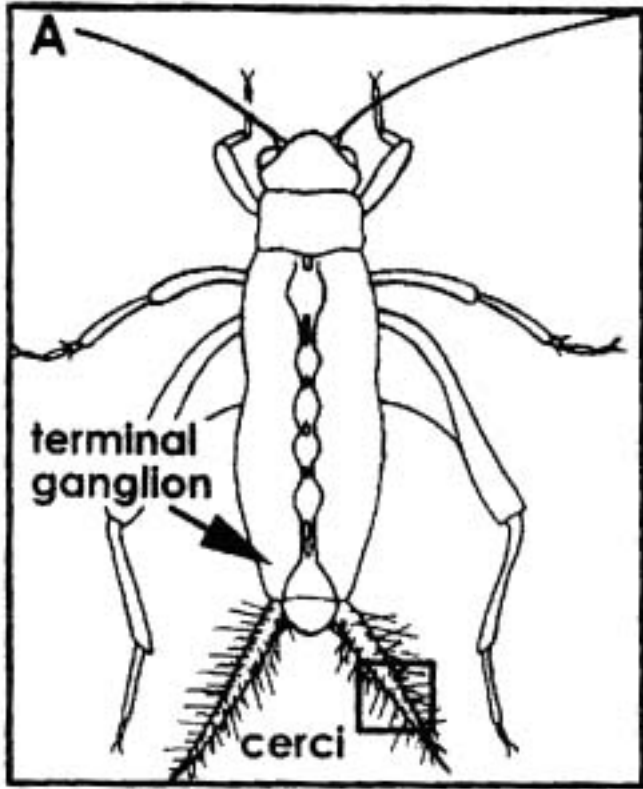
They test positive. What conclusion should we draw?

- $P(d)$       *prior*              has the disease
- $P(t\&d)$     *joint*                    tests positive *and* has the disease
- $P(t|d)$       *likelihood*            tests positive *given* has the disease
- $P(d|t)$       *posterior*              has disease *given* test is positive
- $P(t)$         *evidence*                test is positive (aka “marginal likelihood”)

Bayes' Rule:

$$P(d|t) = \frac{P(d \& t)}{P(t)} = \frac{P(t|d) \cdot P(d)}{P(t)}$$

# Cricket Cercal System Encodes Wind Direction Using Four Sensory Neurons



Max firing rate  $\sim 40$  Hz; baseline 5 Hz. Assume a Poisson spike rate distribution. Bayesian method gives lowest total decoding error.

Error relative to population vector.

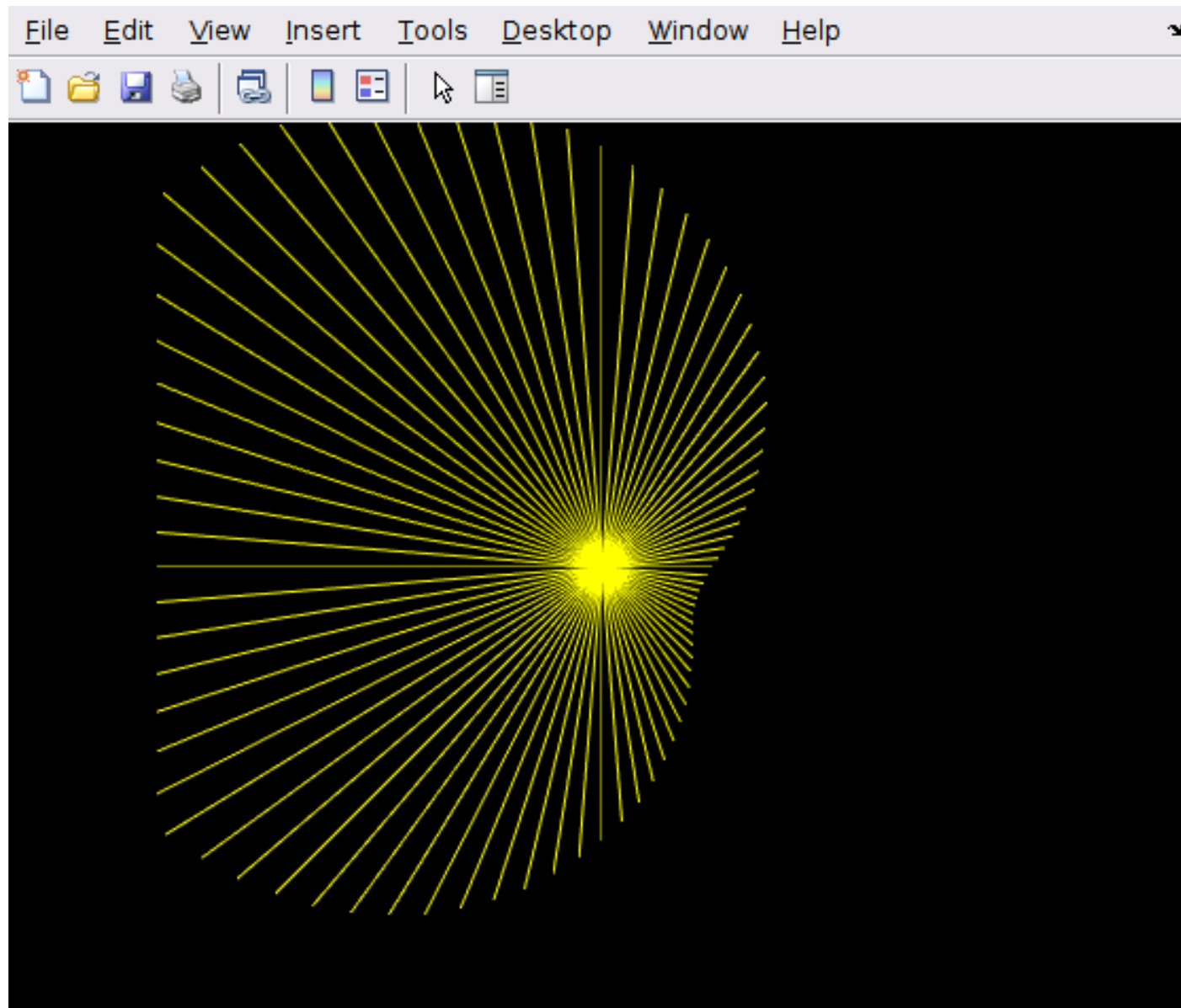
# Population Vector

- Term introduced by Georgopoulos to describe a method of decoding reaching direction in motor cortex.
- Given a set of neurons with preferred direction unit vectors  $\vec{v}_i$  and firing rates  $r_i$ , compute the direction  $V$  encoded by the population as a whole.
- Solution: weight each preferred direction vector by its normalized firing rate  $r_i/r_{max}$ .

$$\vec{V} = \frac{1}{N} \sum_{i=1}^N \frac{r_i}{r_{max}} \cdot \vec{v}_i$$

- This is a simple decoding method, but not optimal when neurons are noisy.

# popvec demo



# Maximum Likelihood Estimator

- MLE uses information about the spike rate distribution to decide how likely is a population spike rate vector  $\mathbf{r}$  given stimulus value  $s$ . For a Poisson spike rate distribution, where  $r_i$  is the spike count for true firing rate  $f_i$ :

$$P[\mathbf{r}|s] = \prod_{i=1}^N \exp[-f_i(s)\Delta t] \cdot (f_i(s)\Delta t)^{r_i\Delta t} \frac{1}{(r_i\Delta t)!}$$

- We can then use Bayes' rule to assign a probability to each possible stimulus value. Assume that all stimulus values are equally likely:  $P[s]$  is uniform. Then:

$$P[s|\mathbf{r}] \approx \frac{P[\mathbf{r}|s]}{P[\mathbf{r}]}$$

# Bayesian Estimator

- If we know something about the distribution of stimulus values  $P[s]$ , we can use this information to derive an even better estimate of the stimulus value.
- For example: the cricket may know that not all wind direction values are equally likely, given the behavior of its predators.
- From Bayes' rule:

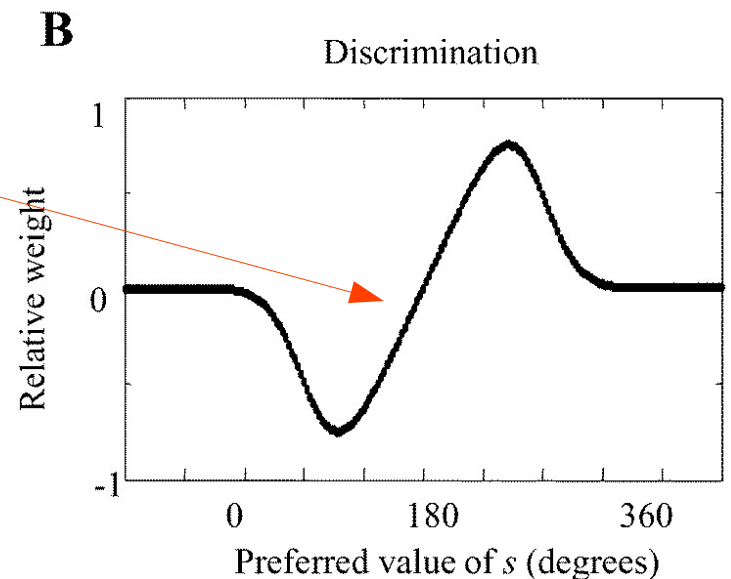
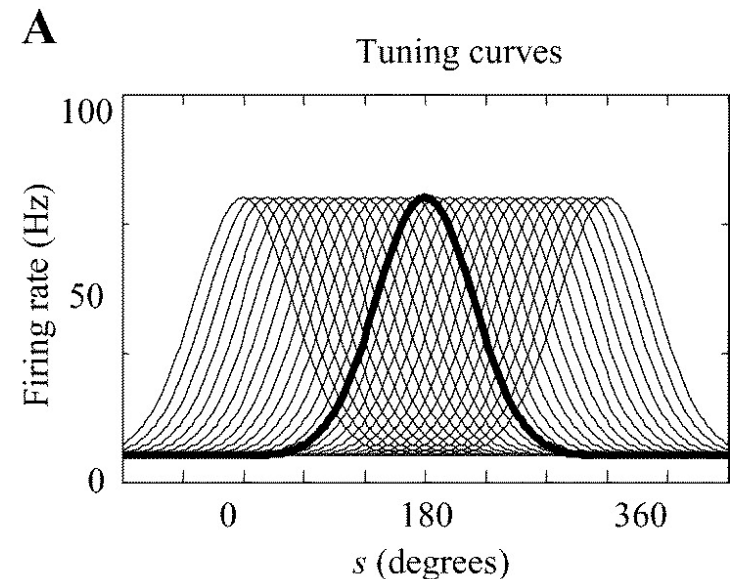
$$P[s|\mathbf{r}] = \frac{P[\mathbf{r}|s] \cdot P[s]}{P[\mathbf{r}]}$$

# Homogeneous Population Code for Orientation in Primary Visual Cortex (V1)

- Gaussian tuning curves with  $\sigma = 15^\circ$ .  
Baseline firing rate = 5 Hz.
- Optimal linear decoder weights to discriminate a stimulus  $s^* - \delta s$  from a stimulus  $s^* + \delta s$ , where  $s^* = 180^\circ$ .  
Note that the weight on the unit coding for  $180^\circ$  is zero.

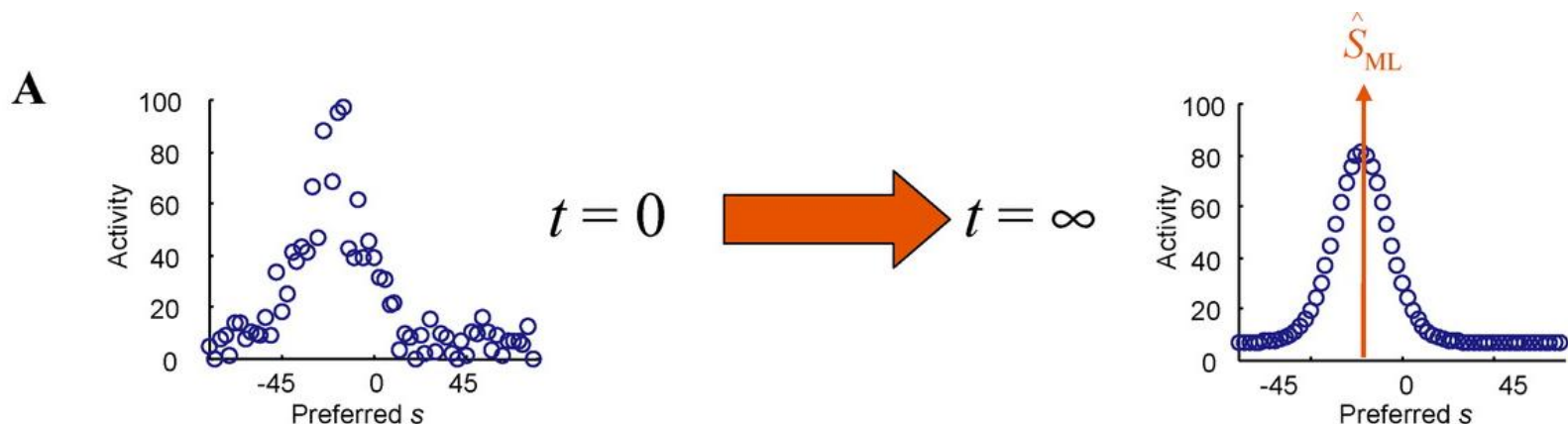
If  $t(\mathbf{r}) > 0$  conclude that stimulus  $> s$ .

$$t(\mathbf{r}) = \sum_i r_i w_i$$



# Cleaning Up Noise With Recurrent Connections

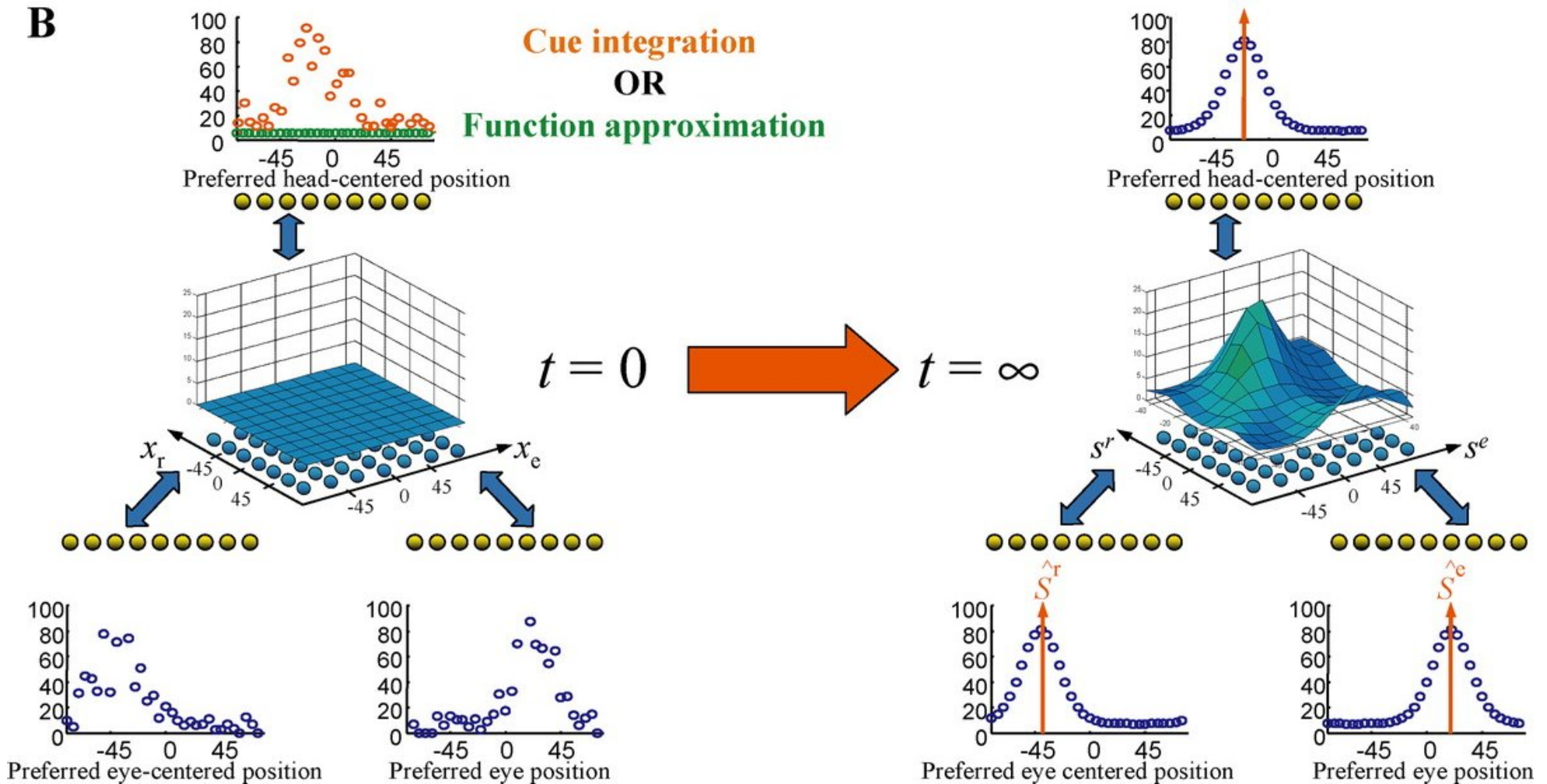
- Construct an attractor network whose attractor states correspond to perfect (noise-free) representations of stimulus values.
  - For a 1D linear variable, this would be a line attractor.
  - For a direction variable like head direction, use a ring attractor.
- The attractor network will map a noisy activity vector  $\mathbf{r}$  into a cleaner vector  $\mathbf{r}^*$  encoding the stimulus value that is most likely being encoded by  $\mathbf{r}$ .



# Basis Functions

- You can think of the neurons' tuning curves as a set of basis functions from which to construct a linear decoding function.
- But instead of decoding, we can also use these basis functions to transform one representation into another.
- Or use them to do arithmetic.
- Example: calculating head-centered coordinates from retinal position plus eye position.

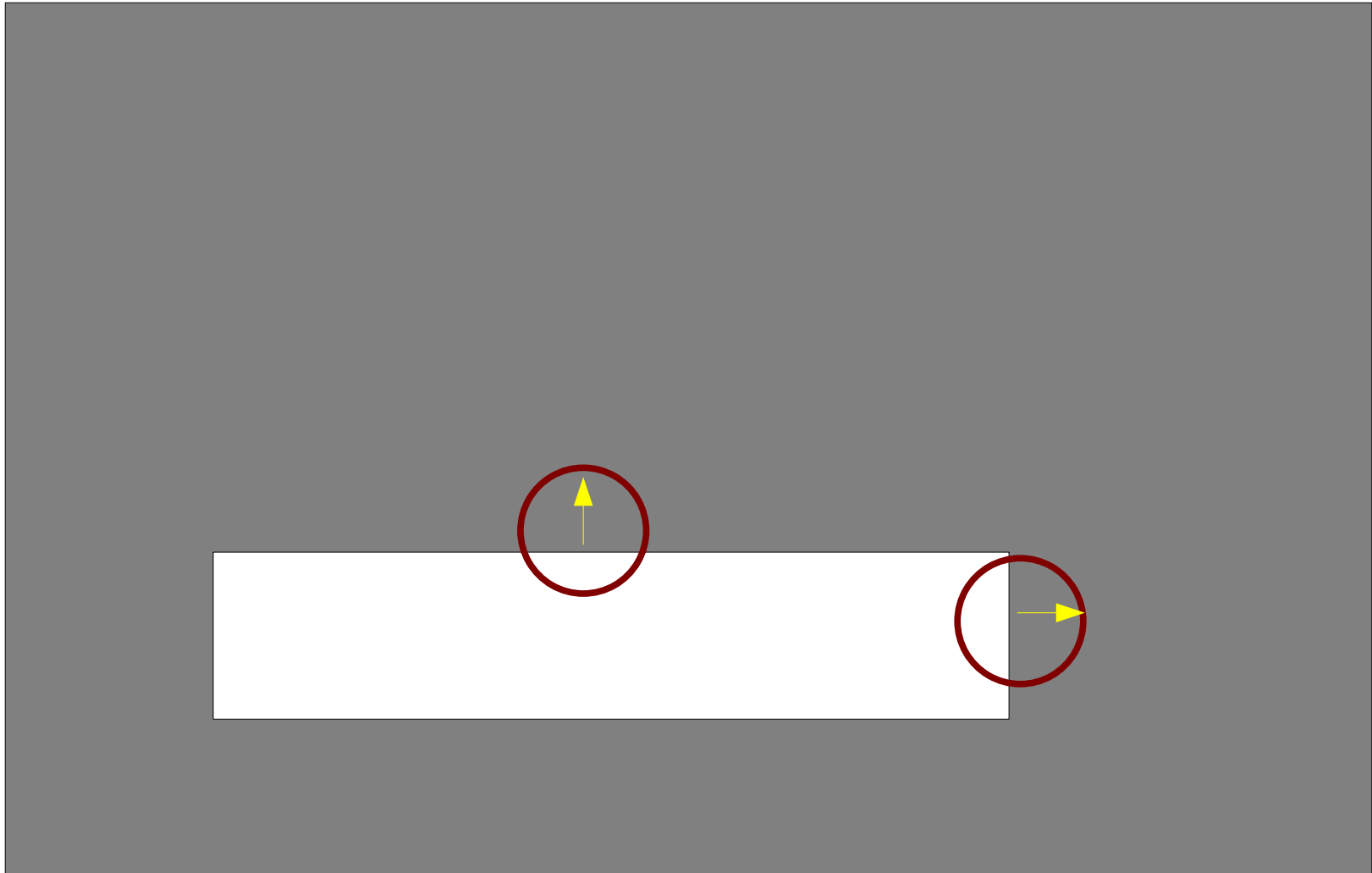
# Recurrent Network Maintains Proper Relationships Between Retinal, Eye, and Head Coordinates



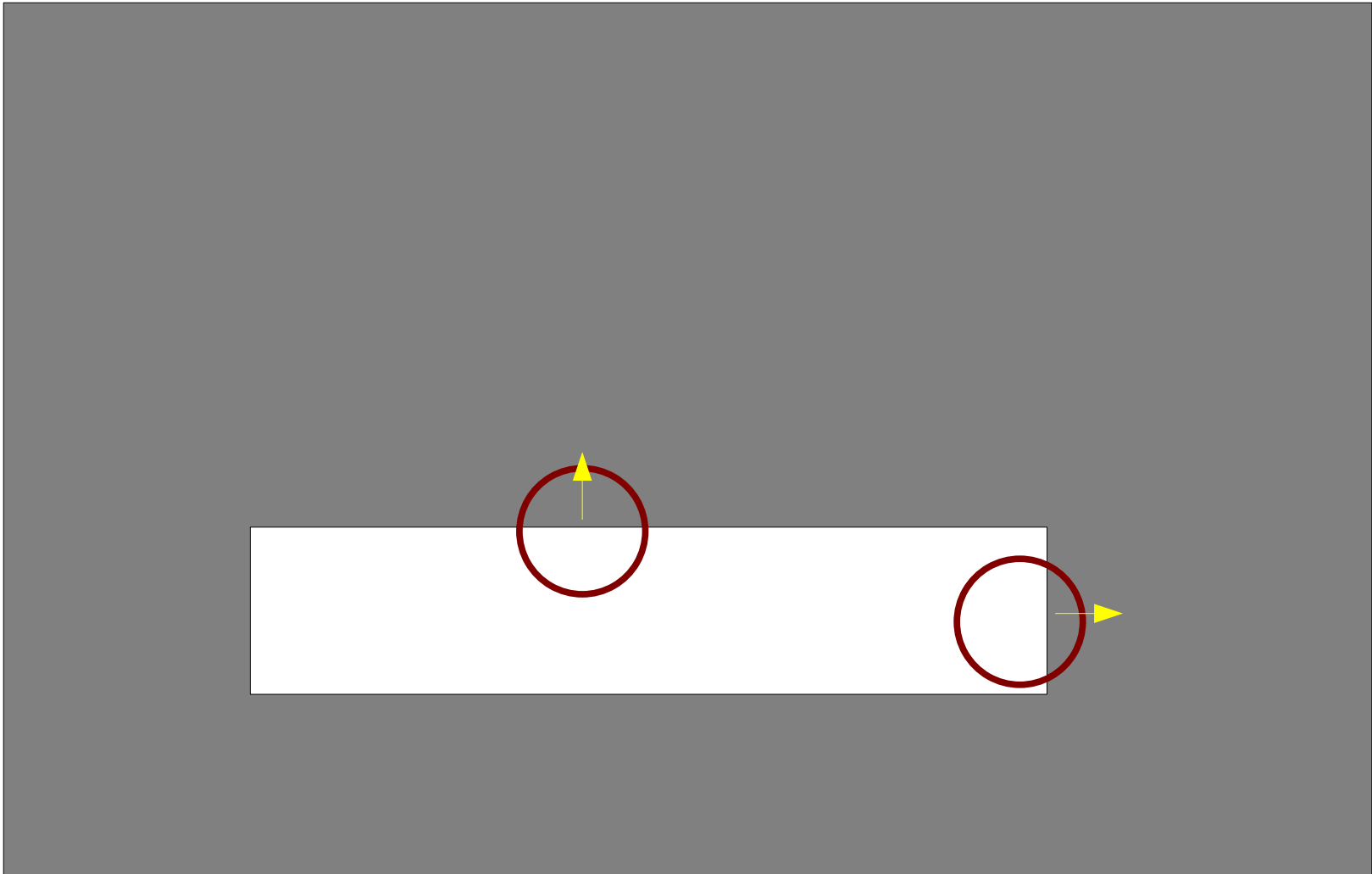
# Encoding Probability Distributions

- The previous decoding exercise assumed that the activity vector was a noisy encoding of a single value.
- What if there were inherent uncertainty as to the value of a variable?
- The brain might want to encode its beliefs about the *distribution* of possible values.
- Hence, population codes might represent probability distributions.

# Aperture Problem: In What Direction Is the Bar Moving?

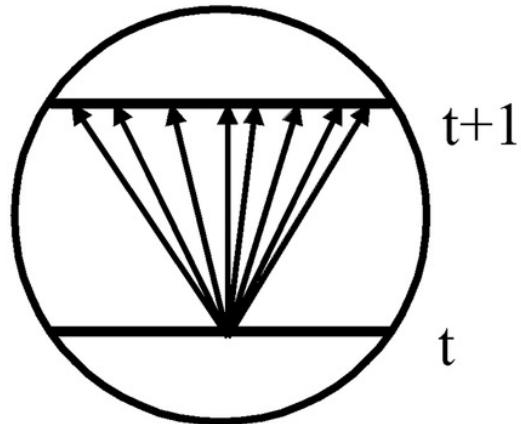


# Aperture Problem: In What Direction Is the Bar Moving?



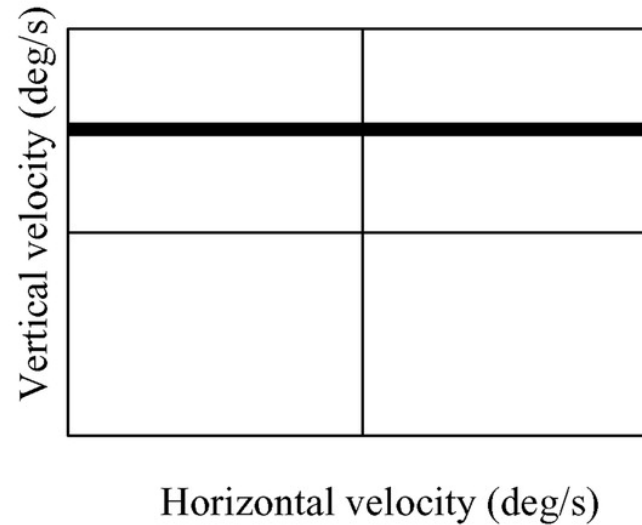
# Horizontal Direction Uniformly Distributed Because No Information Available

**A**

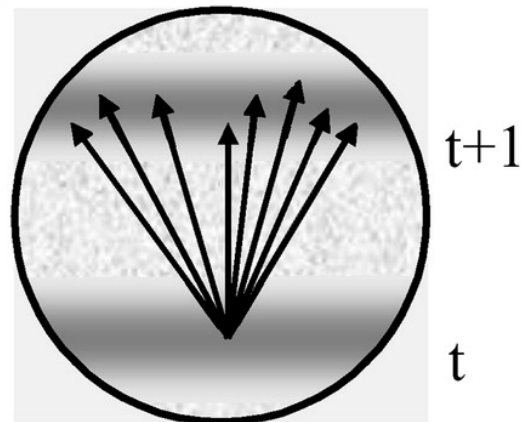


High Contrast

**B**

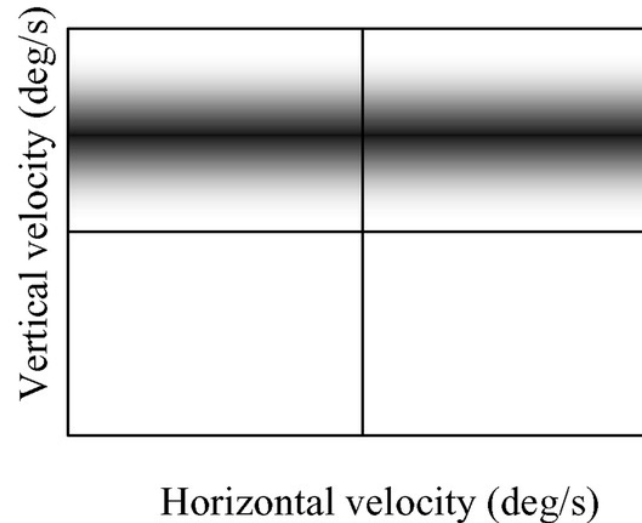


**C**



Low Contrast

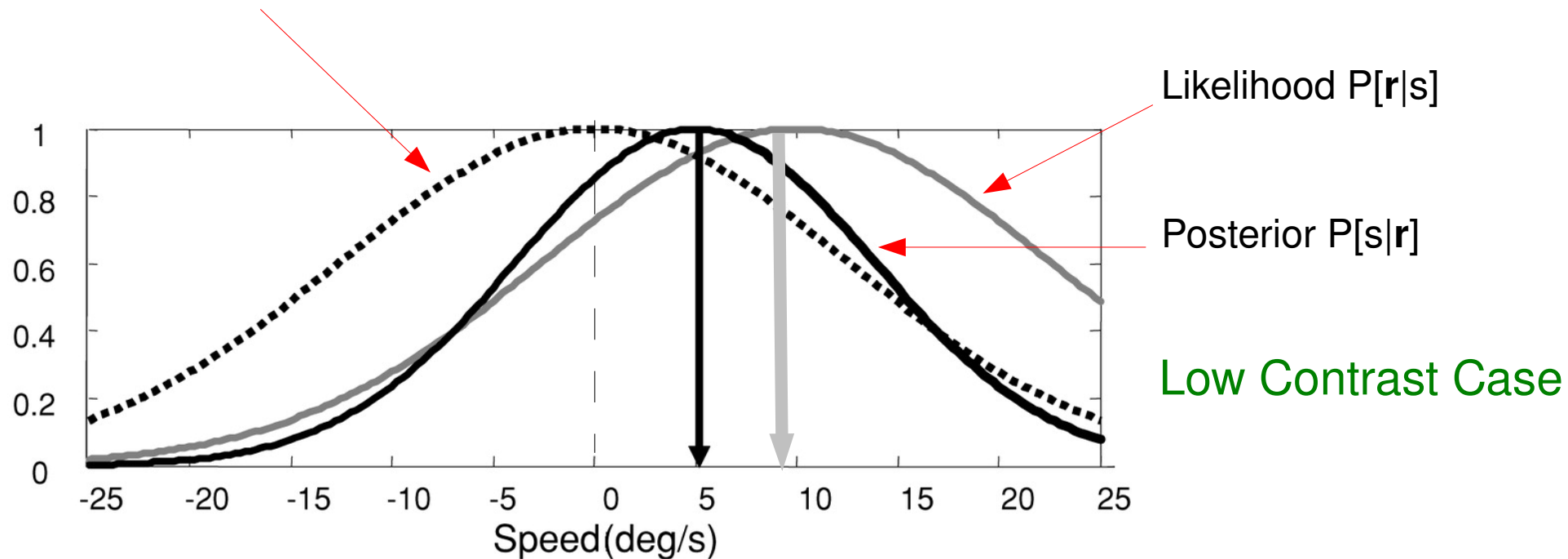
**D**



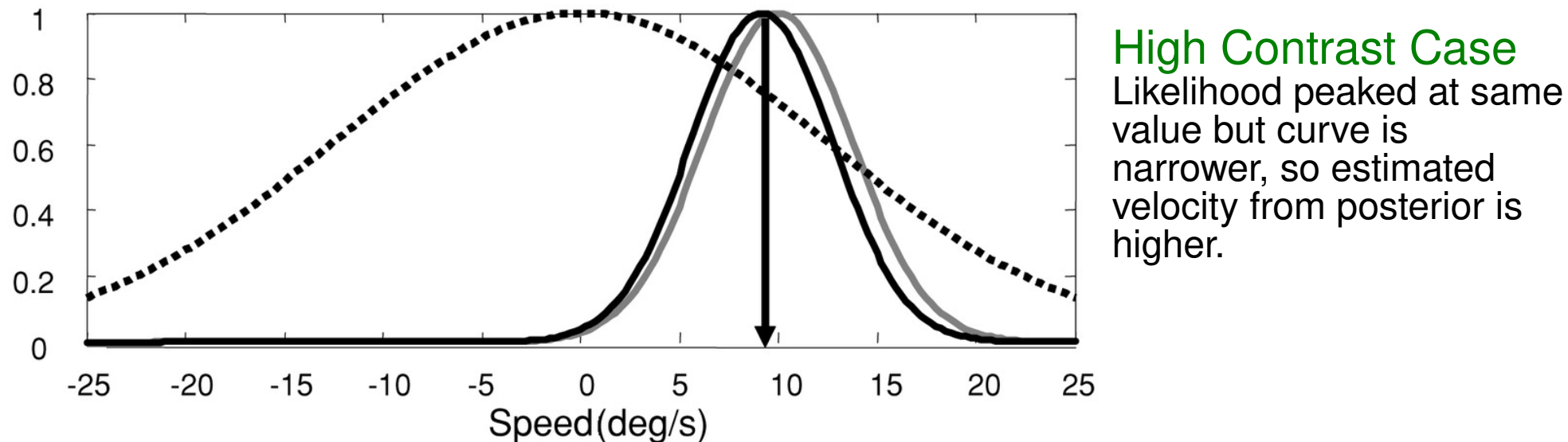
Some uncertainty about vertical velocity yields a distribution of possible values.

# Bayesian Estimation of Velocity: Prior $P(s)$ is a Gaussian Centered on Zero

A



B

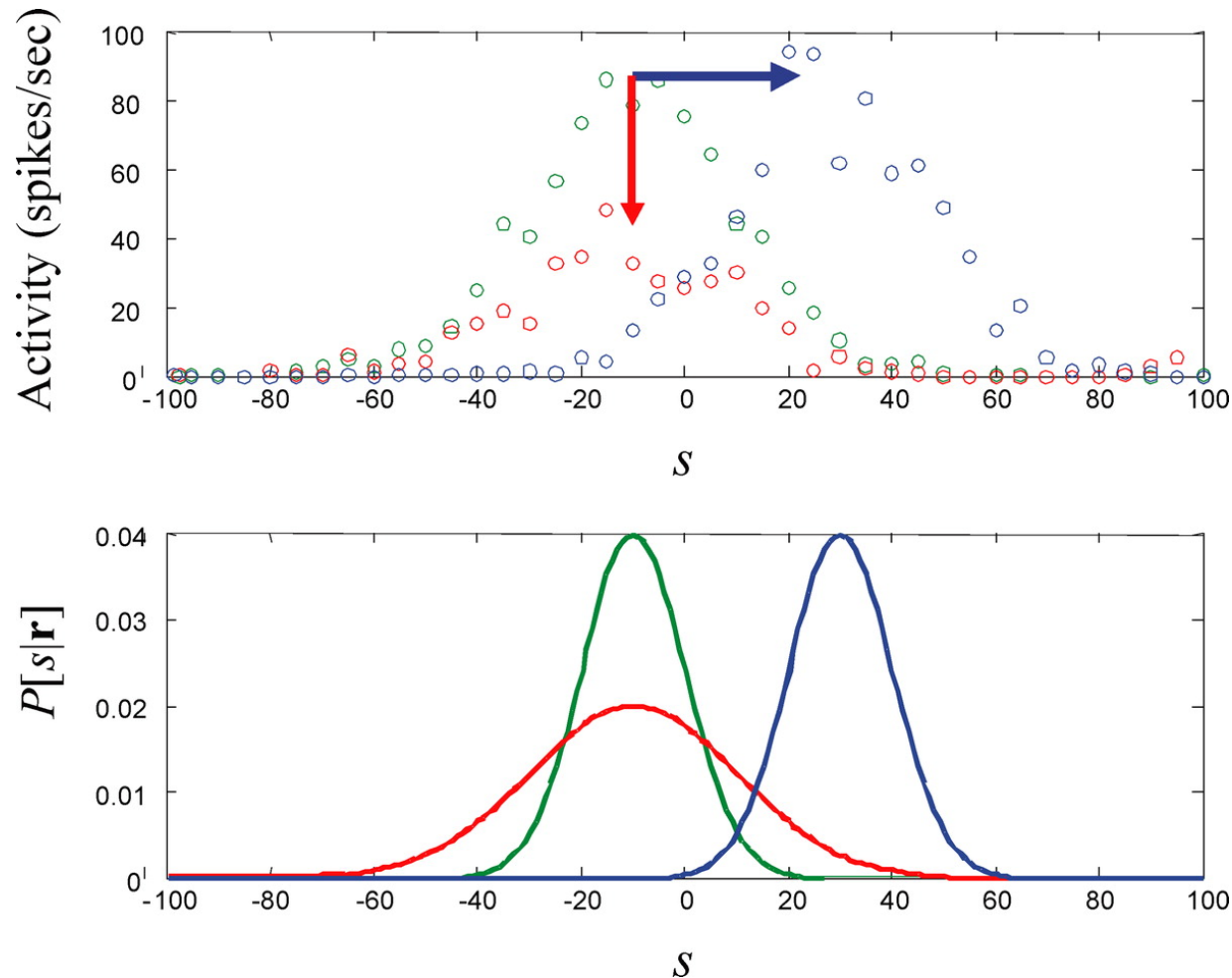


# Psychophysical Argument for Representing Distributions Instead of Expected Values

- People estimate velocities as higher when the contrast is greater. How to account for this?
- The Bayesian estimator produces this effect. Humans behave as predicted by Bayes' law.
- Why does this model work? Because:
  - The width of the likelihood distribution is explicitly represented
- Other psychophysical experiments confirm the view of humans as Bayesian estimators.
- This suggests that the nervous system utilizes probability distribution information, not just expected values.

# Decoding Gaussian Signals with Poisson Noise

- Translation (blue) shifts the probability distribution but does not change the shape from the original (green).
- Scaling down (red) broadens the variance.



# Convolutional Encodings

- For other types of probability distributions we don't want to use uniform Gaussian tuning curves. Instead, convolve the probability distribution with a set of basis functions.
- Fourier encoding (sine wave basis functions):

$$f_i(P[s|\mathbf{r}]) = \int ds \cdot \sin(w_i s + \phi_i) \cdot P[s|\mathbf{r}]$$

- Gaussian kernels:

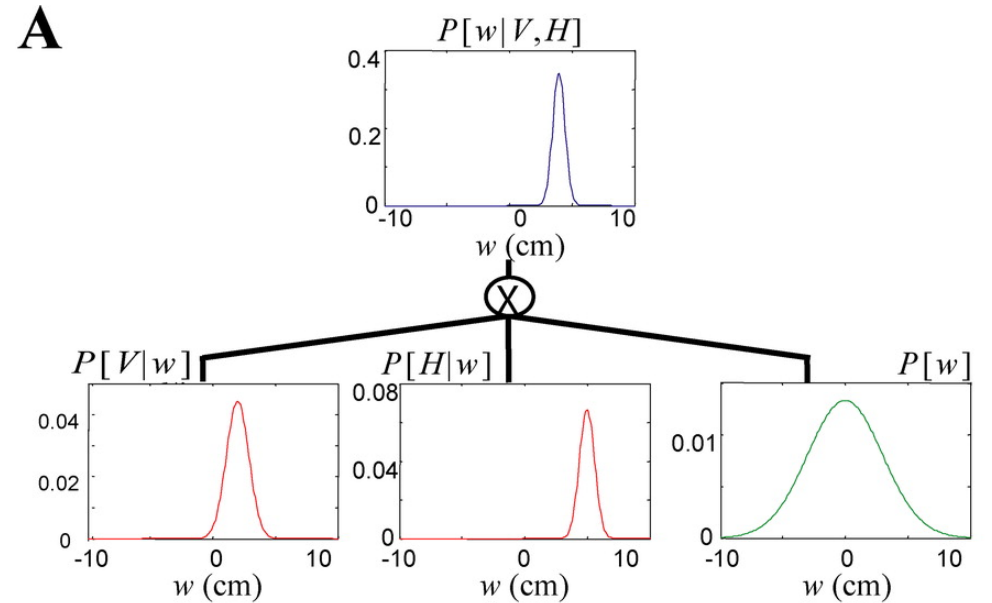
$$f_i(P[s|\mathbf{r}]) = \int ds \cdot \exp\left(-\frac{(s-s_i)^2}{2\sigma_i^2}\right) \cdot P[s|\mathbf{r}]$$

- Decoding of these representations is tricky.

# Ernst & Banks Experiment

Estimating the width of a bar using both visual (V) and haptic (H) cues.

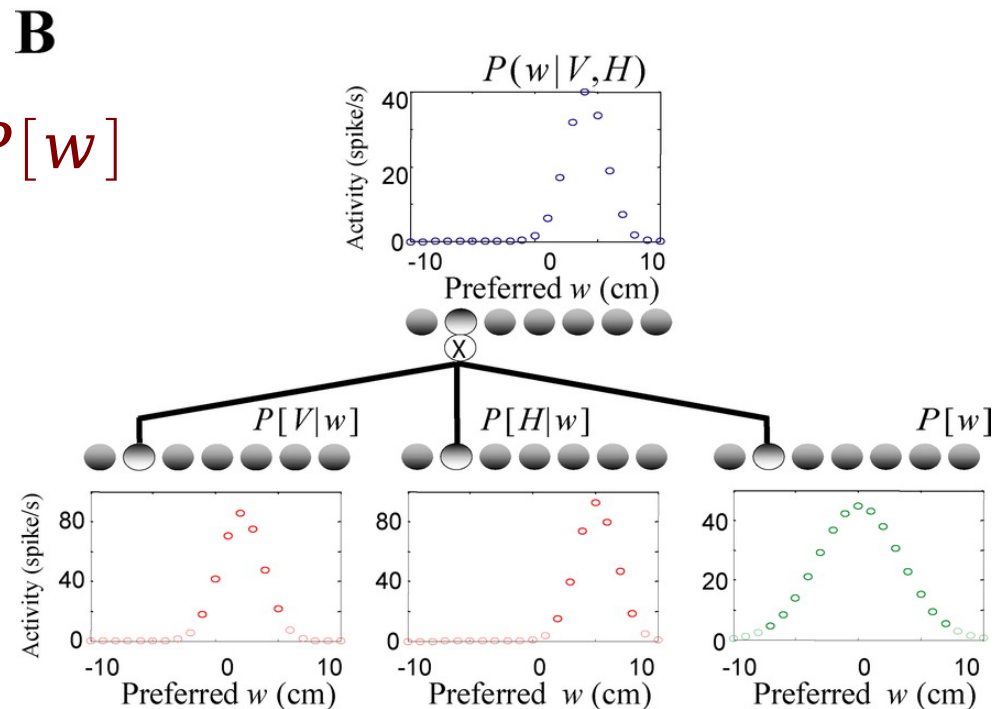
Population codes are computed by convolving with Gaussian kernels.



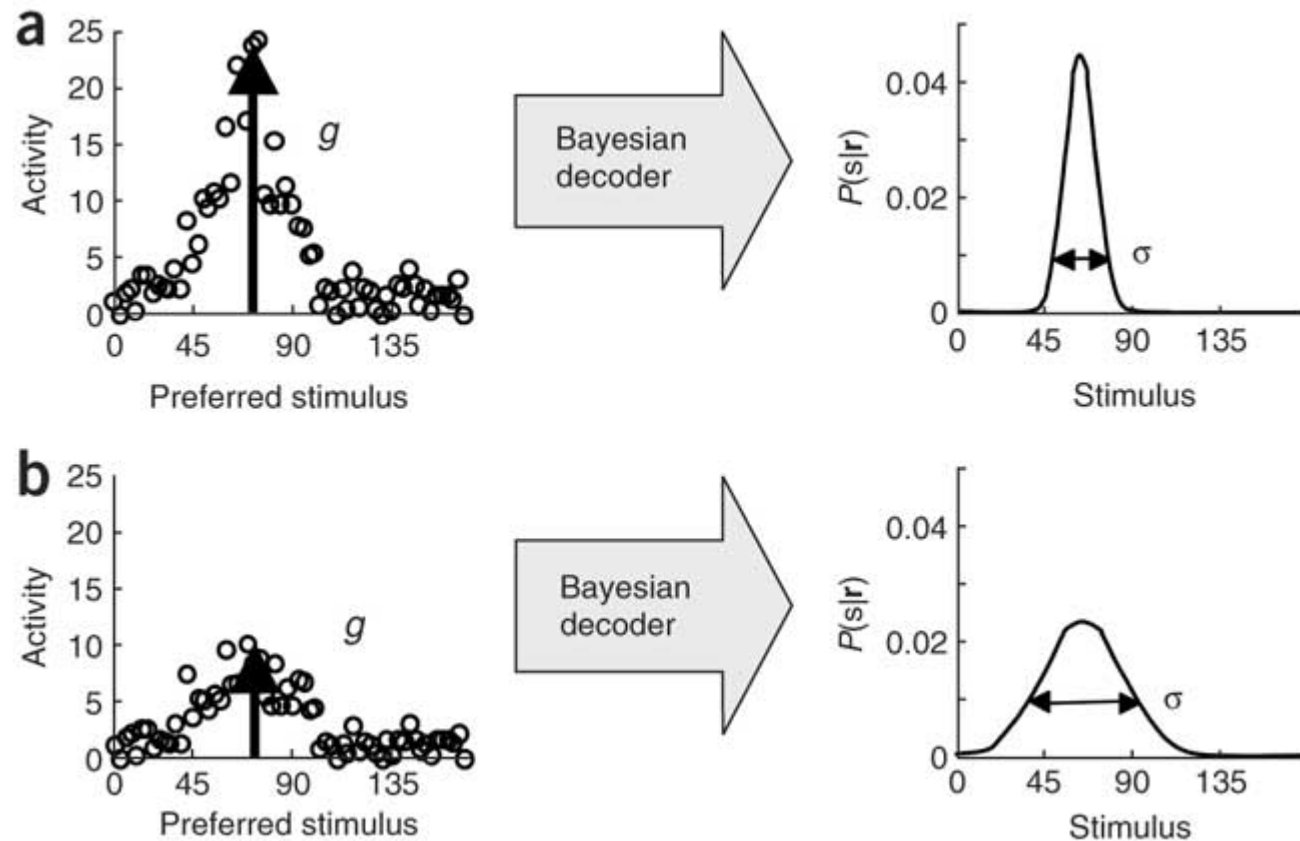
$$P[w|V,H] \propto P[V|w] \cdot P[H|w] \cdot P[w]$$

“Neural” model does three-way element-wise multiplication.

In this way, we can do inference using noisy population codes.

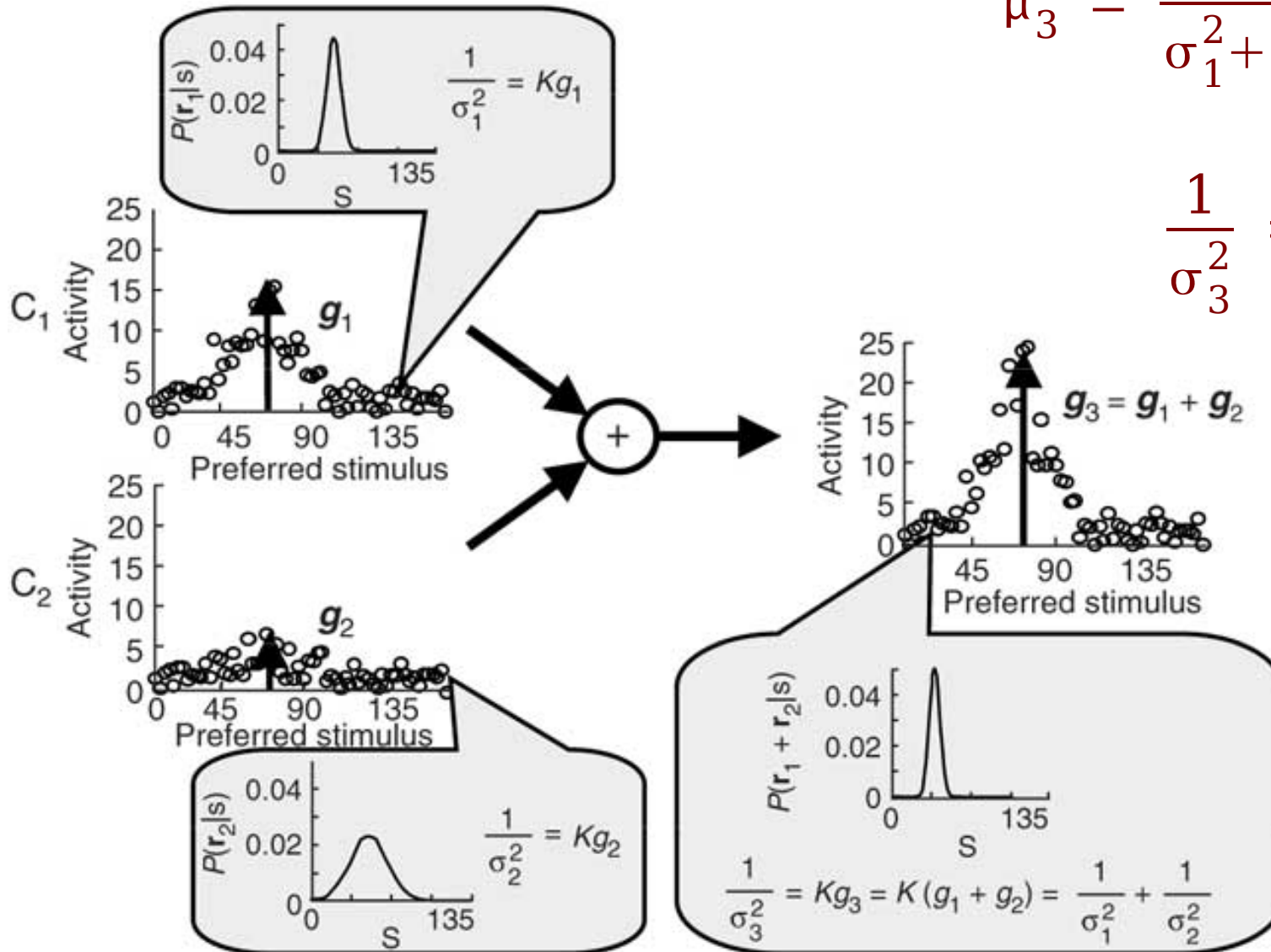


# Ma et al. (2006): Bayesian Inference with Population Codes



Lower amplitude means broader variance.

# Sensory Integration of Gaussians w/Poisson Noise



$$\mu_3 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \mu_2$$

$$\frac{1}{\sigma_3^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

# Generalizing the Approach

- Gaussians with Poisson noise are easy to combine: we can do element-wise addition of firing rates, and the resulting representation is Bayes-optimal.
- Can we generalize to non-Gaussian functions and other types of noise, and retain Bayes-optimality?
- $\mathbf{r}_3 = \mathbf{r}_1 + \mathbf{r}_2$  is Bayes-optimal if  $p(s|\mathbf{r}_3) = p(s|\mathbf{r}_1) p(s|\mathbf{r}_2)$ .
- This doesn't hold for most distributions but it does for some that are “Poisson-like”.

# Poisson-Like Distributions

$$P(\mathbf{r}_k | s, g) = \phi(\mathbf{r}_k, g_k) \cdot \exp(\mathbf{h}^T(s) \mathbf{r}_k)$$

$$\mathbf{h}'(s) = \Sigma_k^{-1}(s, g_k) \mathbf{f}'(s, g_k)$$

$\Sigma_k$  is the covariance matrix of  $\mathbf{r}_k$

$$\text{gain } g_k = K / \sigma_k^2$$

$\mathbf{f}_k(s)$  is the tuning curve function

For identical tuning curves and Poisson noise

$$\mathbf{h}(s) = \log \mathbf{f}(s)$$

$$\phi_k(\mathbf{r}_k, g_k) = \exp(-c g_k) \prod_i \exp(r_{ki} \log g_k) / r_{ki}!$$

# Non-Identical Tuning Curves

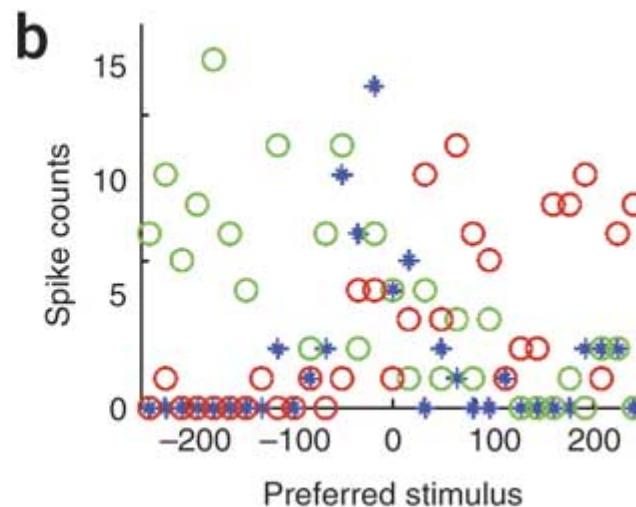
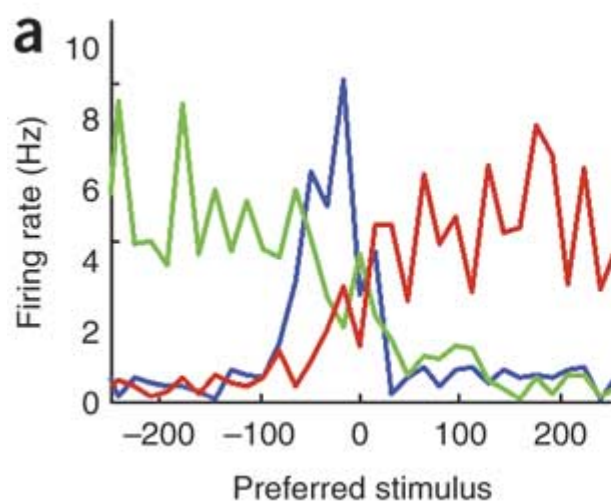
- When tuning curve functions  $\mathbf{f}_k$  are not the same,  $\mathbf{h}(s)$  is not the same for all tuning curves. Simple addition doesn't work.
- But we can still combine tuning curves using linear coefficients  $A_k$ , provided the  $\mathbf{h}_k(s)$  functions are drawn from a common basis set.

$$\mathbf{r}_3 = A_1^T \mathbf{r}_1 + A_2^T \mathbf{r}_2$$

# Combining Three Poisson-Like Populations Using Different Types of Tuning Curves

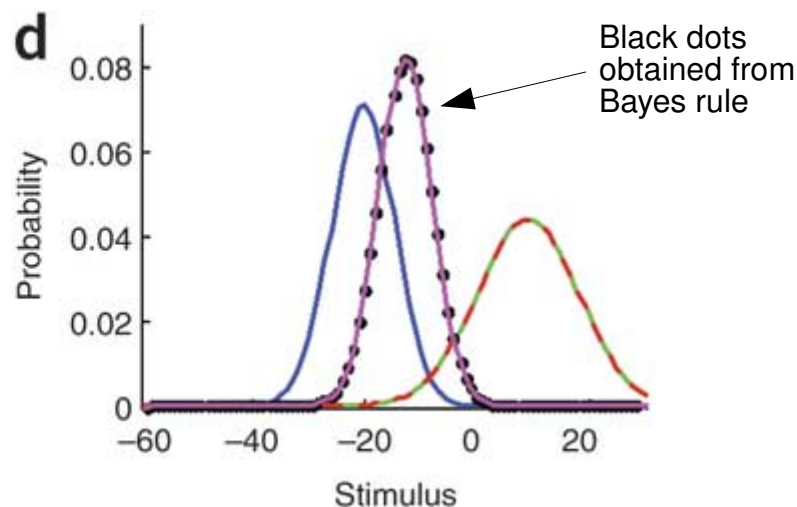
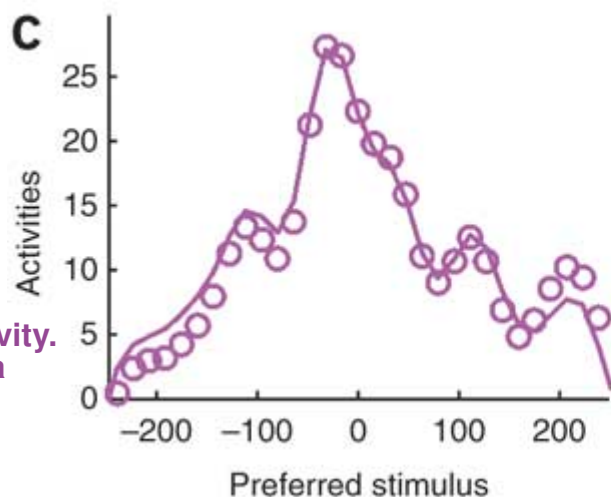
## Inputs:

Blue:  
gaussian  
Red/Green:  
sigmoidal



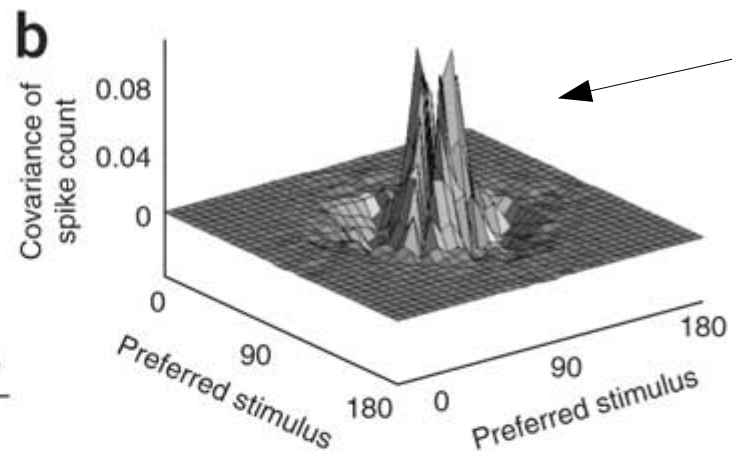
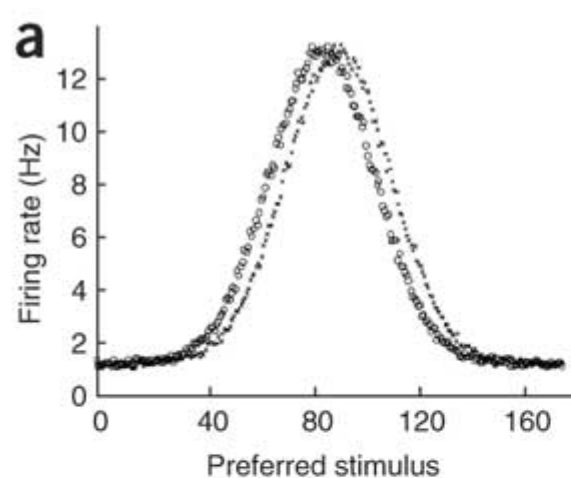
## Outputs:

Solid line: mean activity.  
Circles: activity on a single trial.

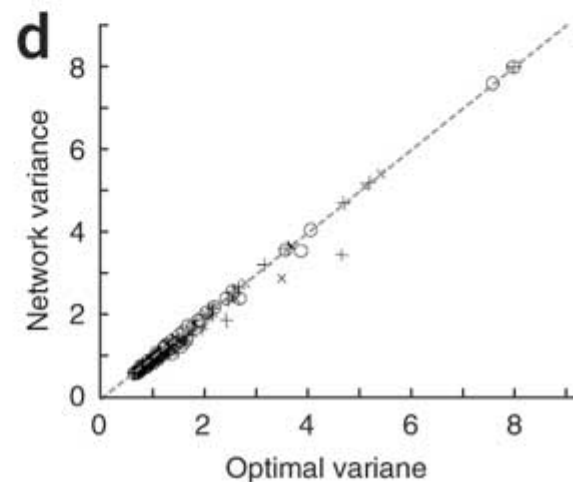
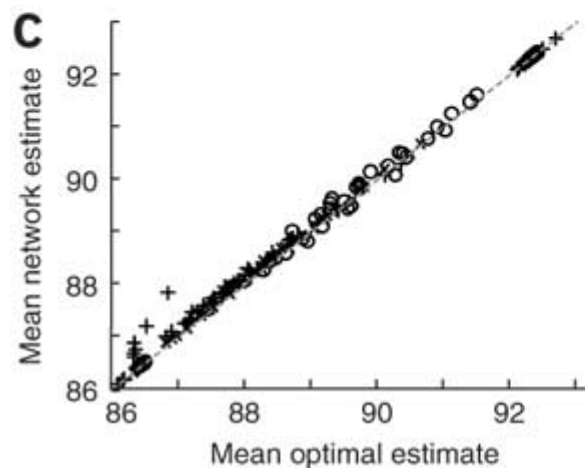


# Simulation with Integrate-and-Fire Neurons

Inputs:  
 $\mu_1 = 86.5$   
 $\mu_2 = 92.5$   
Simulates  
cue conflict.



Output units are correlated. Diagonal is suppressed in this graph.



Combined estimate is Bayes-optimal!

# Summary

- Population codes are widely used in the brain (visual cortex, auditory cortex, motor cortex, head direction system, place codes, grid cells, etc.)
- The brain uses these codes to represent more than just a scalar value. They can encode probability distributions.
- We can do arithmetic on probability distributions if the population code satisfies certain constraints.
  - Codes that are Poisson-like are amenable to this.
- The population code serves as a basis set.
  - Populations can be combined via linear operations, and in the simplest case, element-wise addition.