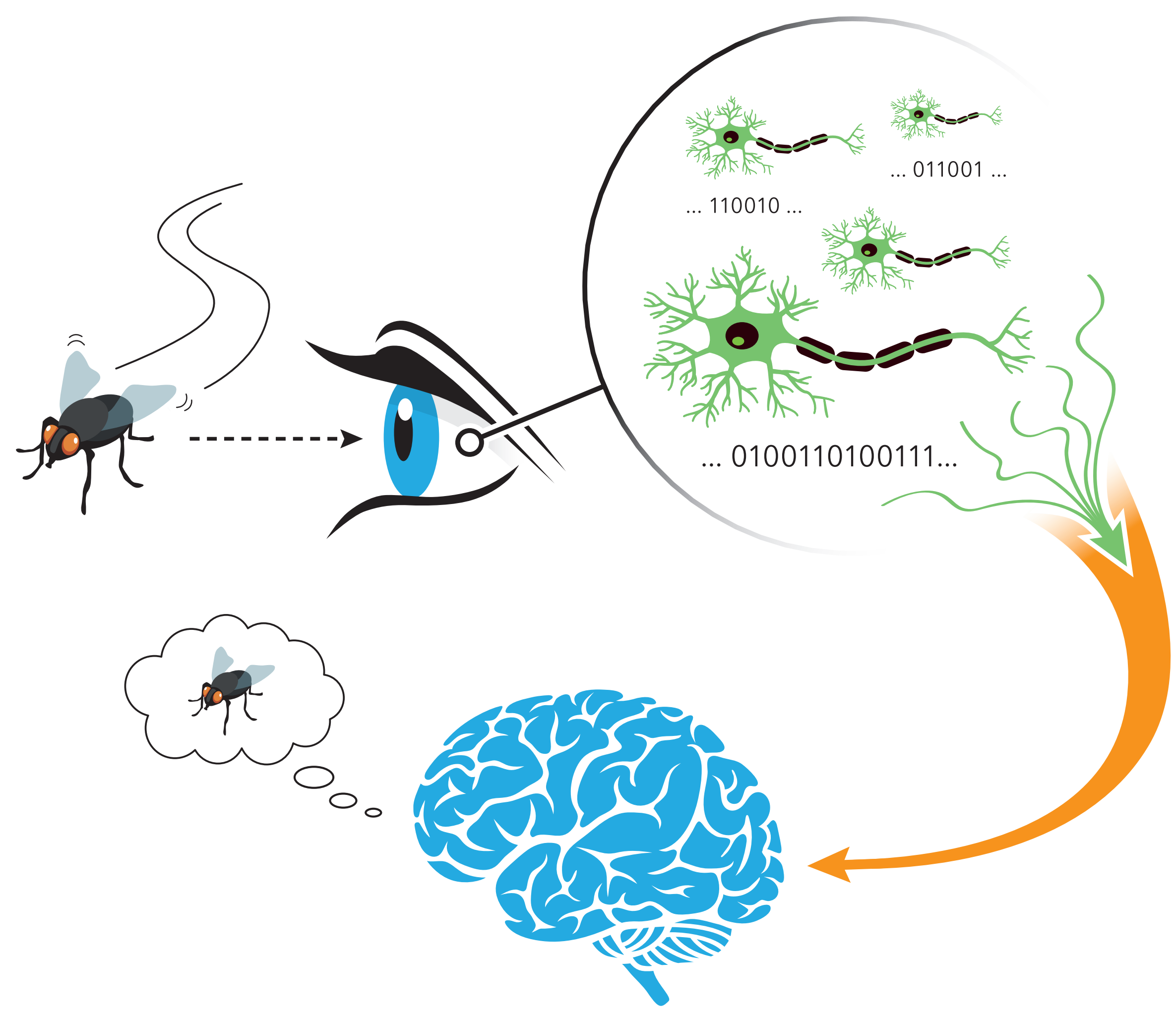
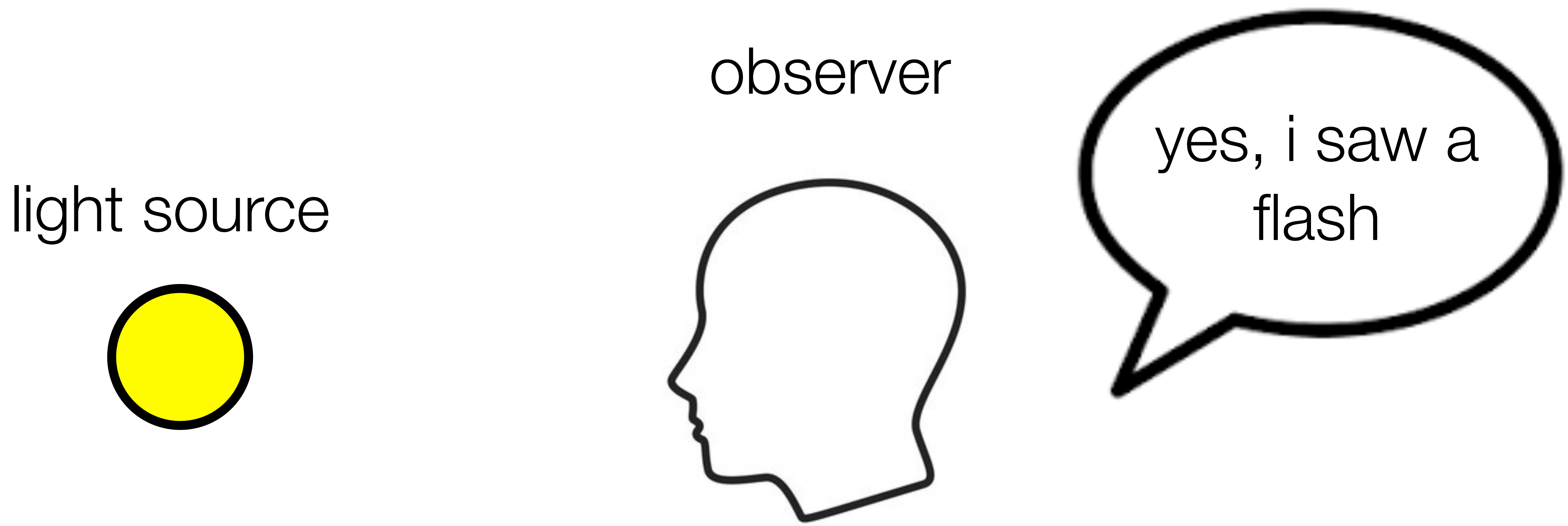
# Information coding by the visual system


Matthew Chalk

Institut de la Vision
Sorbonne Université
Paris, France

# How good is the visual system?



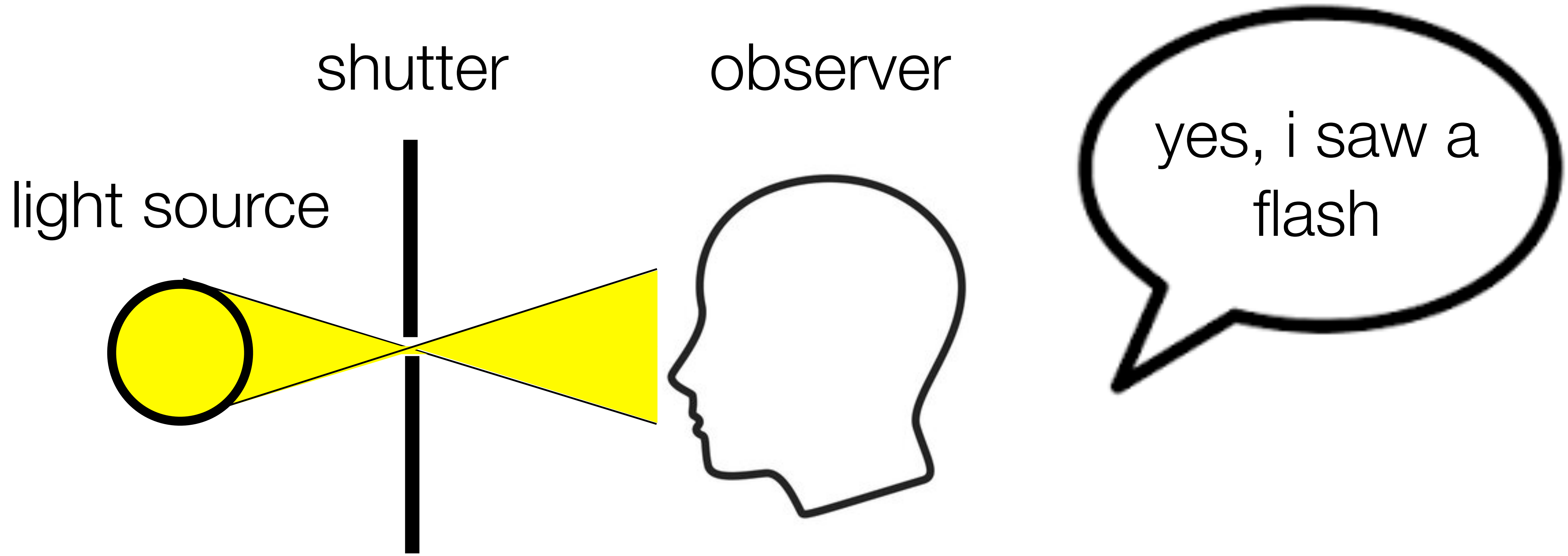

What are the limits of perception?

# Preliminaries

- http://oliviermarre.free.fr/index.php/teaching/
    - Information coding by the visual system: download code, open in matlab

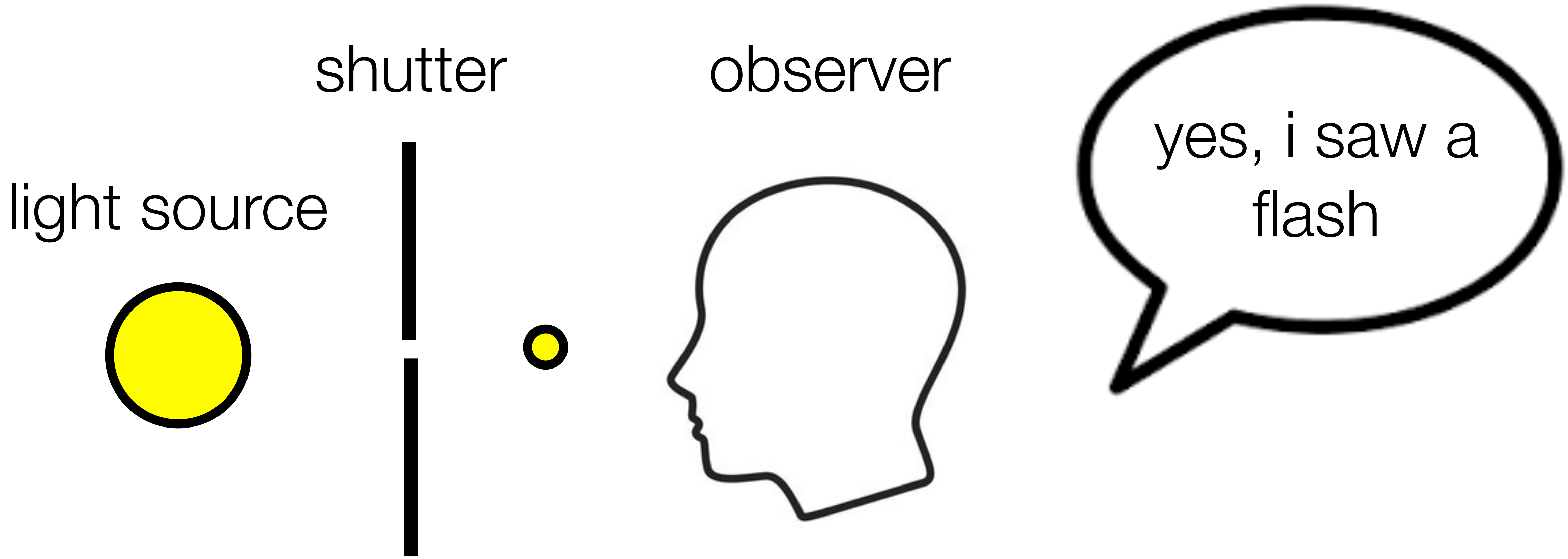
- Please ask questions!

# How good is the visual system?




What are the limits of perception?

# How good is the visual system?




What are the limits of perception?

Can we see a single photon?

Hecht, Shlaer, and Pirenne, 1942

# How good is the visual system

Rod cells in the retina absorb photons, and convert to electrical signal

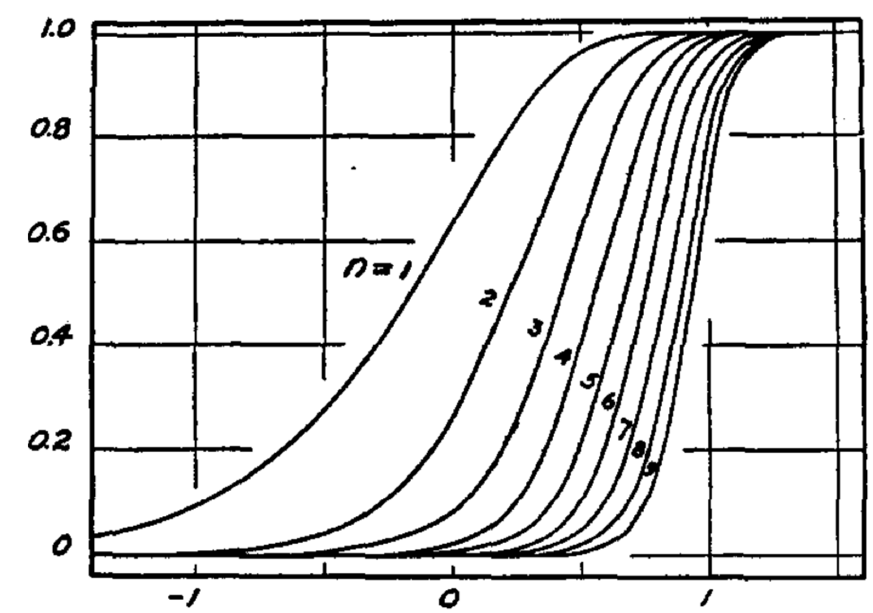
$$\lambda = \text{\% absorbed} \times \text{number photons}$$

absorption events occur individually and at random -> poisson distribution

$$p(k\,|\,\lambda) = \frac{1}{k!}\lambda^k e^{-\lambda}$$

Compute probability that >n photons absorbed

probability n or more photons




log average number of photons

Hecht, Shlaer, and Pirenne, 1942

Tinsley et al. 2016 (Nature Comms)

# How good is the visual system?

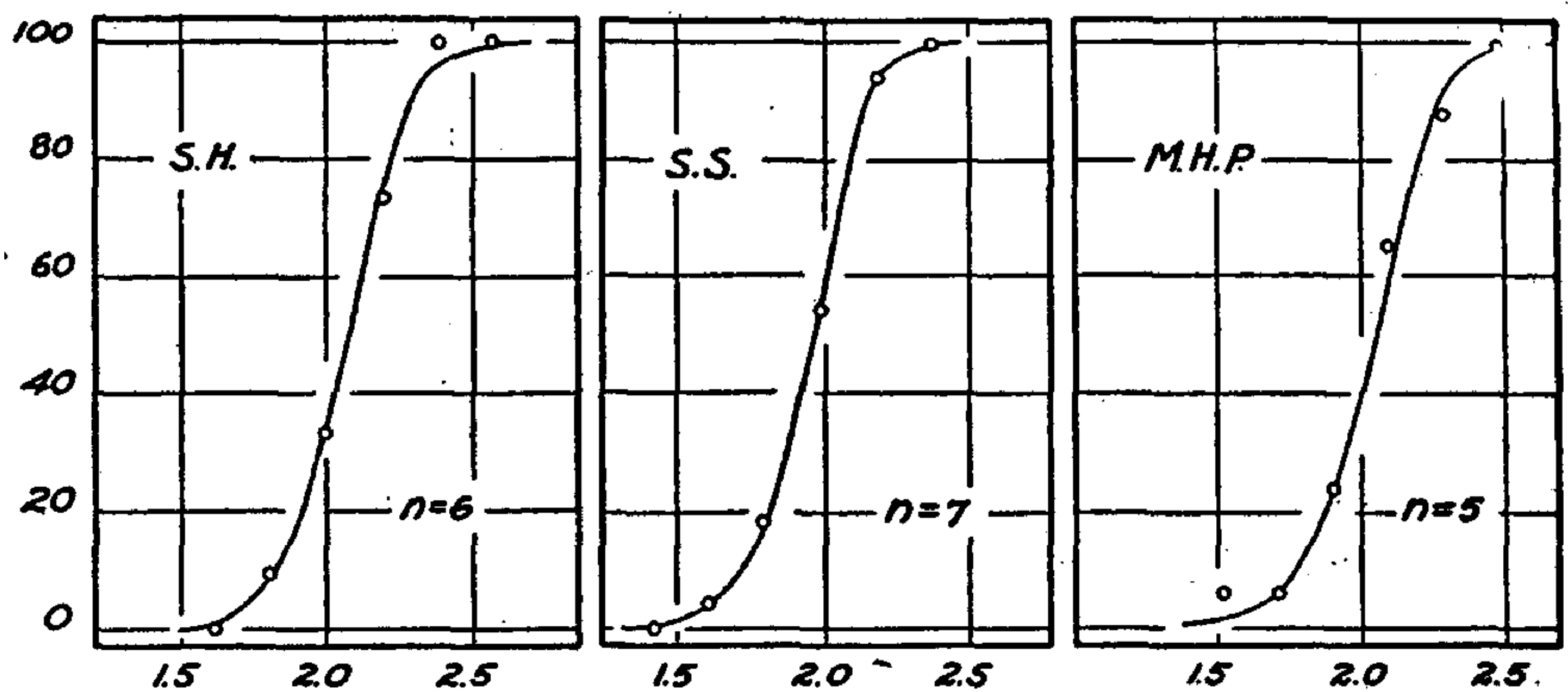
Measure how often people detect flash

$$p(\text{detected} \mid \lambda)$$

-> People need to n or more photons to see flash

subject one subject two subject three

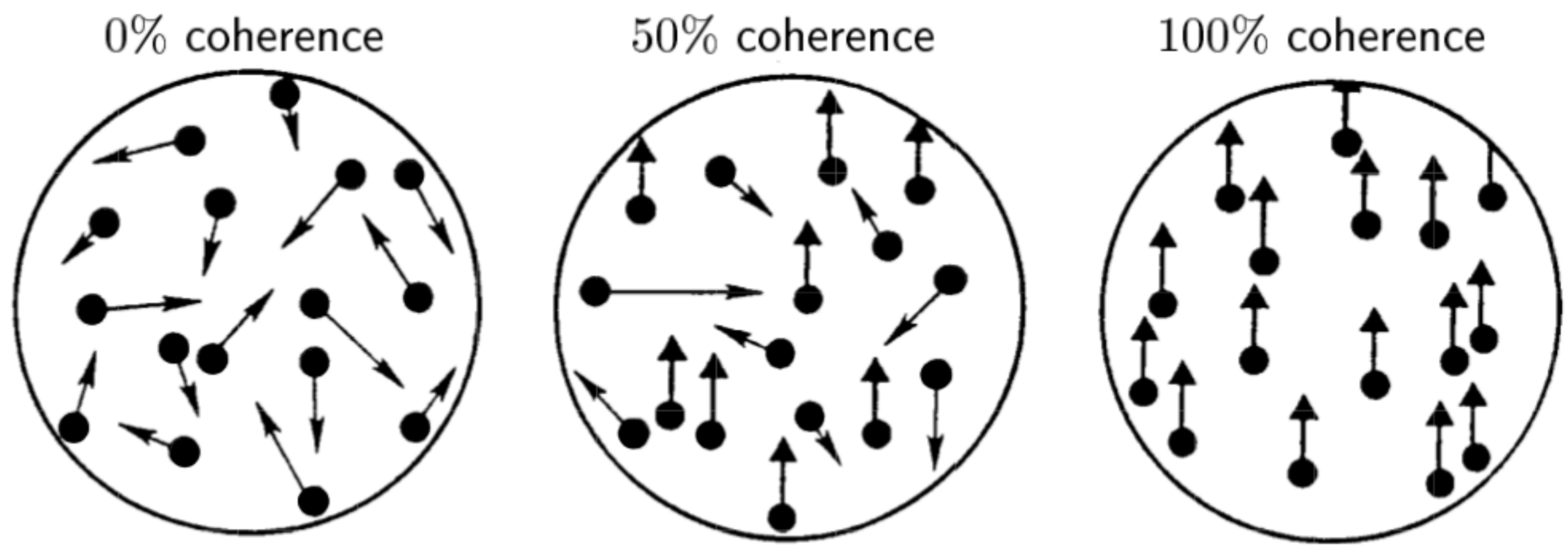
percent detected



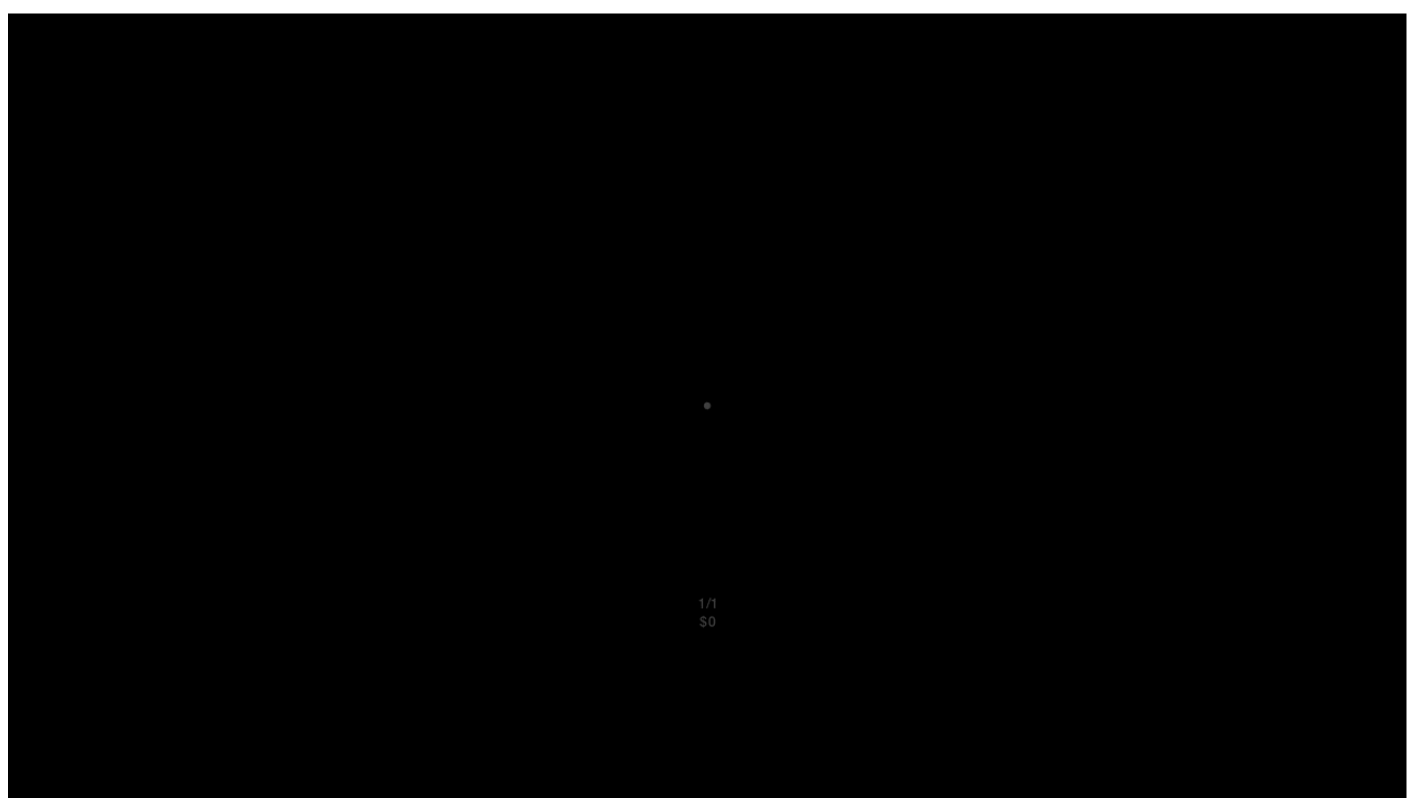

log average number of photons

Britten, Shadlen, Newsome, 1992

# Discriminating with single neurons

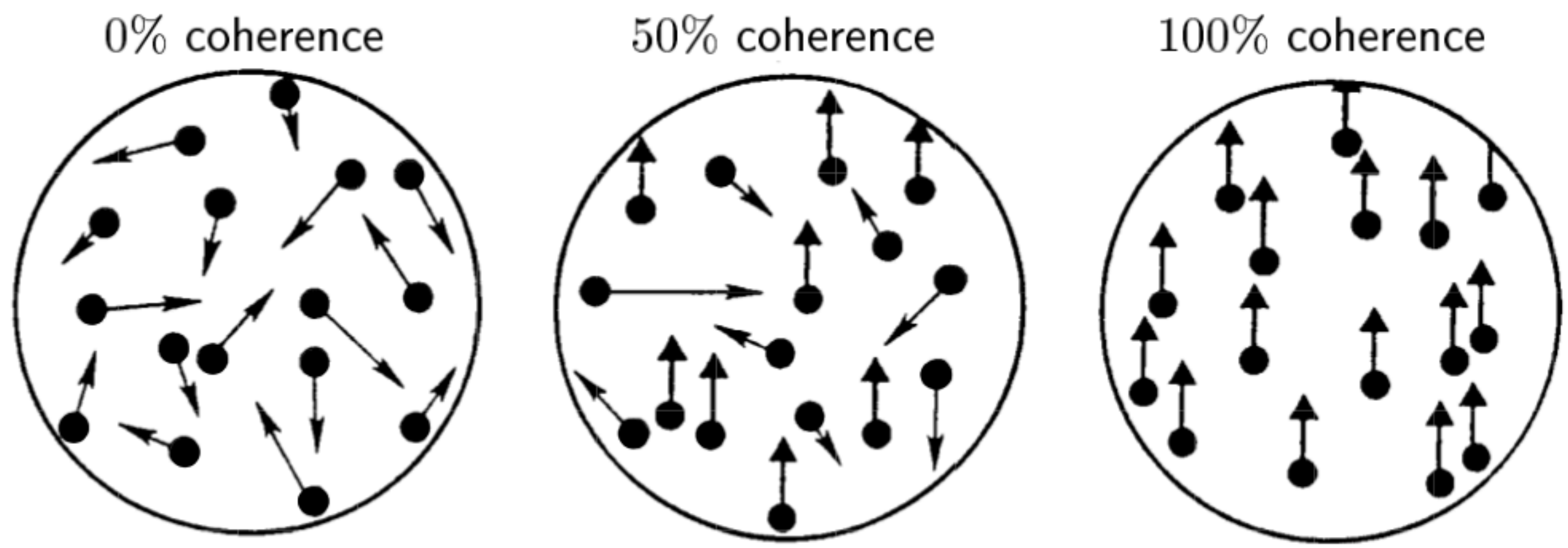



80%
coherence

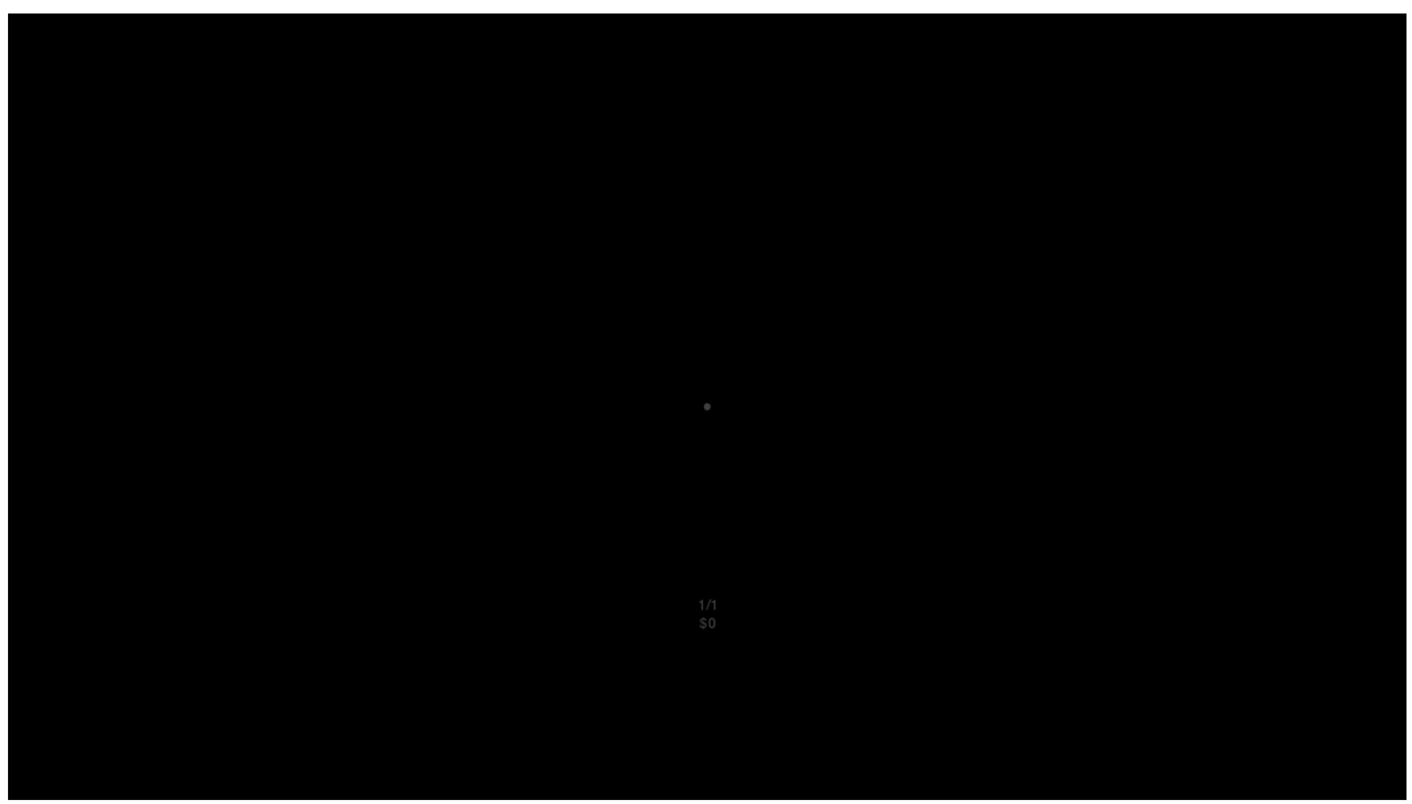


Britten, Shadlen, Newsome, 1992

# Discriminating with single neurons

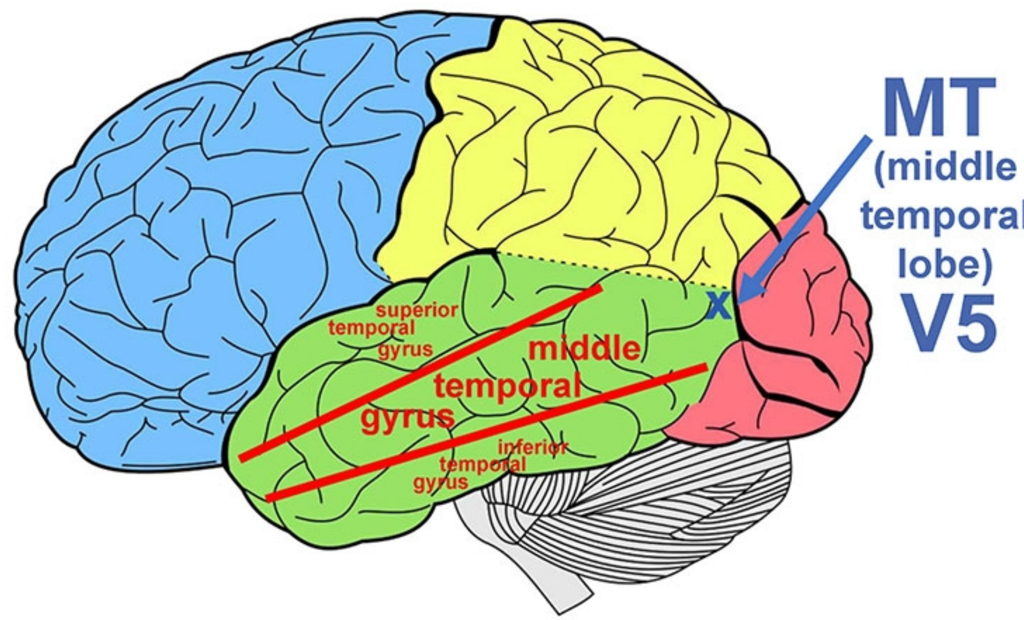



20% coherence



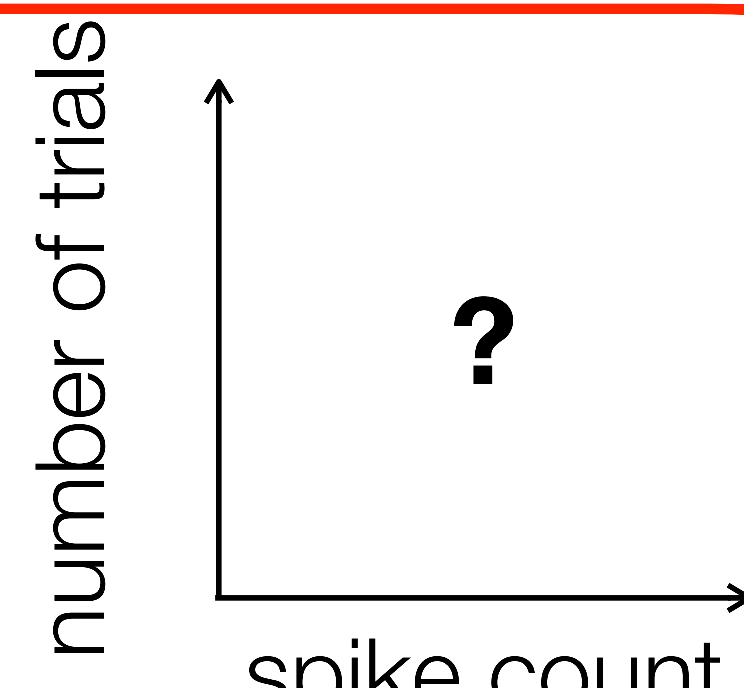
# Discriminating with single neurons

**Visual task**: are dots moving?
**Measure**: number of spikes fired by neurons in visual area MT in both conditions




**TD:** plot distribution of spikes fired by neurons in each condition

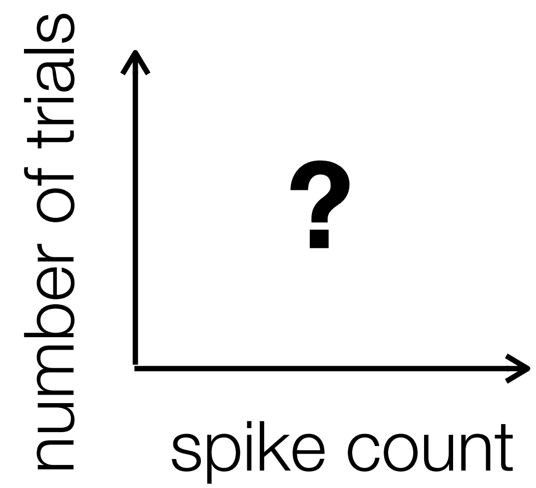
(use ‘histcounts’)



# Challenge

**TD:** plot distribution of spikes fired by neurons in each condition

(use ‘histcounts’)




*r*: spike counts (n. trials X n. stim)

*r0*: spike counts when no motion present (n. trials X 1)

```
%% plot histograms

x = 0:50;                    % different spike counts for histogram
edges =  [x-0.5,x(end)+1]; % bin edges for histogram

%%%%                         % TODO generate histogram of spike counts (0% coherence)

figure('Name', 'firing rate histograms')
for i = 1:nstim
    %%%%                       % TODO generate histogram of spike counts

    subplot(nstim, 1, i)
    bar(x, [n0; n])
    ylabel('trials')
    title(sprintf('coherence = %.1f %%', coherence(i)*100));

end
xlabel('spike count')
```

# Discriminating with single neurons

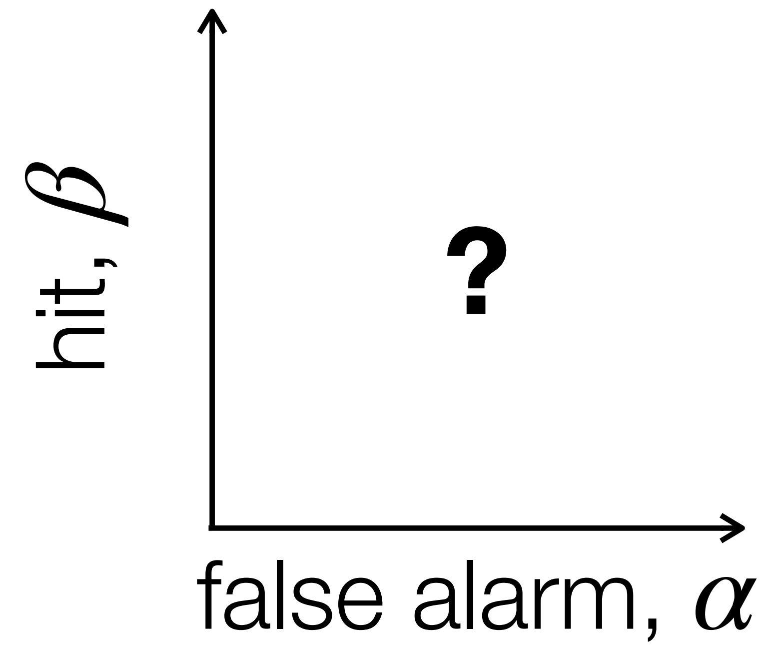
- Test if the response, is greater than a threshold, *z*
- If stimulus is present *x*=1; if stimulus not present, *x=0*
- False alarm: $\alpha(z) = P[r \geq z \,|\, x = 0]$
- Hit: $\beta(z) = P[r \geq z \,|\, x = 1]$

| | probability | |
|---|---|---|
| stimulus | correct | incorrect |
| $x = 1$ | $\beta$ | $1 - \beta$ |
| $x = 0$ | $1 - \alpha$ | $\alpha$ |

# Receiver operator characteristic (ROC) curve

**TD:** plot hit rate ($\beta$) versus the false alarm rate ($\alpha$)

**This is called the receiver operator characteristic (ROC) curve**



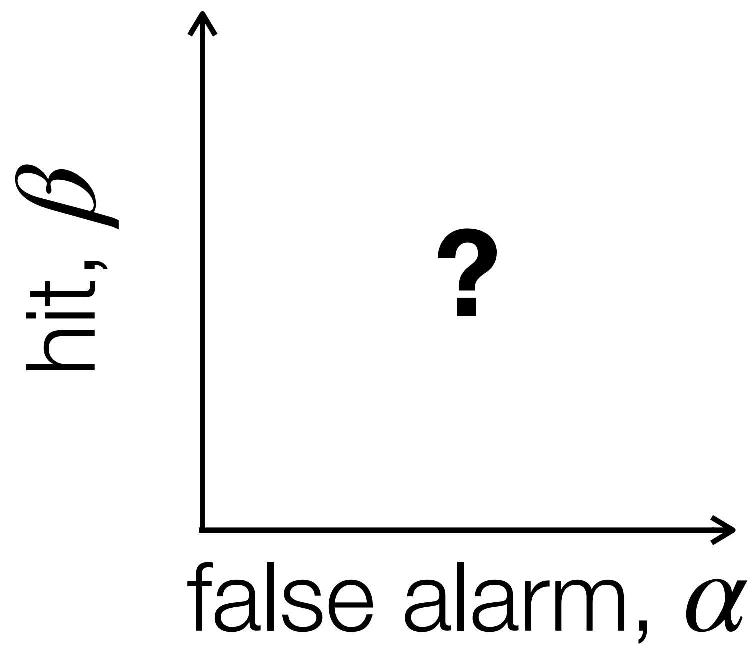

How does the hit rate vary with false alarm as we vary threshold? Is this expected?

How does ROC curve vary as we increase vary the stimulus coherence? Is this expected?

# Challenge

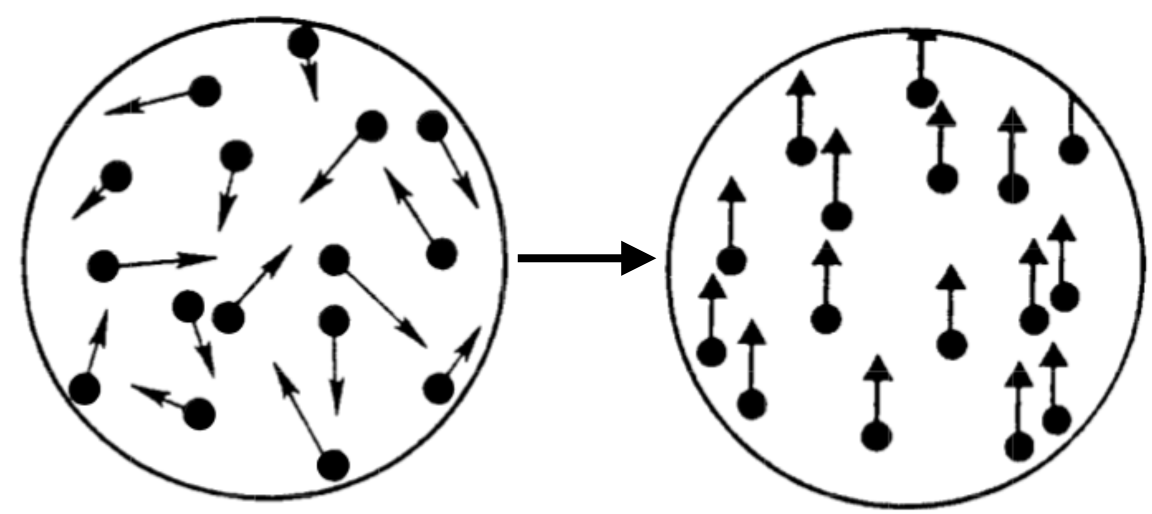
- False alarm: $\alpha(z) = P[r \geq z \,|\, x = 0]$
- Hit: $\beta(z) = P[r \geq z \,|\, x = 1]$

```
%% ROC curves
z = 50:-1:0;                                % thresholds
nz = numel(z);                              % number of thresholds

alpha = zeros(nz, 1);                       % false alarm rate
beta = zeros(nz, nstim);                    % hit rate

% loop over thresholds
for i = 1:nz
    %%%%%%%% alpha(i) =                     TODO false alarm rate
    %%%%%%%% beta(i)  =                     TODO: compute hit rate
end

% plot ROC curve
figure('Name', 'ROC curve')
%%% plot( ... , 'o-')                       % TODO plot ROC curve
plot([0 1], [0 1], 'k--'); hold off
xlabel('alpha')
ylabel('beta')
```



# Two alternative forced choice

Two alternative forced choice (2AFC): present two stimuli in random order. Which was first?



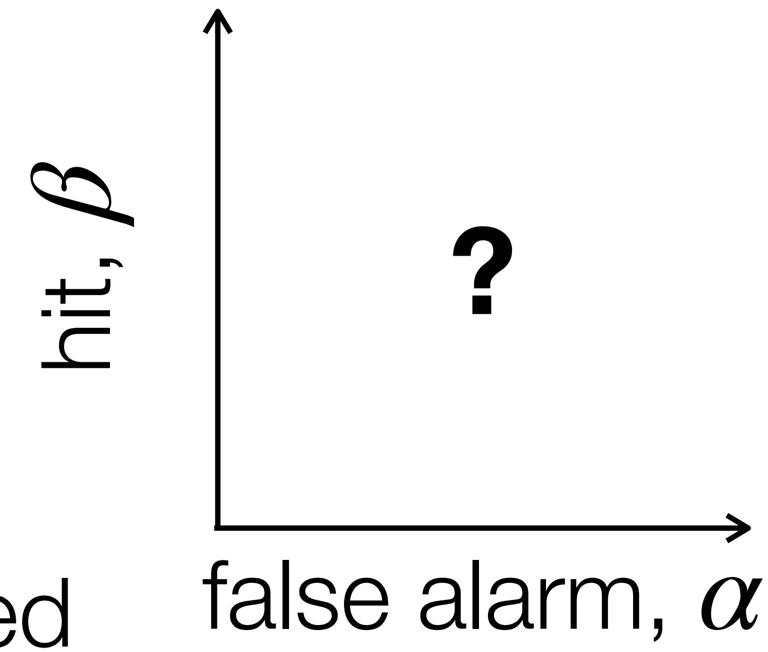
- Readout by comparing and $r_1$, $r_2$

# Two alternative forced choice

- Hit rate: $P[r_2 \geq z \,|\, x_2 = 1] = \beta(z)$, with $z = r_1$
- Probability correct: $P[\text{correct}] = \int_0^\infty \beta(z) p(z \,|\, x_1 = 0) dz$
- False alarm: $\alpha(z) = P[r_2 \geq z \,|\, x_2 = 0] = \int_z^\infty p(r_2 \,|\, x_2 = 0)$
- Taking derivative: $\frac{d\alpha}{dz} = -p[z \,|\, x_2 = 0]$
- Thus: $d\alpha = -p[z \,|\, x_2 = 0] dz$
- Substituting in: $P[\text{correct}] = \int_0^1 \beta d\alpha$

# Two alternative forced choice

$$P[\text{correct}] = \int_0^1 \beta d\alpha$$




**TD:** Use the ROC curve you plotted earlier to estimate $P[\text{correct}]$

**TD:** Estimate $P[\text{correct}]$ directly, by comparing $r_1$ and $r_2$. How does this compare with result estimated from ROC curve?

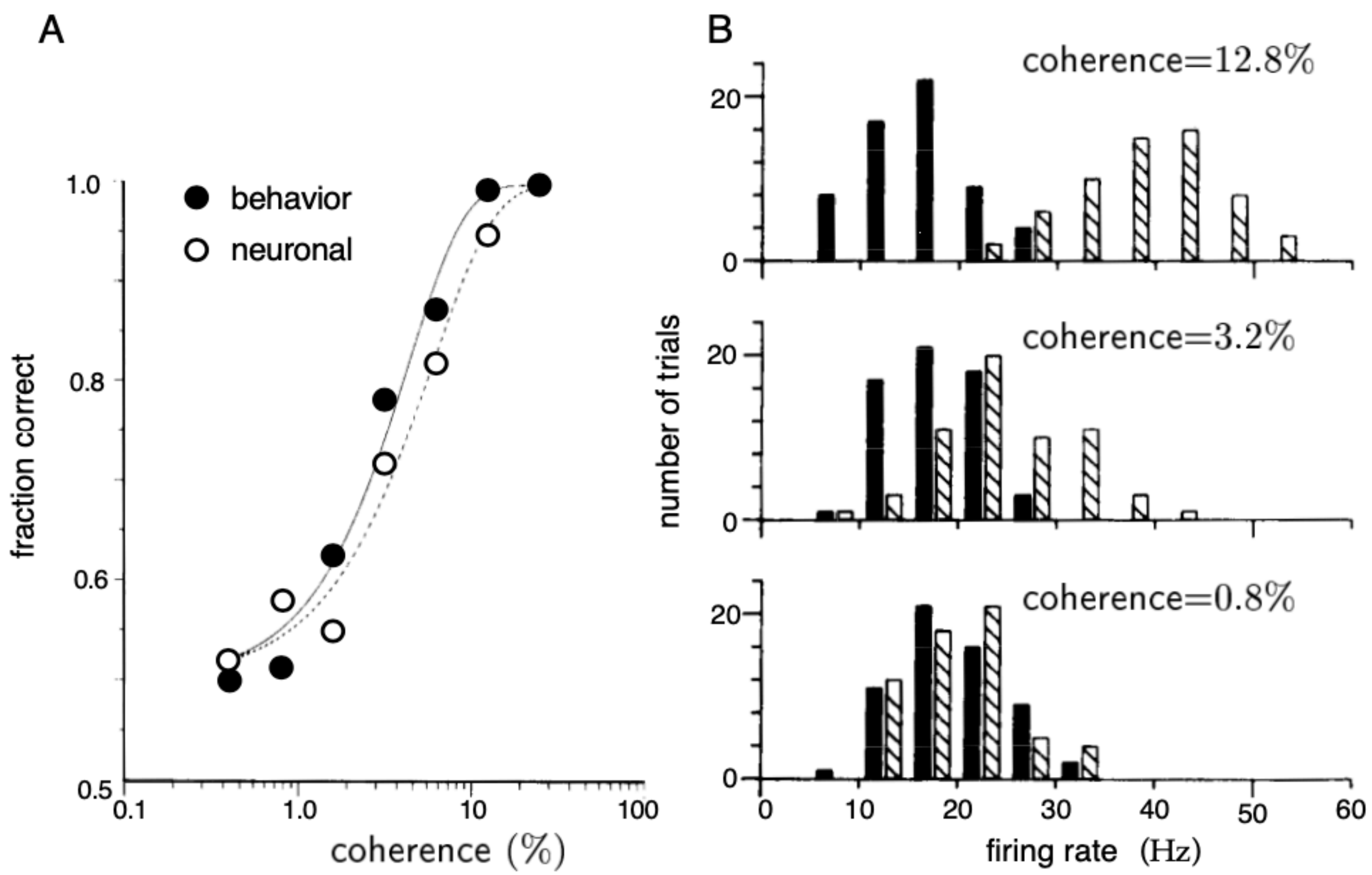
# Challenge

Area under curve (AUC): $P[\text{correct}] = \int_0^1 \beta d\alpha$

```
%%%%%%%%%%                              TODO: compute area under curve
% AUC =

%%%                                     TODO: compute error in 2AFC
% p2AFC =

figure('Name', 'neuronal')
% semilogx(100*coherence, AUC, '-o'); hold on
% semilogx(100*coherence, p2AFC, 'r-o'); hold on
xlabel('log coherence')
ylabel('area under curve')
leg = legend('area under curve', 'probability correct');
set(leg, 'Location', 'SouthEast', 'Box', 'off', 'Fontsize', 12)
```

Britten, Shadlen, Newsome, 1992

# Comparison with a real experiment




Performance of one neuron is similar to monkey!

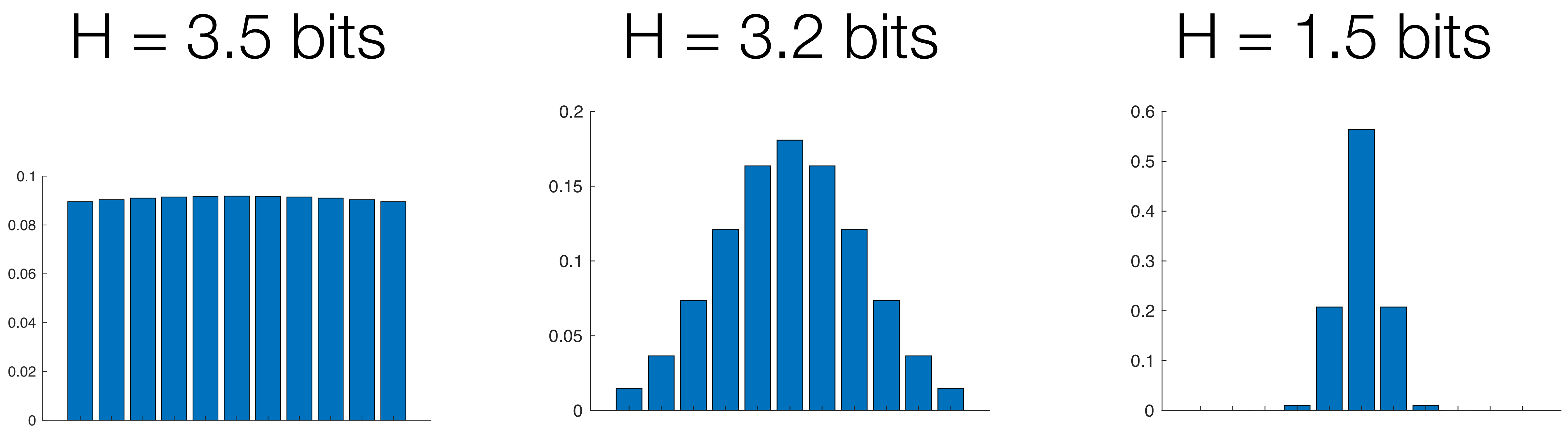
Is this surprising?

# Introduction to information theory

- Can we quantify how much ‘information’ neuron encodes about stimuli?
- First, define measure of surprise: $h(p(x)) = -\log p(x)$
  - Less probable events (small $p(x)$) are more surprising.
  - The surprise from two independent events adds:
    $h(p(x_2)p(x_2)) = -\log(p(x_1)p(x_2)) = -\log p(x_1) - \log p(x_2)$
- Define entropy average surprise: $H = -\sum p(x) \log p(x)$
- If entropy is very large, lots of uncertain events.

# Introduction to information theory

- Entropy ~ uncertainty




- **TD:** In the experiment we studied the stimulus binary (x=0/1). Can you plot how entropy varies as we alter $p(x)$?

# Challenge

- **TD:** In the experiment we studied the stimulus binary (x=0/1). Can you plot how entropy varies as we alter $p(x)$?

- Tip: $H(X) = -\sum_{x} p(x) \log p(x)$

```
% compute probability (that x=1) and entropy

figure('Name', 'entropy of a binary stimulus')
% plot(p, H)
xlabel('p')
ylabel('H')
```

# Introduction to information theory

- Entropy: $$H(X) = -\sum_{x} p(x)\log p(x)$$
  - How uncertain are we about stimulus?
- Conditional entropy: $$H(X|R) = -\sum_{x,r} p(x,r)\log p(x|r)$$
  - How uncertain are we about stimulus, after measuring $r$?
- Mutual information: $I(X;R) = H(X) - H(X|R)$
  - Reduction in uncertainty about $X$ when we measure $R$
  - Commutative: $I(X;R) = H(R) - H(R|X)$

# Introduction to information theory

- **TD:** measure how much information neurons encode about motion coherence?
- How does this change if we change (choose 1)
  - (a) minimum/maximum firing rate
  - (b) stimulus distribution (make some coherences more likely than others)

- **Aids:**

$$p_{poisson}(r \mid \lambda) = \frac{1}{k!}\lambda^k e^{-\lambda}$$

$$\log p_{poisson}(r \mid \lambda) = \log \lambda - \lambda - \log \Gamma(k+1)$$

$$H(R) = -\sum_r p(r) \log p(r)$$

$$I(R;X) = H(R) - H(R \mid X)$$

$$H(R \mid X) = -\sum_r p(r,x) \log p(r \mid x)$$

# Efficient coding hypothesis

**Horace Barlow (1952):**

Neurons have evolved to encode maximal information about natural stimuli, given limited resources

Laughlin et al. (1981) measured response of blowfly LMC neurons versus image contrast

What does efficient coding predict?

# Efficient coding hypothesis

Mutual information: $I(X;R) = H(R) - \cancel{H(R|X)}$

low noise limit

Assume constraint on maximum firing rate: $r \leq r_{max}$

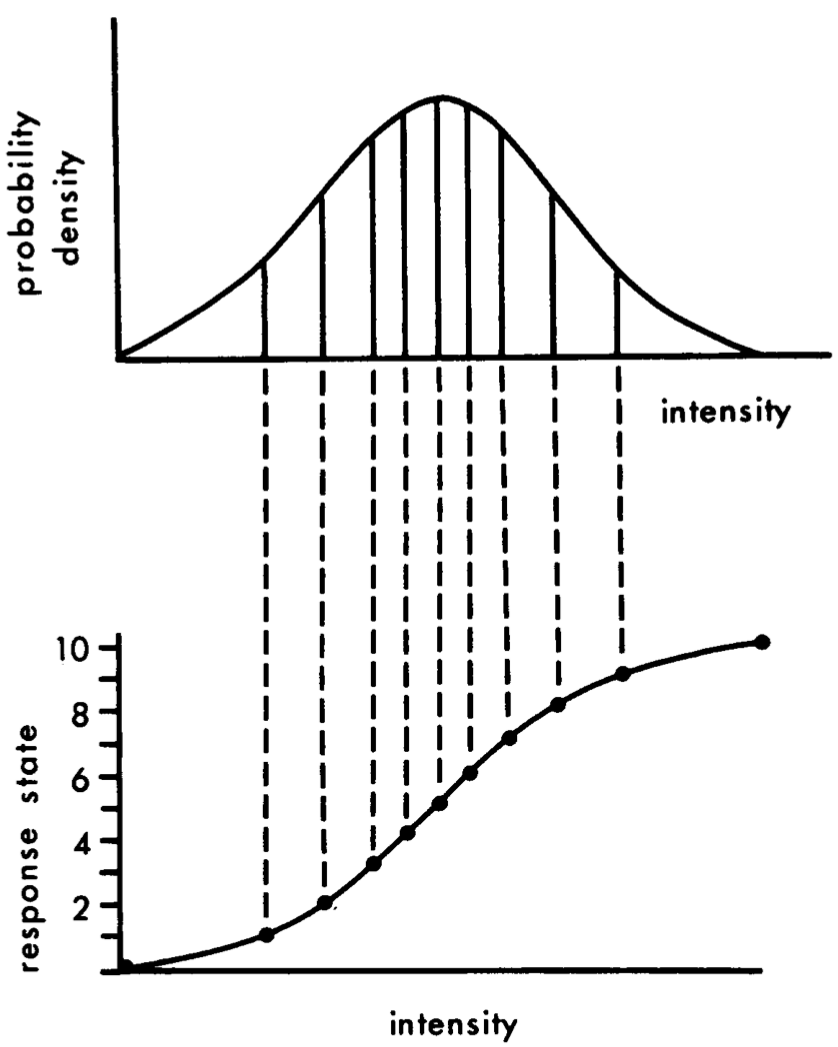
Maximise: $-\int_0^{r_{max}} p(r) \log p(r) dr$ Subject to: $\int_0^{r_{max}} p(r) dr = 1$

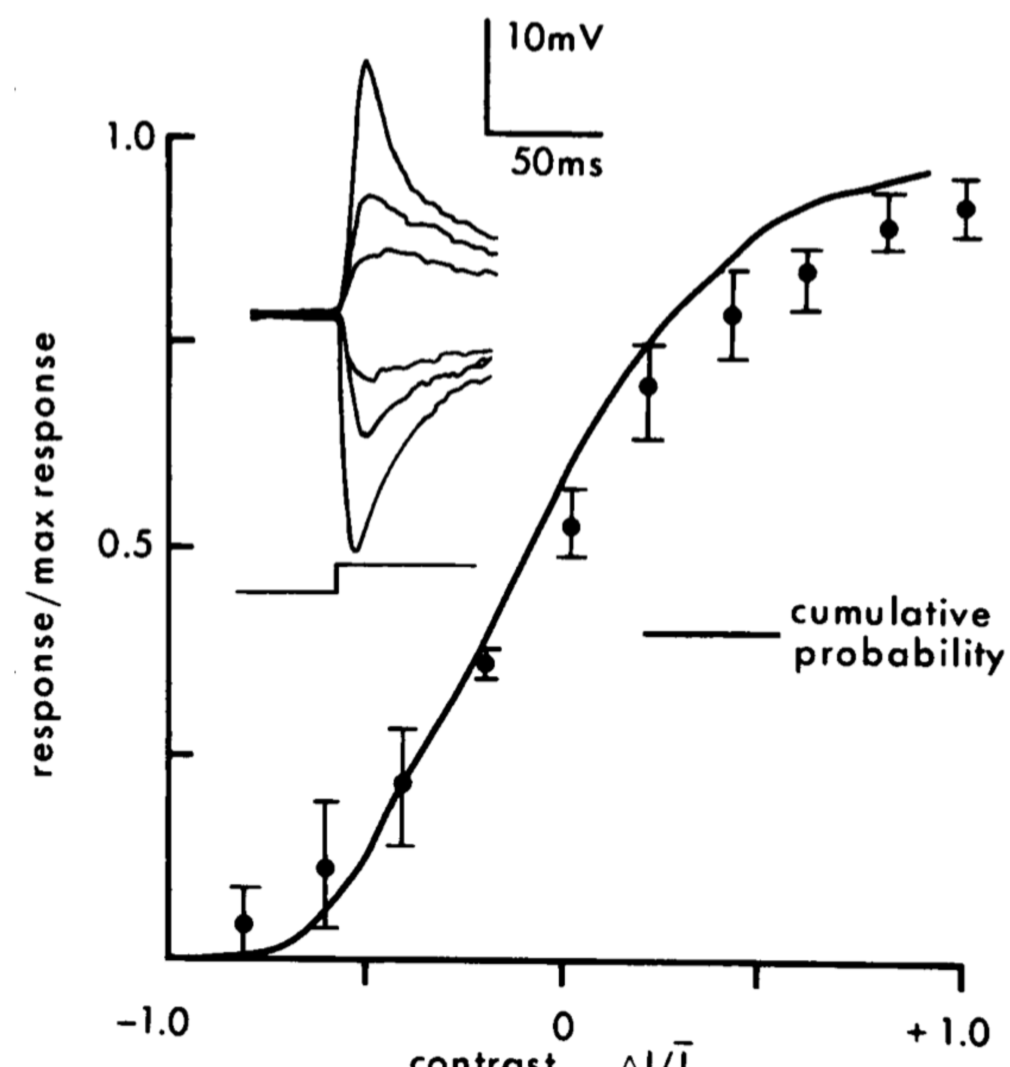
Solution: $p(r) = \dfrac{1}{r_{max}}$ (Histogram equalization)

Suppose: $r = f(x)$, Optimal solution: $f(x) = \int_{x_{min}}^{x} p(x') dx'$

# Efficient coding hypothesis

Suppose: $r = f(x)$, Optimal solution: $f(x) = \int_{x_{min}}^{x} p(x')dx'$



# Decoding stimuli from neural responses

stimulus $\qquad$ response $\qquad$ estimate

$$x \xrightarrow[\text{encoding}]{} r \xrightarrow[\text{decoding}]{} \hat{x}$$

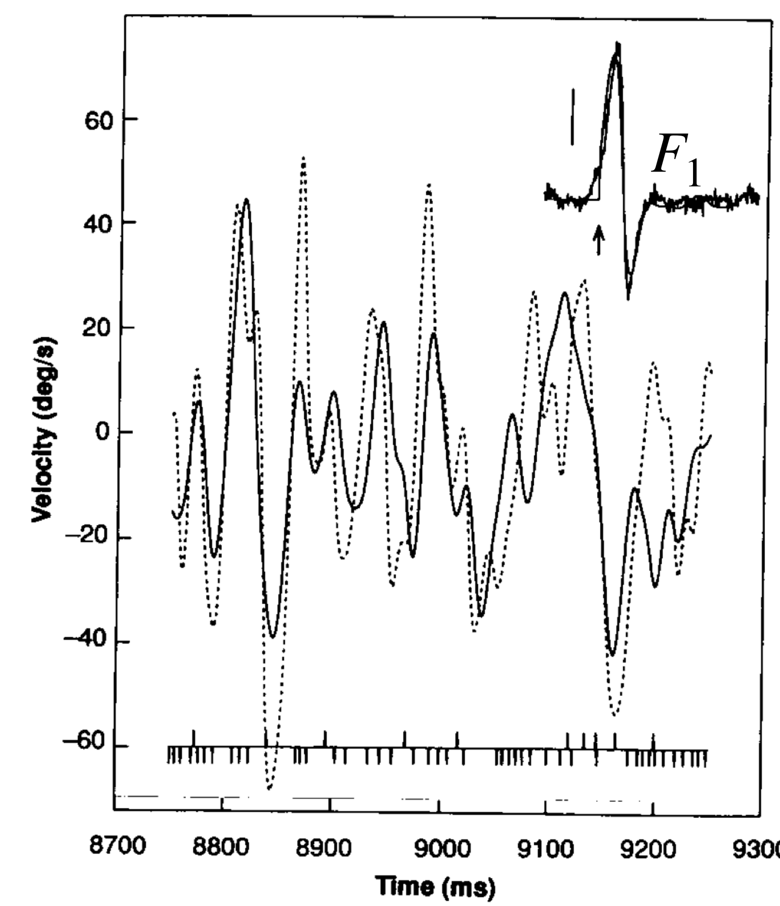
**Bialek et al (Science 1991)**: measured responses of motion sensitive H1 neuron in fly visual system to moving visual pattern

**Decoded motion:**

$$\hat{x}(t) = \sum_i F_1(t - t_i) + \cancel{\sum_{i,j} F_2(t - t_i, t - t_j)} + \ldots$$

**Estimate $F_1$ to minimise:**

$$\chi = \int (x(t) - \hat{x}(t))^2 dt$$



# Decoding stimuli from neural responses

- How good are reconstructions?

- Decomposition of reconstruction into deterministic and random component:

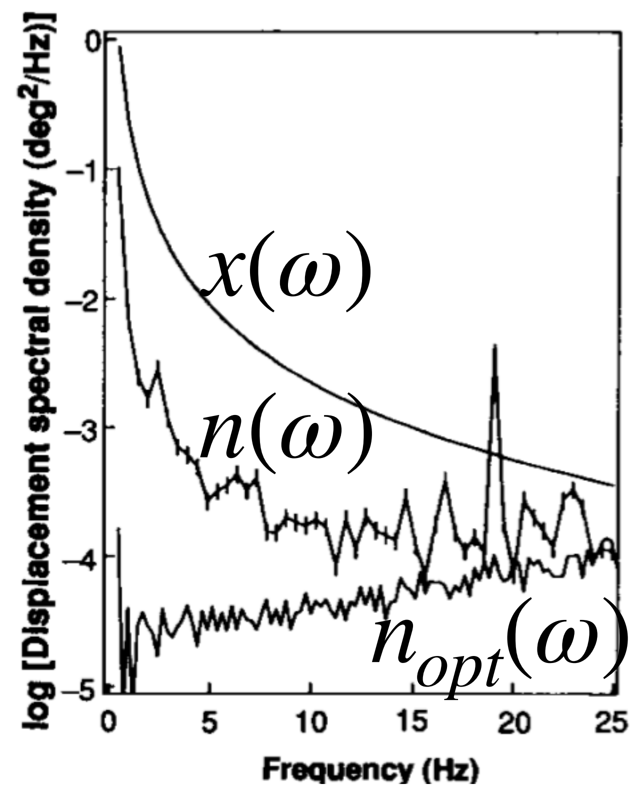
$$\hat{x}(\omega) = g(\omega)(x(\omega) + n(\omega))$$

- Can estimate mutual information:

$$R = \frac{1}{2\pi} \int \log \left( 1 + \frac{\langle x(\omega)^2 \rangle}{\langle n(\omega)^2 \rangle} \right) d\omega \approx 64 \pm 1 \text{ bit}$$

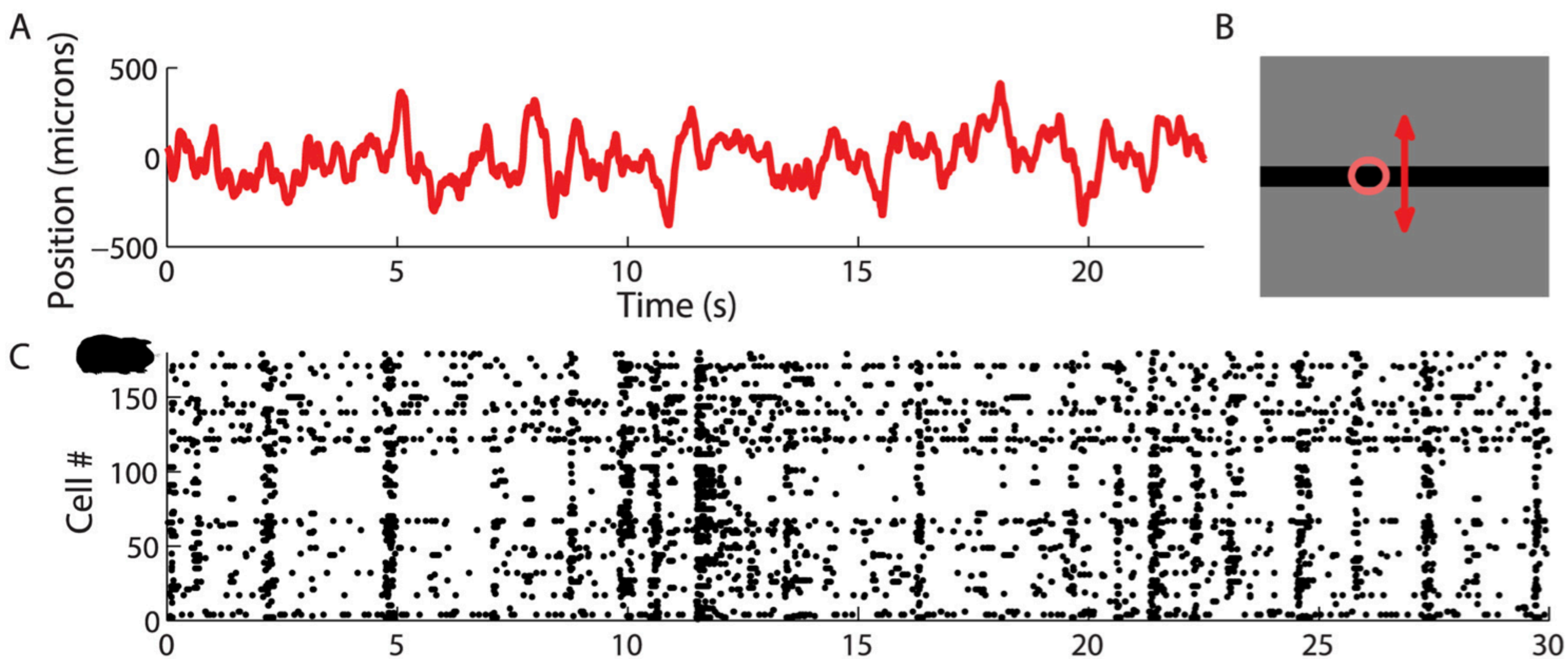
- Compare noise to optimal limit (set by a model of photoreceptors)

  - for frequencies >10 Hz, close to optimal



# Decoding from neural populations

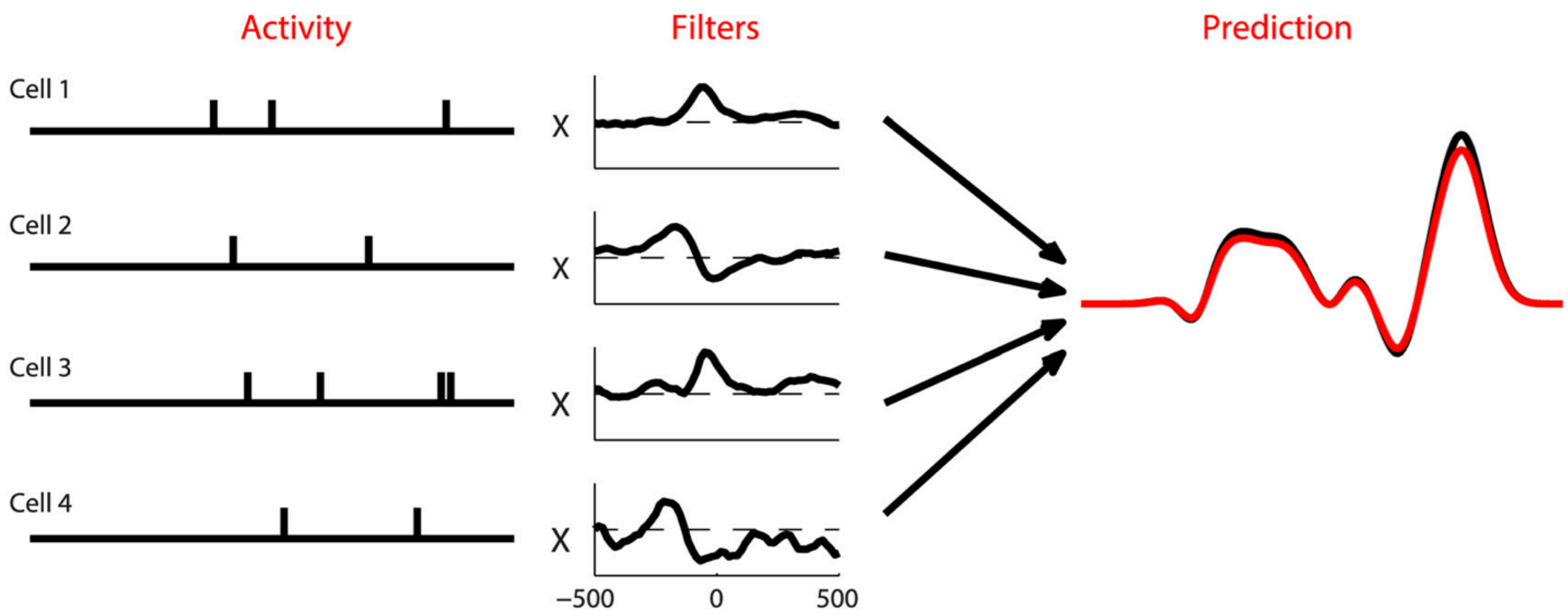
- Marre et al.: recorded response of population of ~200 retinal ganglion cells to moving visual bar



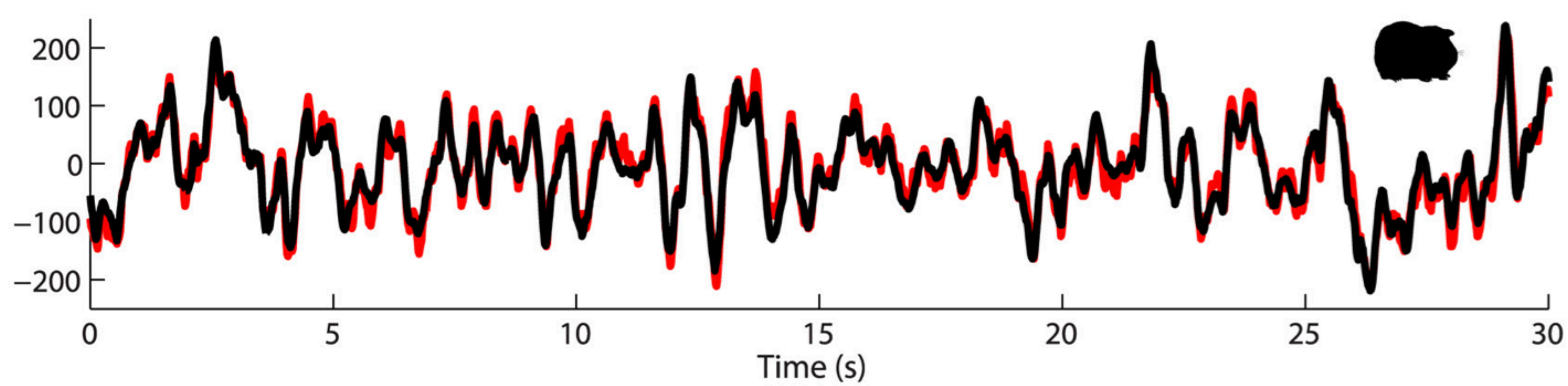

time (s)

# Decoding from neural populations

- Estimate linear filters for each neuron:

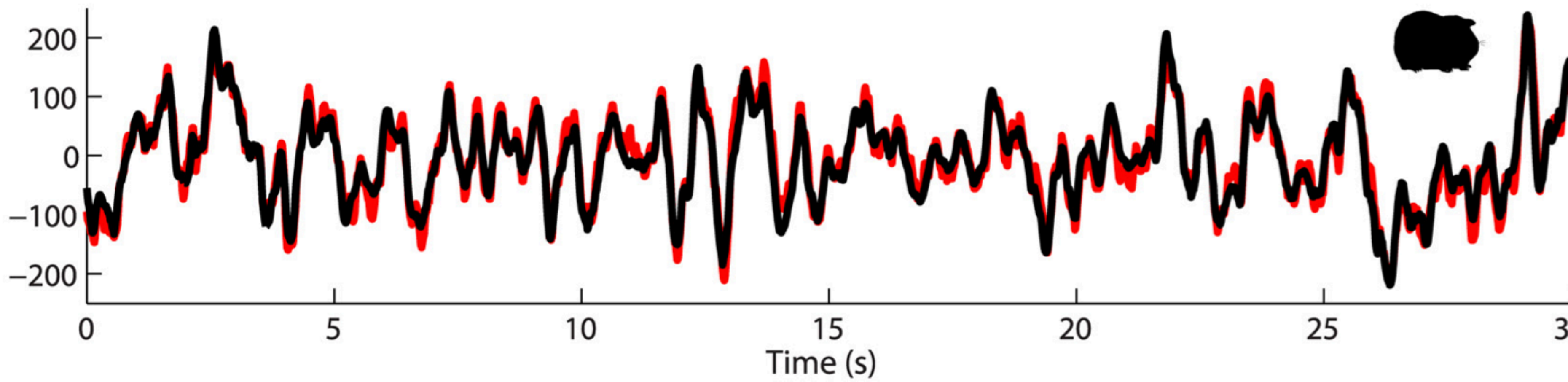



- Visual reconstruction:

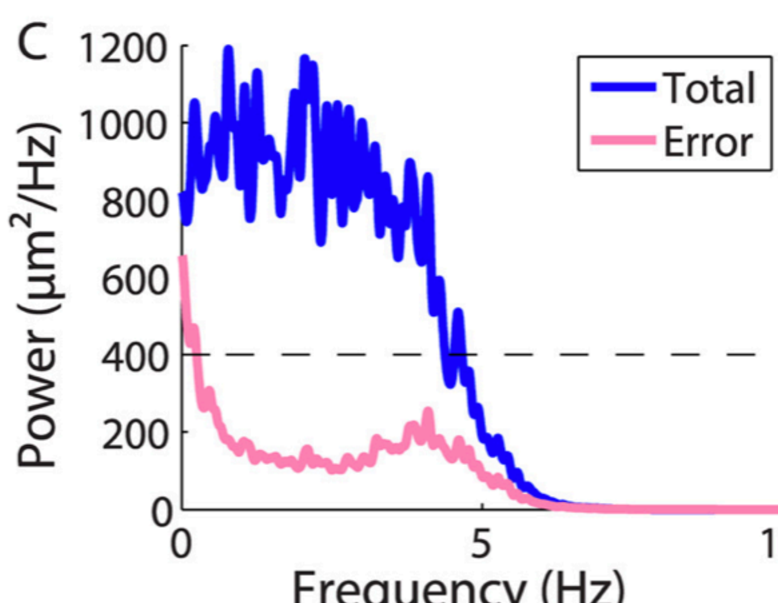


# Decoding from neural populations

- Visual reconstruction:




- Prediction error is in hyperacuity range for salamander (comparable to distance between photoreceptors).




spacing between photoreceptors

# Summary

- Detection of low light levels
- Discriminating motion from single neurons (signal detection theory)
- Introduction to information theory
- Efficient coding
- Decoding continuous stimuli from neural responses