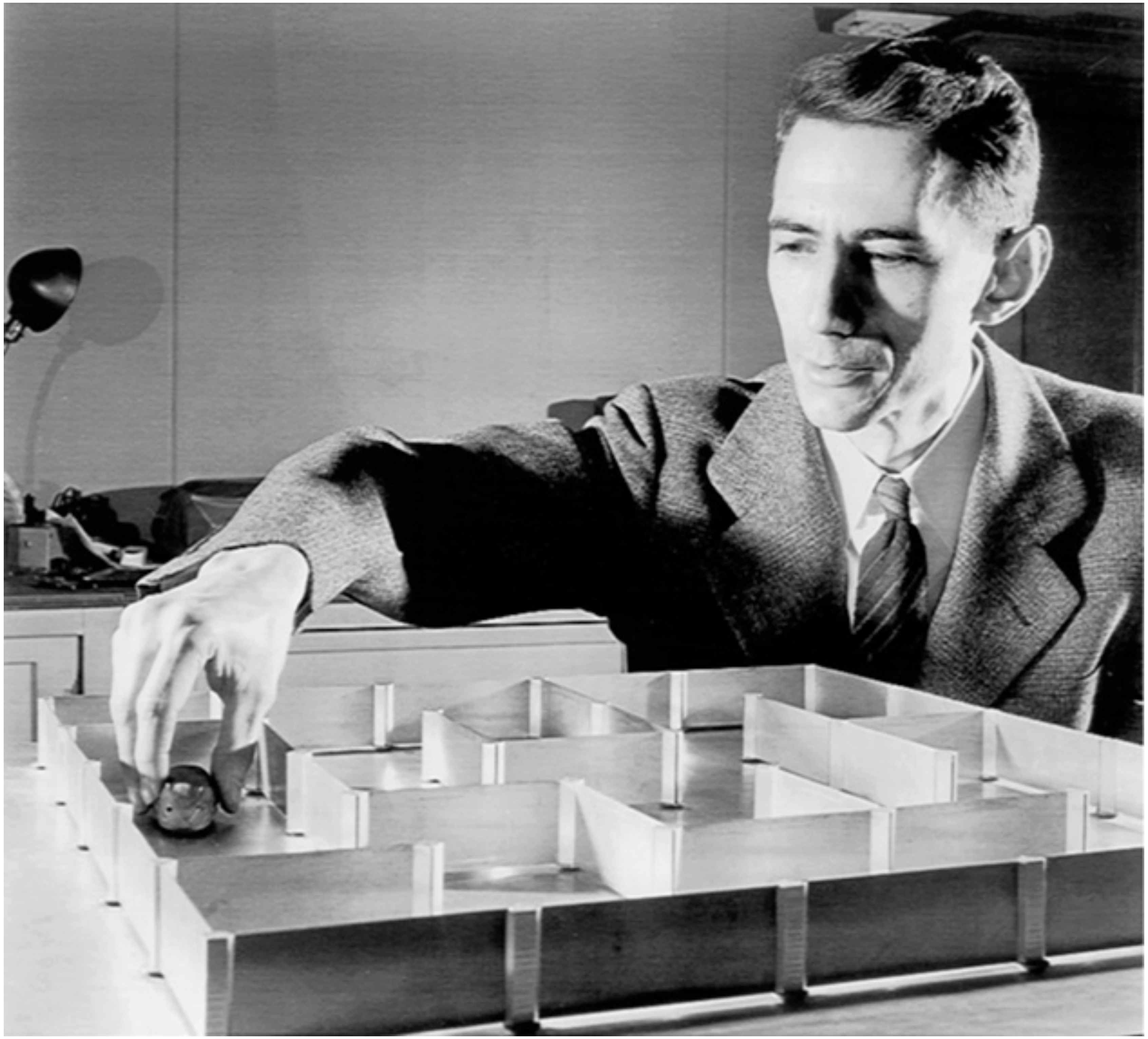
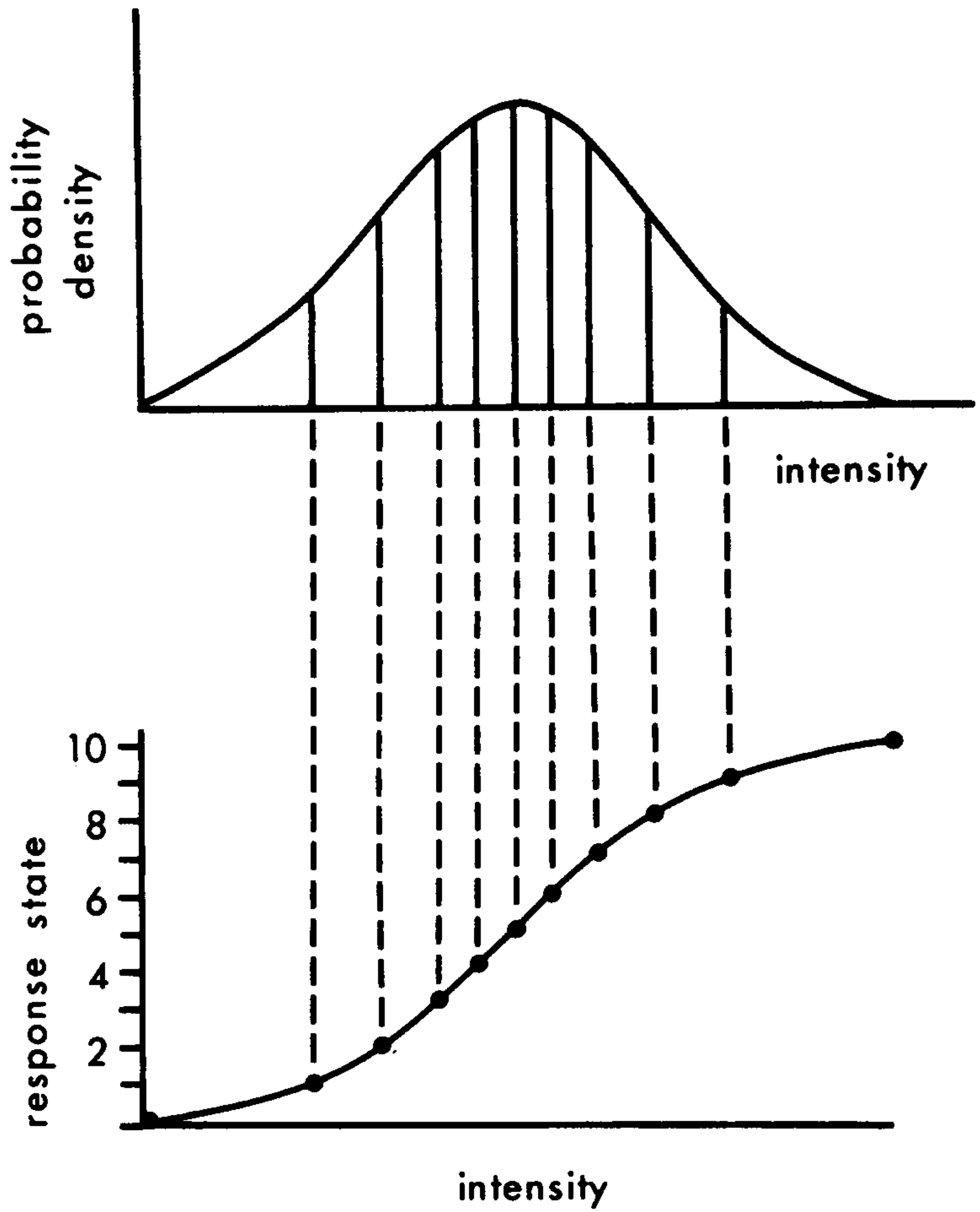
Lecture 8: *Info theory and efficient coding*



Maximum entropy and efficient coding in single neurons



The efficient coding hypothesis, brief history:

- Claude Shannon (1948) *A Mathematical Theory of Communication*
- Fred Attneave (1954) *Some informational aspects of visual perception*
- Horace Barlow (1961) *Possible principles underlying the transformation of sensory messages*

Are sensory systems optimized for information transmission?

Recall: info theory basics

- Stimulus s drawn from $P(s)$;
- $\Rightarrow$ neural response y, $P(y|s)$
- Mutual information:

$$
\begin{aligned}
I &= \int dsdyP(s,y)\log_2\left(\frac{P(s,y)}{P(s)P(y)}\right) \\
&= \int dsP(s)\int dyP(y|s)\log_2\left(\frac{P(y|s)}{P(y)}\right) \\
&= -\int dyP(y)\log_2 P(y) + \int dsP(s)\int dyP(y|s)\log_2 P(y|s)
\end{aligned}
$$

- 1 bit of information reduces the uncertainty about the stimulus by a factor 2.
- n bits of information reduce the uncertainty about the stimulus by a factor 2^n
- Linear Gaussian channel $y = s + z$ where s is Gaussian with variance S^2, z is a Gaussian noise with variance N^2:

$$
I = \frac{1}{2}\log_2\left(1 + \frac{S^2}{N^2}\right)
$$

Recall: info theory basics

- Stimulus s drawn from $P(s)$;
- $\Rightarrow$ neural response y, $P(y|s)$
- Mutual information:

$$
\begin{aligned}
I &= \int dsdyP(s,y)\log_2\left(\frac{P(s,y)}{P(s)P(y)}\right) \\
&= \int dsP(s)\int dyP(y|s)\log_2\left(\frac{P(y|s)}{P(y)}\right) \\
&= -\int dyP(y)\log_2 P(y) + \int dsP(s)\int dyP(y|s)\log_2 P(y|s)
\end{aligned}
$$

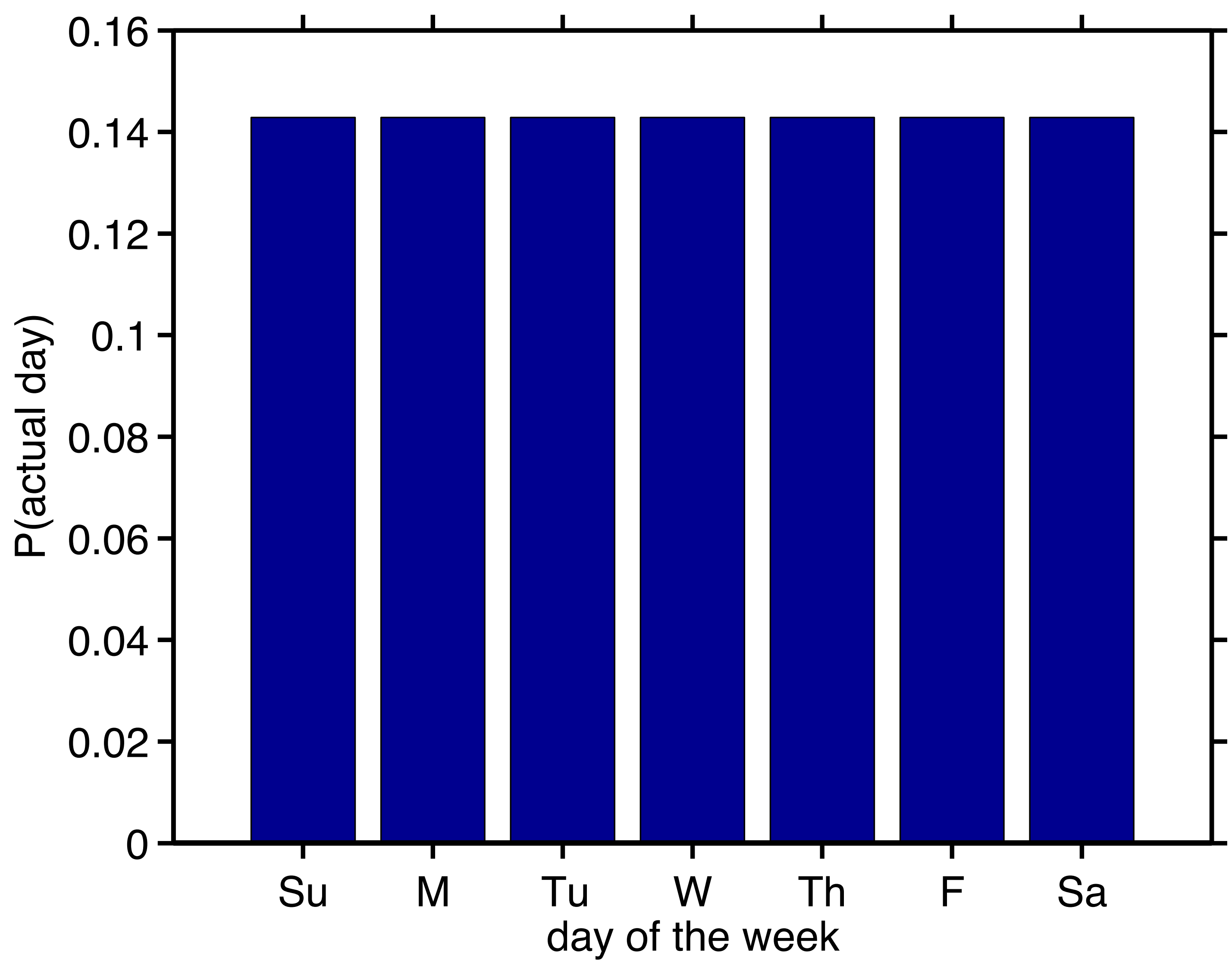
chalkboard interlude

- 1 bit of information reduces the uncertainty about the stimulus by a factor 2.
- n bits of information reduce the uncertainty about the stimulus by a factor 2^n
- Linear Gaussian channel $y = s + z$ where s is Gaussian with variance S^2, z is a Gaussian noise with variance N^2:

$$I = \frac{1}{2}\log_2\left(1 + \frac{S^2}{N^2}\right)$$

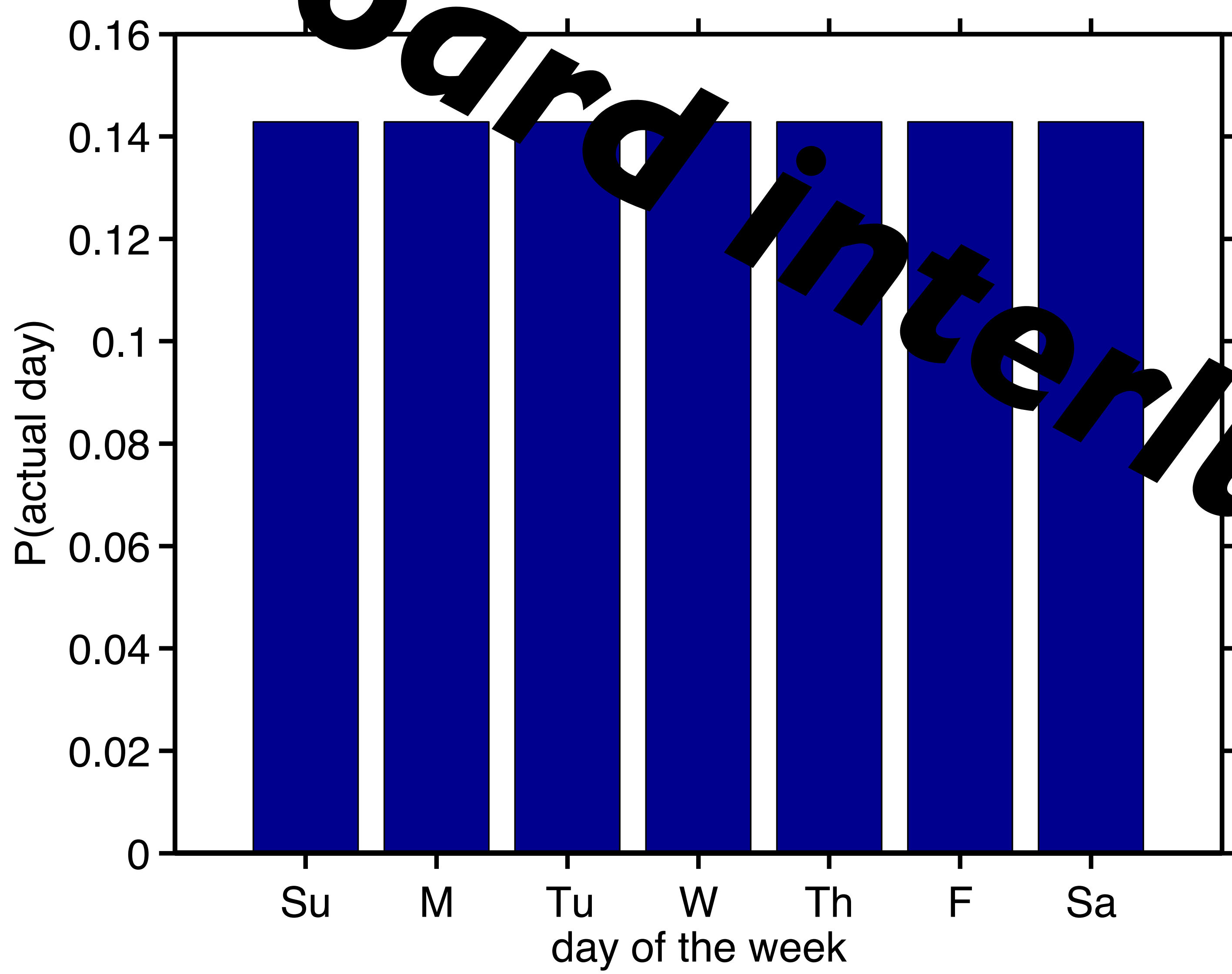
Histogram equalization:

*For discrete variables, the **uniform** distribution has the maximum entropy.*



Histogram equalization:

*For discrete variables, the **uniform** distribution has the maximum entropy.*



chalkboard interlude

Histogram equalization:

$$p[r] = \frac{1}{r_{\max}}$$

$$|f(s + \Delta s) - f(s)|/r_{\max} = p[s]\Delta s.$$

$$\frac{df}{ds} = r_{\max}\, p[s]$$

$$f(s) = r_{\max} \int_{s_{\min}}^{s} ds'\, p[s']$$

Histogram equalization:

$$p[r] = \frac{1}{r_{\max}}$$

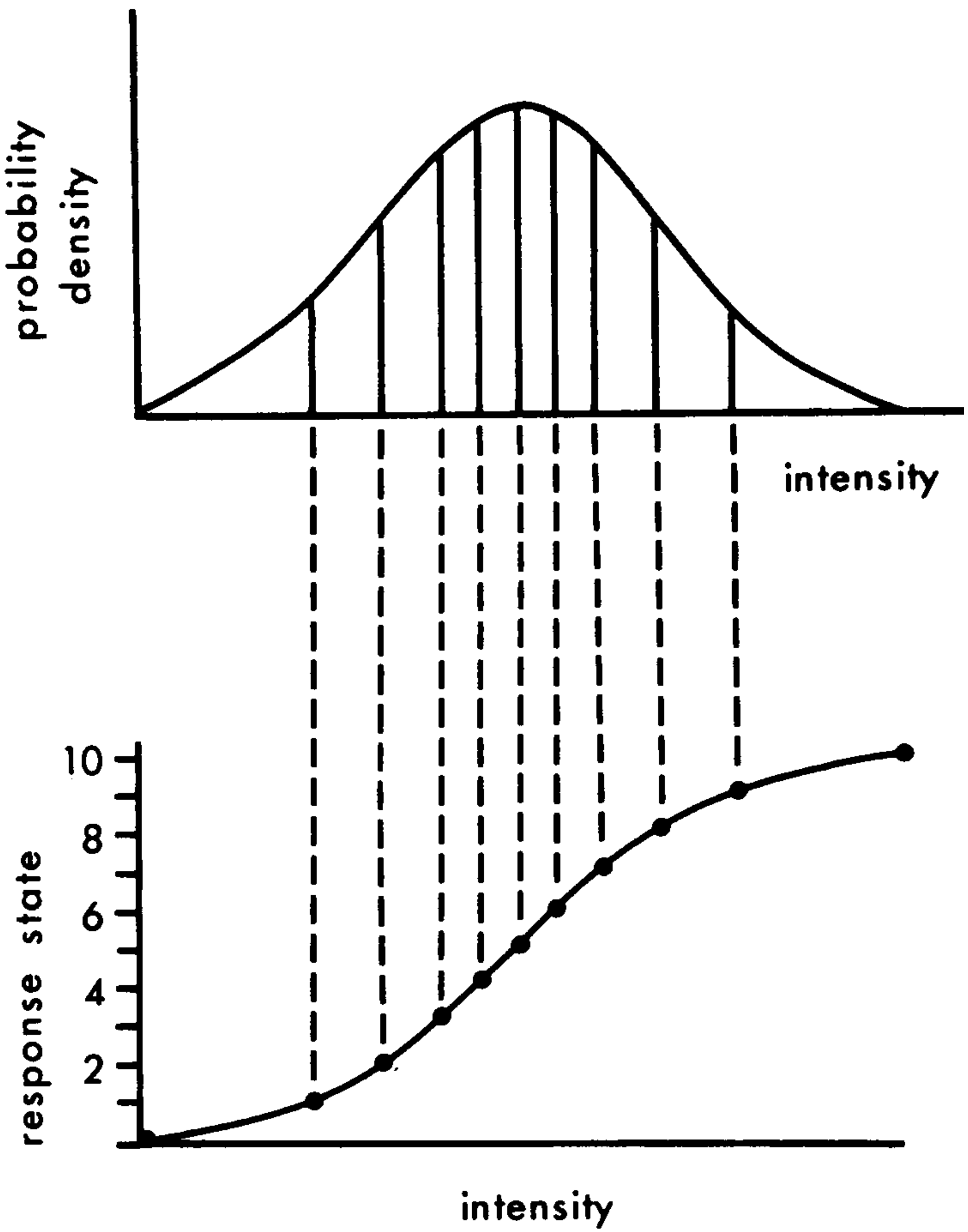
$$|f(s + \Delta s) - f(s)|/r_{\max} = p[s]\Delta s.$$

chalkboard interlude

$$\frac{df}{ds} = r_{\max} p[s]$$

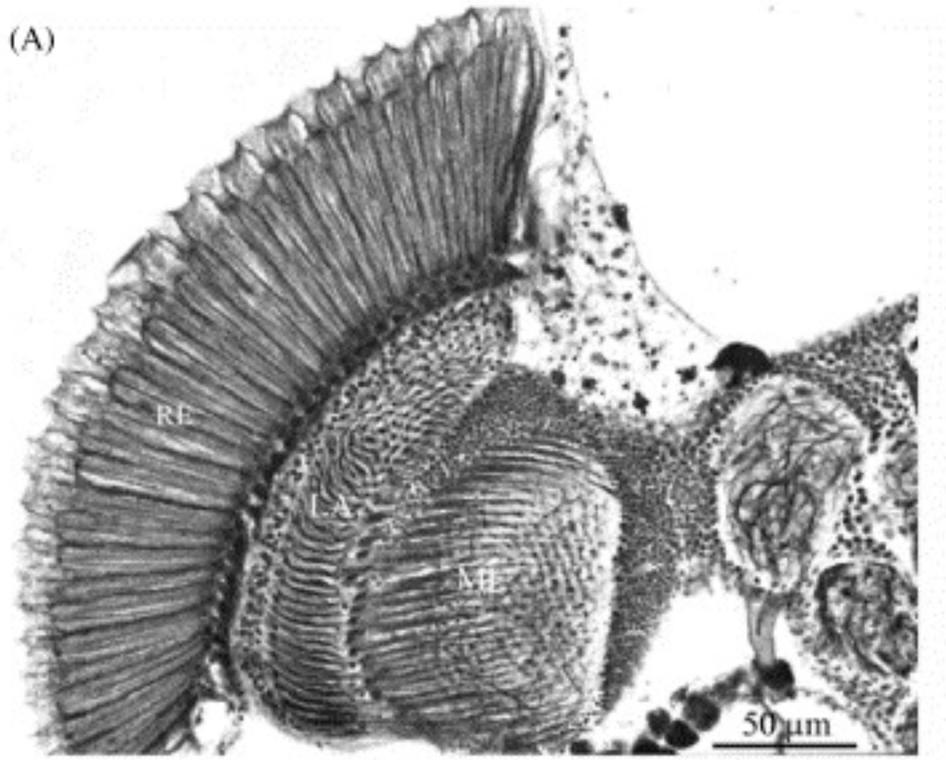
$$f(s) = r_{\max} \int_{s_{\min}}^{s} ds' \, p[s']$$

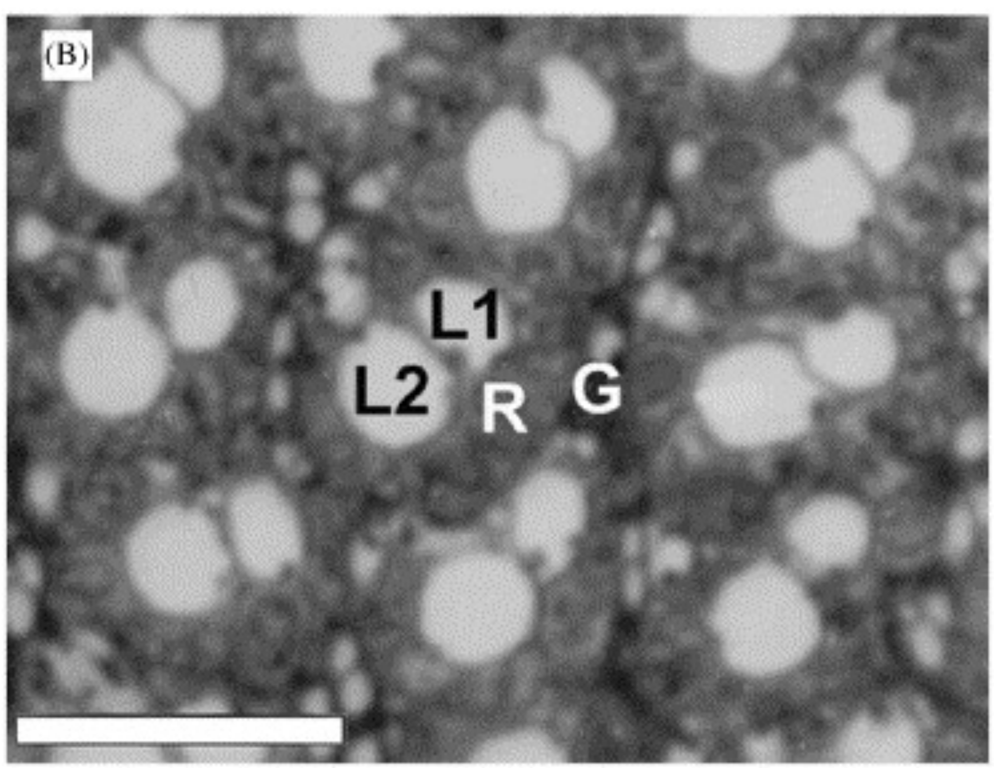
Evidence for entropy maximization in the fly:

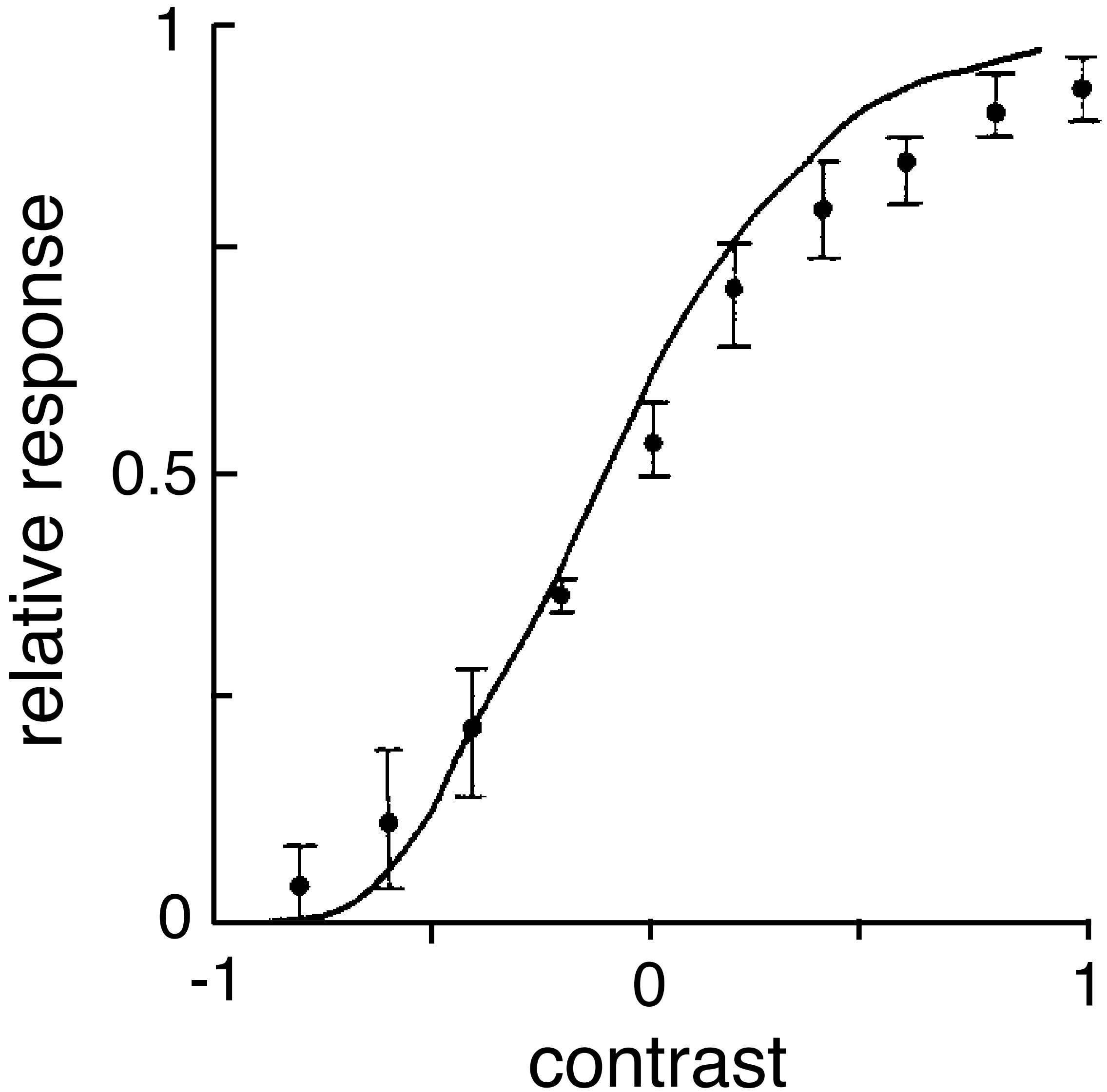


Laughlin, 1981

Evidence for entropy maximization in the fly:



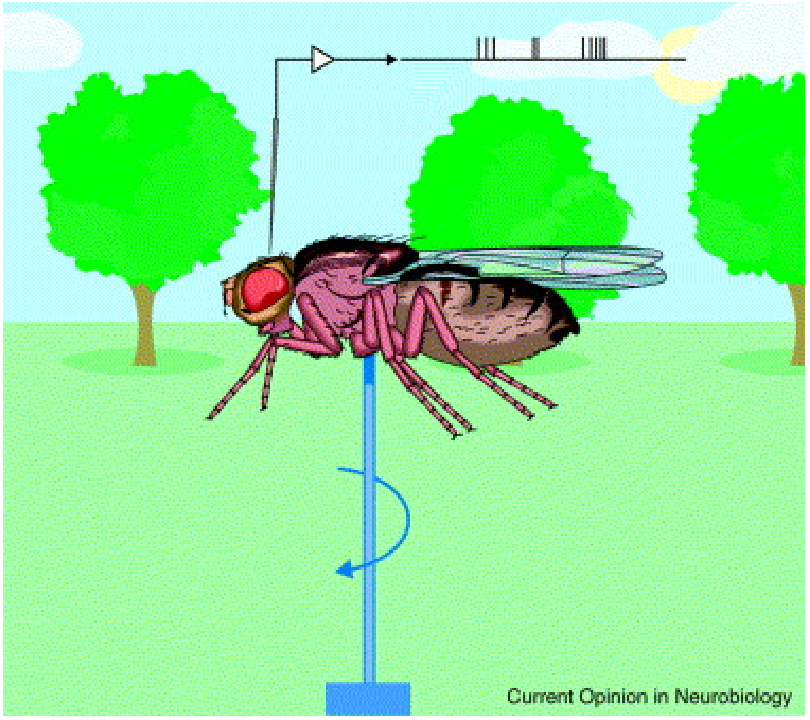


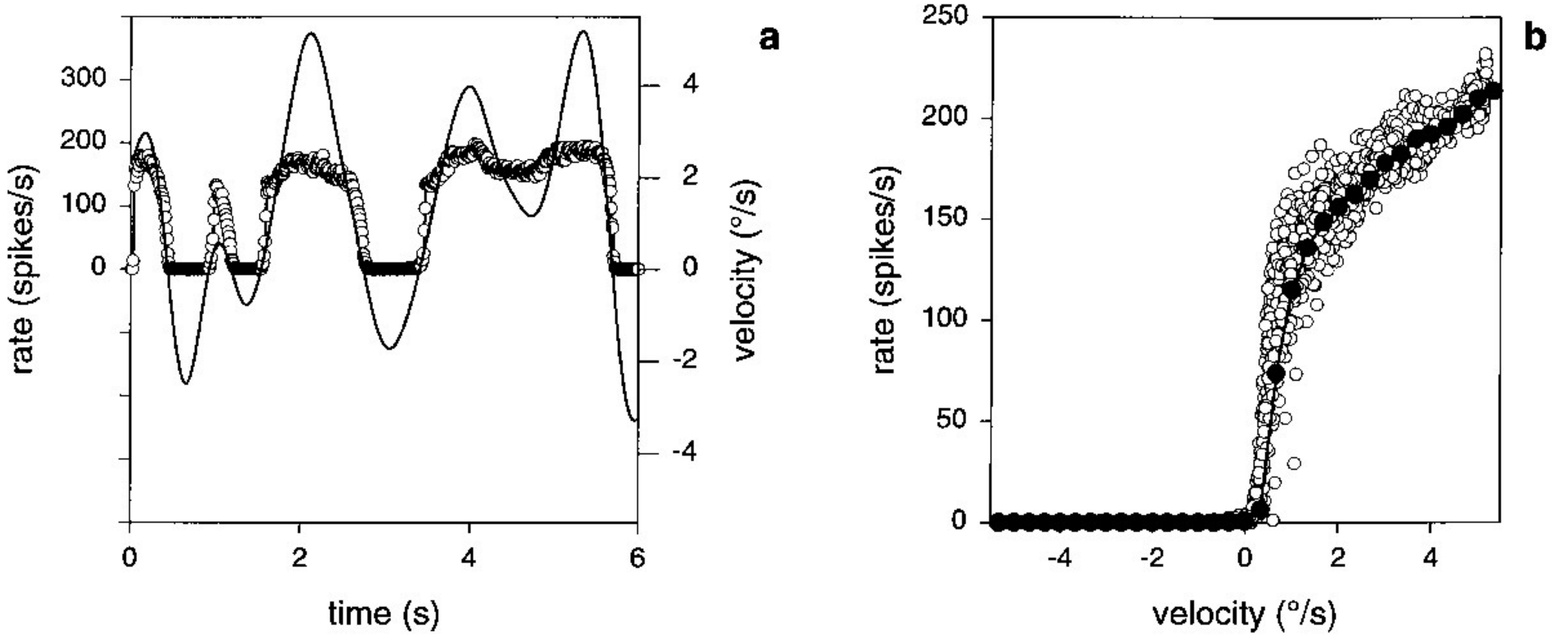


Laughlin, 1981

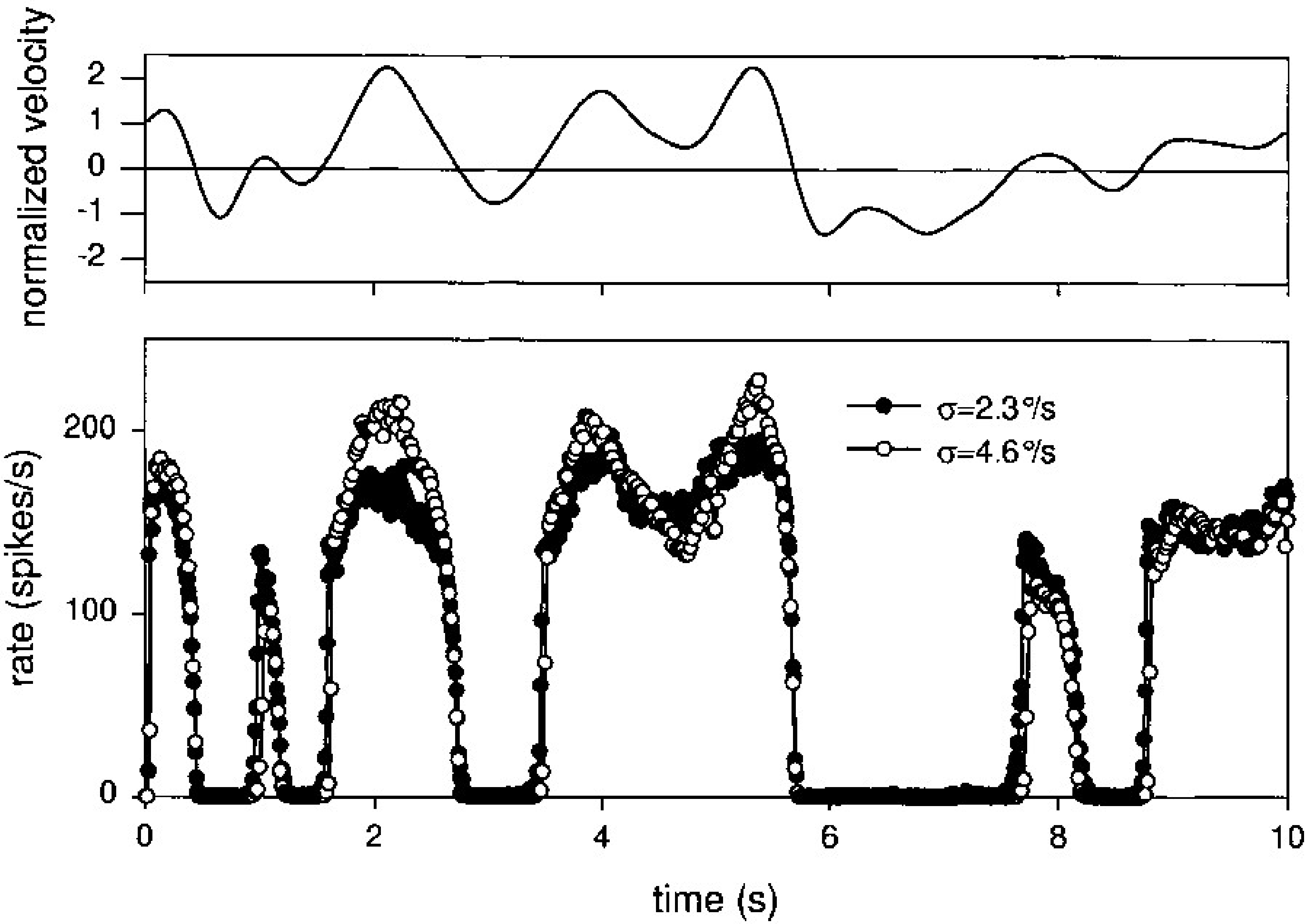
Adaptive rescaling in the fly H1 neuron:

Brenner et al 2000 (H1 neuron, visual system of the blowfly)

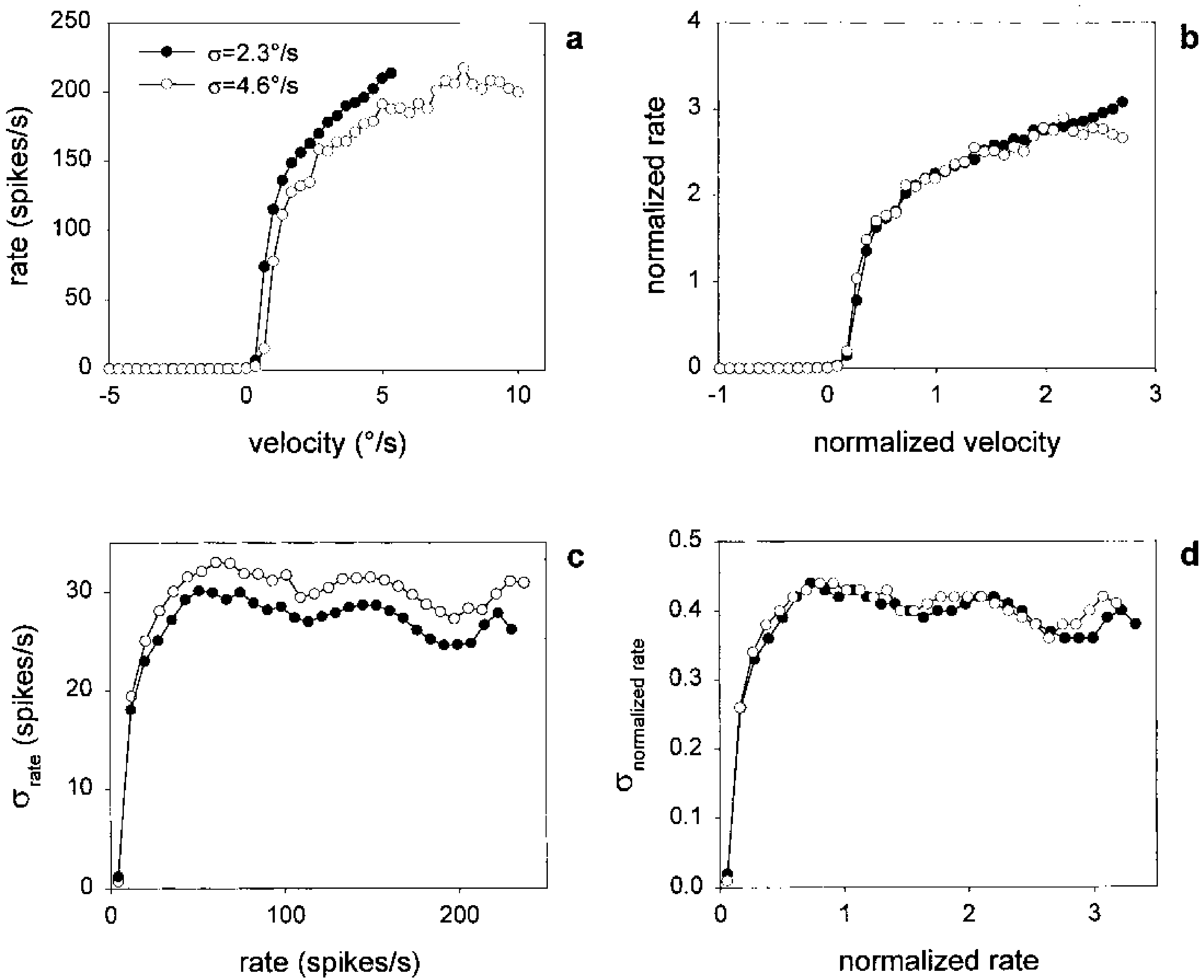




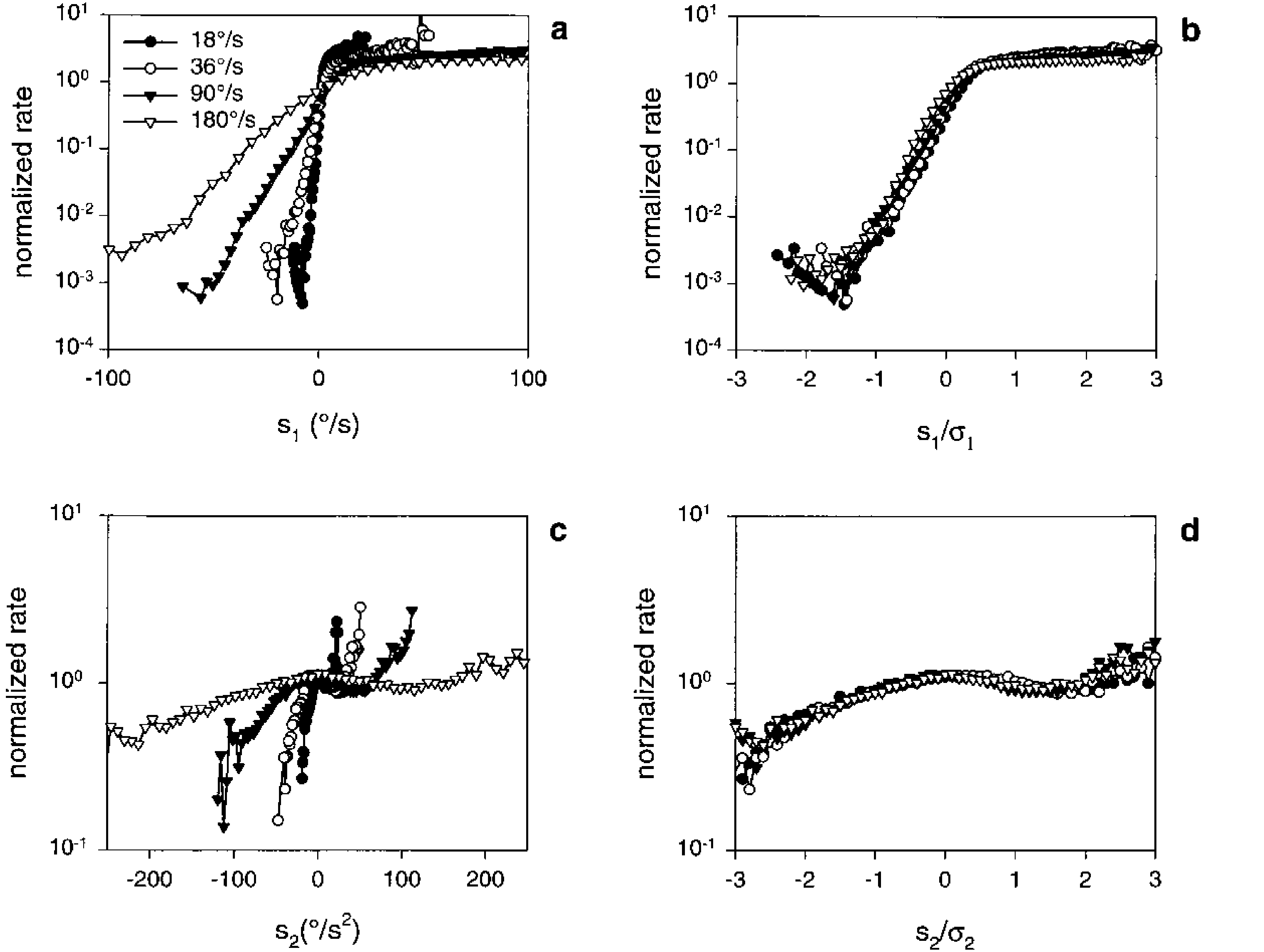
Responses to stimuli with 2x standard deviation change are nearly identical:



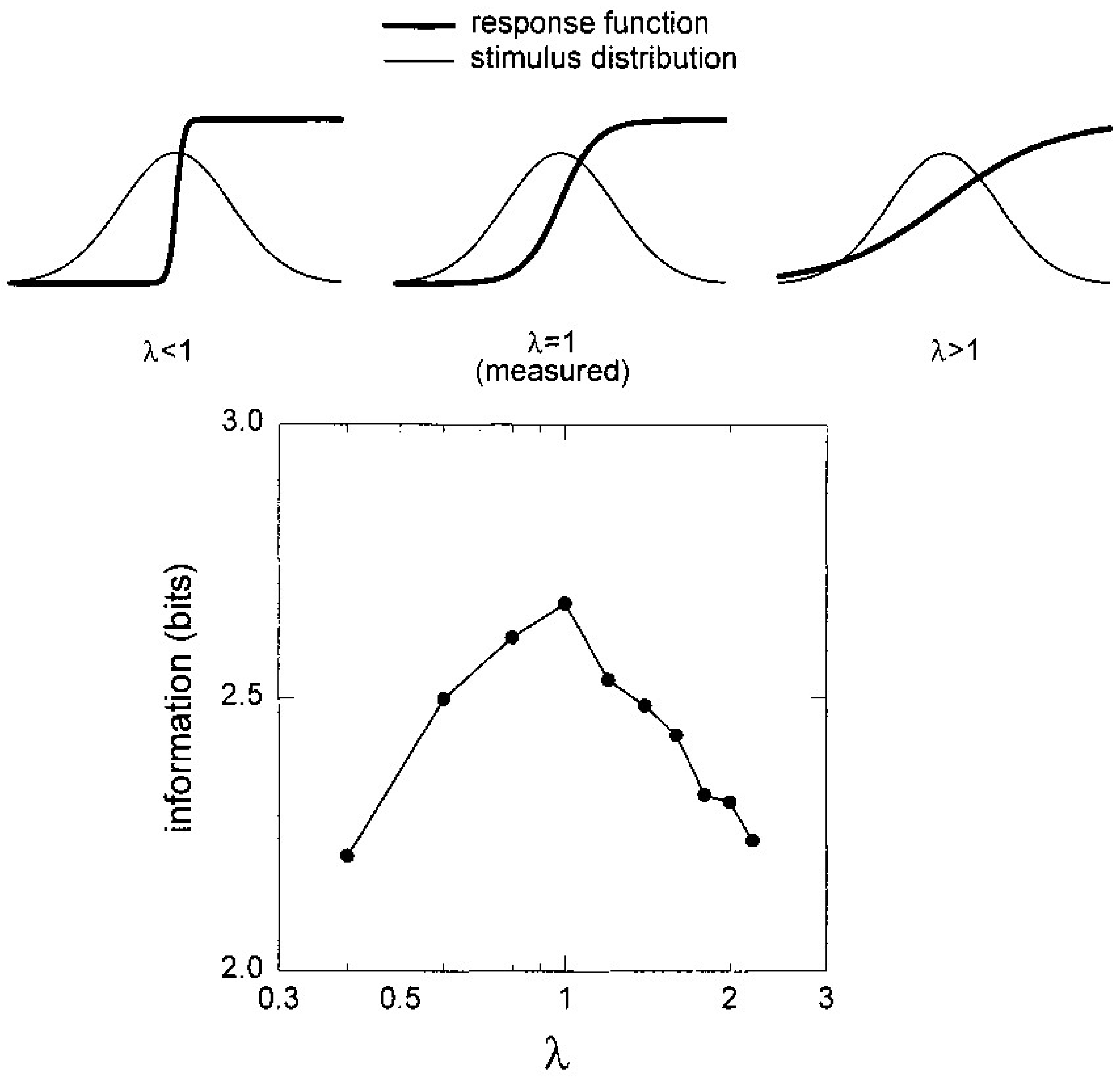
Signal and noise in the two stim conditions:



Adaptive rescaling to fast varying inputs:



Rescaling maximizes information transmission:



Optimal filter: whitening

- Optimize mutual information:

$$I = \frac{1}{2} \int \frac{d\omega}{2\pi} \log_2 \left(1 + \frac{|\tilde{K}(\omega)|^2 S(\omega)}{N(\omega)} \right)$$

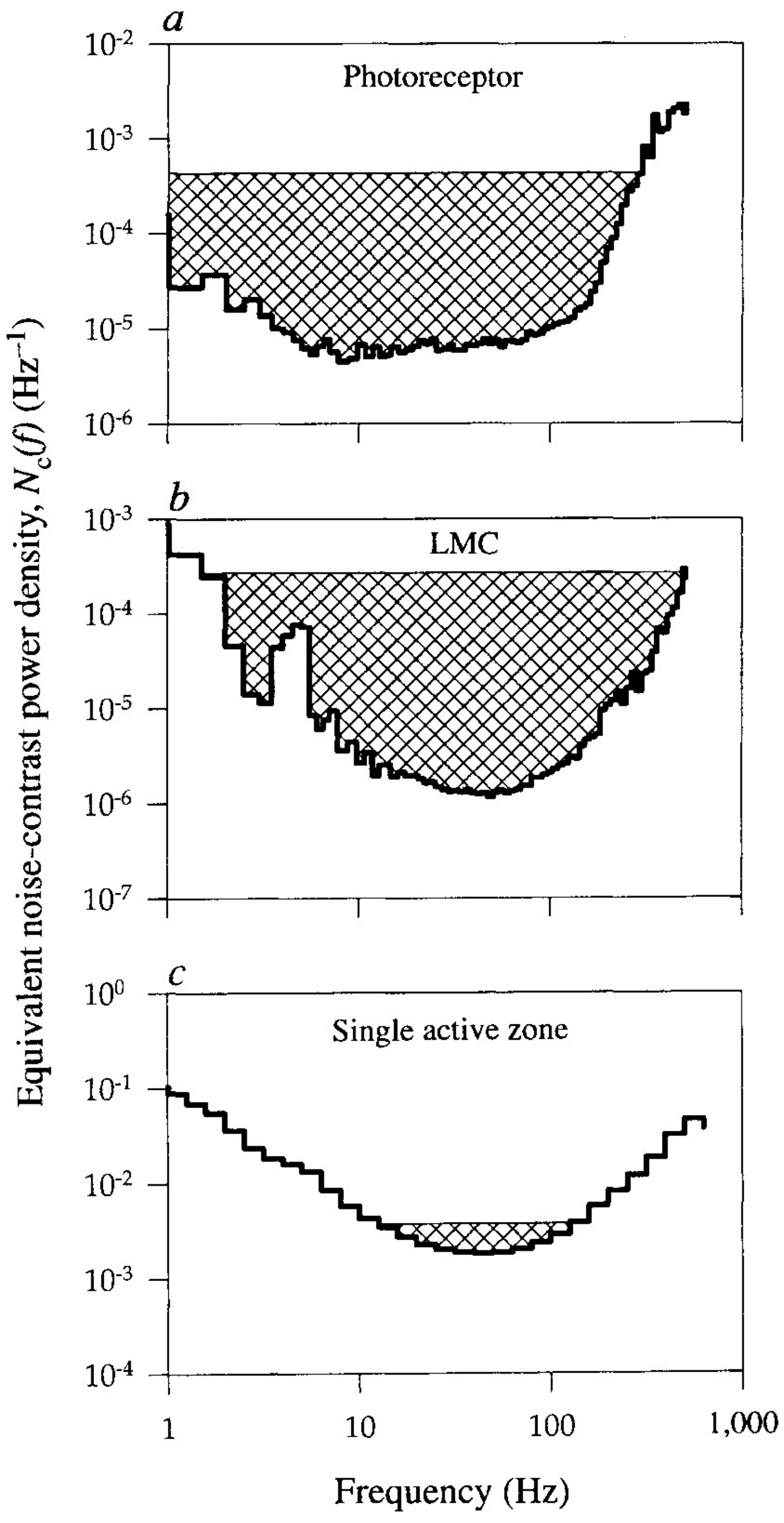
subject to constraint

$$\int d\omega |\tilde{K}(\omega)|^2 S(\omega) = \text{constant}$$

- Solution

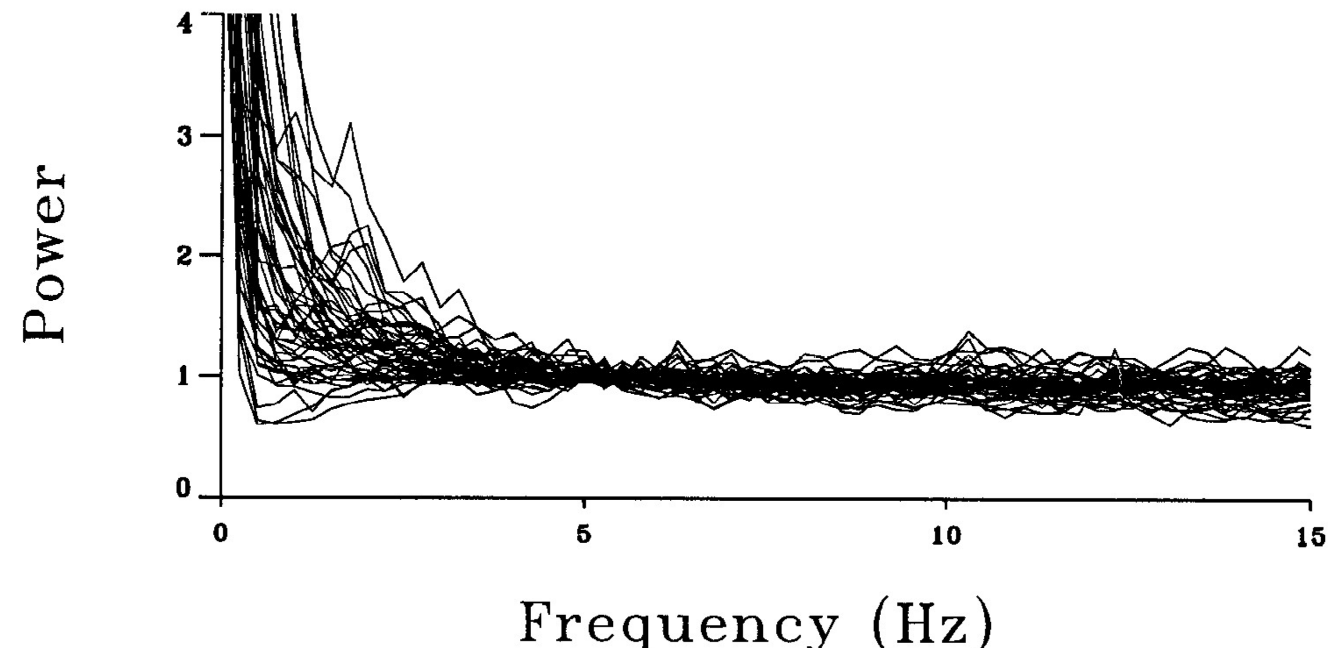
$$|\tilde{K}(\omega)|^2 S(\omega) = [A - N(\omega)]_+$$

- Whitening (water-filling analogy)



Whitening in the LGN (Dan et al 1996)

- Natural stimuli



- White-noise stimuli