NEURAL CODING AND STATISTICAL INFERENCE (LECTURE 1)

Rob Kass

Department of Statistics
Machine Learning Department
Center for the Neural Basis of Cognition
Carnegie Mellon University

BEFORE WE START

Student backgrounds?

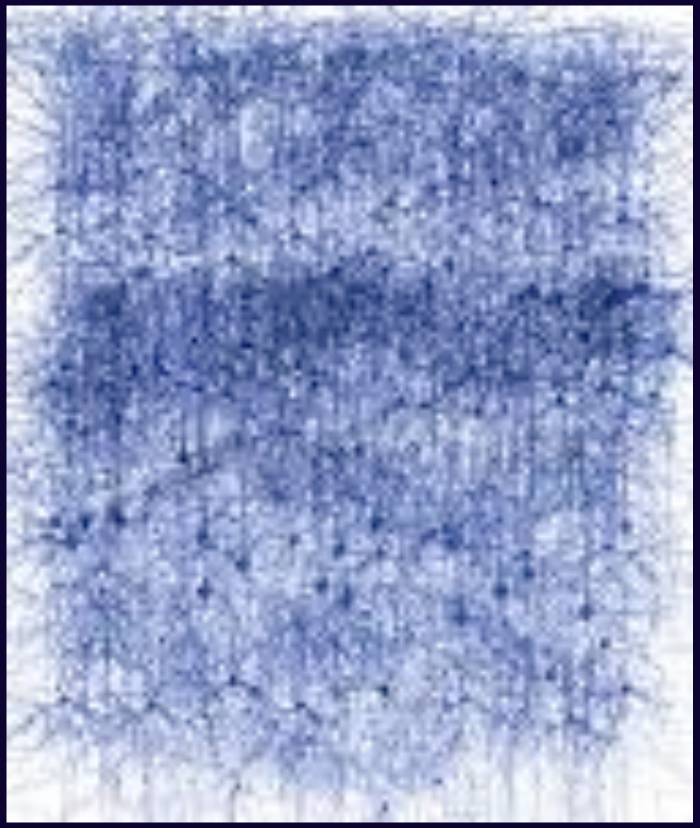
My background

PLAN FOR LECTURES

1. Motivation: Neural coding
2. Statistical concepts
   Shadlen and Newsome (1998), section 1
3. Motivation: More advanced methods
   Statistical thinking
4. Point process analysis via generalized regression

$\sim 10^{10}$ cells

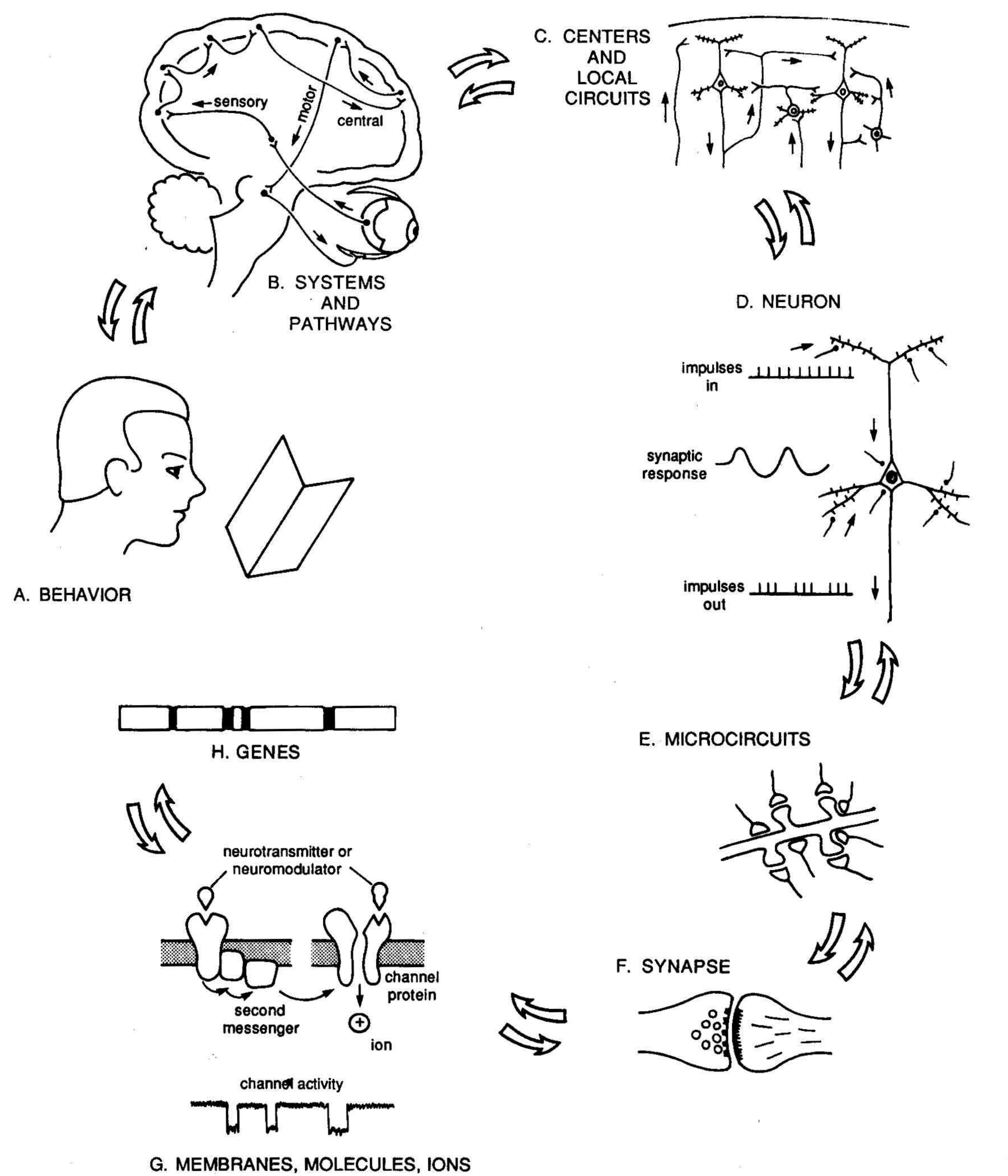
$\sim 10^{10}$ cells



$\sim 10^{10}$ cells

$\sim 10^{14}$ connections

Jennifer Aniston


C. CENTERS AND LOCAL CIRCUITS
sensory
motor
central
B. SYSTEMS AND PATHWAYS
D. NEURON
impulses in
synaptic response
impulses out
A. BEHAVIOR
H. GENES
E. MICROCIRCUITS
neurotransmitter or neuromodulator
channel protein
second messenger
ion
F. SYNAPSE
channel activity
G. MEMBRANES, MOLECULES, IONS

THE PROBLEMS OF BRAIN SCIENCE

- What principles may be used to describe mental processes?
- What principles may be used to describe the brain?
- How does neural activity produce behavior?

THE PROBLEMS OF BRAIN SCIENCE

- What principles may be used to describe mental processes?
- What principles may be used to describe the brain?
- How does neural activity produce behavior?

The more specific problem of neural coding:

THE PROBLEMS OF BRAIN SCIENCE

- What principles may be used to describe mental processes?
- What principles may be used to describe the brain?
- How does neural activity produce behavior?

The more specific problem of neural coding:

To elucidate "the representation and transmission of information in the nervous system"

Perkel and Bullock (1968)

THE PROBLEMS OF BRAIN SCIENCE

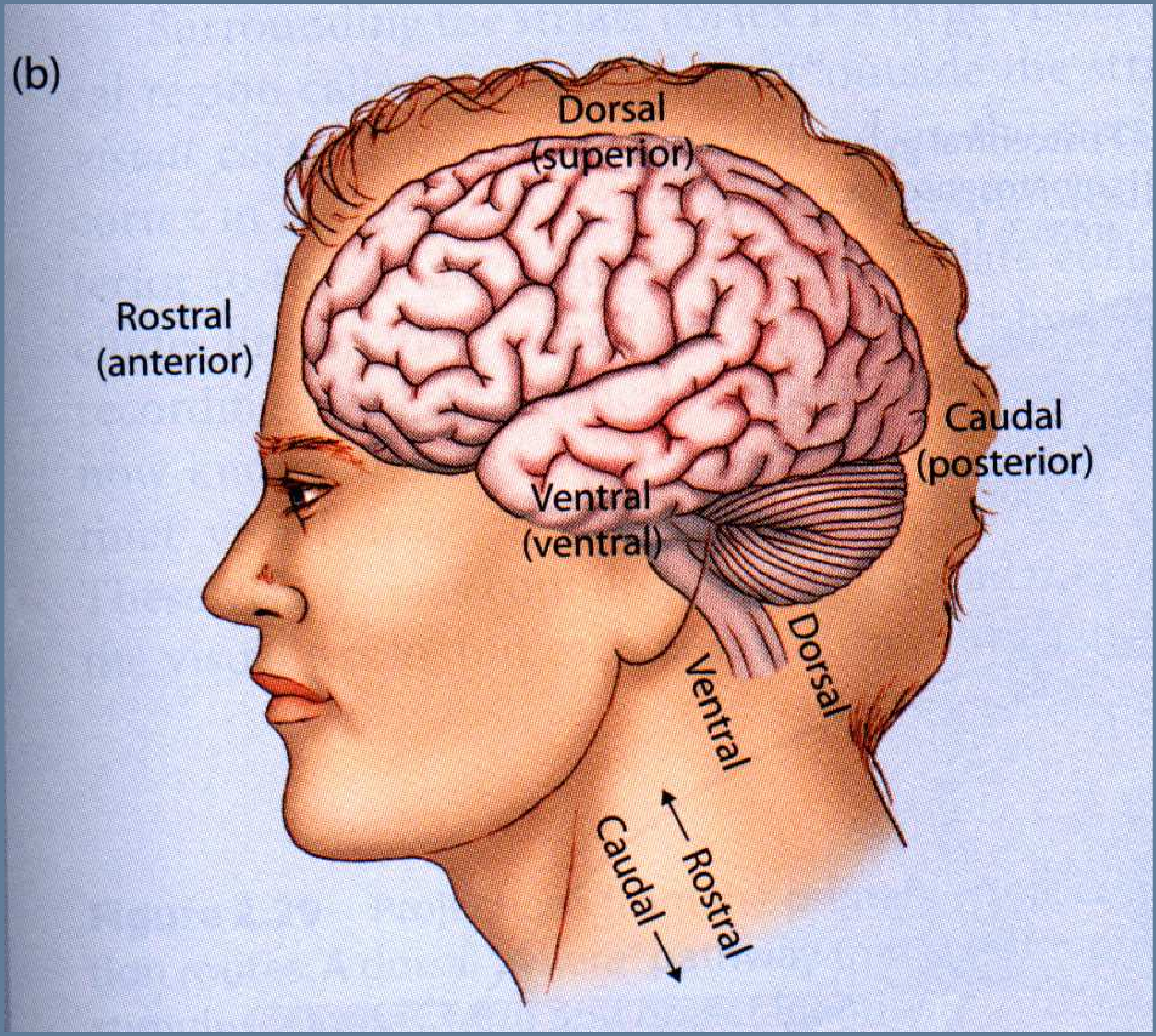
- What principles may be used to describe mental processes?
- What principles may be used to describe the brain?
- How does neural activity produce behavior?

The more specific problem of neural coding:

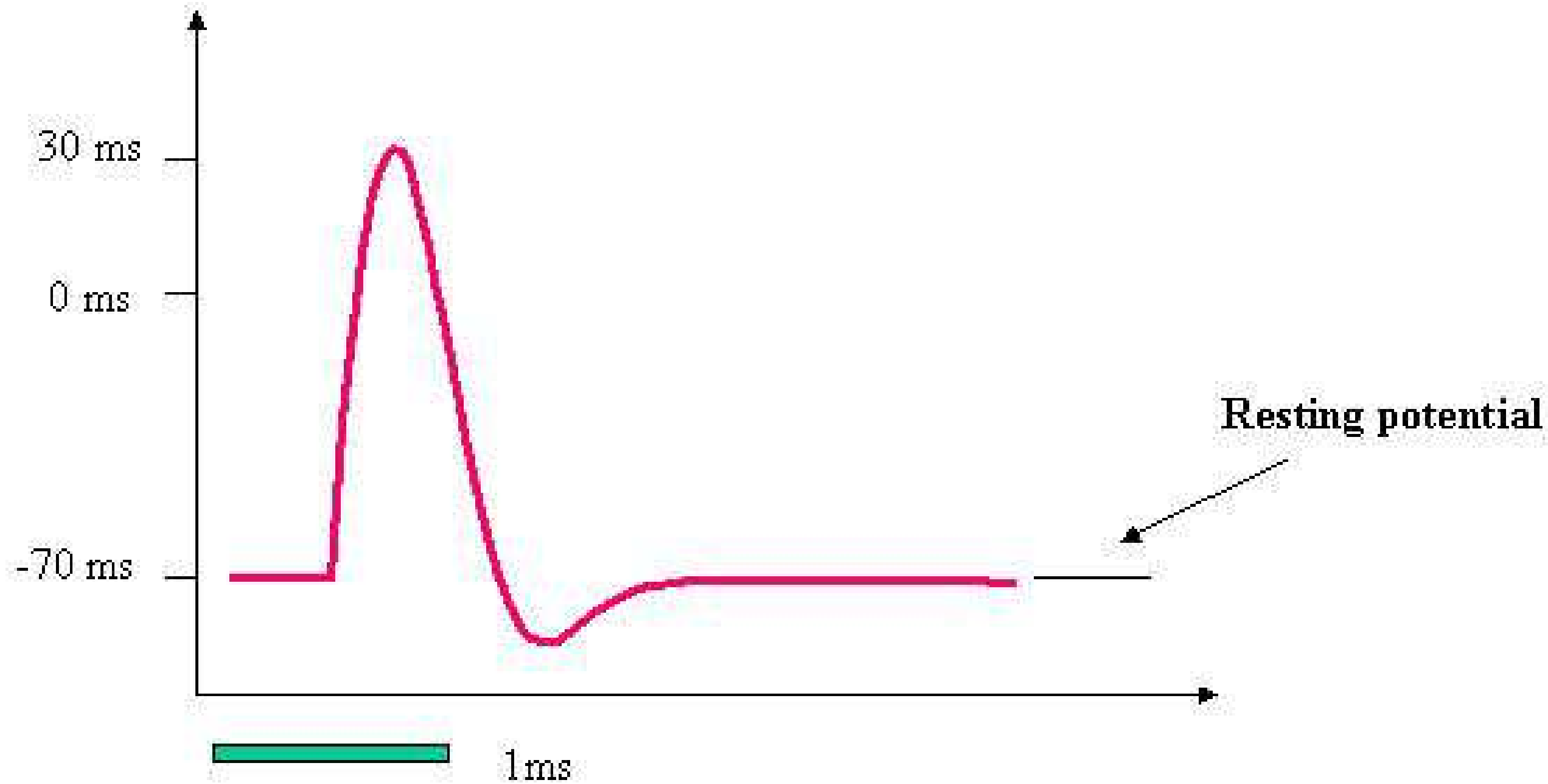
To elucidate "the representation and transmission of information in the nervous system"

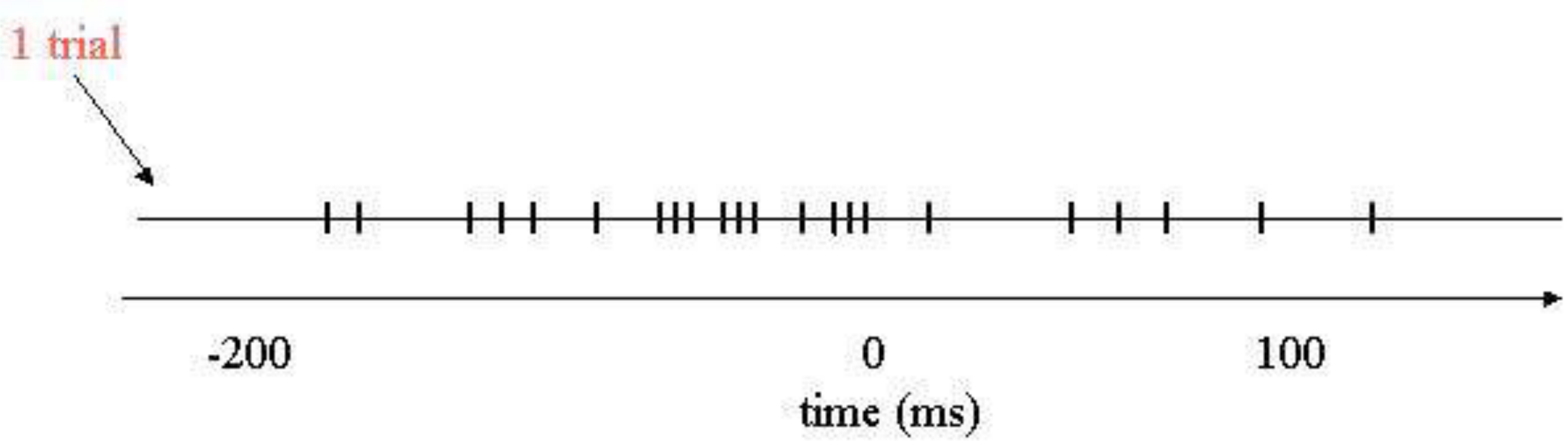
Perkel and Bullock (1968)

To begin: some background

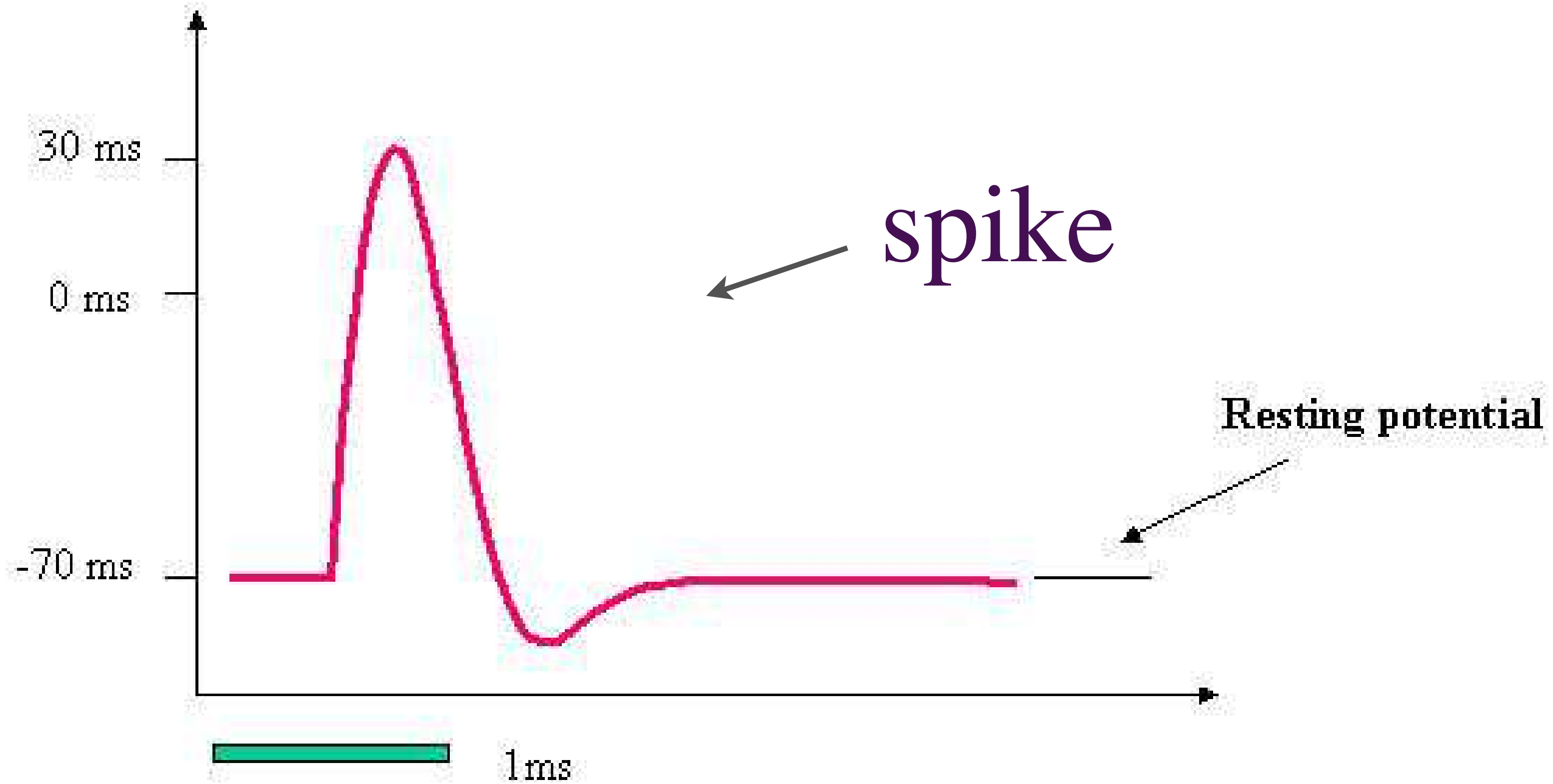

(b)
Dorsal
(superior)
Rostral
(anterior)
Caudal
(posterior)
Ventral
(ventral)
Dorsal
Ventral
← Rostral
Caudal →

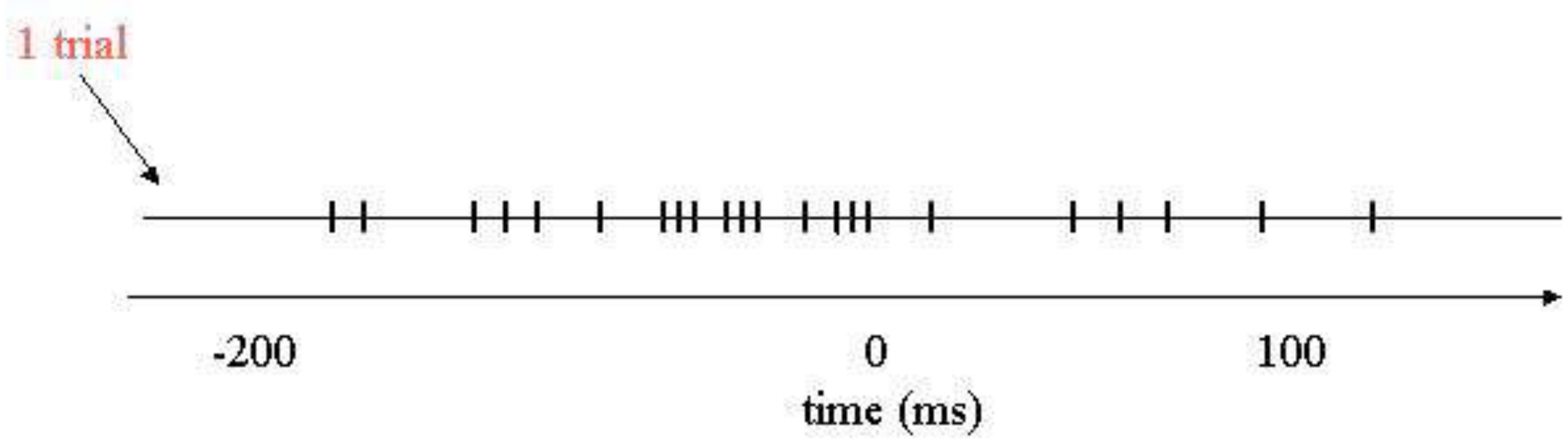
Action potential and spike train



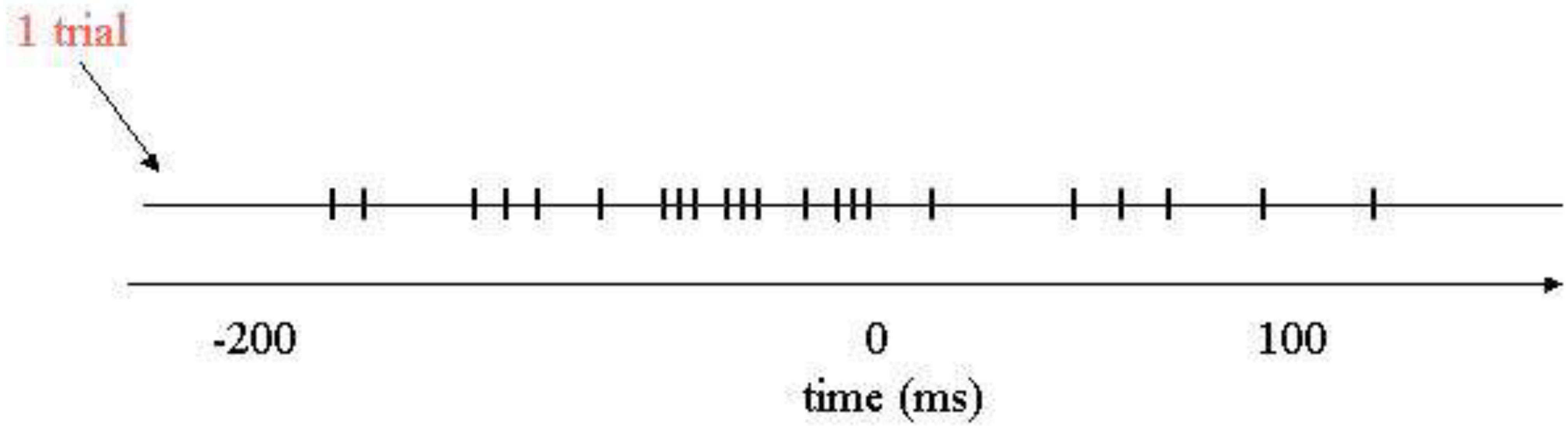


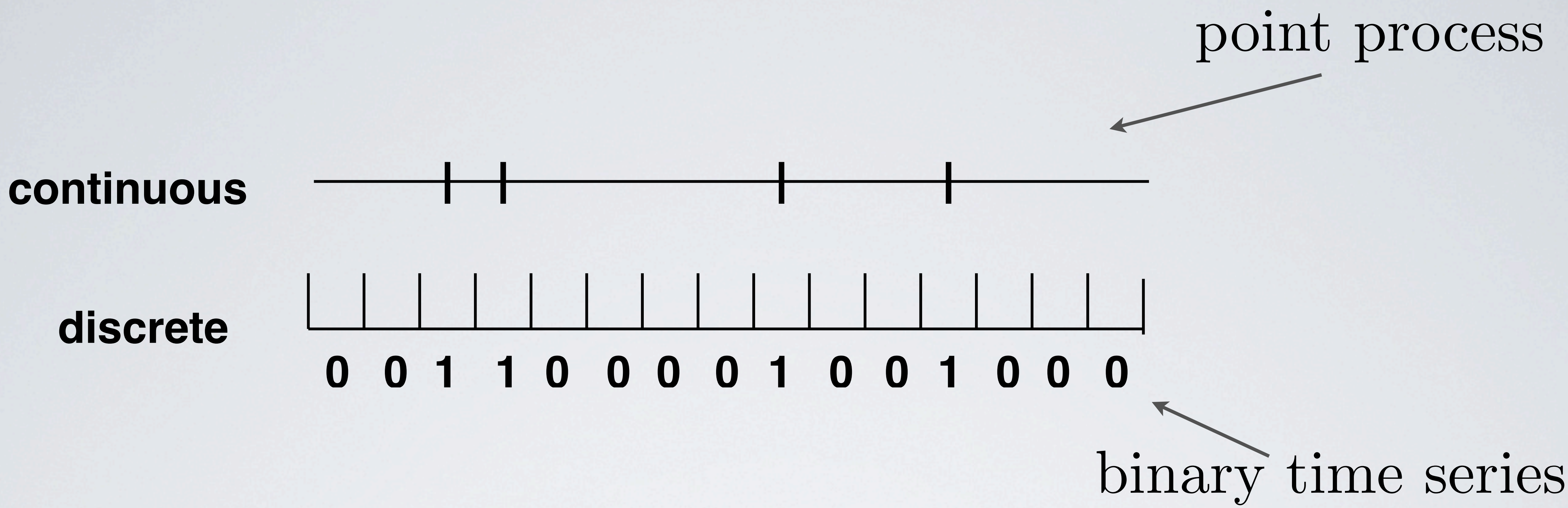
Action potential and spike train





spike train





Two Representations of Spike Trains

How do neurons represent and transmit information?

How do neurons represent and transmit information?

First answer: By modulating firing rate

Adrian (1926), Hartline (1938), Hubel and Wiesel (1963)

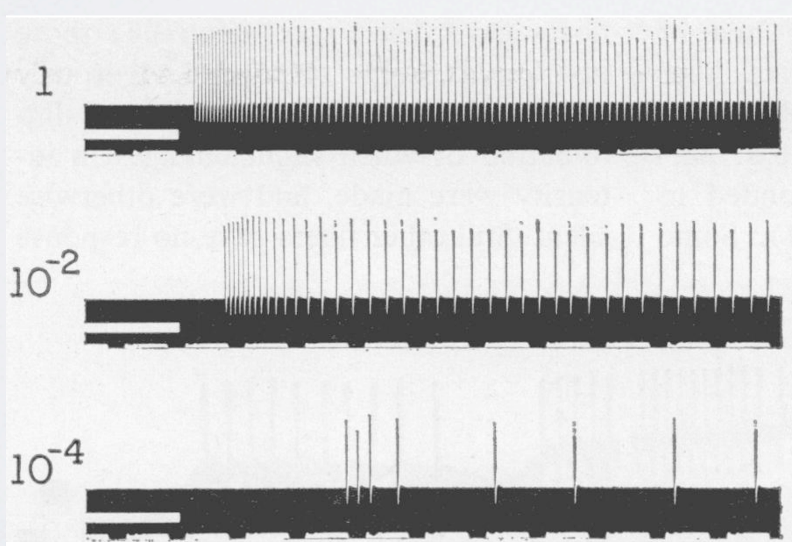
light on

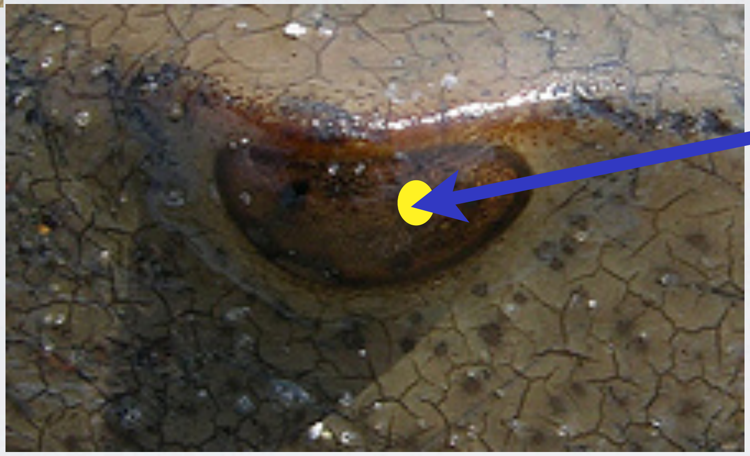
spike trains
from 1 neuron

bright light

medium light

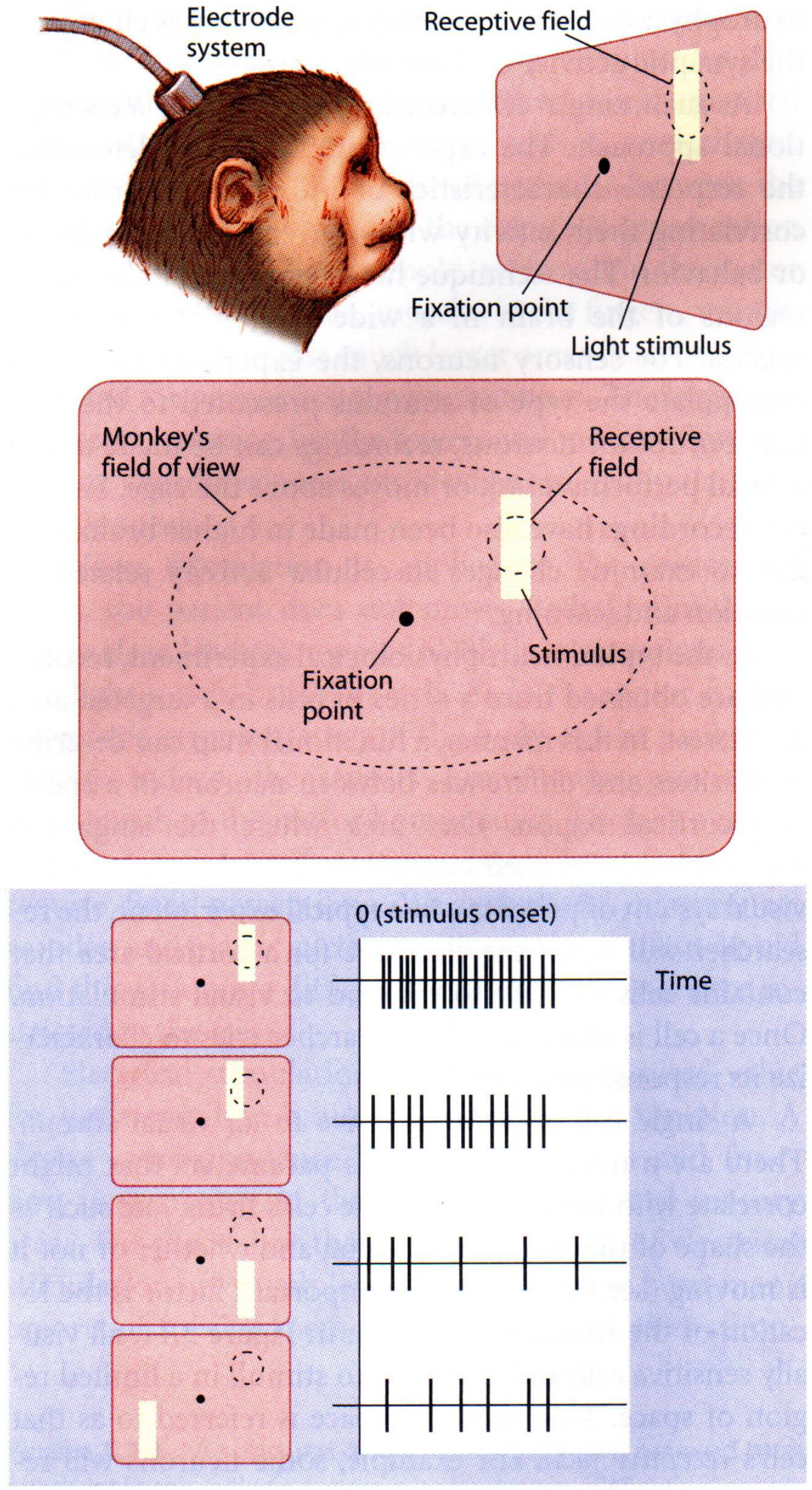
dim light



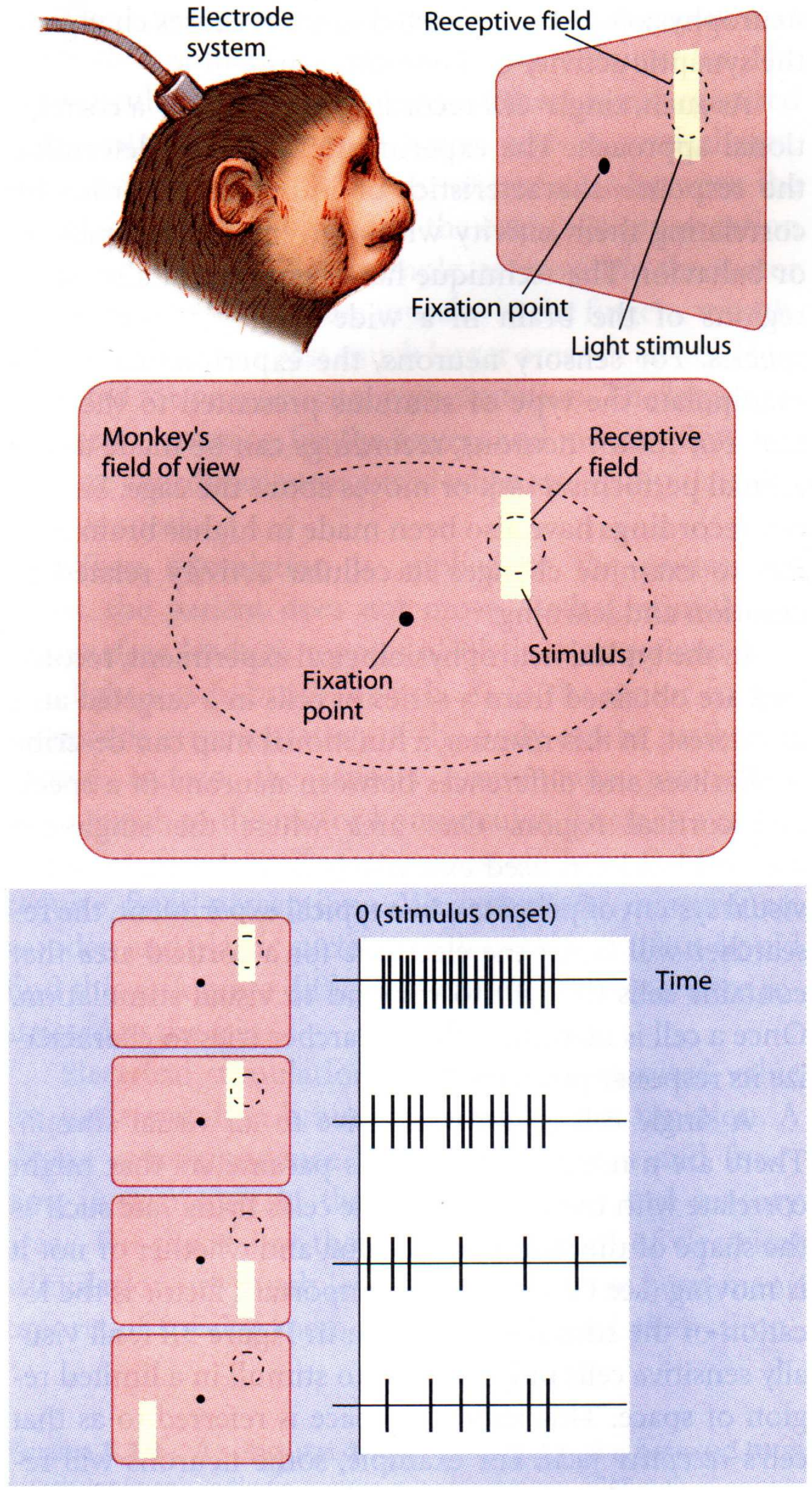


receptive field
for
that neuron

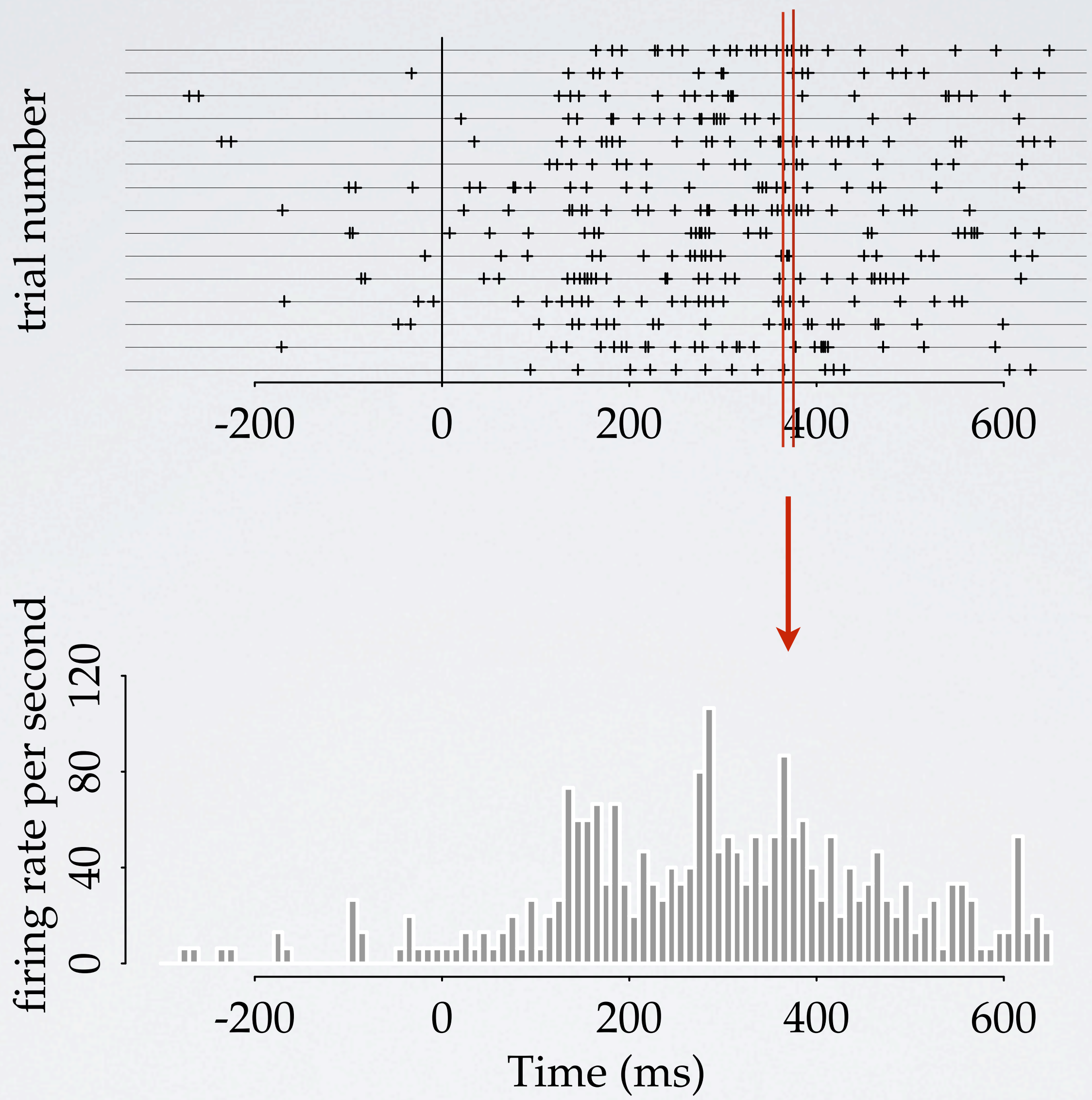
Firing rate modulation in the brain



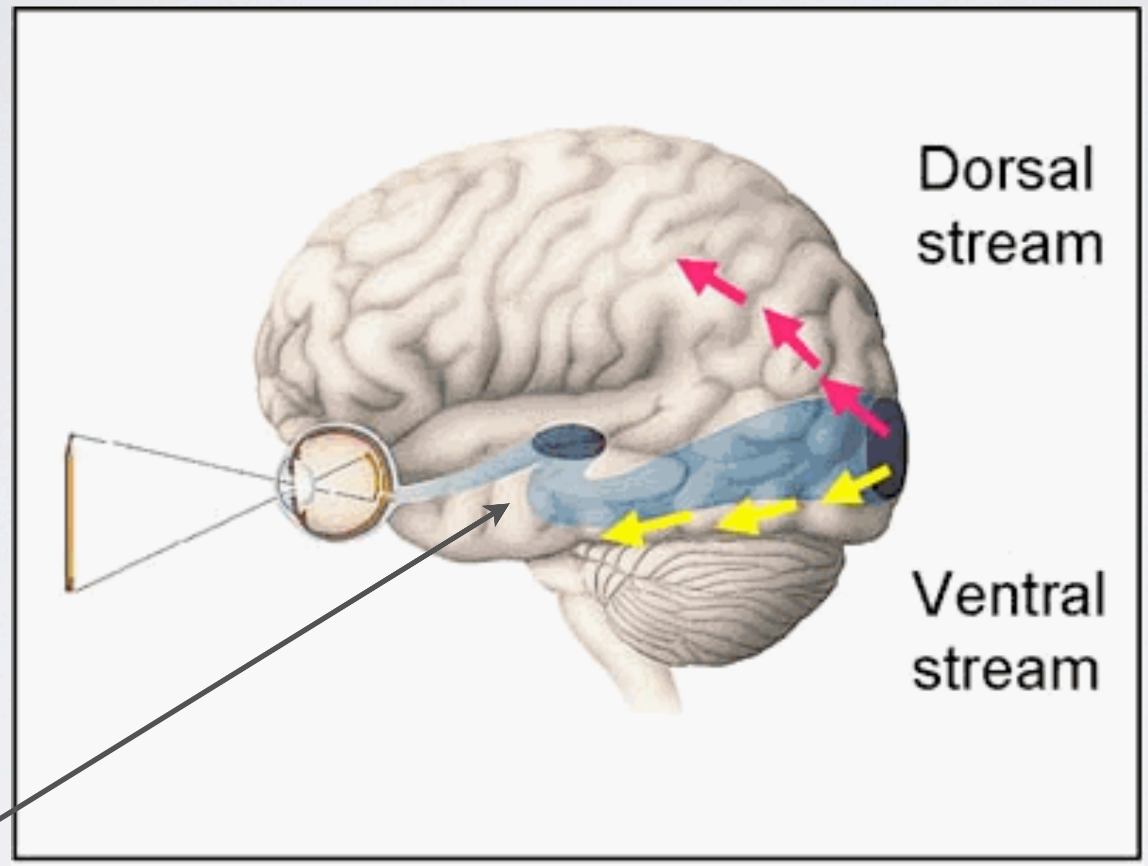
Firing rate modulation in the brain



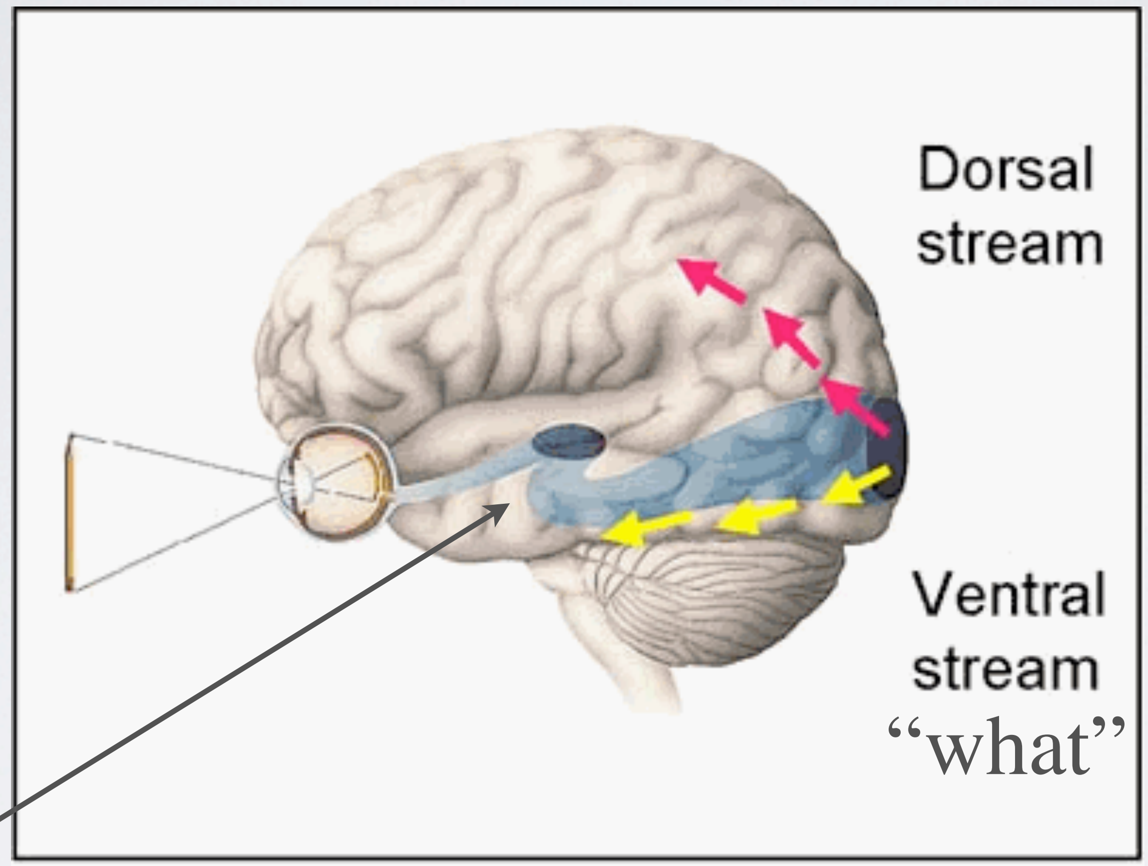
15 repetitions (trials)



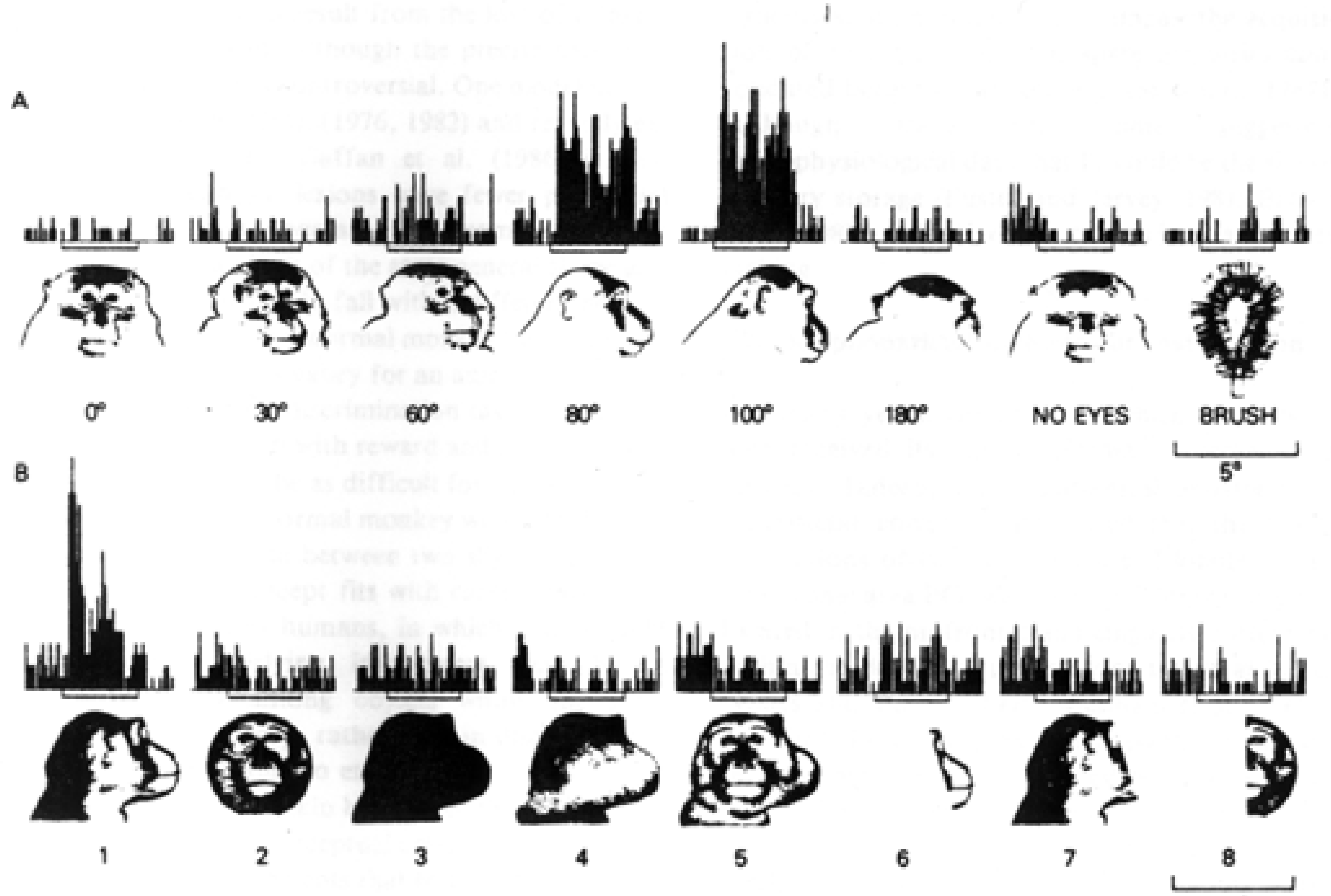
peristimulus-time histogram



inferotemporal region



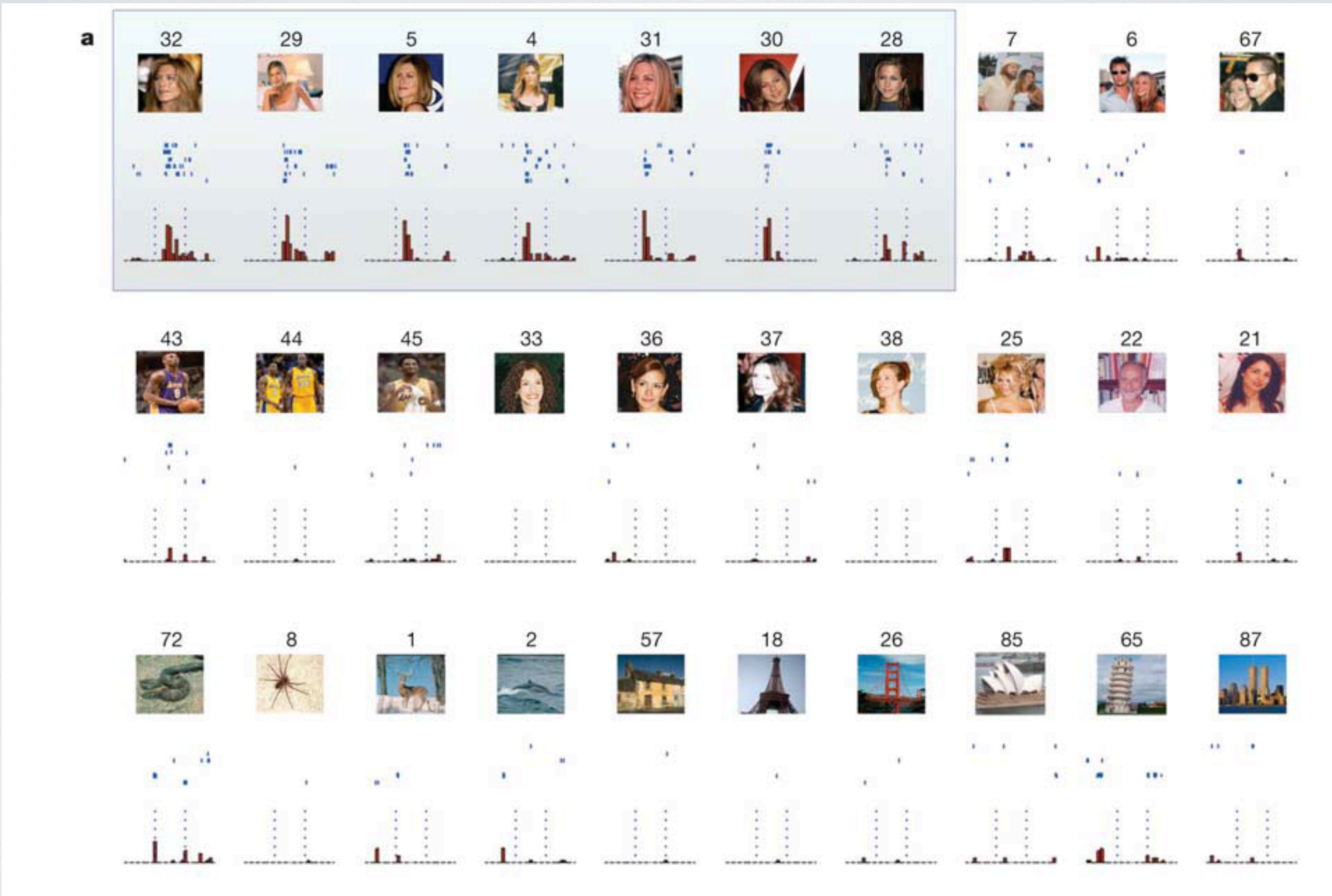
inferotemporal region


A
0°
30°
60°
80°
100°
180°
NO EYES
BRUSH
5°
B
1
2
3
4
5
6
7
8

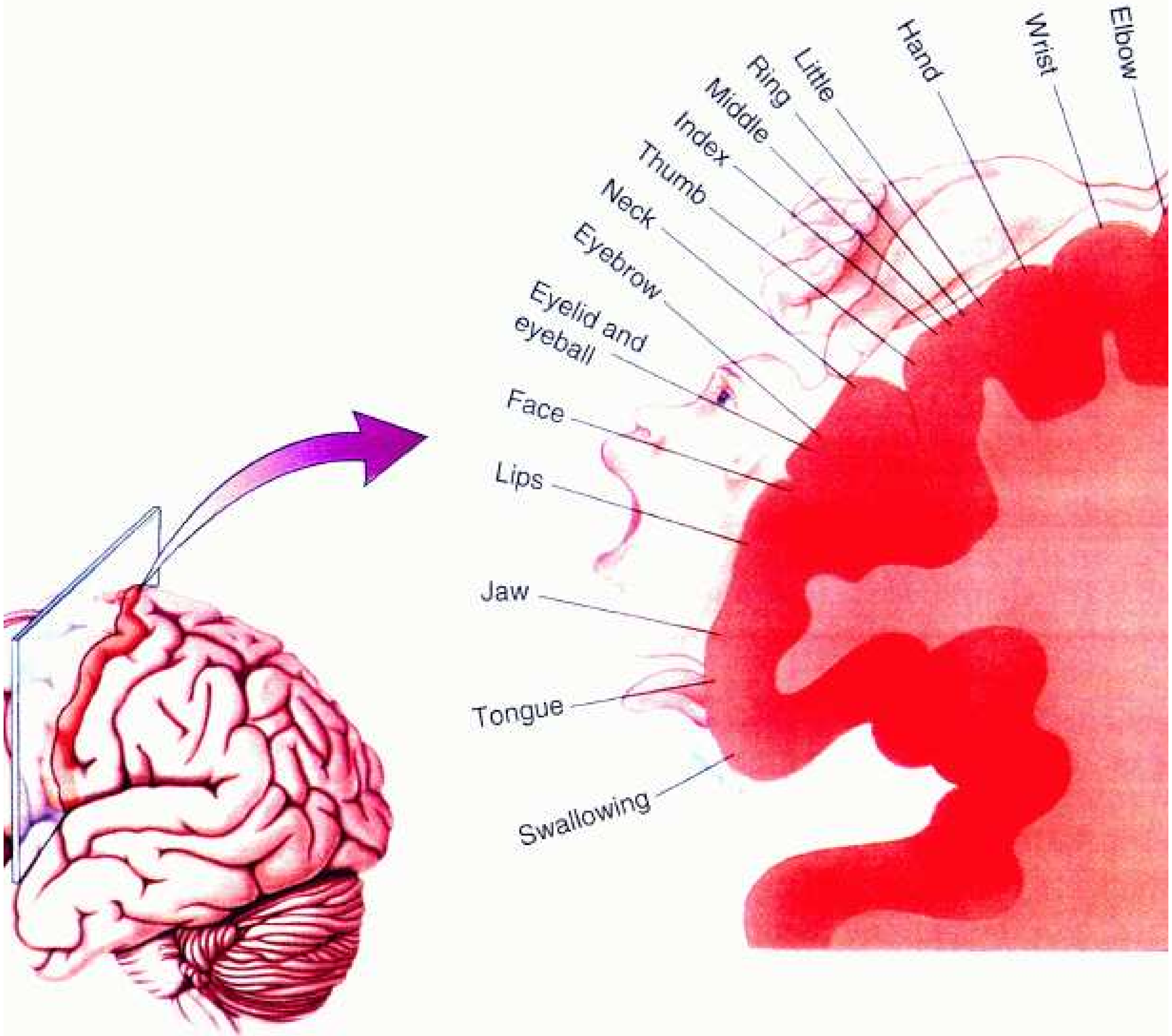
Nature (2005)

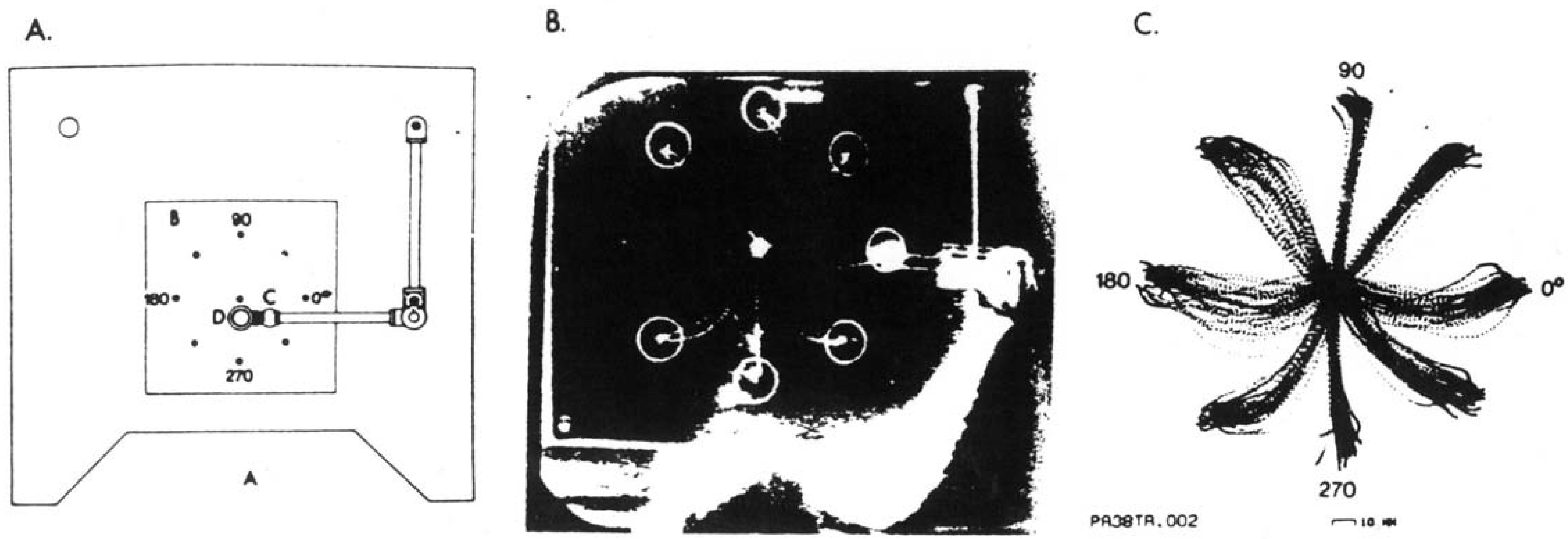
Invariant visual representation by single neurons in the human brain

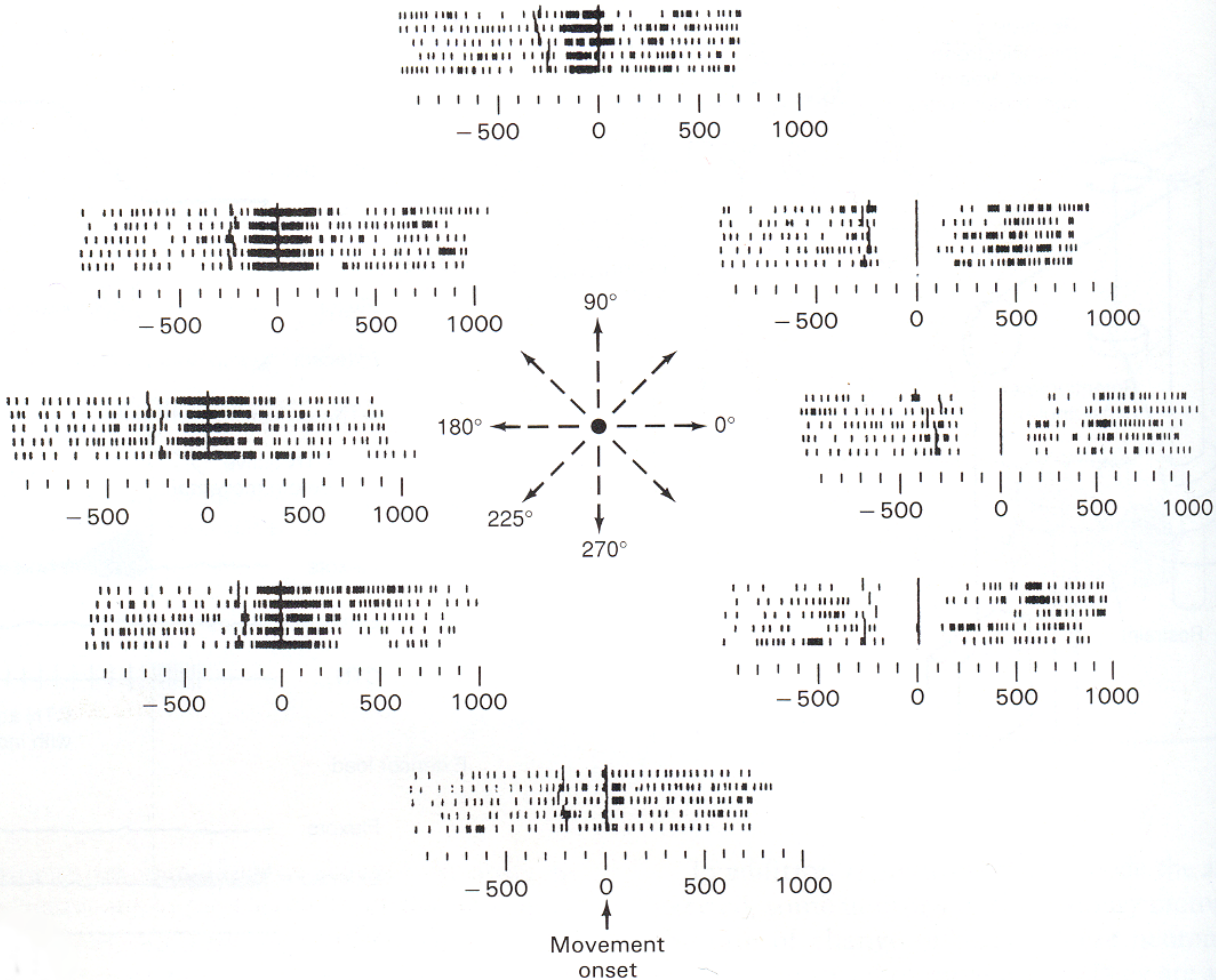
R. Quian Quiroga[1,2]†, L. Reddy[1], G. Kreiman[3], C. Koch[1] & I. Fried[2,4]



Firing rate higher for images of Jennifer Anniston


Elbow
Wrist
Hand
Little
Ring
Middle
Index
Thumb
Neck
Eyebrow
Eyelid and
eyeball
Face
Lips
Jaw
Tongue
Swallowing


A.
B.
C.
B
90
180
C
0°
D
270
A
90
180
0°
270
PA38TR.002
10 mm



90°
0°
180°
225°
270°
−500
0
500
1000
Movement onset

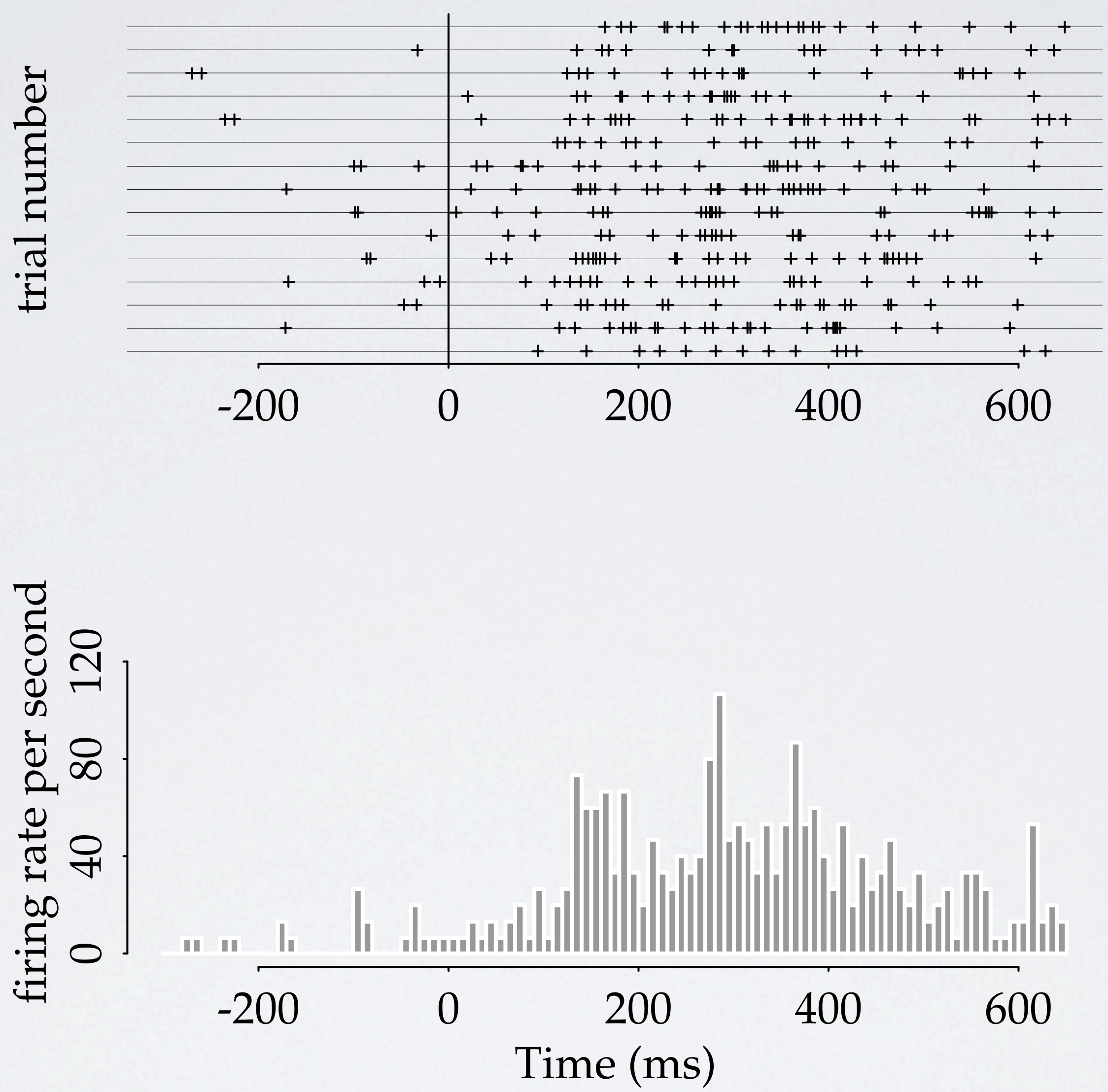
Neural Coding Questions

Neural Coding Questions

- What stimuli, or actions, drive neurons (in a particular part of the brain) to fire rapidly?

Neural Coding Questions

- What stimuli, or actions, drive neurons (in a particular part of the brain) to fire rapidly?
- How is information represented in ways that go beyond firing rates---or spike counts---in individual neurons?


trial number
-200
0
200
400
600
firing rate per second
0
40
80
120
Time (ms)

General Questions:

Which features of spike trains are “signal” and which are “noise”?

Does the PSTH from a single neuron represent well the signal from a population of similar neurons?
In what ways are population signals carried that are not apparent from responses of individual neurons?

What time scales are relevant to neural coding?

What do we mean by firing rate?

Firing rate as spike count per unit time

$$\frac{\text{number of spikes}}{\Delta t}$$

What do we mean by firing rate?

Firing rate as spike count per unit time

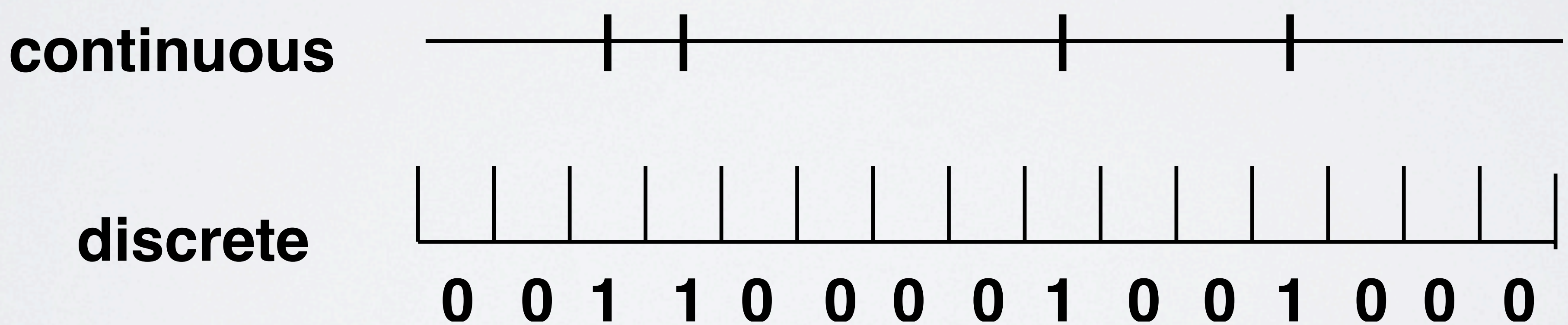
$$\frac{\text{number of spikes}}{\Delta t}$$

---> only works for large intervals

---> extreme case of 1 spike in very small interval: firing rate becomes infinite as interval width decreases

What do we mean by firing rate?

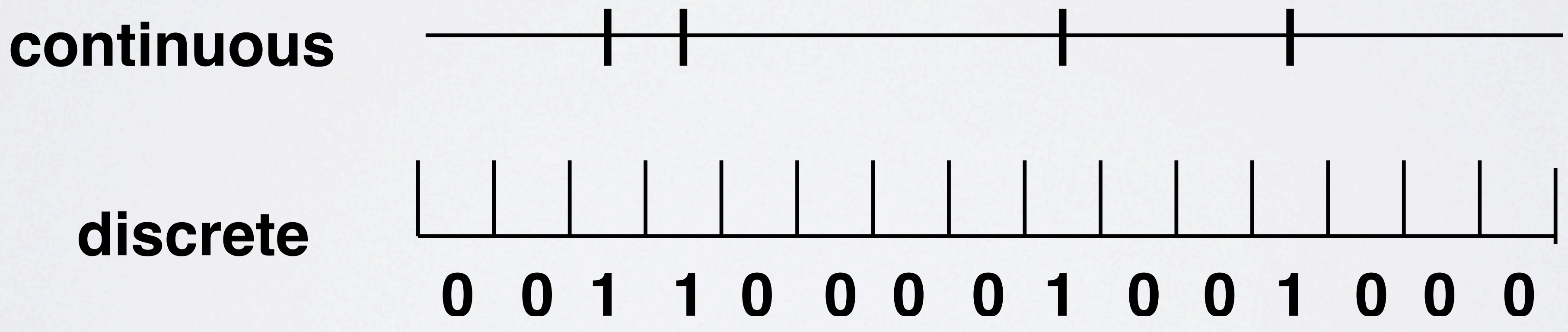
used with *point processes*



What do we mean by firing rate?

$\textit{Theoretical}$ instantaneous firing rate $\lambda(t)$ (or $\lambda(t|H_t)$)

used with *point processes*



observation vs. *random variable* X

sample mean vs. (theoretical) *mean* or *expectation* of X

$$\mu_X = E(X) = \sum_x x \cdot p(x)$$

sample variance

$$s^2 = \frac{1}{n-1} \sum_x (x - \bar{x})^2$$

vs. *variance* of X

$$\sigma_X^2 = V(X) = \sum_x (x - \mu_X)^2 \cdot p(x)$$

homogeneous Poisson:

$$P(\text{event in } (t, t+dt]) = \lambda dt.$$

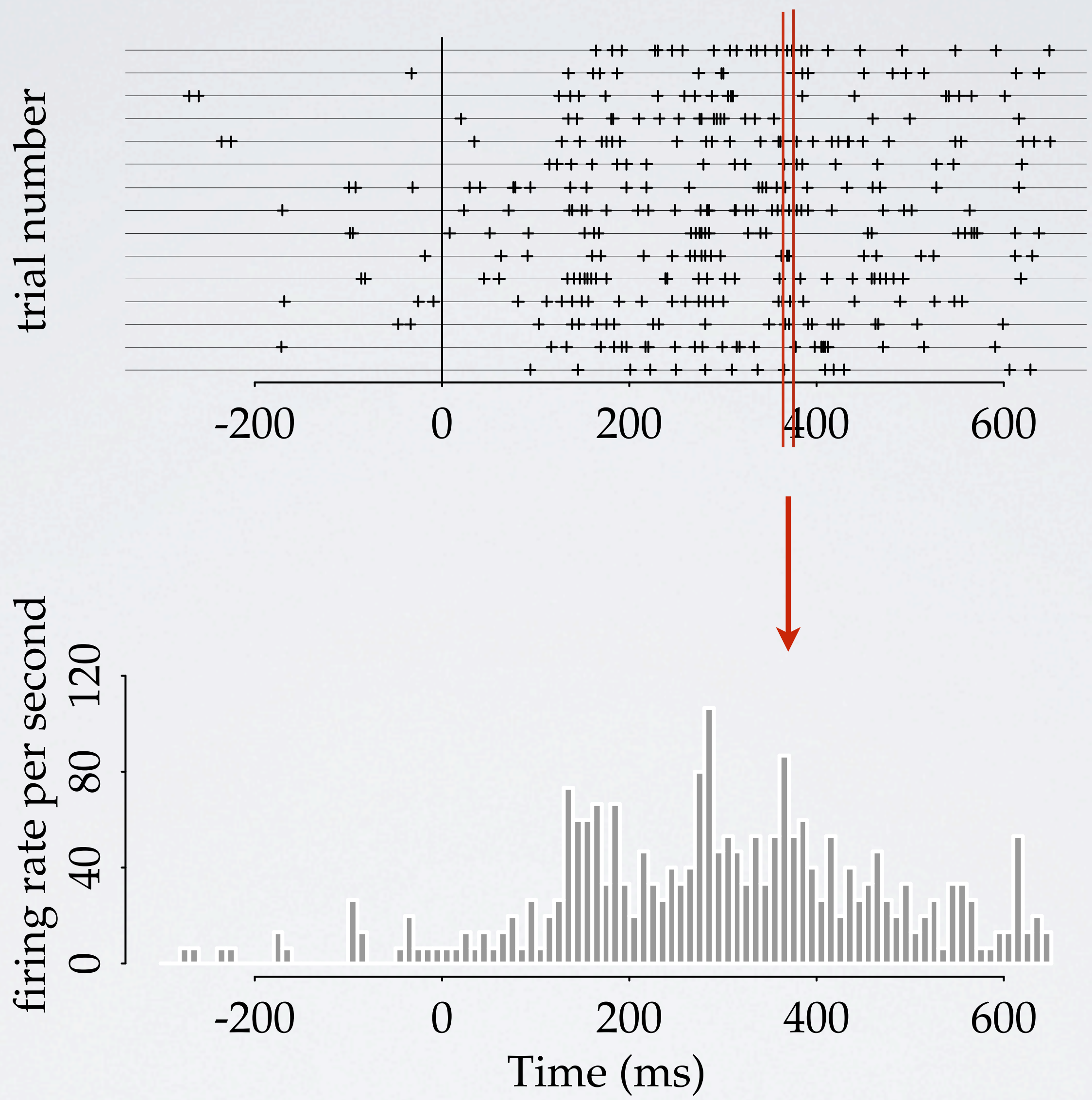
inhomogeneous Poisson:

$$P(\text{event in } (t, t+dt]) = \lambda(t) dt.$$

general:

$$P(\text{event in } (t, t+dt] | H_t) = \lambda(t|H_t) dt,$$

$$(\text{for Poisson: } \lambda(t|H_t) = \lambda(t))$$

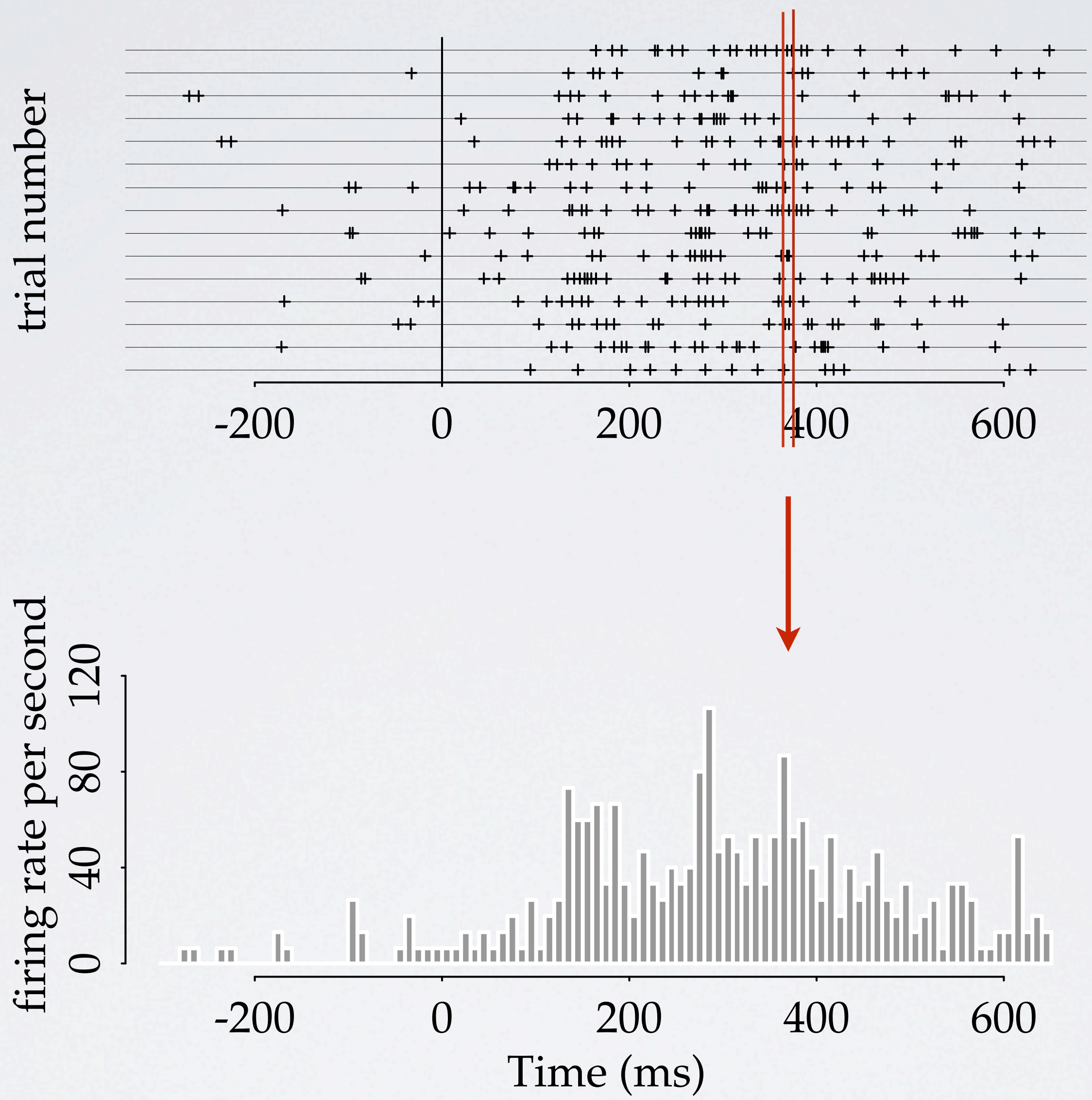


15 repetitions (trials)

peristimulus-time histogram (PSTH)

15 repetitions (trials)

peristimulus-time histogram (PSTH)



smooth the PSTH to get estimated intensity

trial-averaged (marginal)
firing rate $FR = \lambda(t)$ $\longleftarrow$ marginal intensity

trial-averaged (marginal)
firing rate $FR = \lambda(t)$ $\longleftarrow$ marginal intensity

within-trial (conditional)
firing rate $FR = \lambda(t|H_t)$ $\longleftarrow$ conditional intensity
where H_t is spiking history

trial-averaged (marginal)
firing rate $FR = \lambda(t)$ $\longleftarrow$ marginal intensity

within-trial (conditional)
firing rate $FR = \lambda(t|H_t)$ $\longleftarrow$ conditional intensity
where H_t is spiking history

If we model spike trains using the marginal intensity we are assuming they follow a Poisson process---otherwise, we must use the conditional intensity.

Neural coding issue: if downstream neurons simply count spikes (integrate) then marginal intensity is all that matters

and the nervous system essentially assumes that neural spike trains are Poisson processes.

Neural coding issue: if downstream neurons simply count spikes (integrate) then marginal intensity is all that matters and the nervous system essentially assumes that neural spike trains are Poisson processes.

Otherwise, conditional intensity relevant and we should see neural coding effects of non-Poisson-ness main message ~10 years ago:

analyses based on Poisson processes may be generalized by modeling the conditional intensity function

Kass and Ventura, 2001; Barbieri, ..., Brown, 2001; Brillinger, 1988

Difficulty: $\lambda(t|H_t) = \lambda(t|s_1, s_2, \ldots, s_{n(t)})$ depends on all the previous spike times. Need some simplifying assumption.

Difficulty: $\lambda(t|H_t) = \lambda(t|s_1, s_2, \ldots, s_{n(t)})$ depends on all the previous spike times. Need some simplifying assumption.

A first-pass (pretty good) solution: Markov model

$$\lambda(t|H_t) = \lambda(t, t - s_*(t))$$

where $s_*(t)$ is time of last spike preceding t

Difficulty: $\lambda(t|H_t) = \lambda(t|s_1, s_2, \ldots, s_{n(t)})$ depends on all the previous spike times. Need some simplifying assumption.

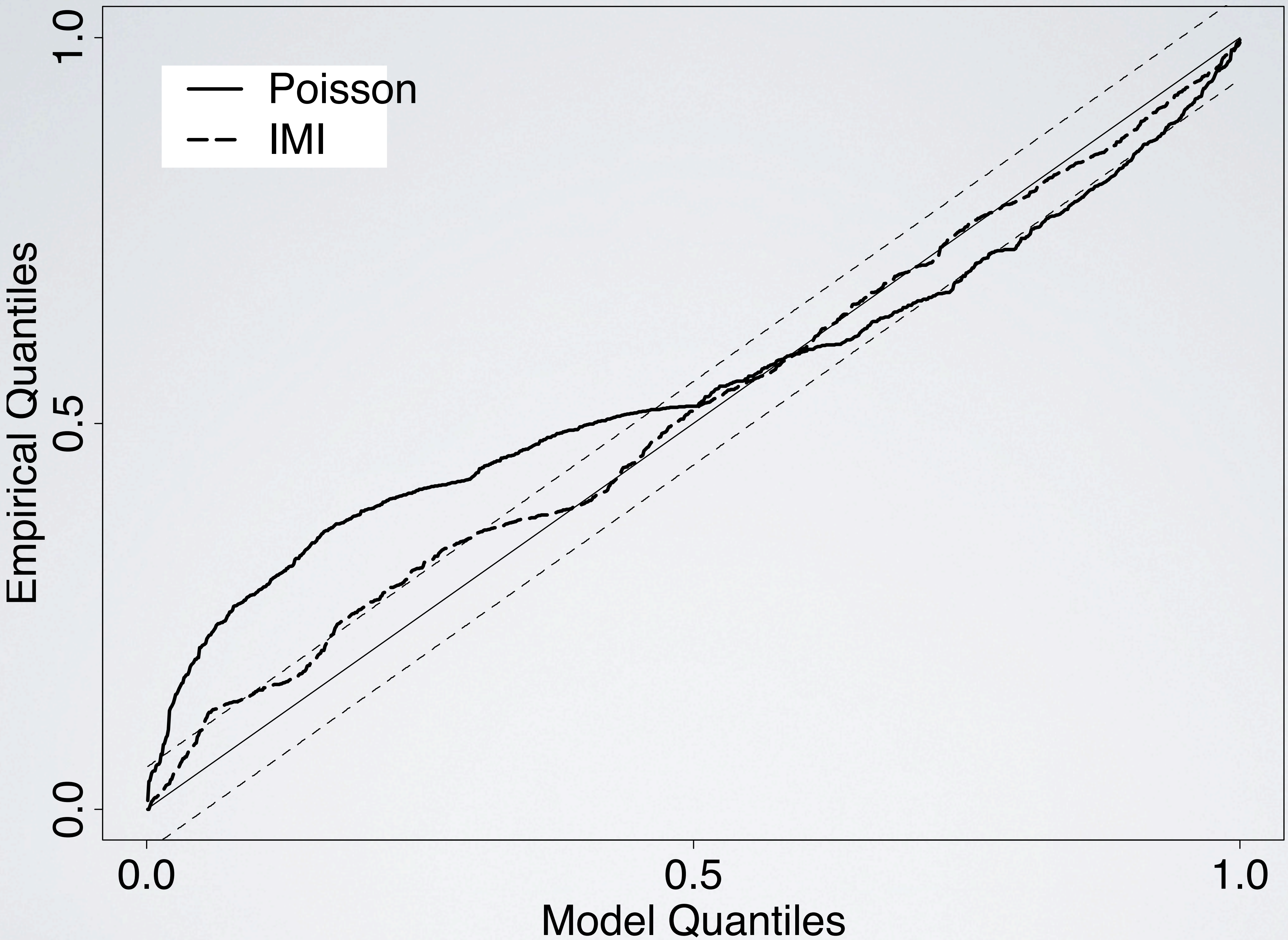
A first-pass (pretty good) solution: Markov model

$$\lambda(t|H_t) = \lambda(t, t - s_*(t))$$

where $s_*(t)$ is time of last spike preceding t

Multiplicative special case:

$$\lambda(t, s_*(t)) = \lambda_1(t) g_1(t - s_*(t))$$



Poisson
IMI
Empirical Quantiles
Model Quantiles
0.0
0.5
1.0
0.0
0.5
1.0

Spike train modeling within trials

conditional intensity

$$FR = \lambda(t|H_t)$$

Modeling starting point: $\log \lambda(t|H_t) = f(t, H_t, x)$

Right-hand side typically involves several additive terms

Jacobs et al. experiment (2009, *PNAS*)

Neural coding question: Are spike counts enough to decode stimulus? Is marginal firing rate (Poisson process) enough? Or is relative spike timing needed (non-Poisson process)?

Jacobs et al. experiment (2009, *PNAS*)

Neural coding question: Are spike counts enough to decode stimulus? Is marginal firing rate (Poisson process) enough? Or is relative spike timing needed (non-Poisson process)?

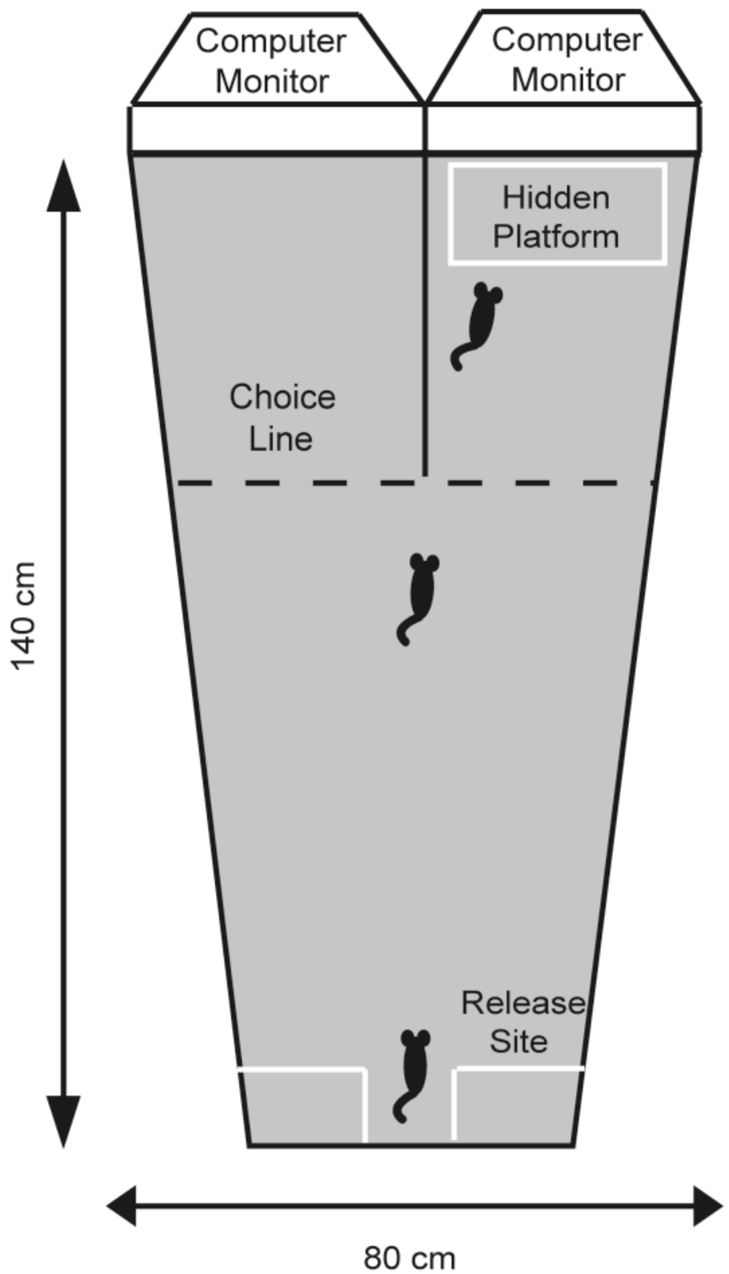
Idea: compare retinal output *in vitro* to animal behavior

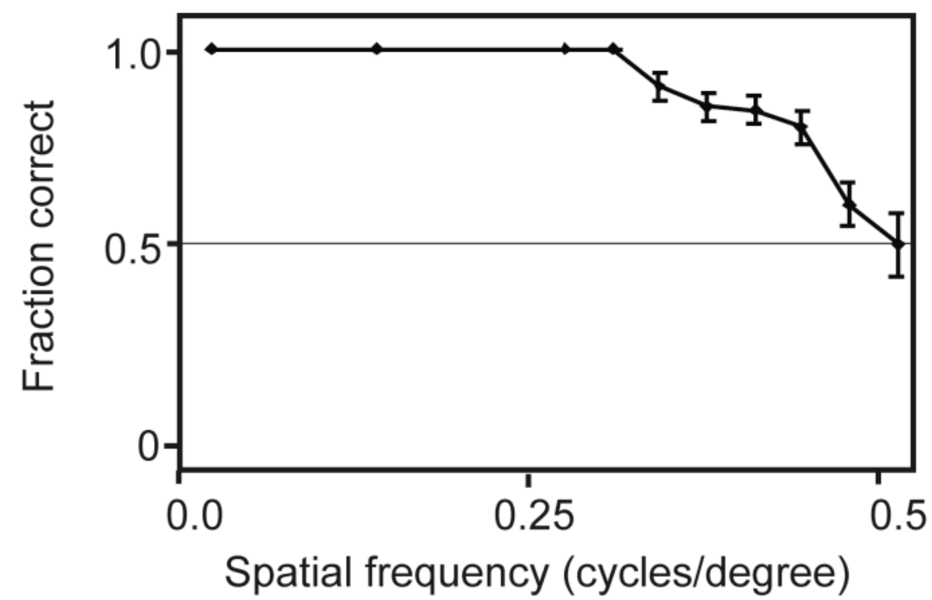
Jacobs et al. experiment (2009, *PNAS*)

Neural coding question: Are spike counts enough to decode stimulus? Is marginal firing rate (Poisson process) enough? Or is relative spike timing needed (non-Poisson process)?

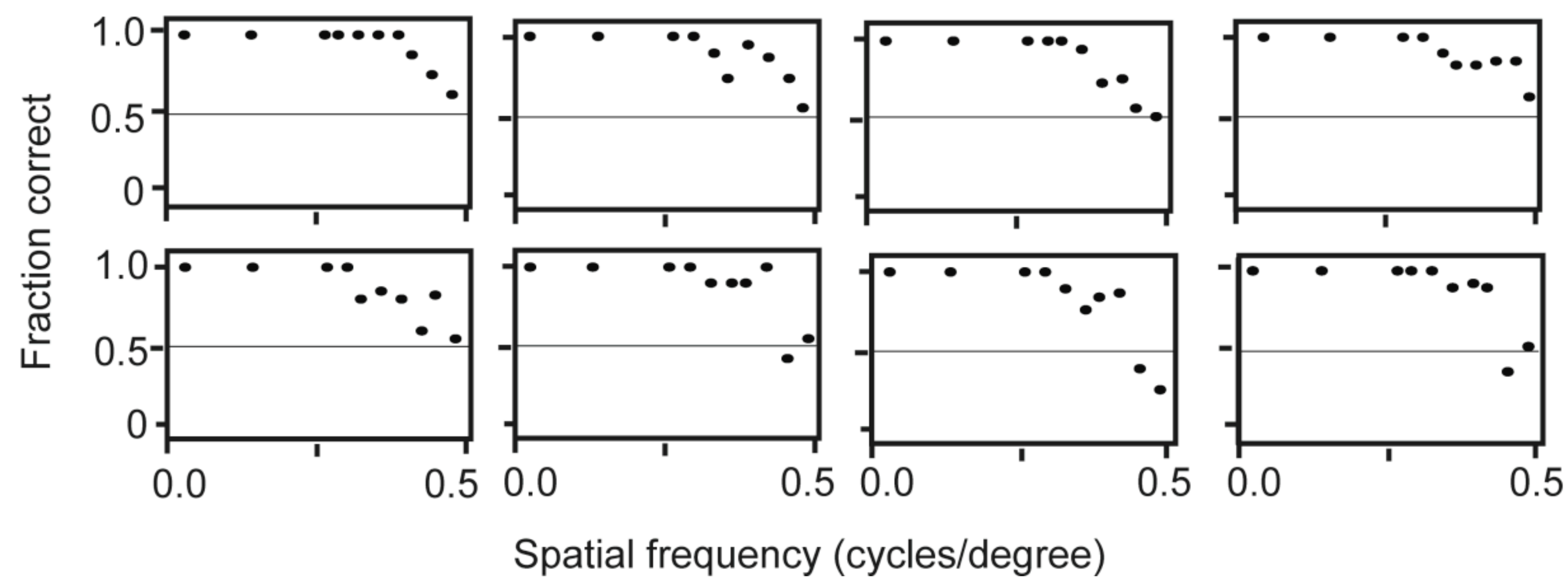
Idea: compare retinal output *in vitro* to animal behavior

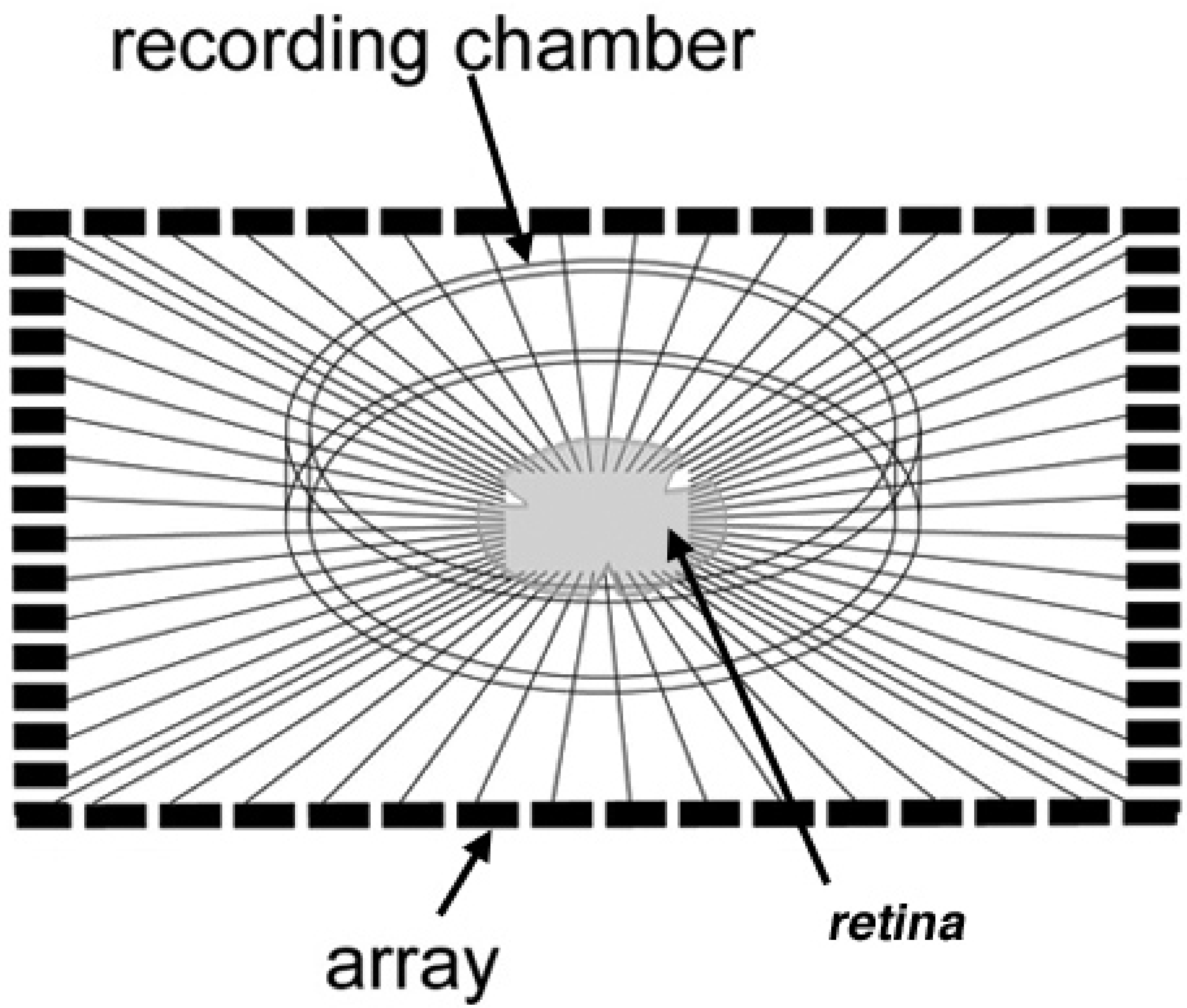
--> examine stimulus-related information contained in retinal spike trains, see how it accounts for behavior


Computer Monitor
Computer Monitor
Hidden Platform
Choice Line
Release Site
140 cm
80 cm


1.0
0.5
0
Fraction correct
0.0
0.25
0.5
Spatial frequency (cycles/degree)

C


1.0
0.5
0
Fraction correct
0.0
0.5
Spatial frequency (cycles/degree)


recording chamber
array
retina

Logic:

1. If all relevant retinal output is recorded, then all stimulus-related information must be contained in the retinal spike trains.

2. If spike counts suffice then behavioral response should be recoverable by a Bayes classifier based on spike counts.

Aside on Bayesian inference and Bayesian decoding

Bayes' Theorem

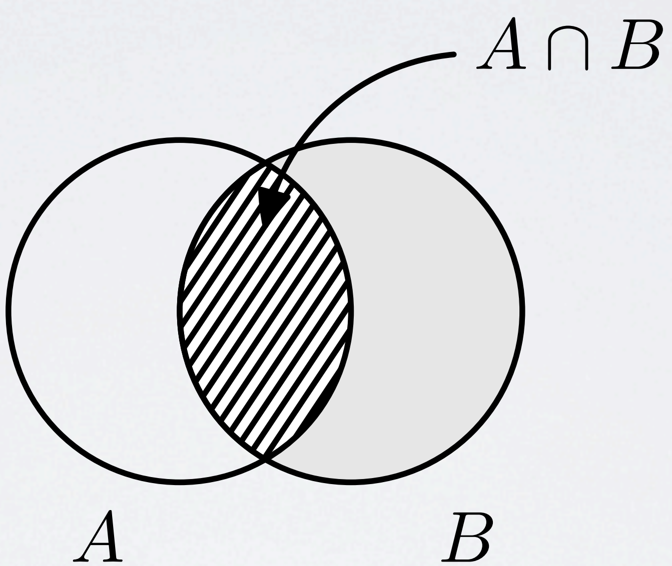
$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

becomes

$$P(H|\text{data}) = \frac{\mathrm{P(data|H)P(H)}}{\mathrm{P(data|H)P(H) + P(data|H^c)P(H^c)}}$$

Definition: Conditional Probability Assume $P(B > 0)$. The conditional probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$



Definition: Independence Two events A and B are independent if and only if $P(A \cap B) = P(A) \cdot P(B)$.

If independent then $P(A|B) = P(A)$

Theorem: Law of Total Probability For events A and B we have

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c).$$

Proof: We begin by decomposing B into two pieces: $B = (B \cap A) \cup (B \cap A^c)$. Because A and A^c are disjoint, $(B \cap A)$ and $(B \cap A^c)$ are disjoint. We then have $P(B) = P(B \cap A) + P(B \cap A^c)$. Applying the multiplication rule to $P(B \cap A)$ and $P(B \cap A^c)$ gives the result. $\square$

Bayes' Theorem in the Simplest Case If $P(B) > 0$ then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}. \tag{3.1}$$

Aside on Bayesian inference and Bayesian decoding

Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

becomes

$$P(H|\text{data}) = \frac{\mathrm{P(data|H)P(H)}}{\mathrm{P(data|H)P(H) + P(data|H^c)P(H^c)}}$$

Uses of Probability

- To describe variation
  - using alleatory or phenomenological or descriptive probability
  - analogous to saying probability of rolling a 3 with an apparently fair die is 1/6.
- To quantify uncertain knowledge
  - by making epistemic or inferential statements
  - analogous to saying you are 90% sure that the capital of Louisiana is Baton Rouge.

Example 1.2.1 Blindsight in patient P.S. Marshall and Halligan (1988, *Nature*, 336: 766–767) reported an interesting neuropsychological finding from a particular patient, identified as P.S. This patient was a 49 year-old woman who had suffered damage to her right parietal cortex that reduced her capacity to process visual information coming from the left side of her visual space. For example, she would frequently read words incorrectly by omitting left-most letters ("smile" became "mile") and when asked to copy simple line drawings, she accurately drew the right-hand side of the figures but omitted the left-hand side without any conscious awareness of her error. To show that she could actually see what was on the left but was simply neglecting it—a phenomenon known as *blindsight*—the examiners presented P.S. with a pair of cards showing identical green line drawings of a house, except that on one of the cards bright red flames were depicted on the left side of the house. They presented to P.S. both cards, one above the other (the one placed above being selected at random), and asked her to choose which house she would prefer to live in. She thought this was silly "because they're the same" but when forced to make a response chose the non-burning house on 14 out of 17 trials. This would seem to indicate that she did, in fact, see the left side of the drawings but was unable to fully process the information. But how convincing is it that she chose the non-burning house on 14 out of 17 trials? Might she have been guessing?

Example 1.2.1 (Continued) Let $Y \sim B(n,p)$ and note that the usual estimator of p is sample proportion $T = \hat{p} = Y/n$. Because $V(Y) = np(1-p)$ we have $V(T) = p(1-p)/n^2$. Thus, we have the formula

$$\sigma_T = \sqrt{\frac{p(1-p)}{n}}. \tag{7.2}$$

The formula in Equation (7.2) quantifies the variation we can associate with the observed proportion $\hat{p} = 14/17 = .824$. However, we can not compute a numerical value for σ_T from Equation (7.2) because we do not know what value of p to use. The obvious solution is to substitute $\hat{p}$ for p in Equation (7.2). When we do this we obtain the *standard error* for the binomial proportion

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}. \tag{7.3}$$

Applying this to the data from P.S. we get

$$SE = \sqrt{\frac{\frac{14}{17}(1 - \frac{14}{17})}{17}} = .092.$$

We then typically write the estimate in the form $.824 \pm .092$, with the $\pm$ indicating that the likely variability in the estimate is .092. $\square$

Result If $Y \sim B(n,p)$ then p may be estimated by $\hat{p} = Y/n$ with standard error $\mathrm{S}E(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. For large n, an approximate 95% CI is given by

$$\hat{p} \pm 2 \cdot SE(\hat{p}),$$

meaning that for n sufficiently large we have

$$P(\hat{p} - 2 \cdot SE(\hat{p}) \leq p \leq \hat{p} + 2 \cdot SE(\hat{p})) \approx .95. \qquad (7.14)$$

For P.S. data get approximate 95% CI: (.64,1.0)

Here are two interpretations of the confidence interval found for the propensity p of P.S. to choose the non-burning house:

> *Interpretation A:* If p were the true value, then the probability that the interval given by (7.14) would contain p is approximately 95%. Based on the data from P.S., the CI is (.64,1.0).
>
> *Interpretation B:* Based on the data from P.S., the probability that (.64,1.0) contains p is approximately 95%.

Bayesian Inference

- Founded on an inferential principle of equivalence: there is only one kind of probability for *both* descriptive and inferential purposes, with inferential statements being obtained from descriptive ones merely by applying the laws of probability.
- Has provably good theoretical properties; tends to work well in practice when applied carefully.
- Generally provides optimal classification, if probabilities can be calculated.

Recall Bayes' Theorem for random variables and vectors: for continuous random variables or vectors U and V we have

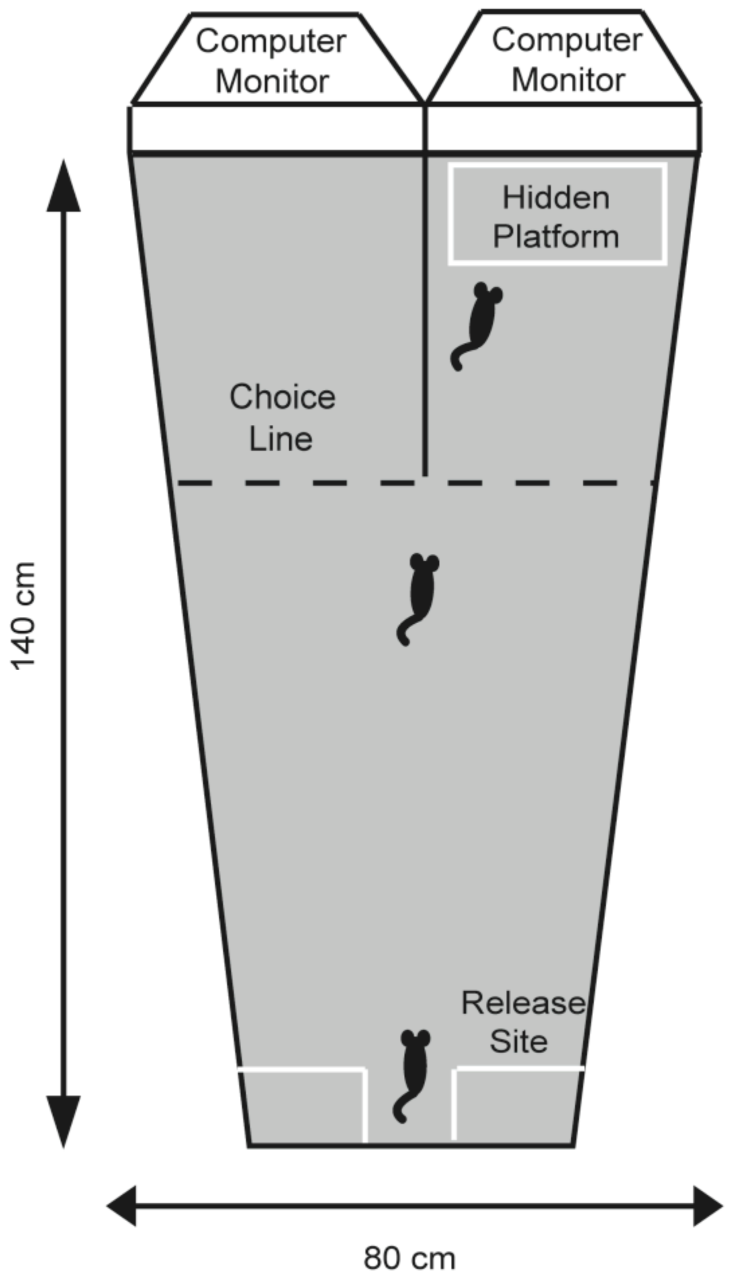
$$f_{U|V}(u|v) = \frac{f_{V|U}(v|u) f_U(u)}{\int f_{V|U}(v|u) f_U(u) du}. \tag{7.17}$$

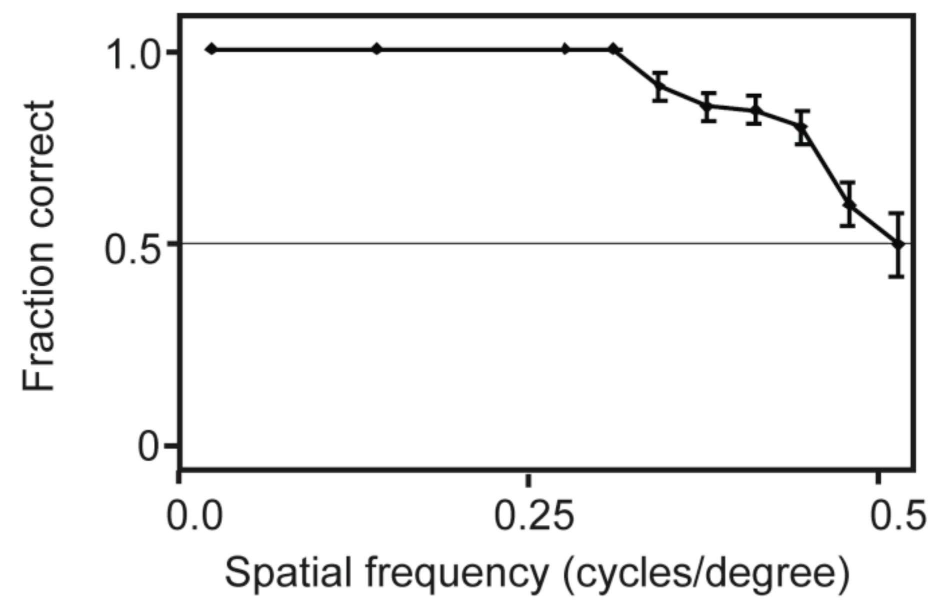
Logic:

1. If all relevant retinal output is recorded, then all stimulus-related information must be contained in the retinal spike trains.

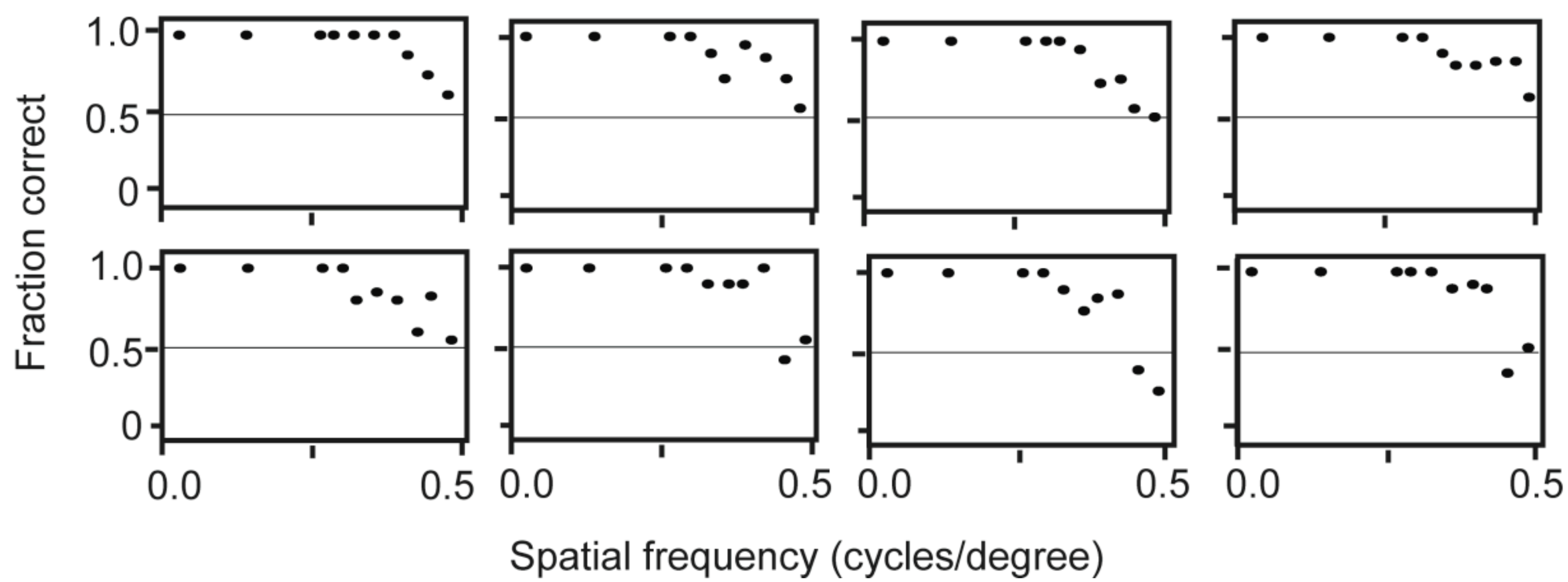
2. If spike counts suffice then behavioral response should be recoverable by a Bayes classifier based on spike counts.

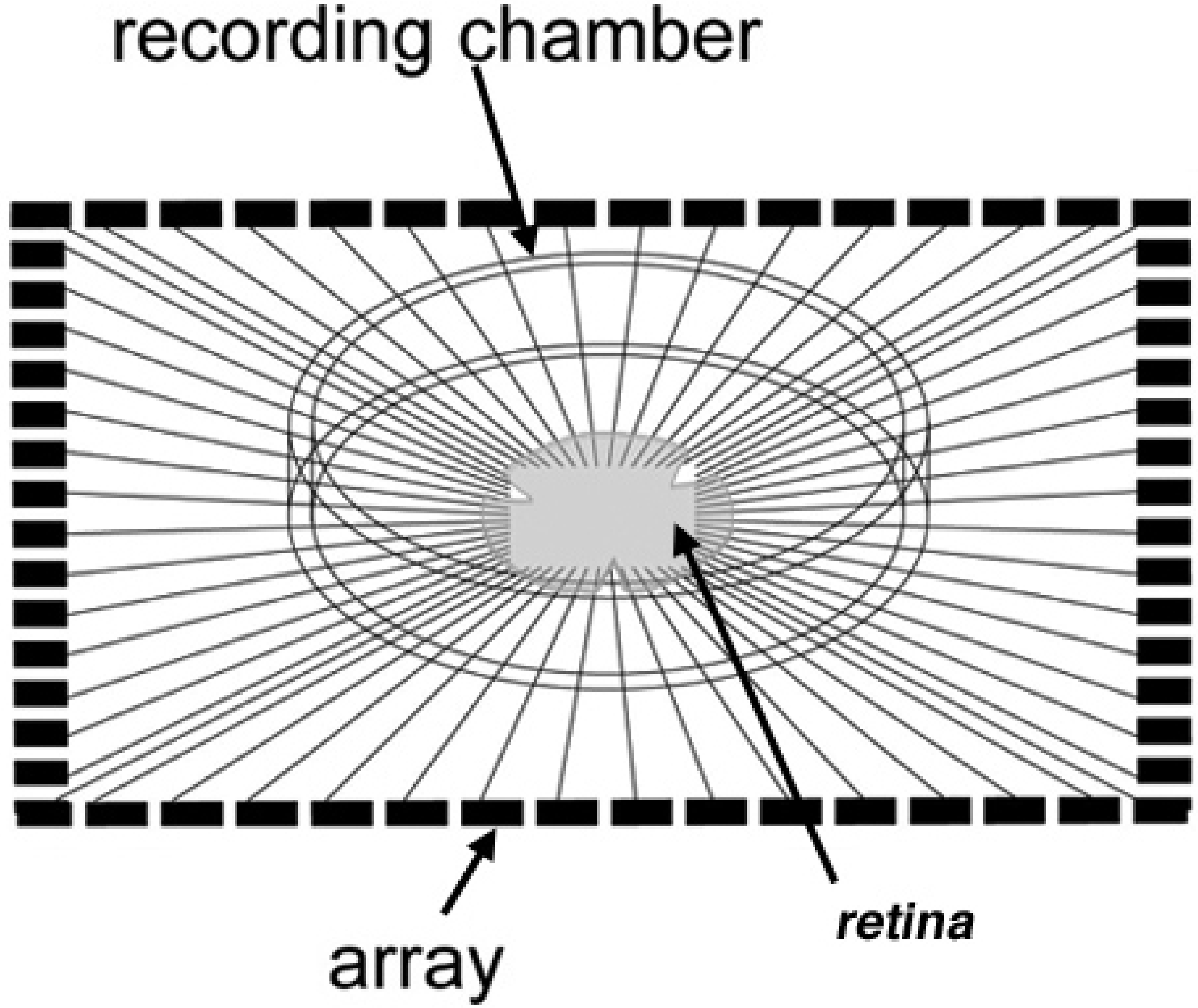
--> the nervous system could be less efficient than a Bayes classifier but it can't possibly be more efficient


Computer Monitor
Computer Monitor
Hidden Platform
Choice Line
Release Site
140 cm
80 cm


1.0
0.5
0
Fraction correct
0.0
0.25
0.5
Spatial frequency (cycles/degree)

C


1.0
0.5
0
Fraction correct
0.0
0.5
Spatial frequency (cycles/degree)


recording chamber
array
retina

Logic:

1. If all relevant retinal output is recorded, then all stimulus-related information must be contained in the retinal spike trains.

2. If spike counts suffice then behavioral response should be recoverable by a Bayes classifier based on spike counts.

3. If Bayes classifier based on spike counts *fails* to reproduce behavior, then nervous system must use some form of spike timing information.

Logic:

1. If all relevant retinal output is recorded, then all stimulus-related information must be contained in the retinal spike trains.

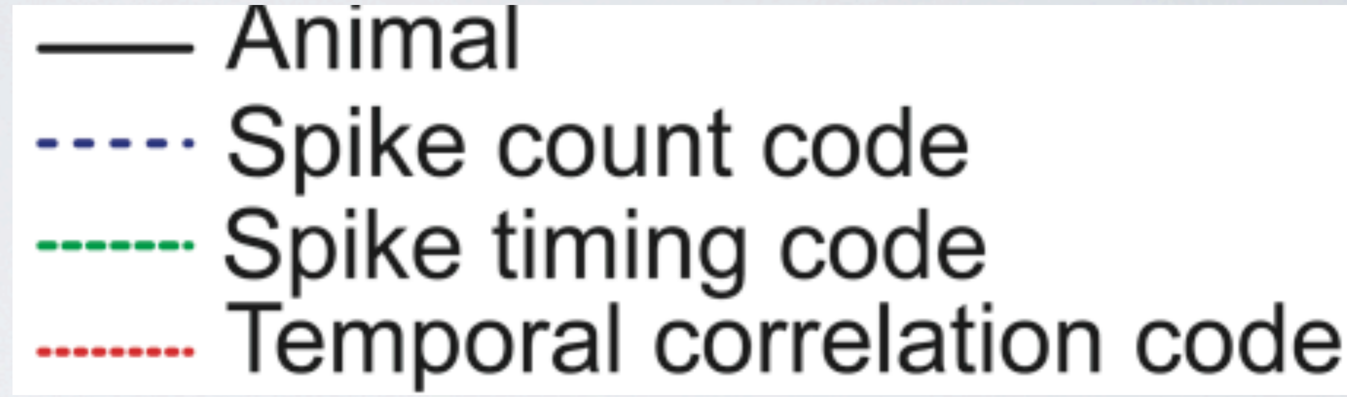
2. If spike counts suffice then behavioral response should be recoverable by a Bayes classifier based on spike counts.

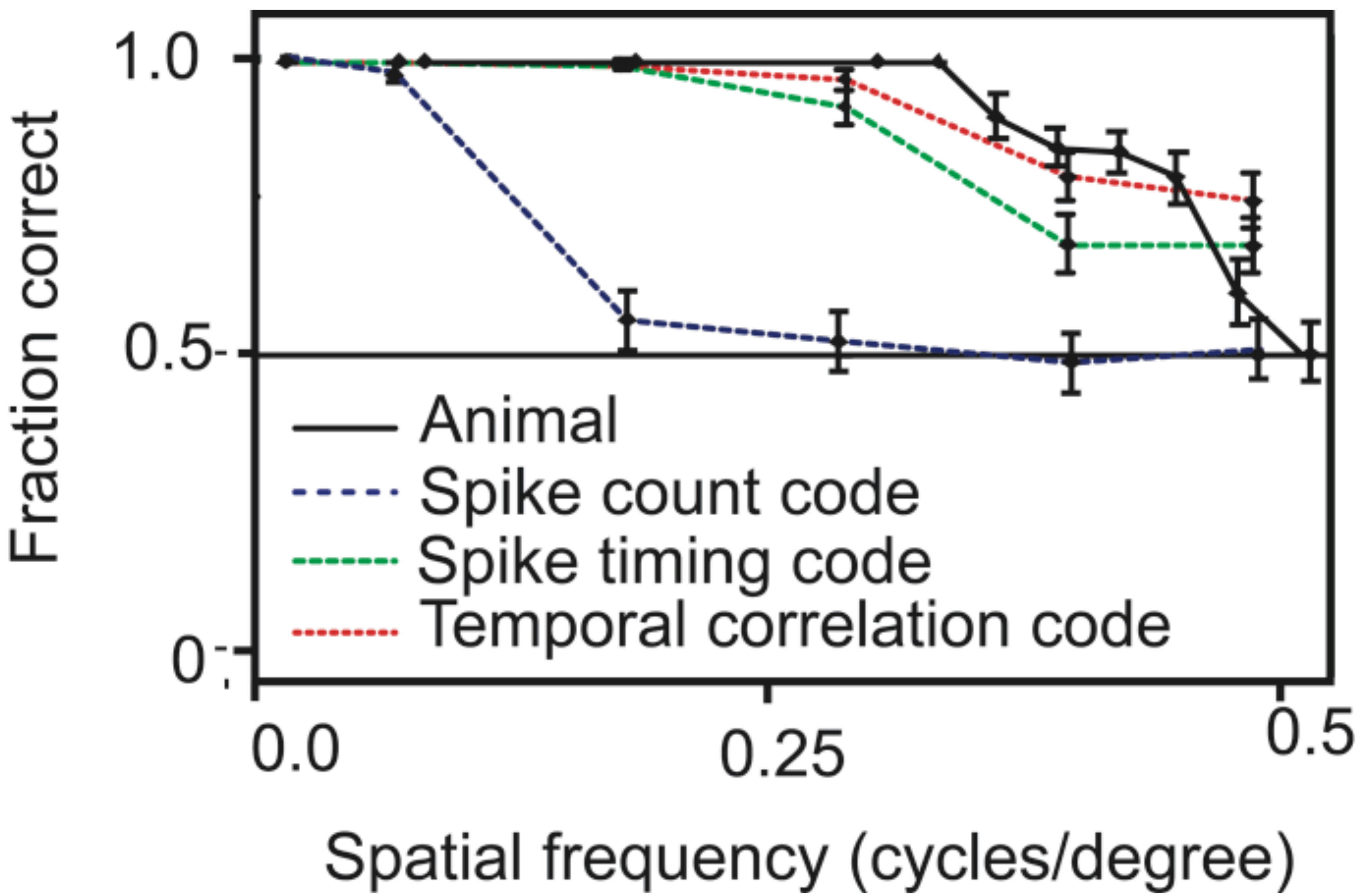
3. If Bayes classifier based on spike counts *fails* to reproduce behavior, then nervous system must use some form of spike timing information.

4. Similarly, if Bayes classifier based on marginal firing rate fails, nervous system must use relative timing of successive spikes.

inhomogeneous Poisson →

IMI model →




1.0
0.5
0
Fraction correct
0.0
0.25
0.5
Spatial frequency (cycles/degree)
Animal
Spike count code
Spike timing code
Temporal correlation code

Statistical model of Poisson spiking fails, while non-Poisson spiking succeeds.

Conclusion: nervous system *must* be relying on non-Poisson coding of retinal output.

End lecture 1. Next overview lecture: more elaborate recordings, and more on statistical models and inferences [preview this slide and next]