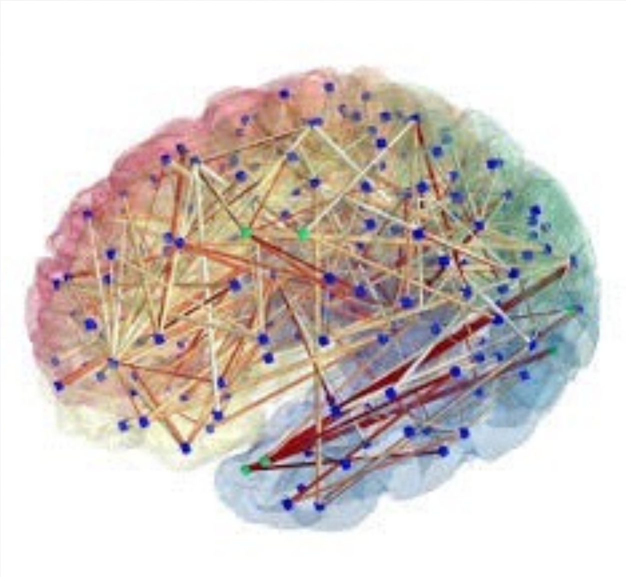
STATISTICAL CONSIDERATIONS IN MAKING INFERENCES ABOUT NEURAL NETWORKS: THE CASE OF SYNCHRONY DETECTION

Rob Kass

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

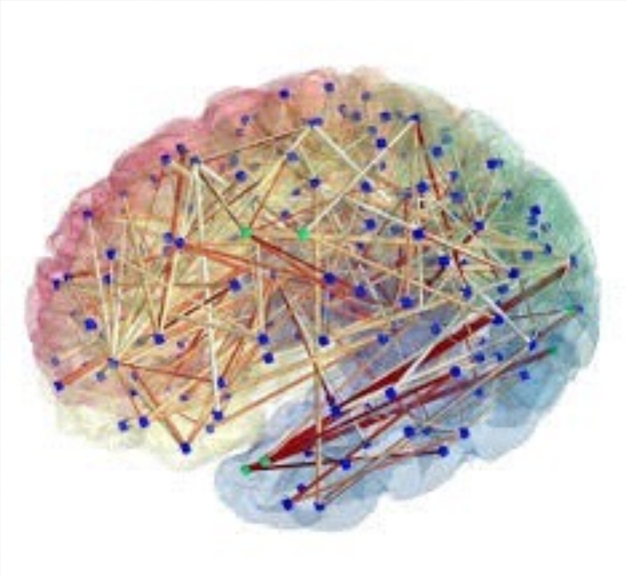
Common statistical problem in noisy networks:

infer "connectivity" among lots of neural signals based on spike trains, LFP, MEG, etc.



Common statistical problem in noisy networks:

infer “connectivity” among lots of neural signals based on spike trains, LFP, MEG, etc.



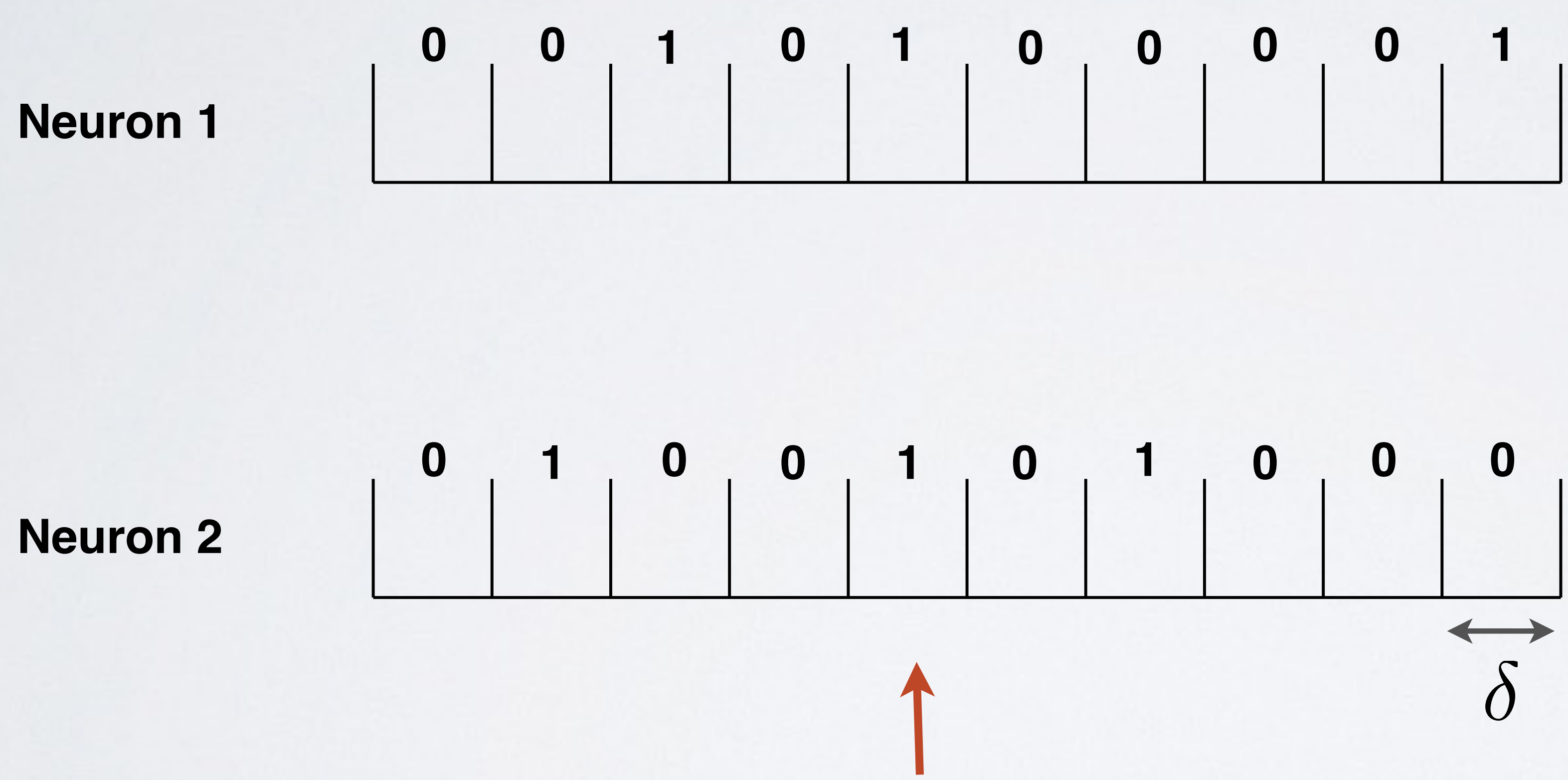
Context here: pairwise spike synchrony across many neurons

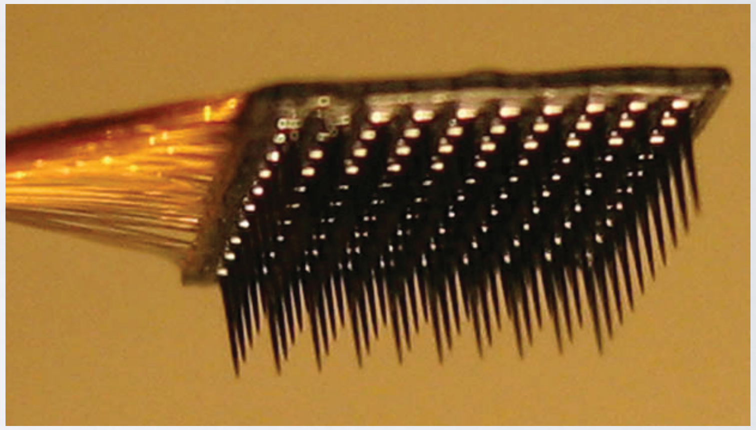
infer “connectivity” among lots of neural signals
based on spike trains, LFP, MEG, etc.

“infer” means
-- get good estimates of parameters
-- get assessments of uncertainty
and/or significance tests

Synchronous firing among neurons

two or more neurons fire nearly at the same time
i.e. within a bin of width δ



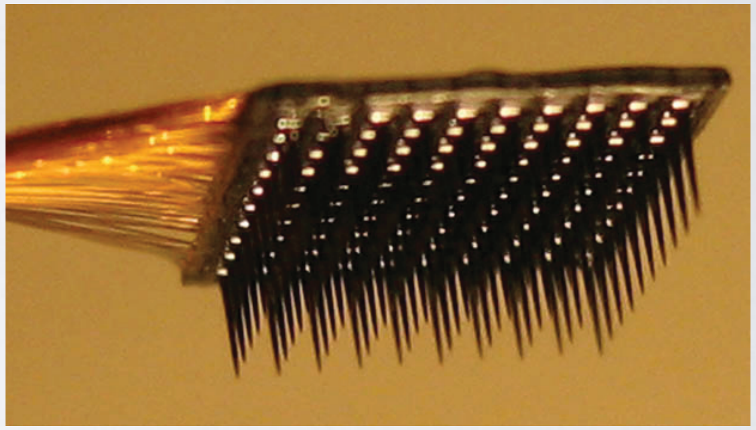


128 neurons

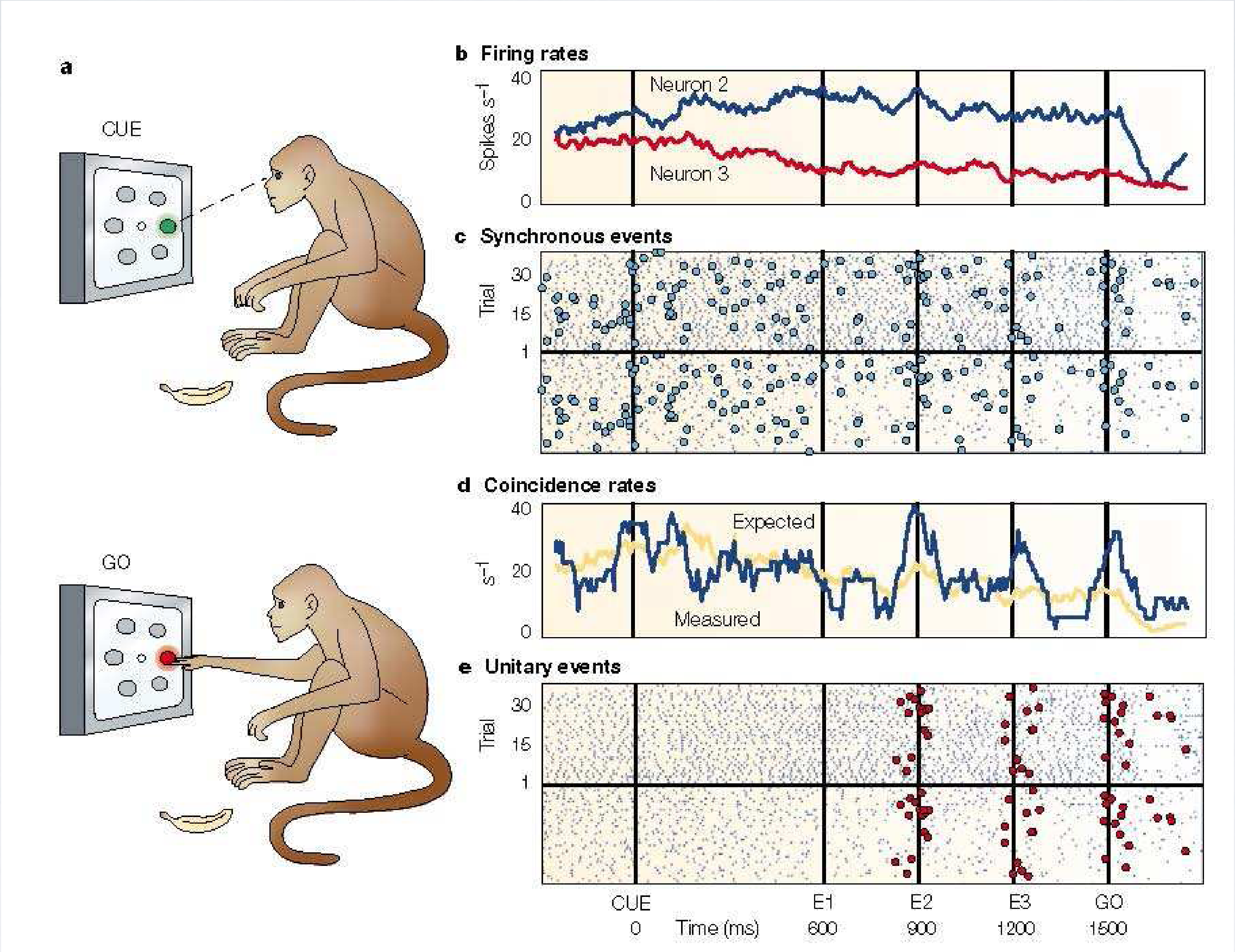
Synchronous firing among neurons

two or more neurons fire nearly at the same time
i.e. within a bin of width δ

here $\delta = 5$ ms



128 neurons



Riehle … Gruen … (1997, *Science*)

My view

There need not be any cortical mechanism for synchrony detection.

Synchrony is a feature of spike train data that needs to be explained.

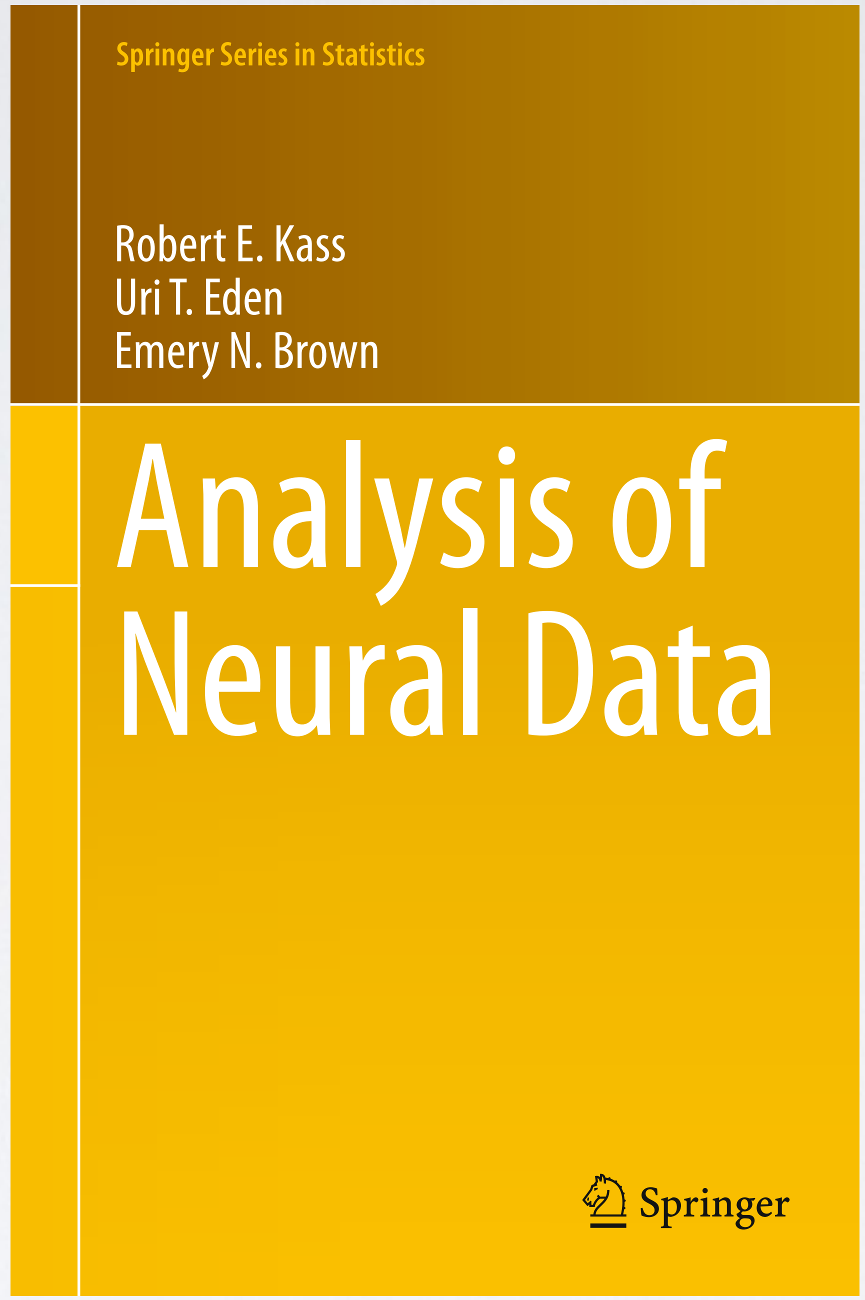
Can be subtle:

Series of papers by Geman, Amarasingham, Harrison …
see Harrison, Amarasingham, Kass (2013) review

We devised a methodology based on point-process regression (a.k.a. GLMs)

Kass, Kelly, Loh (2011, *Ann. Applied Statistics*)
Kelly, Smith, Kass, Lee (2010, *J. Comput. Neurosci.*)
Kelly, Smith, Kass, Lee (2010, *NIPS*)
Kelly and Kass (2012, *Neural Comput.*)
Scott, Kelly, Smith, Zhou, Kass (2015, *J. Amer. Statist. Assoc.*)

Background material may be found in 2014 book



would like to discuss off-line
the close connection of GLM and LIF modeling

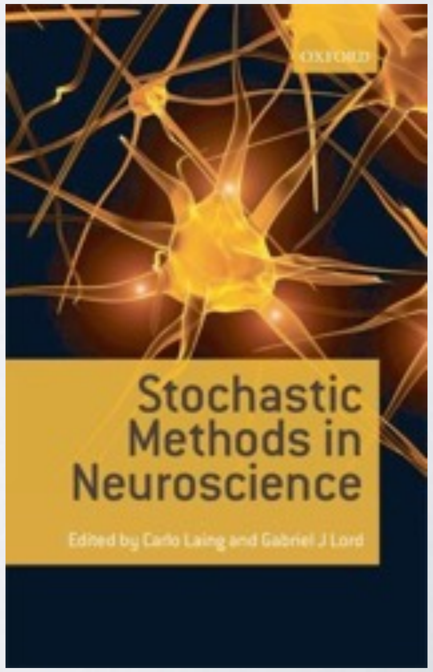
GLM:

$$\log \lambda(t|H_t, I_t) = \int_0^\infty f_1(s) I(t-s)\,ds + f_2(t - s_*(t))$$

Integrated form of LIF:

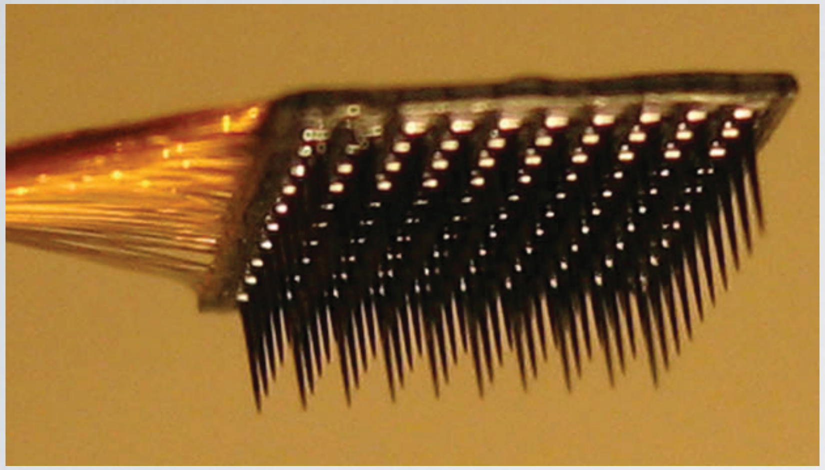
$$V(t) = V_{\text{rest}} + \int_0^\infty f_1(s) I(t-s)\,ds + f_2(t - s_*(t))$$

Paninski, Brown, Iyengar, Kass (2010)



OUTLINE

- Quick summary of previous results
- The problem of false discoveries
- The Bayesian approach to hypothesis testing
- Bayesian protection against false discoveries
- Comments on reproducibility



slide from
Ryan Kelly (Google)

V1 data

128 neurons

We used the statistic

$$\hat{\zeta} = \frac{\text{number of observed simultaneous spikes}}{\text{number of predicted simultaneous spikes}}$$

$$\hat{\zeta} = \frac{\text{number of observed simultaneous spikes}}{\text{number of predicted simultaneous spikes}}$$

prediction from fitted GLM

$$\hat{\zeta} = \frac{\text{number of observed simultaneous spikes}}{\text{number of predicted simultaneous spikes}}$$

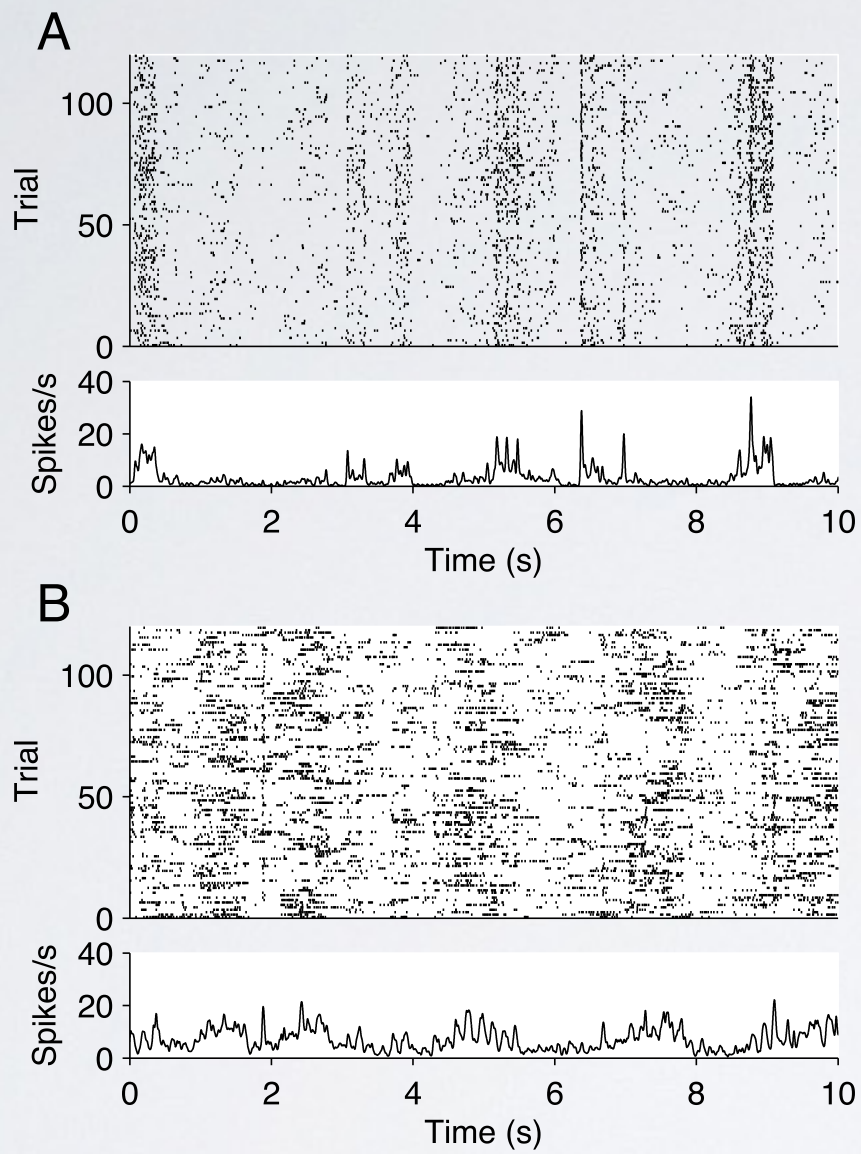
prediction from fitted GLM

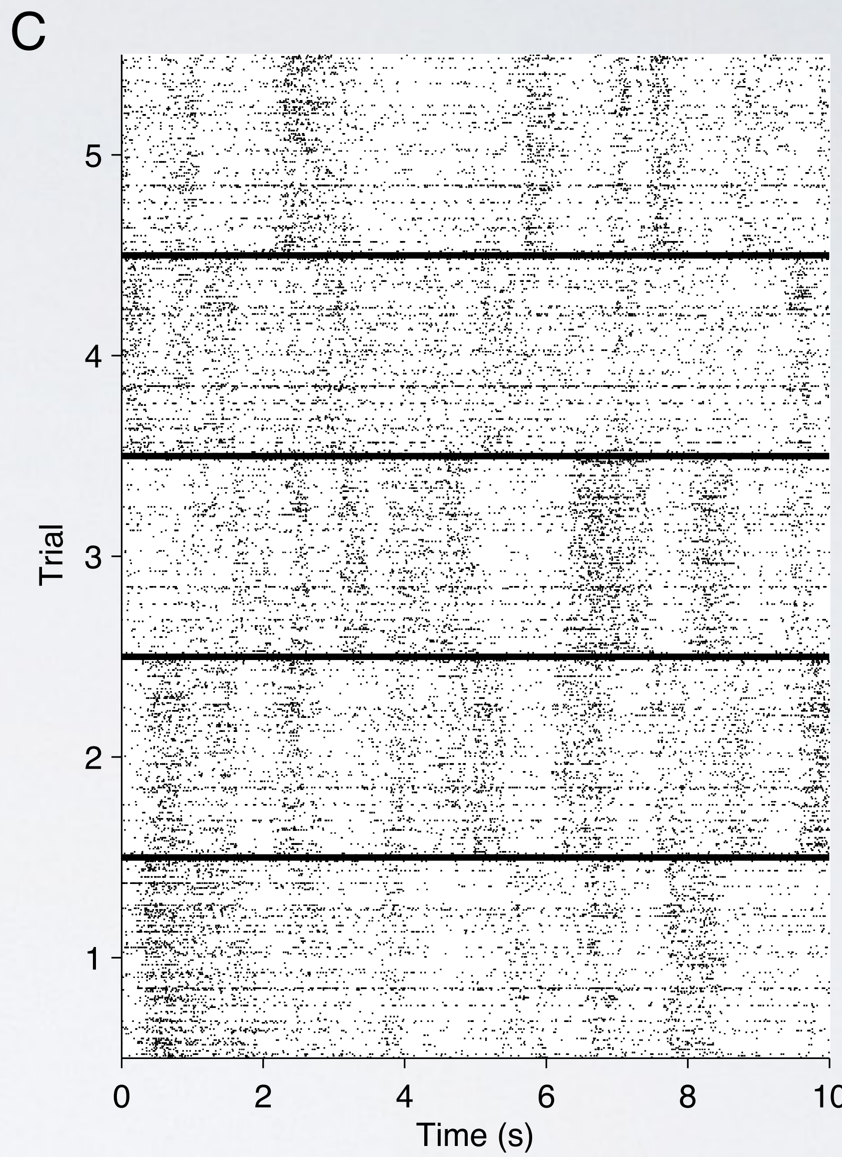
We test H_0: $\zeta = 1$

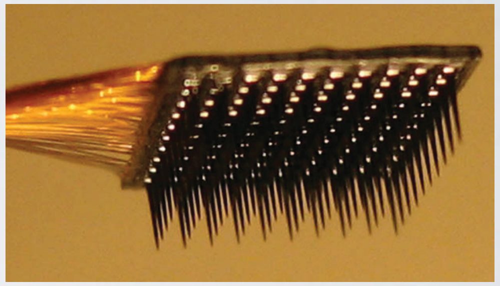
To what extent is excess synchrony, above chance, due to slow-wave oscillation?

1 neuron

a different neuron

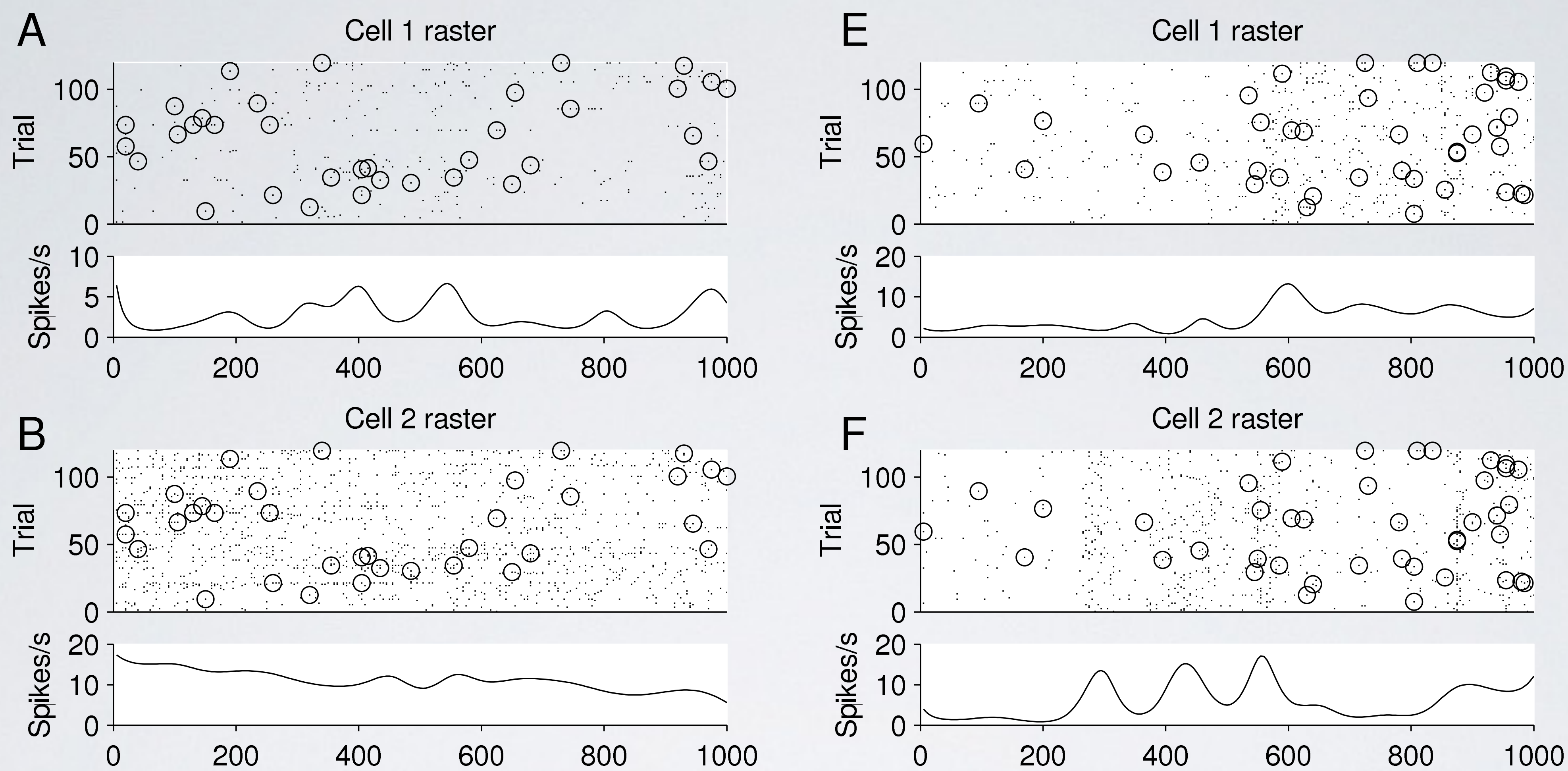






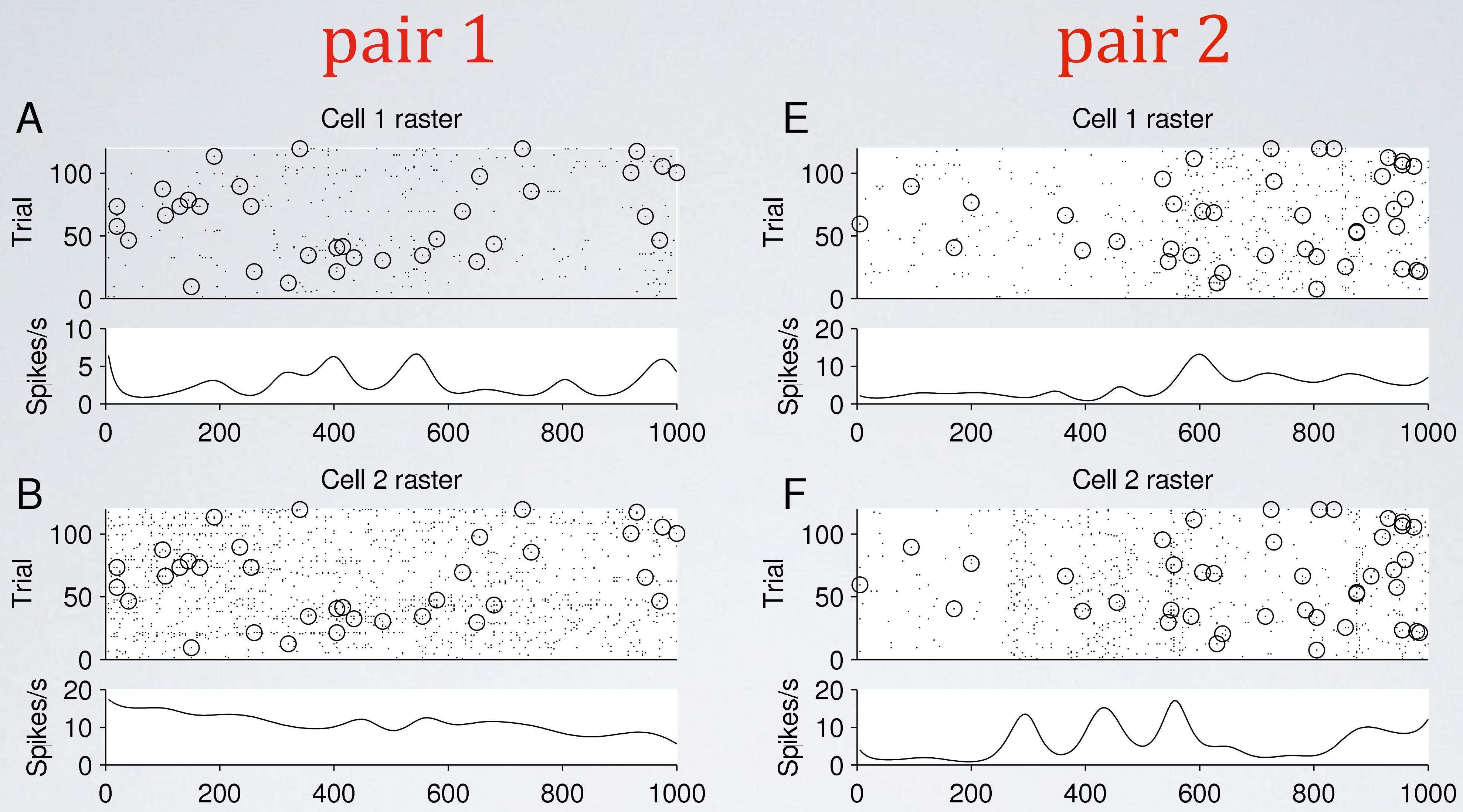
128 neurons

pair 1 pair 2



Both pairs: excess synchrony

next: introduce covariate for slow wave activity



Both pairs: excess synchrony

left pair: synchrony due to slow waves

right pair: not due to slow waves

Here, I am discussing only pairwise synchrony

(though see Kelly and Kass, 2012)

Statistical Issue:

Large number of possible interactions

$$\binom{128}{2} = 8128$$

Statistical Issue:

Large number of possible interactions

$$\binom{128}{2} = 8128$$

huge opportunity for false discoveries

problem of false discoveries among n independent tests

if $P(\text{significant } |H_0) = .05$ then
$P(\text{at least 1 of } n \text{ significant } |H_0) = 1 - (1 - .05)^n$

e.g., $n = 14$ then
$$1 - (1 - .05)^n = .51$$

problem of false discoveries among n independent tests

$$\text{if } P(\text{significant } | H_0) = .05 \text{ then}$$
$$P(\text{at least 1 of } n \text{ significant } | H_0) = 1 - (1 - .05)^n$$

$$\text{e.g., } n = 14 \text{ then}$$
$$1 - (1 - .05)^n = .51$$

$$n = 100 \text{ then}$$
$$1 - (1 - .05)^n = .994$$

Statistical Issue:

Large number of possible interactions

$$\binom{128}{2} = 8128$$

huge opportunity for false discoveries

soluble problem

Bayes' Theorem:

$$P(H_0|data) = \frac{P(data|H_0)P(H_0)}{P(data|H_0)P(H_0) + P(data|H_A)P(H_A)}$$

$$\begin{aligned} H_0 &== \text{ null hypothesis} \\ H_A &== \text{ alternative hypothesis} \\ P(H_0|data) &== \text{ posterior probability of } H_0 \end{aligned}$$

$$P(H_0|data) = \frac{P(data|H_0)P(H_0)}{P(data|H_0)P(H_0) + P(data|H_A)P(H_A)}$$

$$\text{Bayes factor } = \frac{P(data|H_0)}{P(data|H_A)}$$

$$P(H_0|data) = \frac{P(data|H_0)P(H_0)}{P(data|H_0)P(H_0) + P(data|H_A)P(H_A)}$$

Kass and Raftery (1995)

$$P(H_0|data) = \frac{P(data|H_0)P(H_0)}{P(data|H_0)P(H_0) + P(data|H_A)P(H_A)}$$

$P(data|H_A)$ can be difficult to specify

$$P(H_0|data) = \frac{P(data|H_0)P(H_0)}{P(data|H_0)P(H_0) + P(data|H_A)P(H_A)}$$

wonderful (optimal) BUT requires that
we know the null and alternative distributions

—> can use rich network data to estimate them

We start with 8128 (standardized) synchrony coefficients

$$Z \text{ is distributed as } \pi \cdot f_0(z) + (1 - \pi) \cdot f_1(z)$$

standardized
synchrony coefficient

proportion of null cases

$$Z \text{ is distributed as } \pi \cdot f_0(z) + (1 - \pi) \cdot f_1(z)$$

standardized
synchrony coefficient

null pdf

alternative pdf

proportion of null cases

$$Z \text{ is distributed as } \pi \cdot f_0(z) + (1 - \pi) \cdot f_1(z)$$

standardized
synchrony coefficient

null pdf

alternative pdf

mixture of two distributions

proportion of null cases

$$Z \text{ is distributed as } \pi \cdot f_0(z) + (1-\pi) \cdot f_1(z)$$

standardized
synchrony coefficient

null pdf

alternative pdf

mixture of two distributions

can estimate $f_1(z)$ from data

$$P(data|H_A)$$

for networks we can compute

$$P(H_0|data) = \frac{P(data|H_0)P(H_0)}{P(data|H_0)P(H_0) + P(data|H_A)P(H_A)}$$

related to FDR, see Efron (2008, *Statistical Science*) for networks we can compute

$$P(H_0|data) = \frac{P(data|H_0)P(H_0)}{P(data|H_0)P(H_0) + P(data|H_A)P(H_A)}$$

ALSO: CAN INCORPORATE COVARIATES

Scott, Kelly, Smith, Zhou, Kass (2015, *J. Amer. Statist. Assoc.*)

proportion of null cases

$$Z \text{ is distributed as } \pi(x) \cdot f_0(z) + (1 - \pi(x)) \cdot f_1(z)$$

standardized
synchrony coefficient

null pdf

alternative pdf

ALSO: CAN INCORPORATE COVARIATES

prop. null cases depends on distance
and tuning curve correlation

proportion of null cases

$$Z \text{ is distributed as } \pi(x) \cdot f_0(z) + (1 - \pi(x)) \cdot f_1(z)$$

standardized synchrony coefficient

null pdf

alternative pdf

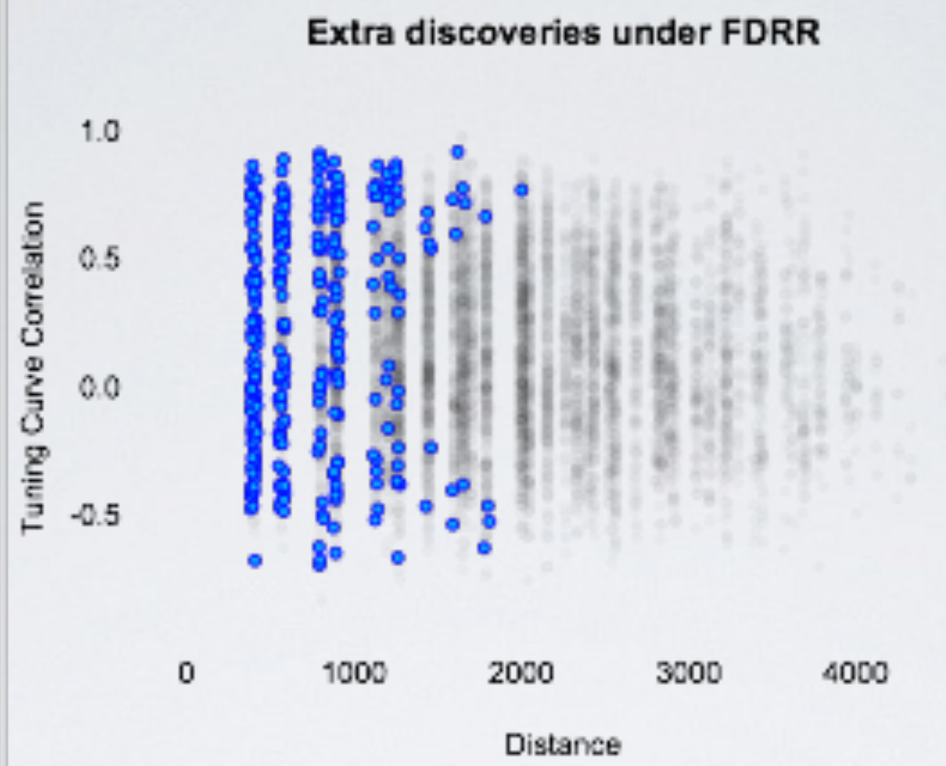
- probability of null depends on covariates
- use "empirical null"
- requires care
- implemented both EB (EM) and Bayes (MCMC)

False Discovery Rate Regression (FDRR)

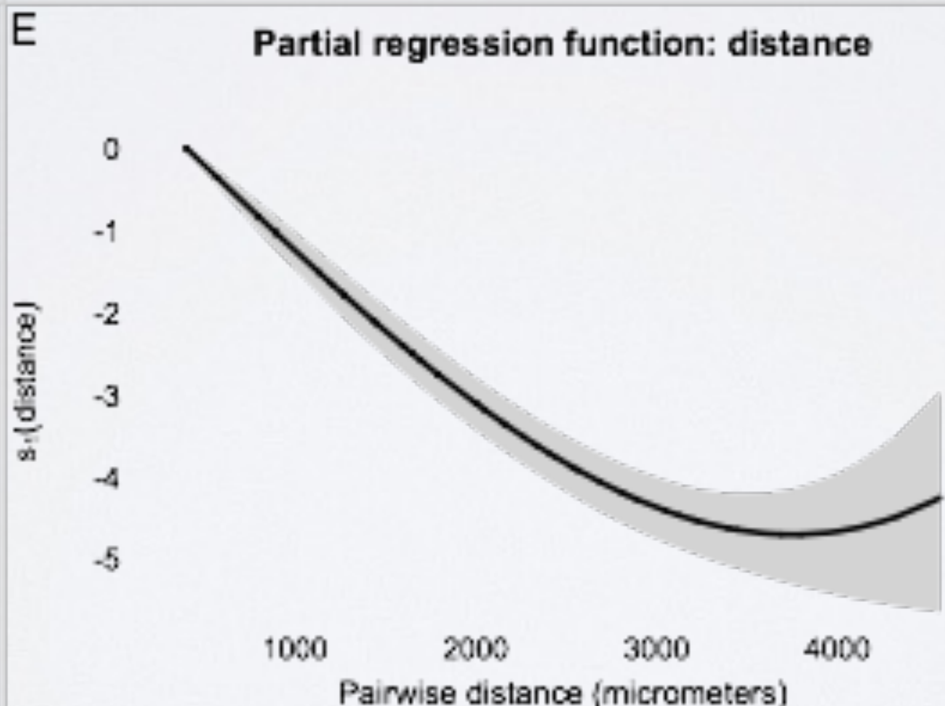
Simulation results:

- pretty robust (see paper)
- get many more discoveries than vanilla FDR

D


Extra discoveries under FDRR
Tuning Curve Correlation
1.0
0.5
0.0
-0.5
0
1000
2000
3000
4000
Distance

E


Partial regression function: distance
s_1(distance)
0
-1
-2
-3
-4
-5
1000
2000
3000
4000
Pairwise distance (micrometers)

But this is still subject to error
(some non-trivial fraction will be false discoveries)

consider those pairs that have
high probability of synchrony

--> use second set of data!
use fitted alternative to find

$$P(data|H_A)$$

and then compute posterior probabilities

Summary:

using Bayesian protection against false discovery
AND
data splitting

can get strong statistical inferences for large numbers of interactions

Summary:

using Bayesian protection against false discovery
AND
data splitting, with BFs for new data

can get strong statistical inferences for large numbers of interactions

The crisis in reproducibility

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For A

News & Comment > Research Highlights > Social Selection > 2015 > March > Article

NATURE | RESEARCH HIGHLIGHTS: SOCIAL SELECTION

A blueprint to boost reproducibility of results

Online commenters show support for a call to shake up science.

Chris Woolston

29 October 2014

Rights & Permissions

Nearly a decade after writing a scathing critique of biomedical research, 'Why Most Published Research Findings Are False', health-policy researcher John Ioannidis has published a follow-up.

Ioannidis, at Stanford University in California, suggests a blueprint for making scientific results more reliable, including increasing the statistical certainty of discoveries, giving more weight to negative results and changing how researchers earn kudos[1].

Altmetric

Based on data from Altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For

Archive > Volume 515 > Issue 7525 > Editorial > Article

NATURE | EDITORIAL

Journals unite for reproducibility

Consensus on reporting principles aims to improve quality control in biomedical research and encourage public trust in science.

05 November 2014

nature International weekly journal of science

NATURE | NEWS

Weak statistical standards implicated in scientific irreproducibility

One-quarter of studies that meet commonly used statistical cutoff may be false.

Erika Check Hayden

11 November 2013

nature International weekly journal of science

NATURE | NEWS FEATURE

عربي

Scientific method: Statistical errors

P **values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume.**

Regina Nuzzo

12 February 2014

nature International weekly journal of science

NATURE | RESEARCH HIGHLIGHTS: SOCIAL SELECTION

Psychology journal bans *P* values

Test for reliability of results 'too easy to pass', say editors.

Chris Woolston

26 February 2015

p-value fallacy:
interpretation of p-value as

$$P(H_0|\ \mathrm{data})$$

p-value fallacy:

interpretation of p-value as

$$P(H_0|\text{ data})$$

instead: could compute this using Bayes’ theorem
(via *Bayes factors*)

$$P(H_0|data) = \frac{P(data|H_0)P(H_0)}{P(data|H_0)P(H_0) + P(data|H_A)P(H_A)}$$

$P(data|H_A)$ can be difficult to specify

$$P(H_0|data) = \frac{P(data|H_0)P(H_0)}{P(data|H_0)P(H_0) + P(data|H_A)P(H_A)}$$

$P(data|H_A)$ can be difficult to specify

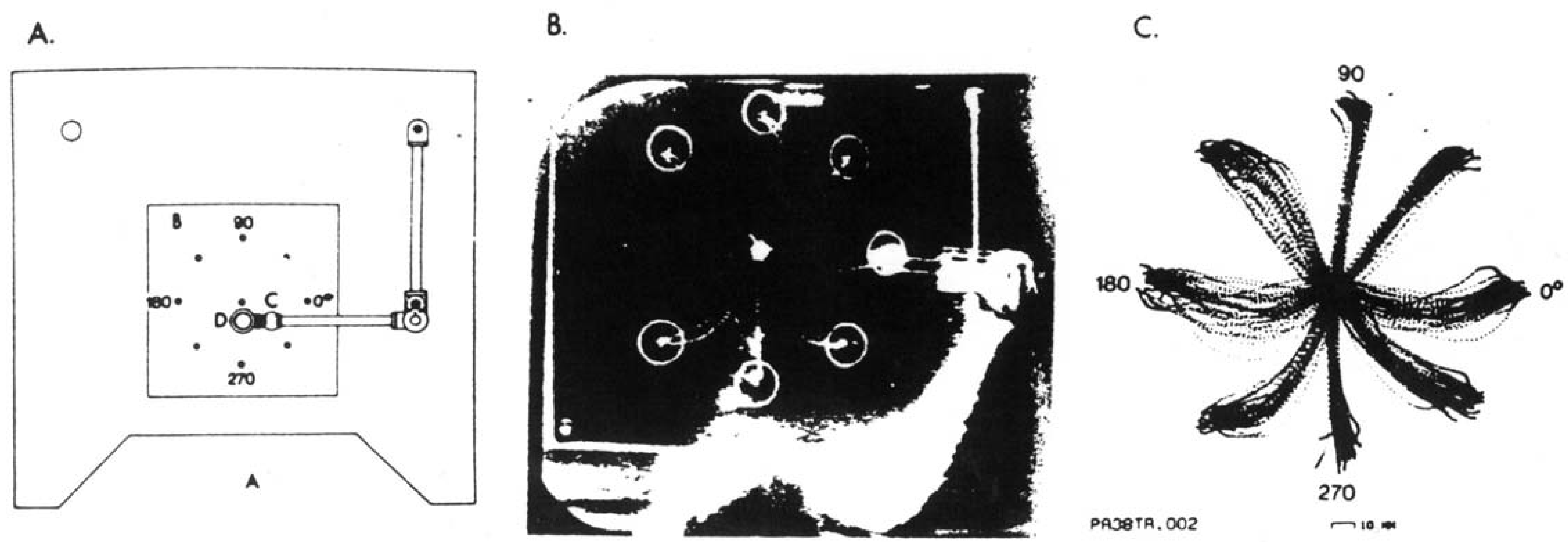
perhaps this indicates a defect in our scientific process

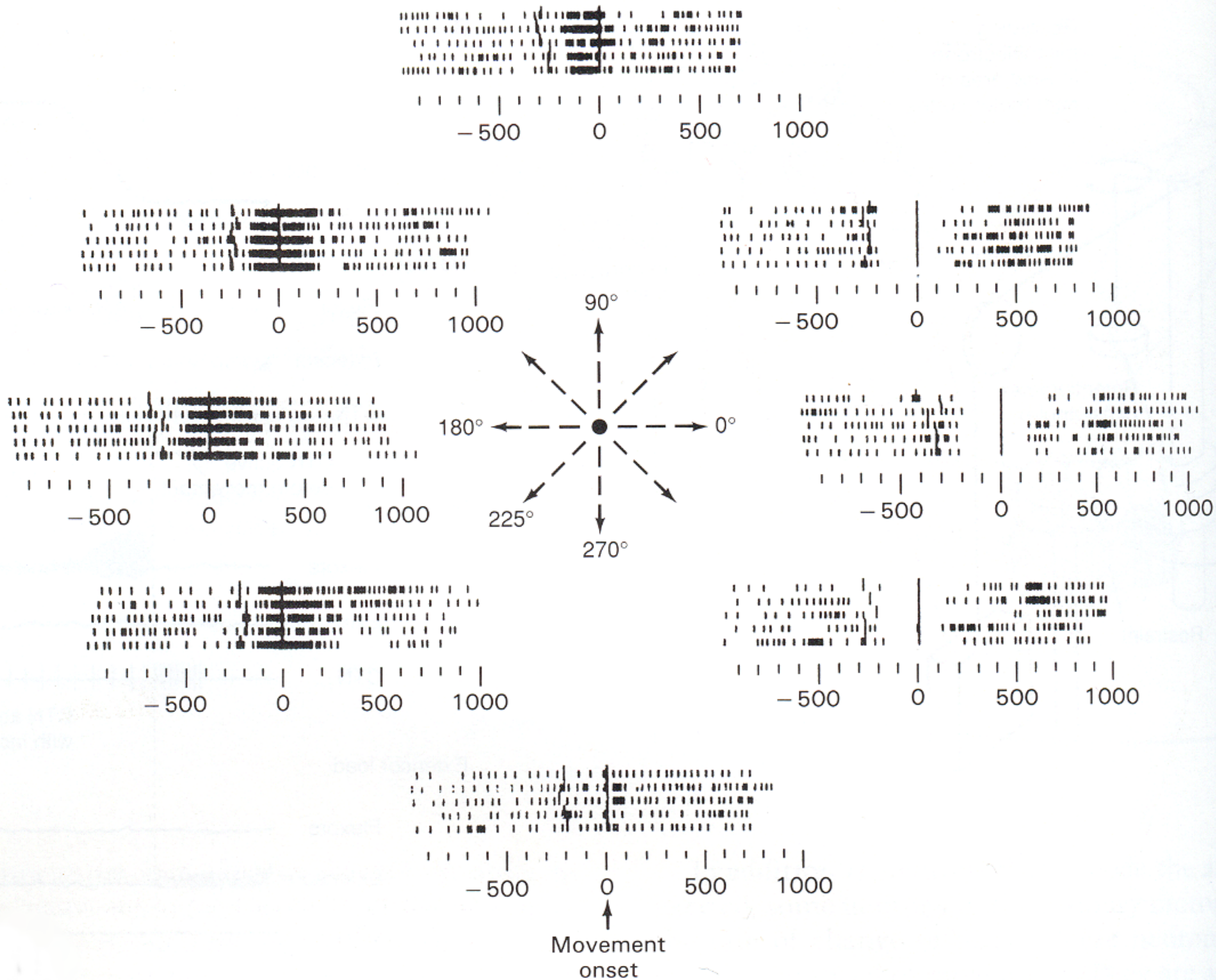
Truism: We believe those scientific results that are repeated across animals (sets of subjects), and across experiments.

Truism: We believe those scientific results that are repeated across animals (sets of subjects), and across experiments.

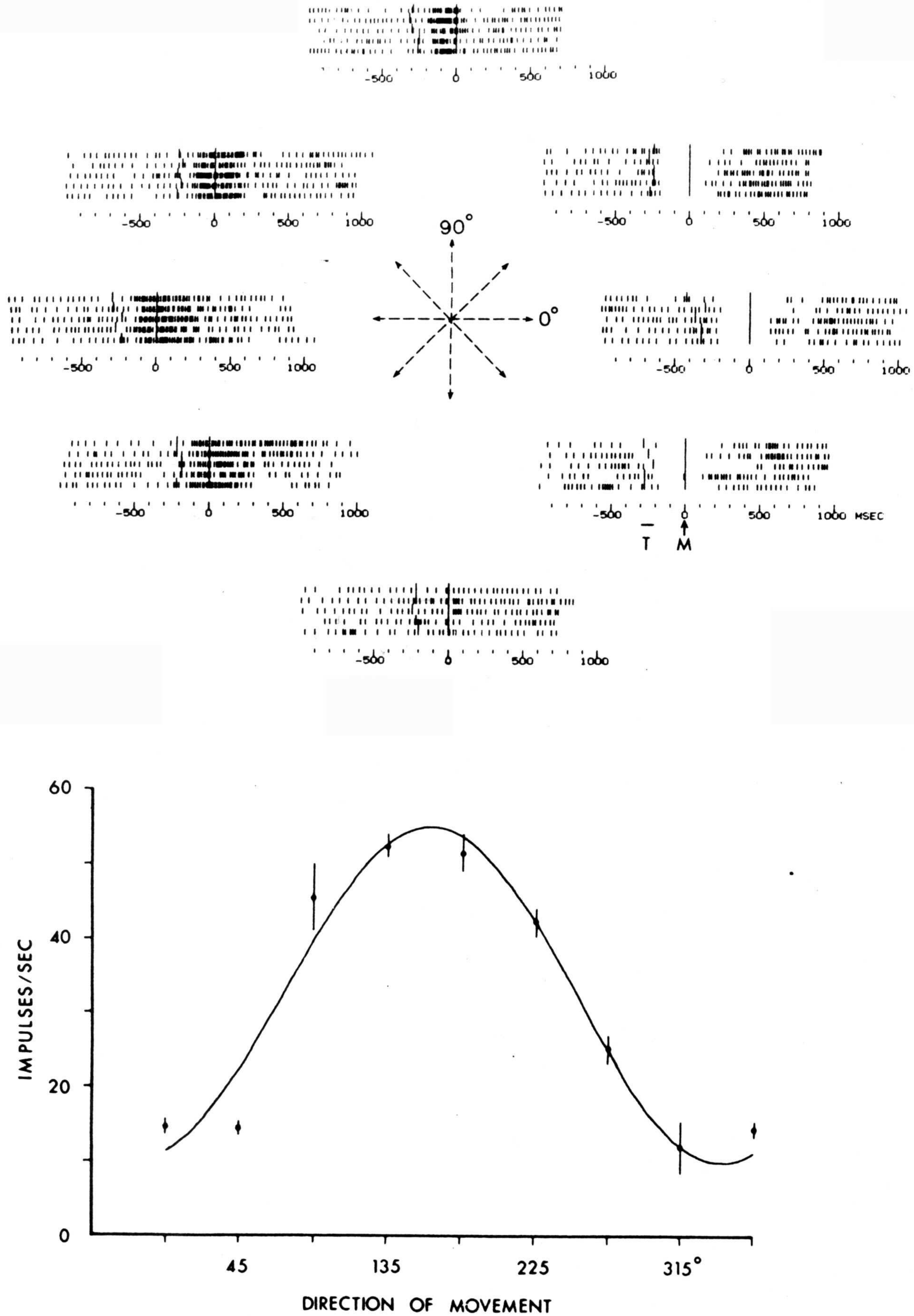
An example from my own experience

Directional tuning of neurons in primary motor cortex

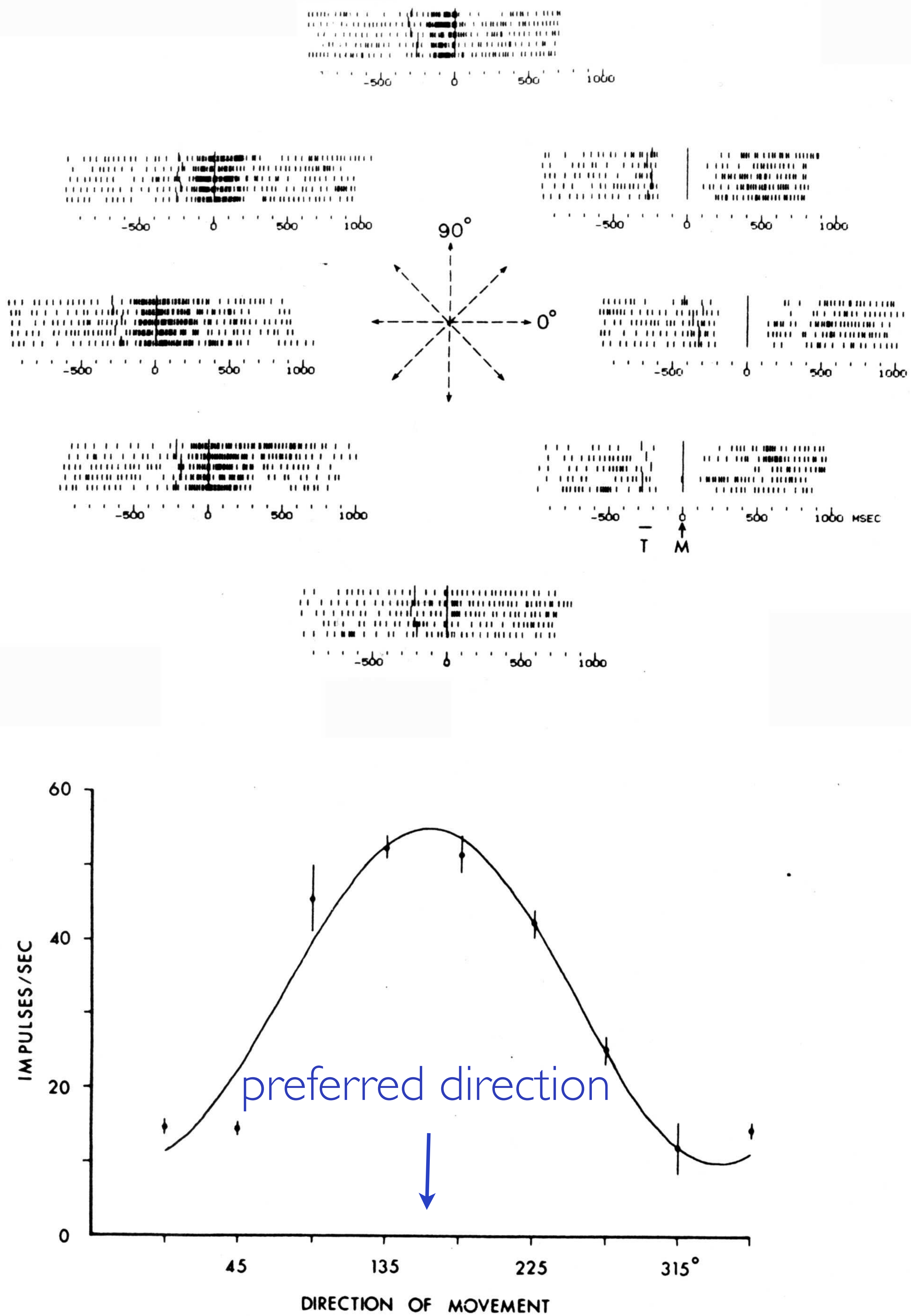




90°
0°
180°
225°
270°
–500
0
500
1000
Movement onset

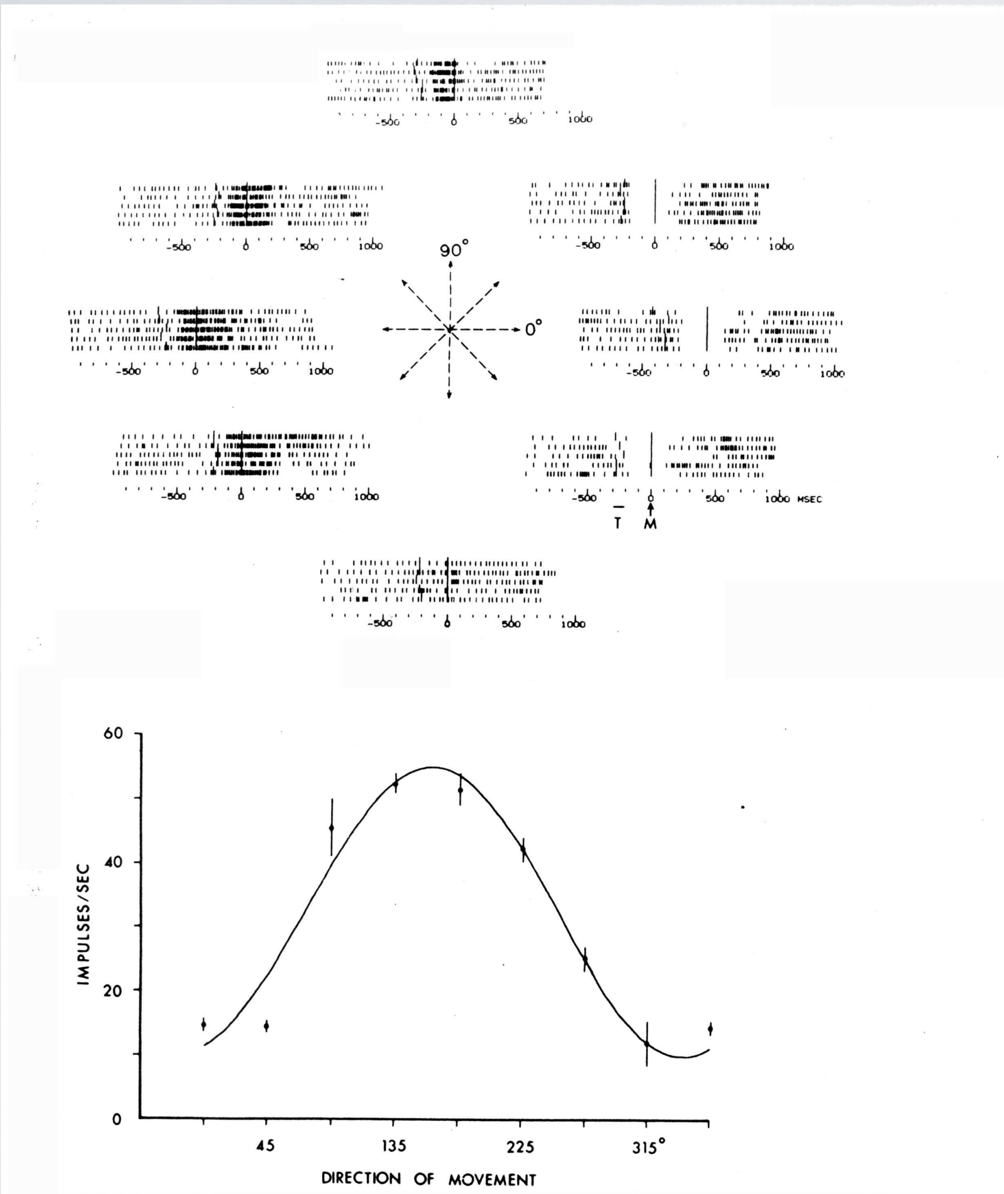


-500 0 500 1000
90°
0°
1000 MSEC
T
M
60
40
20
0
IMPULSES/SEC
45
135
225
315°
DIRECTION OF MOVEMENT

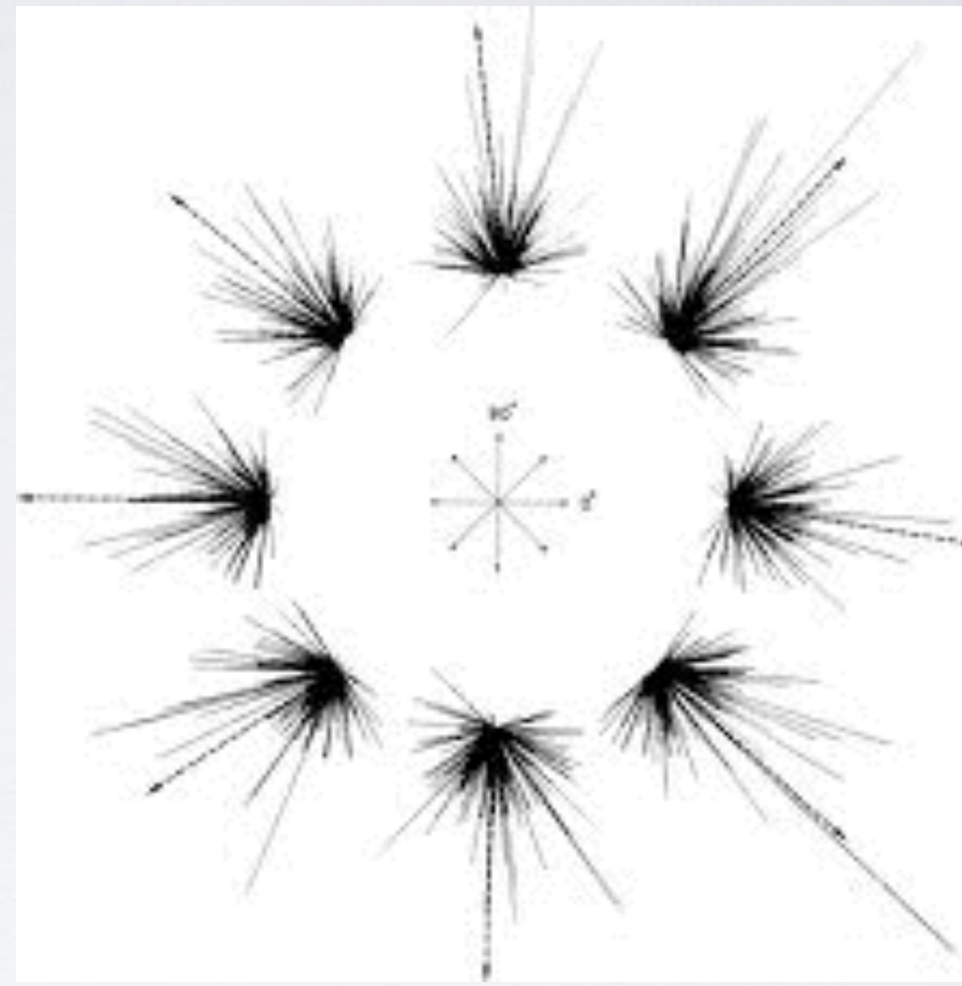


90°
0°
T
M
1000 MSEC
IMPULSES/SEC
60
40
20
0
45
135
225
315°
DIRECTION OF MOVEMENT
preferred direction

directional tuning and the population vector



"population coding"



$$\vec{P} = \sum w_i \, \vec{D}_i$$

DIRECTIONAL TUNING MAY BE CAPTURED TO CREATE A PROSTHETIC DEVICE

Motor cortical neurons are broadly velocity-tuned

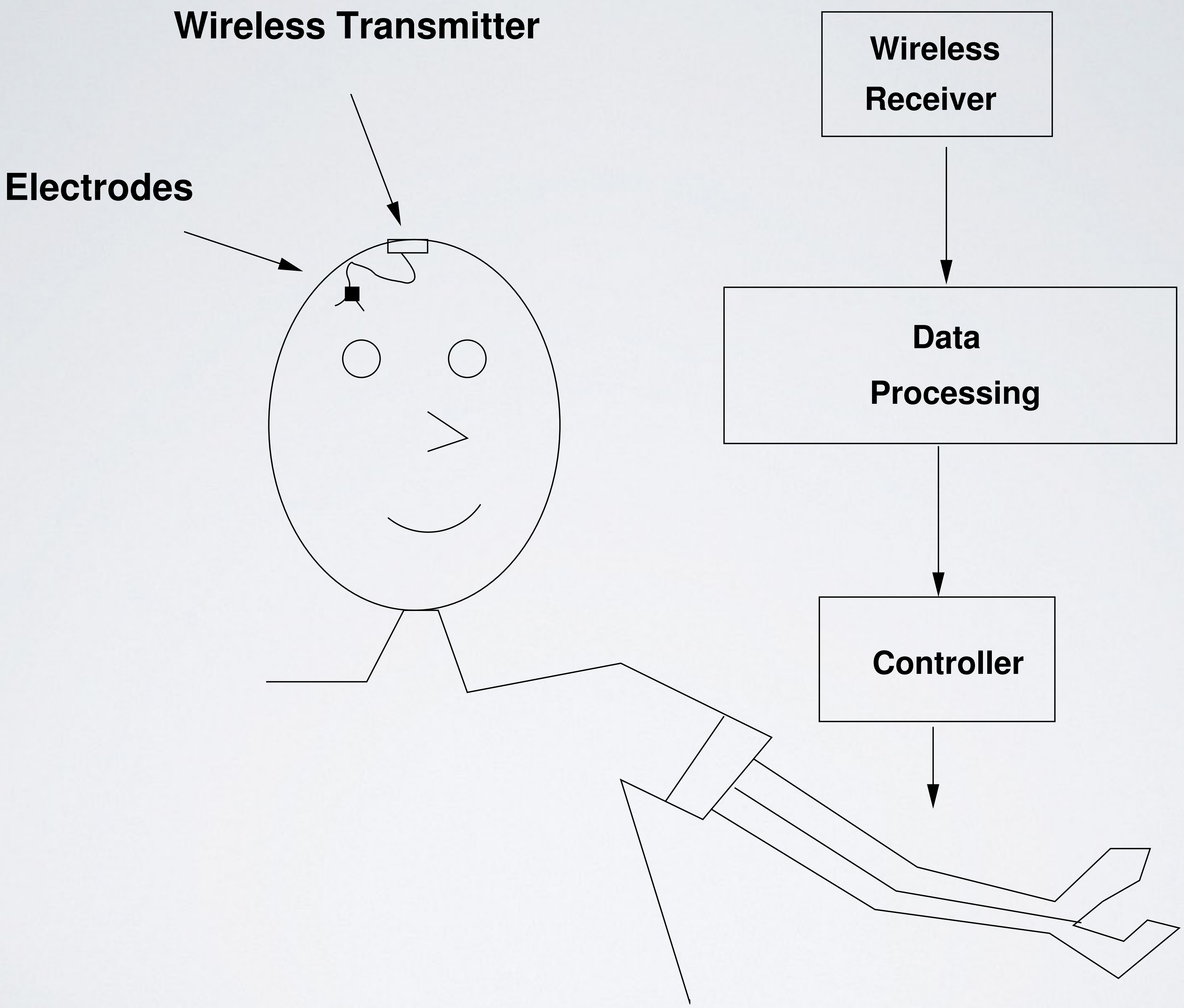
By combining activity of dozens of neurons, hand movement can be predicted

DIRECTIONAL TUNING MAY BE CAPTURED TO CREATE A PROSTHETIC DEVICE

Motor cortical neurons are broadly velocity-tuned

By combining activity of dozens of neurons,
hand movement can be predicted

BCI == Brain Computer Interface



Wireless Transmitter
Electrodes
Wireless
Receiver
Data
Processing
Controller

2) Nov. 30, 2012 - Jan feeds herself a chocolate bar

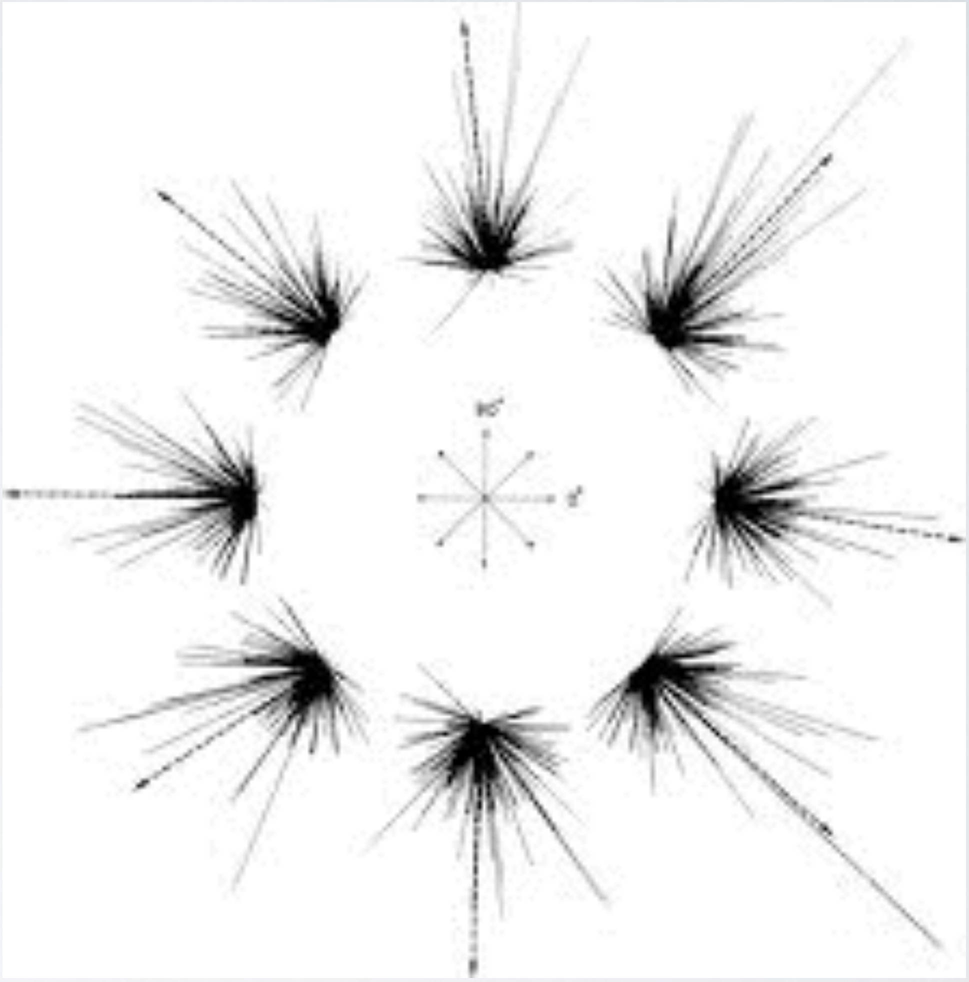
Can also use BCI to *study adaptation,* by altering the mapping between neural output and robotic controller.

Functional network reorganization during learning in a brain-computer interface paradigm

Beata Jarosiewicz[a,b,1,2], Steven M. Chase[a,b,c,1], George W. Fraser[a,b], Meel Velliste[a,b], Robert E. Kass[b,c], and Andrew B. Schwartz[a,b,3]

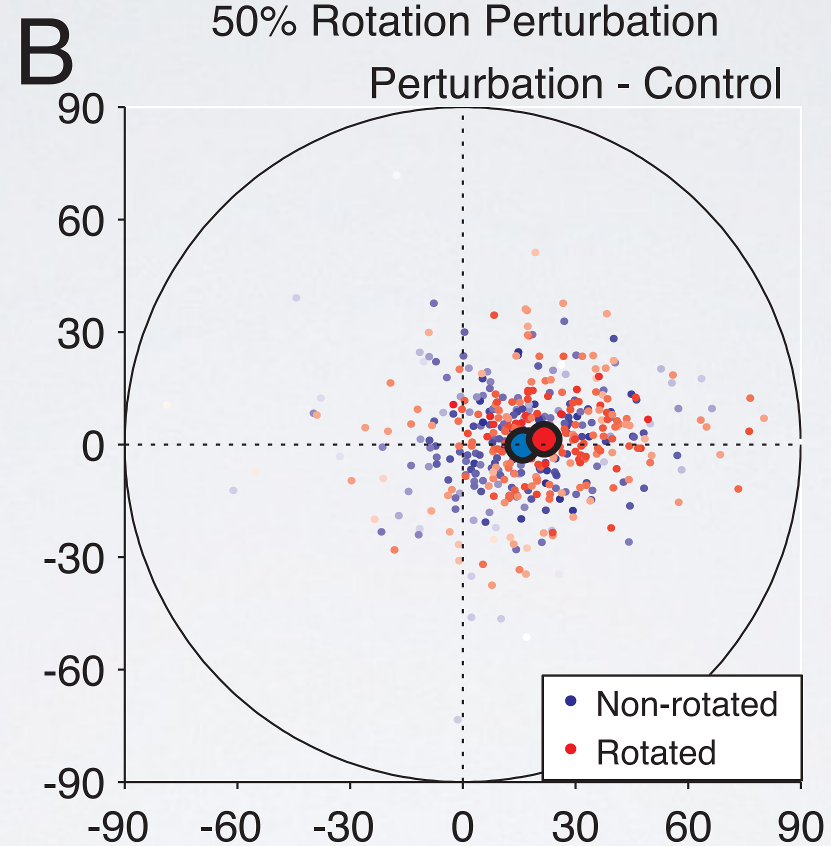
PNAS | December 9, 2008 | vol. 105

idea: in control algorithm rotate vectors, which rotates cursor on screen, and see how neural firing adapts

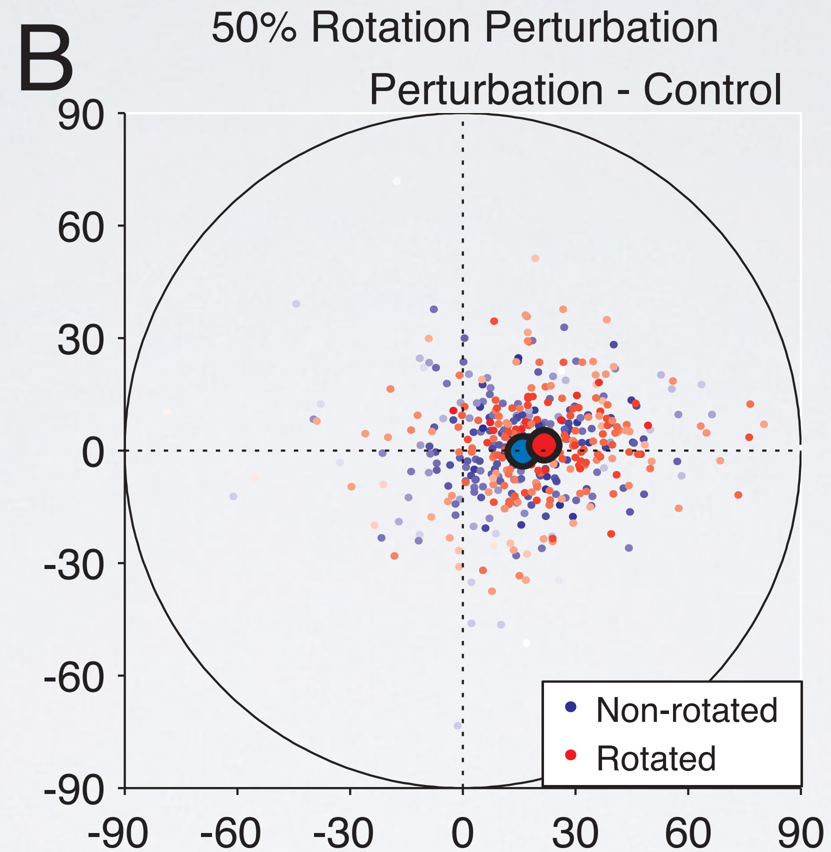


3D reaching experiment

surprising result: predicted direction is different among perturbed (rotated) vs. non-perturbed neurons --- highly significant, but small effect

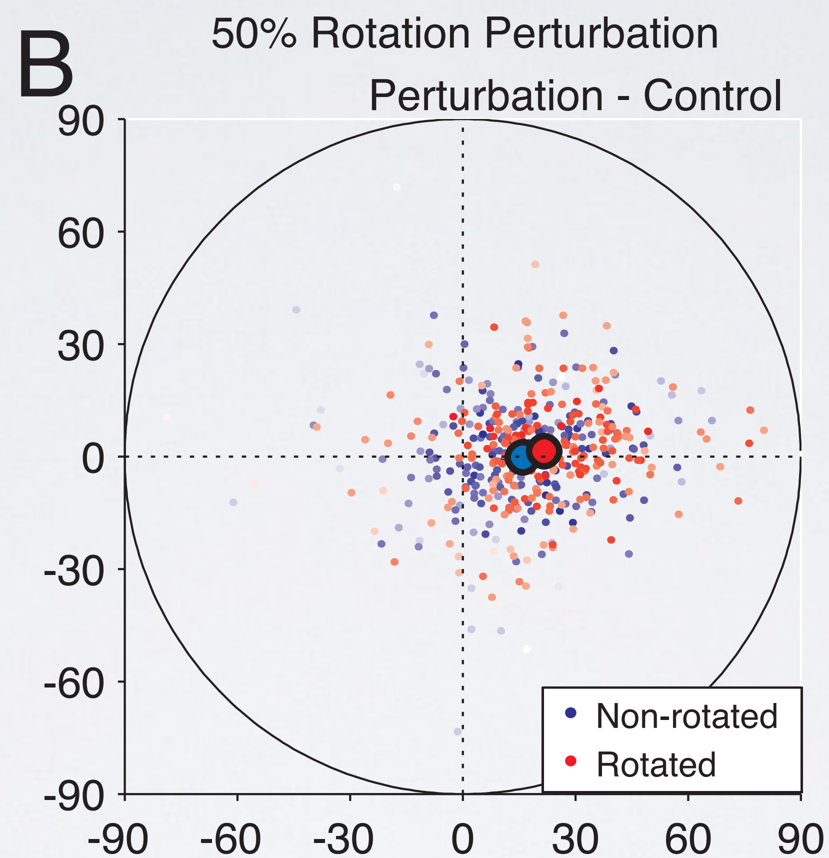


surprising result: predicted direction is different among perturbed (rotated) vs. non-perturbed neurons --- highly significant, but small effect



---> ran 3rd monkey

surprising result: predicted direction is different among perturbed (rotated) vs. non-perturbed neurons --- highly significant, but small effect



---> ran 3rd monkey

---> repeated in a 2D rather than 3D reaching task

Bayes' Theorem:

$$P(H_0|data) = \frac{P(data|H_0)P(H_0)}{P(data|H_0)P(H_0) + P(data|H_A)P(H_A)}$$

wonderful (optimal) BUT requires that
we know the null and alternative distributions

REPLICATE: using first experiment to form probability distributions for second experiment

REPLICATE: using first experiment to form probability distributions for second experiment

—> get extremely large Bayes factor: “overwhelming evidence” ideally replication would occur across *labs*

Summary

Bayes' theorem is (*Bayes factors* are) powerful, but delicate.

Network data can provide the information needed to implement Bayes' theorem.

We should be replicating more often, *and this should allow us to apply Bayes theorem.*

thank you!