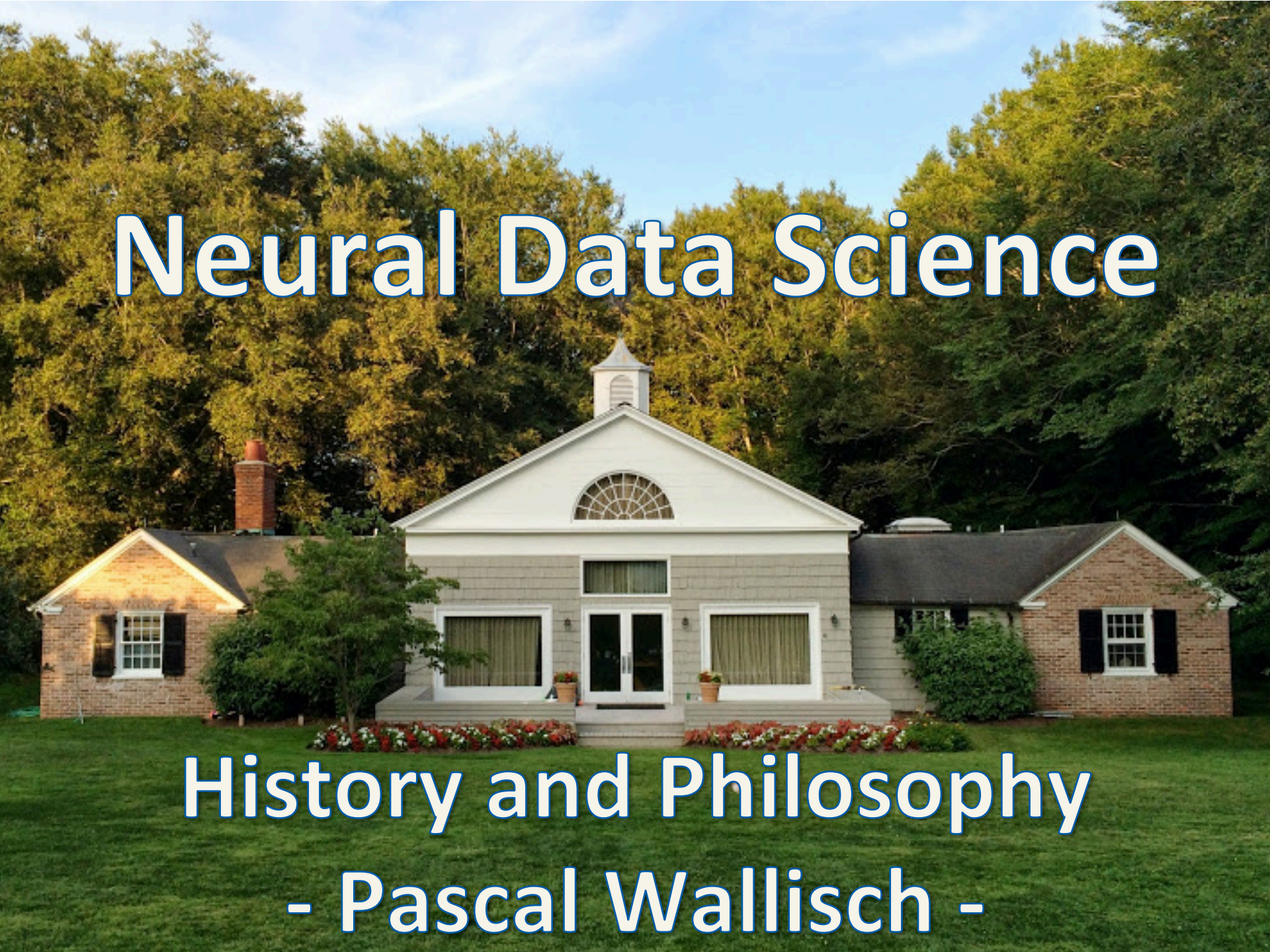
Smallest font

Calibration slide

Smallest font

Neural Data Science



History and Philosophy

- Pascal Wallisch -

Neural data science

- What is it?

- Why are we doing it?

The historical perspective – a woodcut

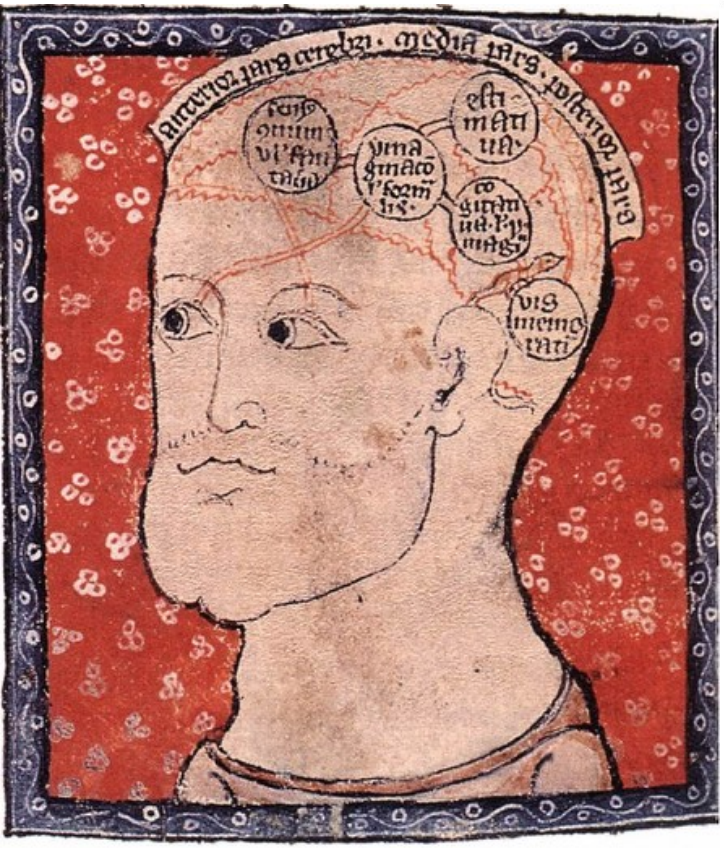
The history of neuroscience

- Neuroscience (defined as the **recording**, **analysis** and **interpretation** of **physiological** signals that originate from **neurons**) has a long past, but only a very short history.
- **Long past**, mostly involving a macroscopic treatment of brain structures: Neuroanatomy, Neurosurgery, early microstimulation sensu Fritsch & Hitzig, philosophical speculation, etc.
- **Short history**: Beginning in early 1900s with early electrophysiology.

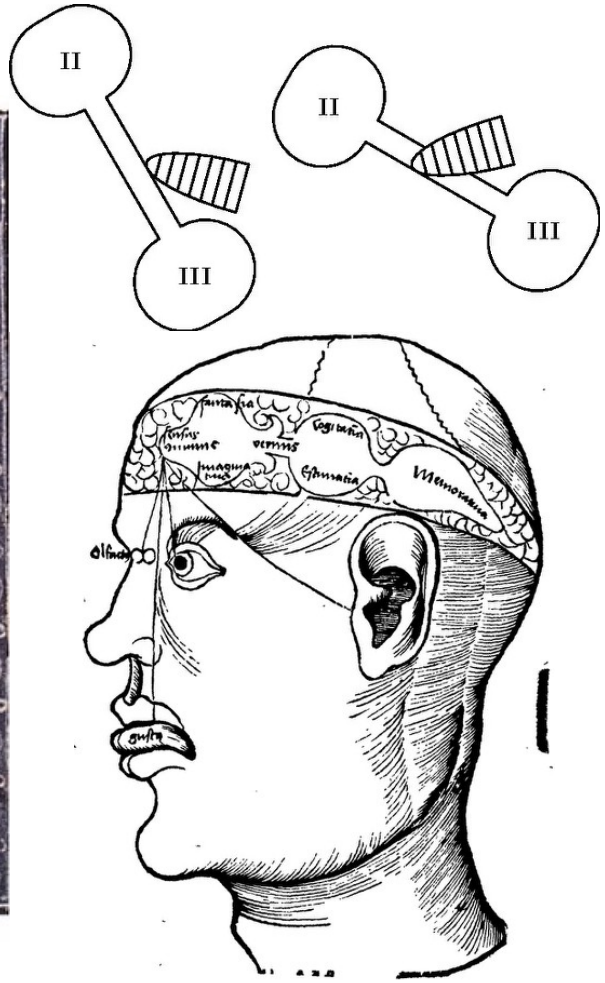
Prehistory: Aristotle's book of problems

- **Why is the brain white?** 1. Because it is cold, and coldness is the mother of white. 2. Because it may receive the similitude and likeness of all colours, which the white colour can best do, because it is most simple.
- **Why is the brain moist?** Because it may easily receive an impression, which moisture can best do, as it appeareth in wax, which doth easily receive the print of the seal when soft.
- **Why is the brain cold?** 1. Because that by this coldness it may clear the understanding of man and make it subtle. 2. That by the coldness of the brain, the heat of the heart may be tempered

The past: Not much progress for ~2k years, but wild speculation about cognition on the basis of loose anatomical observations



Qusta ibn Luqa
(864-923)



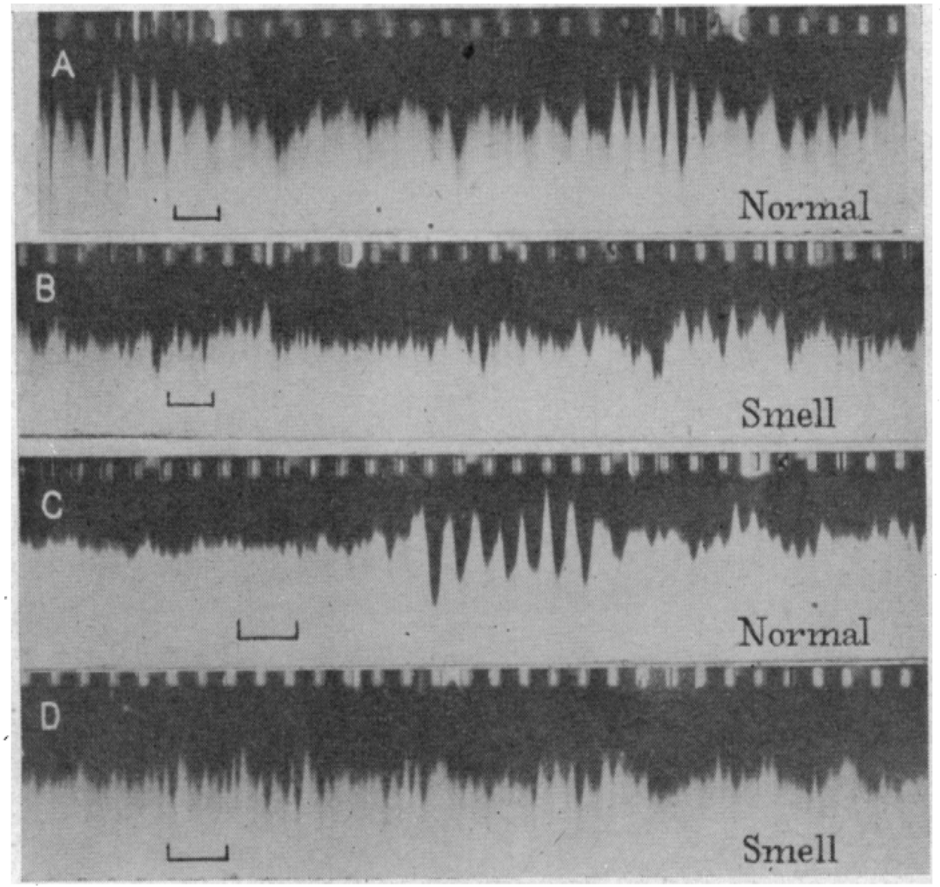
Mondino dei Luzzi
(1270-1326)

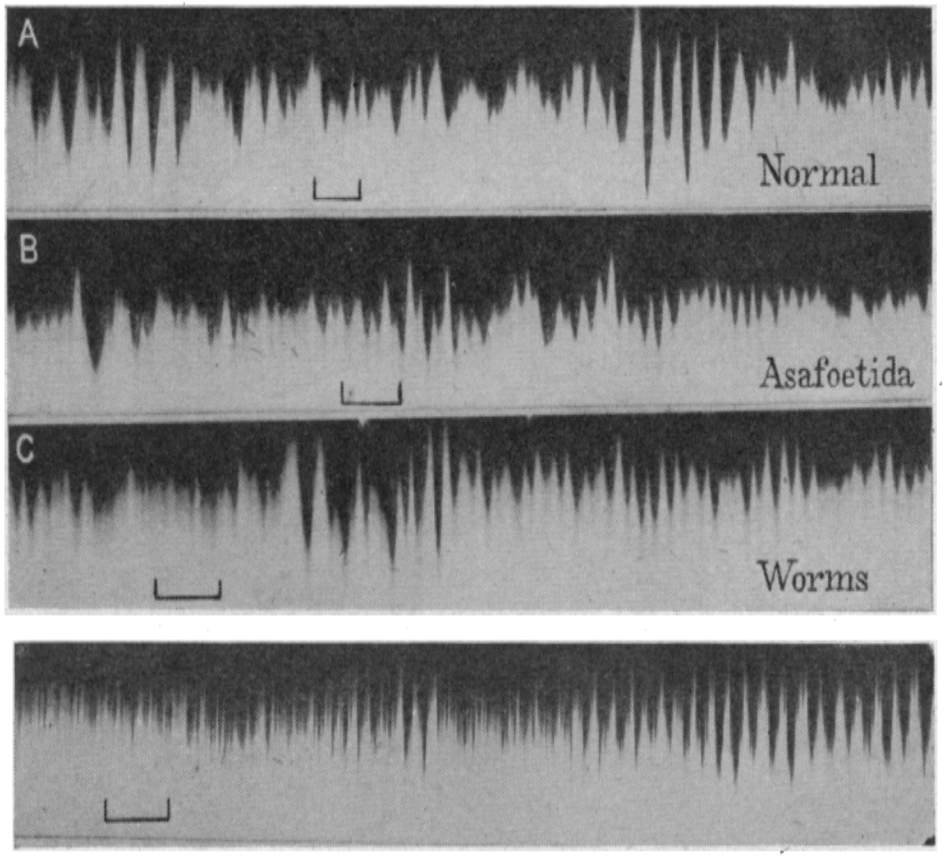
Franz Gall
(1758-1828)

History – 1: The pioneering period

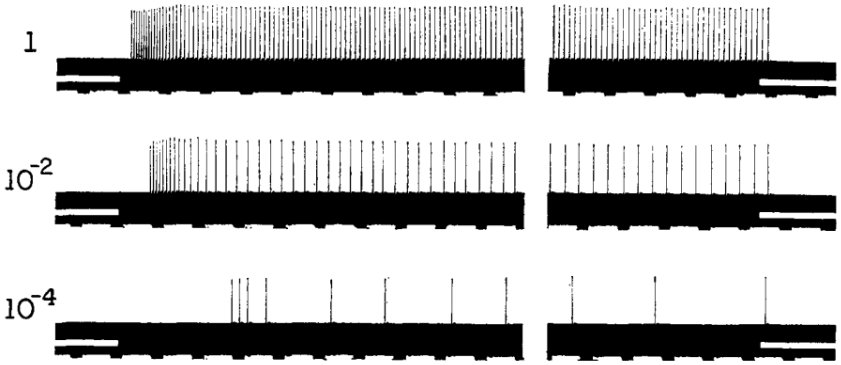
- First half of 1900s. (~1920s-~1950s)
- Early **single** electrode electrophysiology.
- After Cajal had shown with Golgi's **method** (they shared a Nobel in 1906) that neurons are the basic structural unit in the brain, people were curious about the characteristics of these units.
- In addition to morphology and distribution, that included function - how they respond to stimuli and how they transmit signals (which was made possible by technical advances in the amplification of electrical signals in the early 1900s).
- "**Analysis**": Descriptive. Showing the raw physiological traces, mostly.

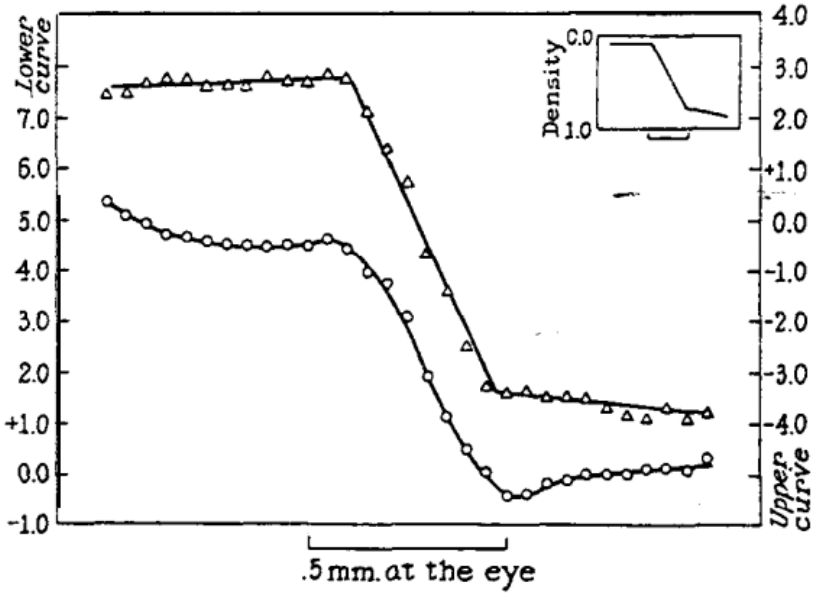
Example: Edgar Adrian (Nobel 1932) on olfactory responses in hedgehogs





Example: H Keffer Hartline on visual responses in limulus (Nobel 1967)

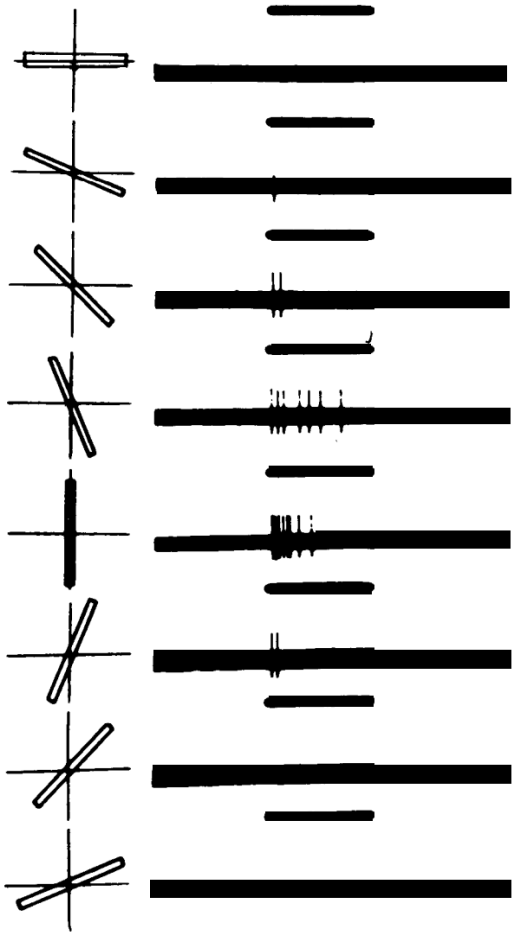


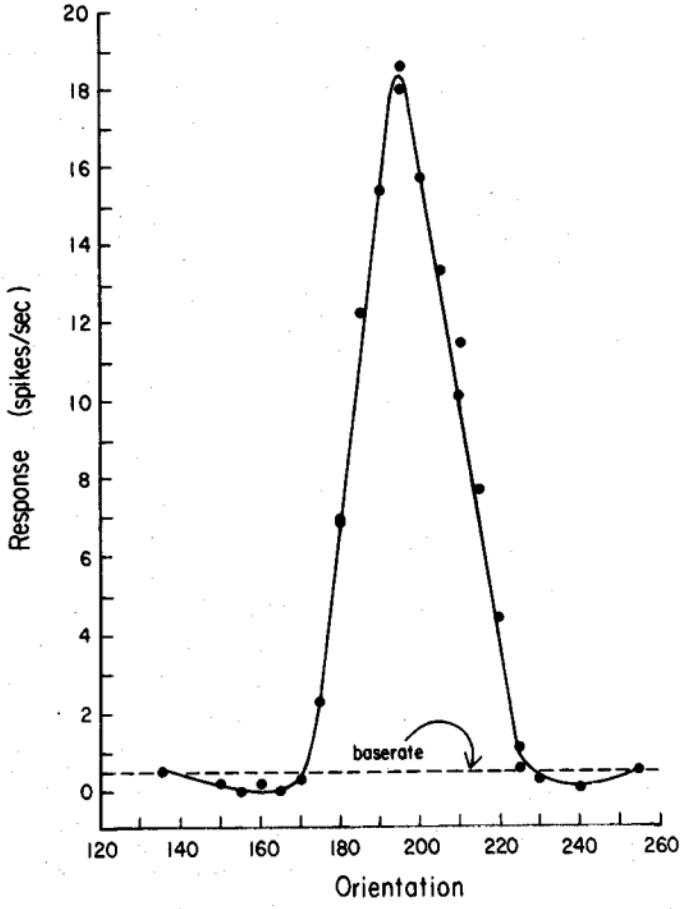


History – 2: The classical period

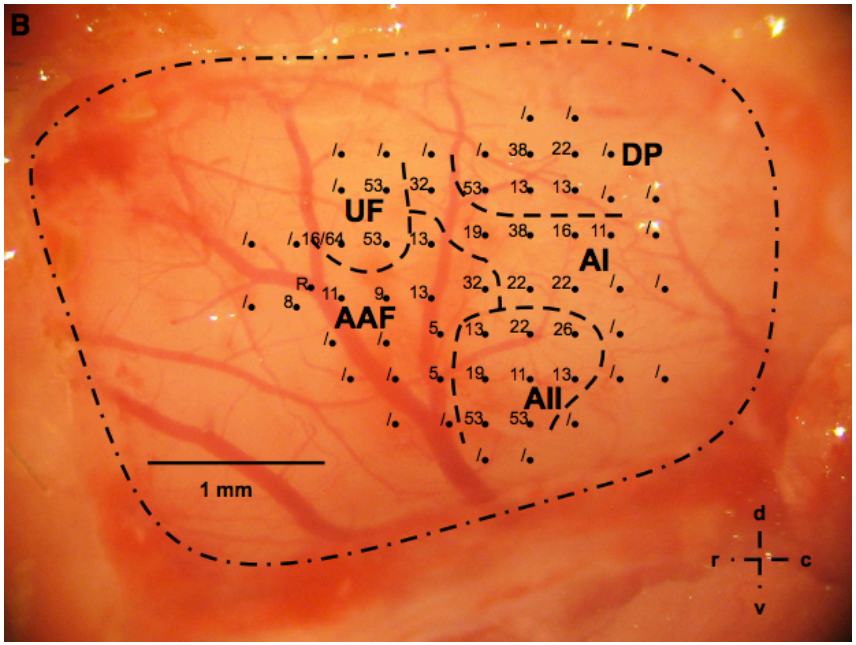
- Second half of 1900s. (~Late 1950s-~1990s)
- Single electrode electrophysiology.
- Now recording from cortical neurons with metal (usually tungsten) microelectrodes from a small set of “standard” animals (mice, rats, cats, monkeys).
- Towards the end of the period also in awake animals.
- Analysis is relatively straightforward – counting spikes from a “single” neuron, repeating stimulation, averaging over conditions. Plotting that. Later “modeling”, usually curve-fitting.

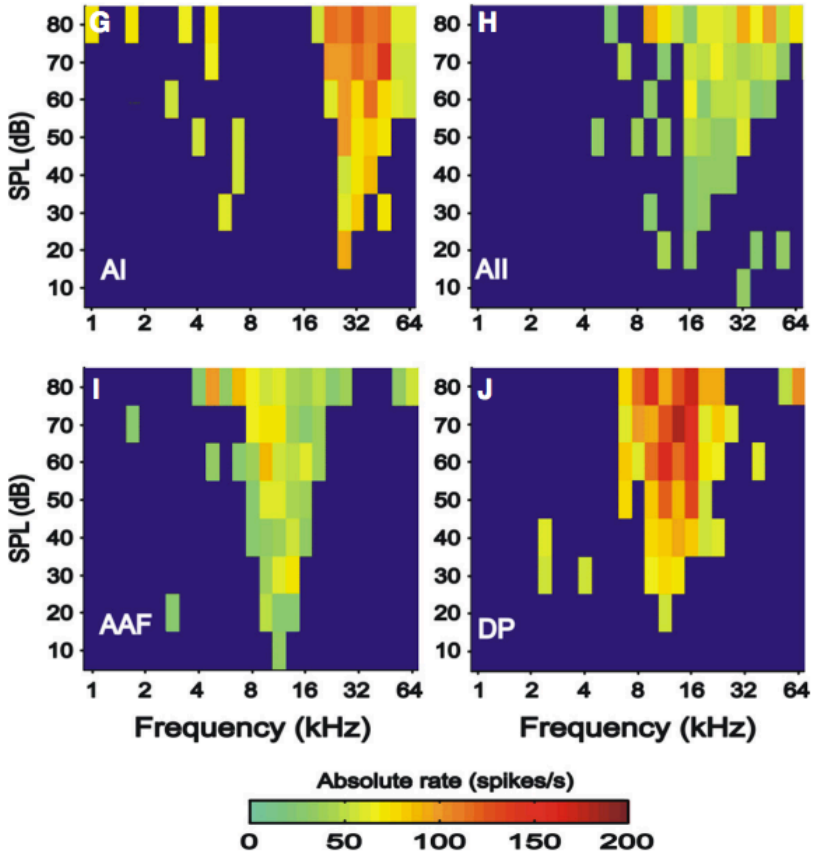
Example: Hubel & Wiesel (Nobel 1981)





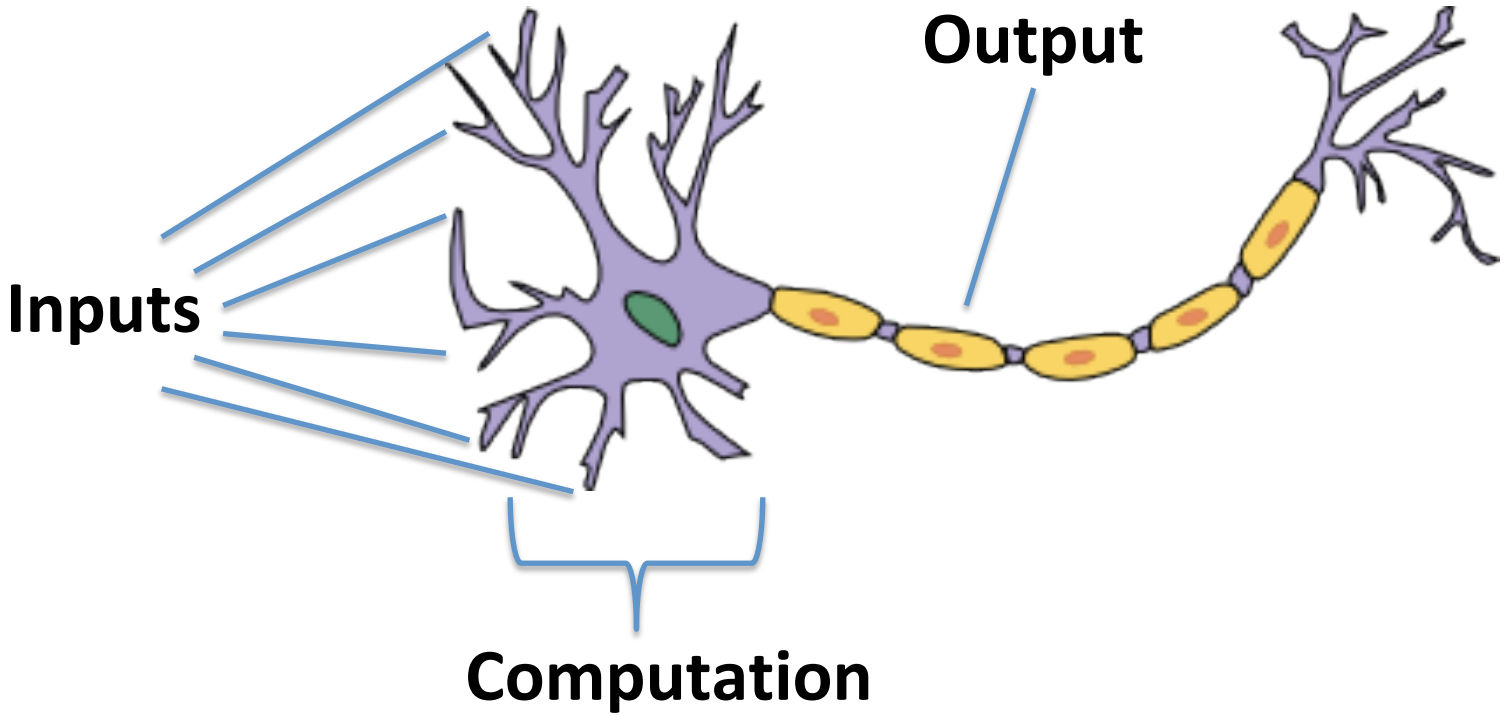
Not just vision in monkeys:



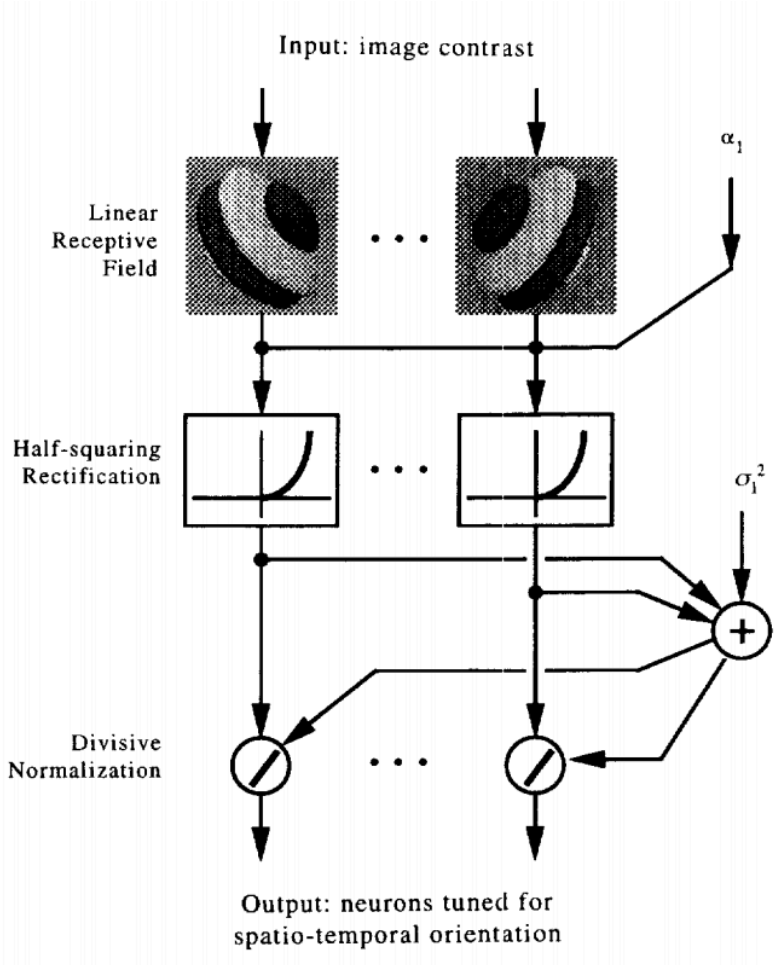


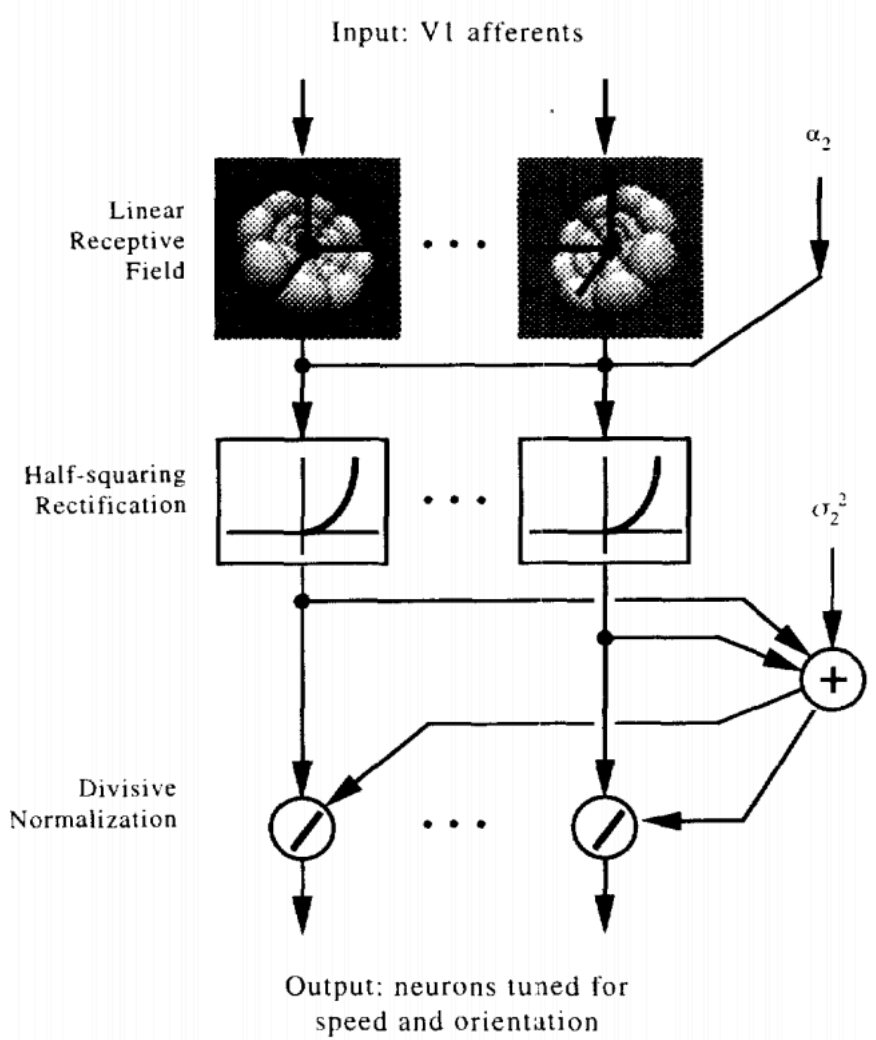
Joachimsthaler et al., 2014

Classic computational neuroscience



Example: Physiologically plausible motion energy models





Simoncelli & Heeger, 1998

Did that get us there?

- Not everyone was on board with this approach:

Hubel and Wiesel (Nobel 1981 – on troubling developments in the field, page 707 of their book):

"A third problem represents to us an example of illnesses that scientific fields can be subject to. In neurophysiology this is the increasing popularity of theory, sometimes called ***computation***"

Their concern

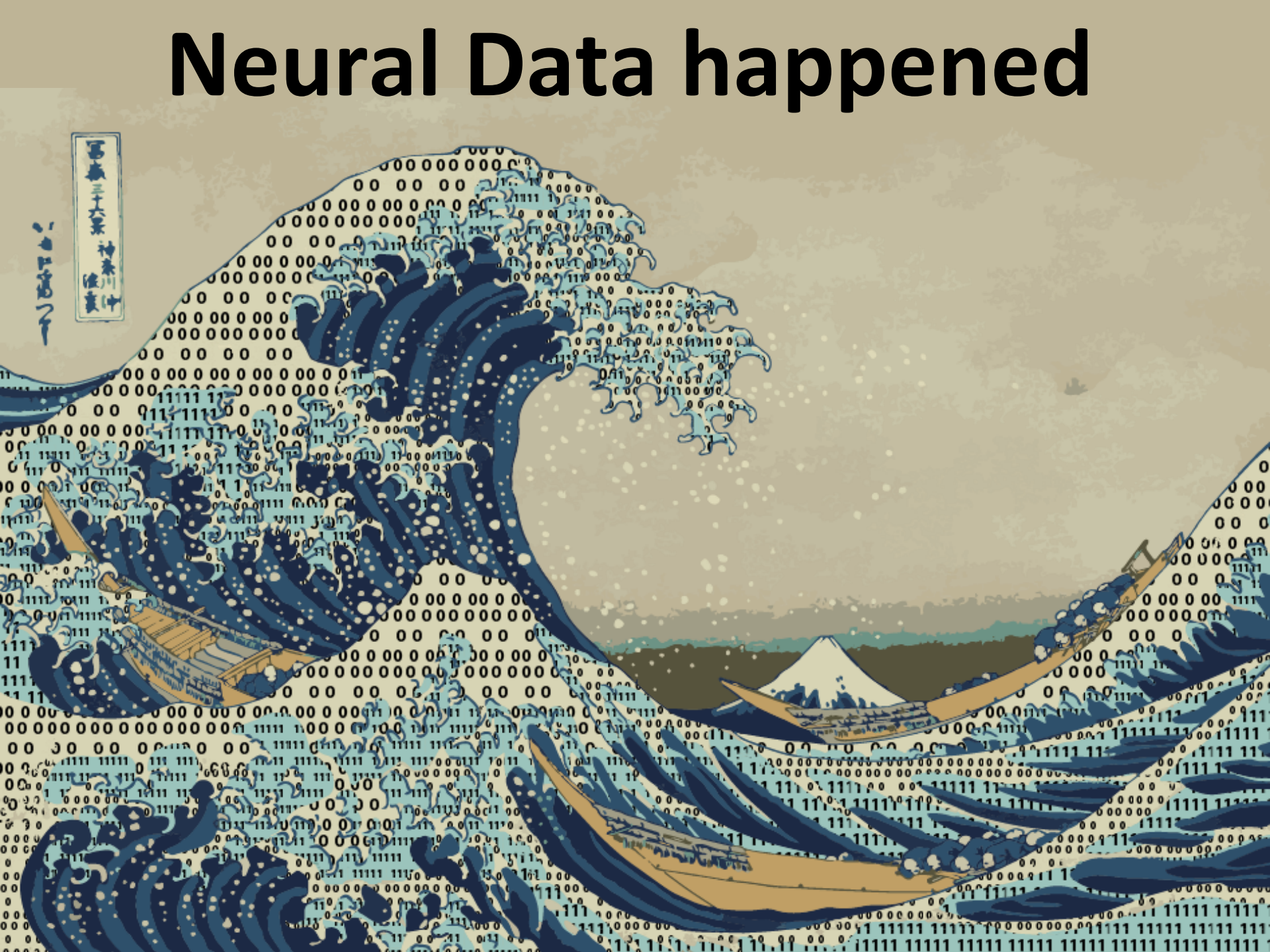
Is about a **hostile takeover** of the field by physicists and mathematicians, using inadequate methods for biology due to their training:

*"The main proponents of **theory** in neurophysiology may be scientists who have trained in mathematics and, having gone into biology, are reluctant to give up the mathematics… in the case of such subjects as "linear systems analysis", the emphasis seems to be either puzzling or wrong. No one with much experience with cortical cells could think of them in any real sense as linear."*

So…

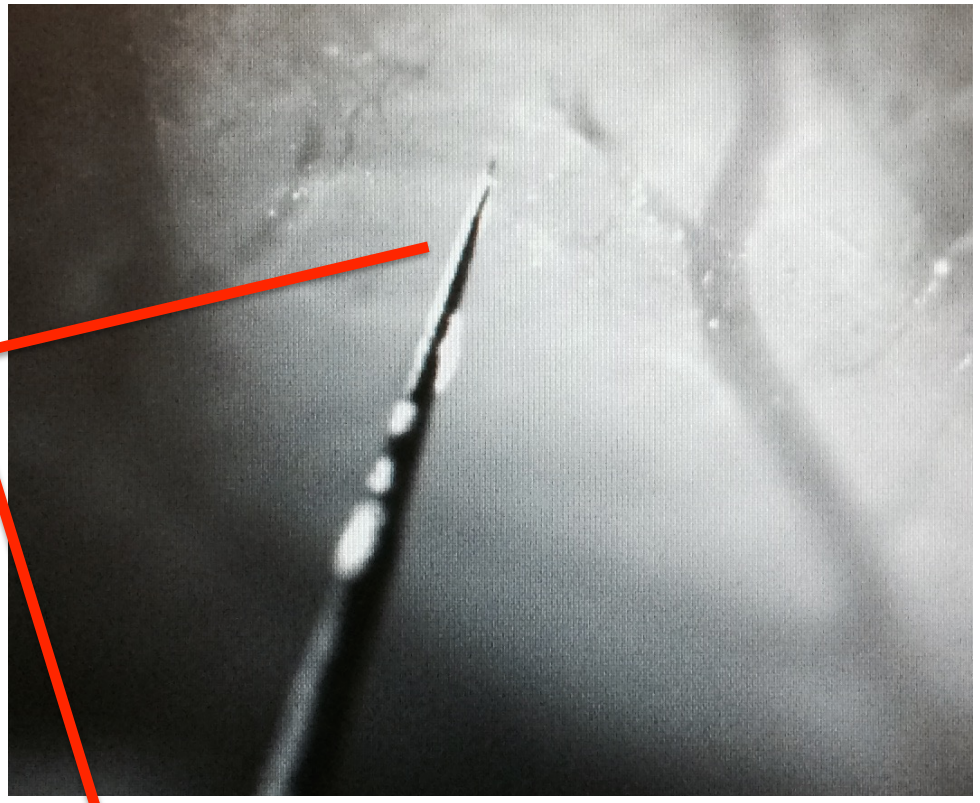
- Where are we now?
- What happened since?
- Where do we want to go?

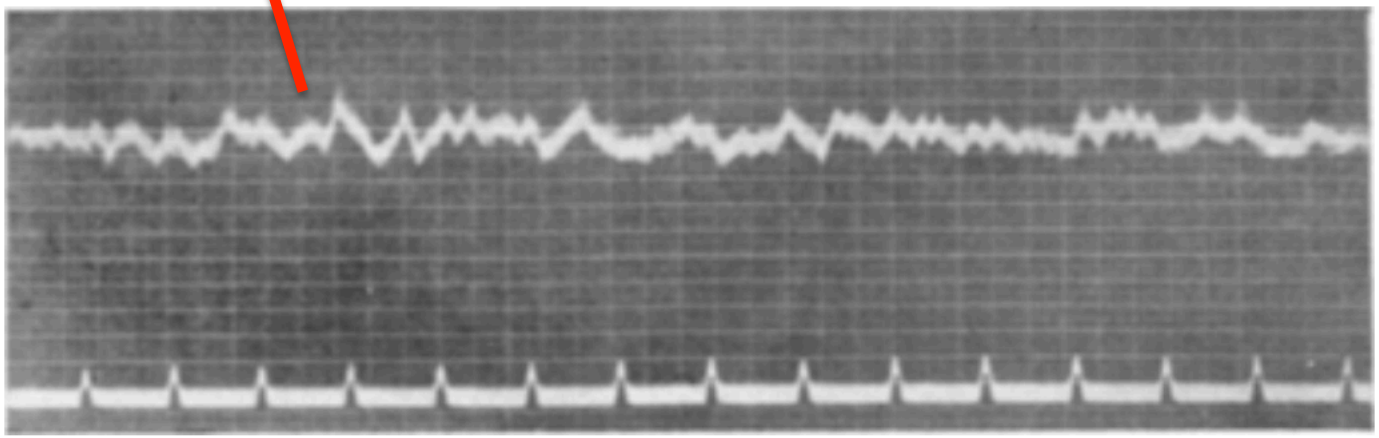
Neural Data happened



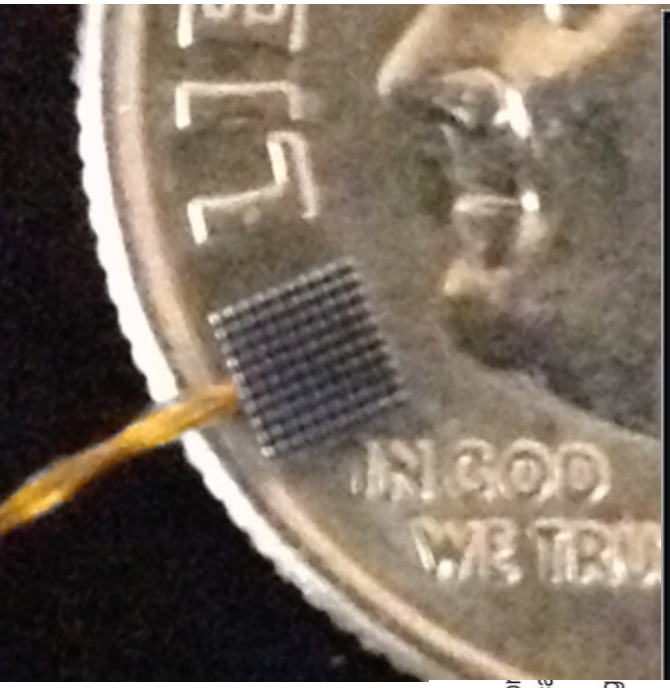
Classic measurements of neural phenomena: Scarcity

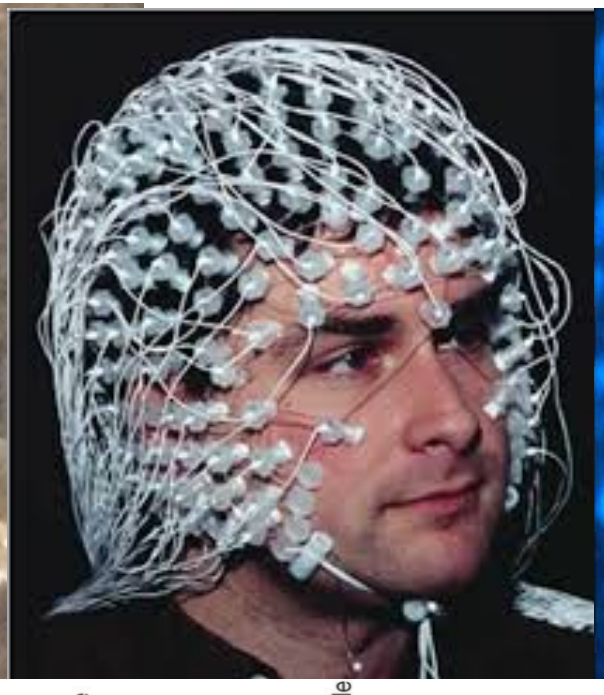
1 channel

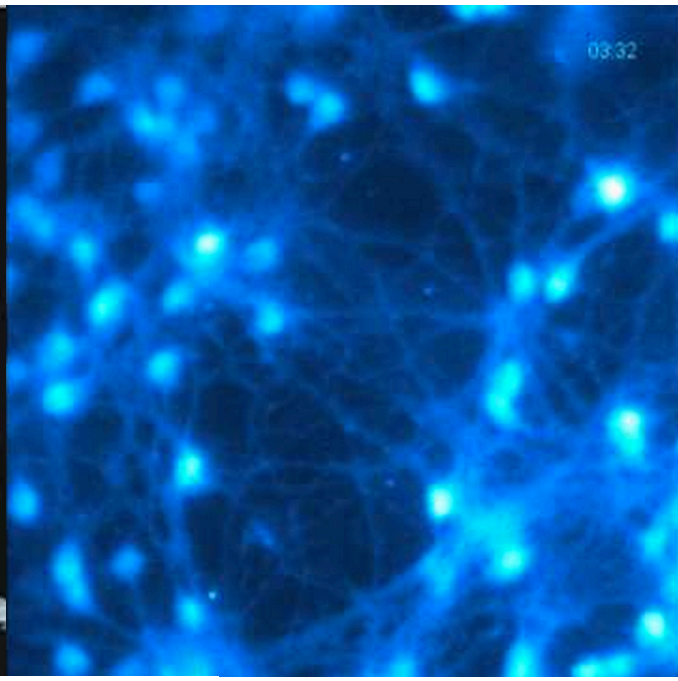


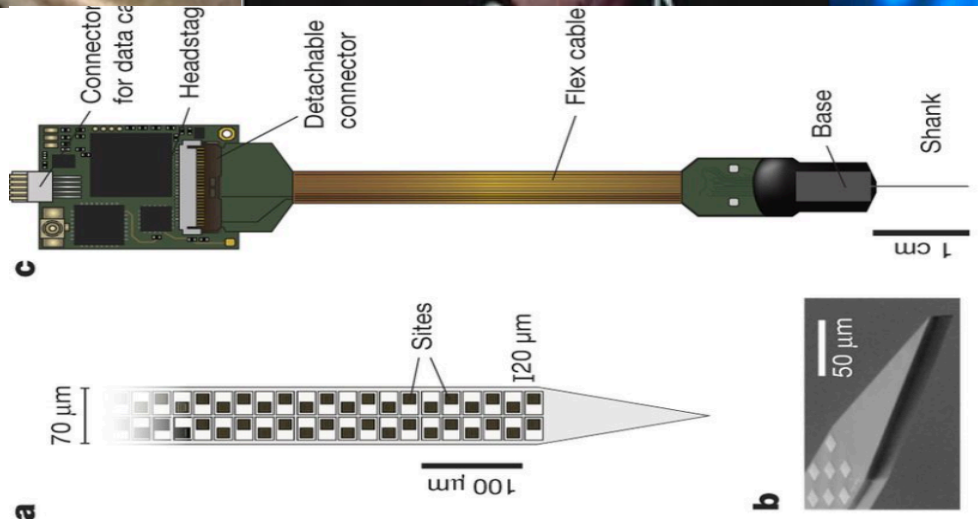


Current measurements of neural phenomena: Abundance / Cornucopia









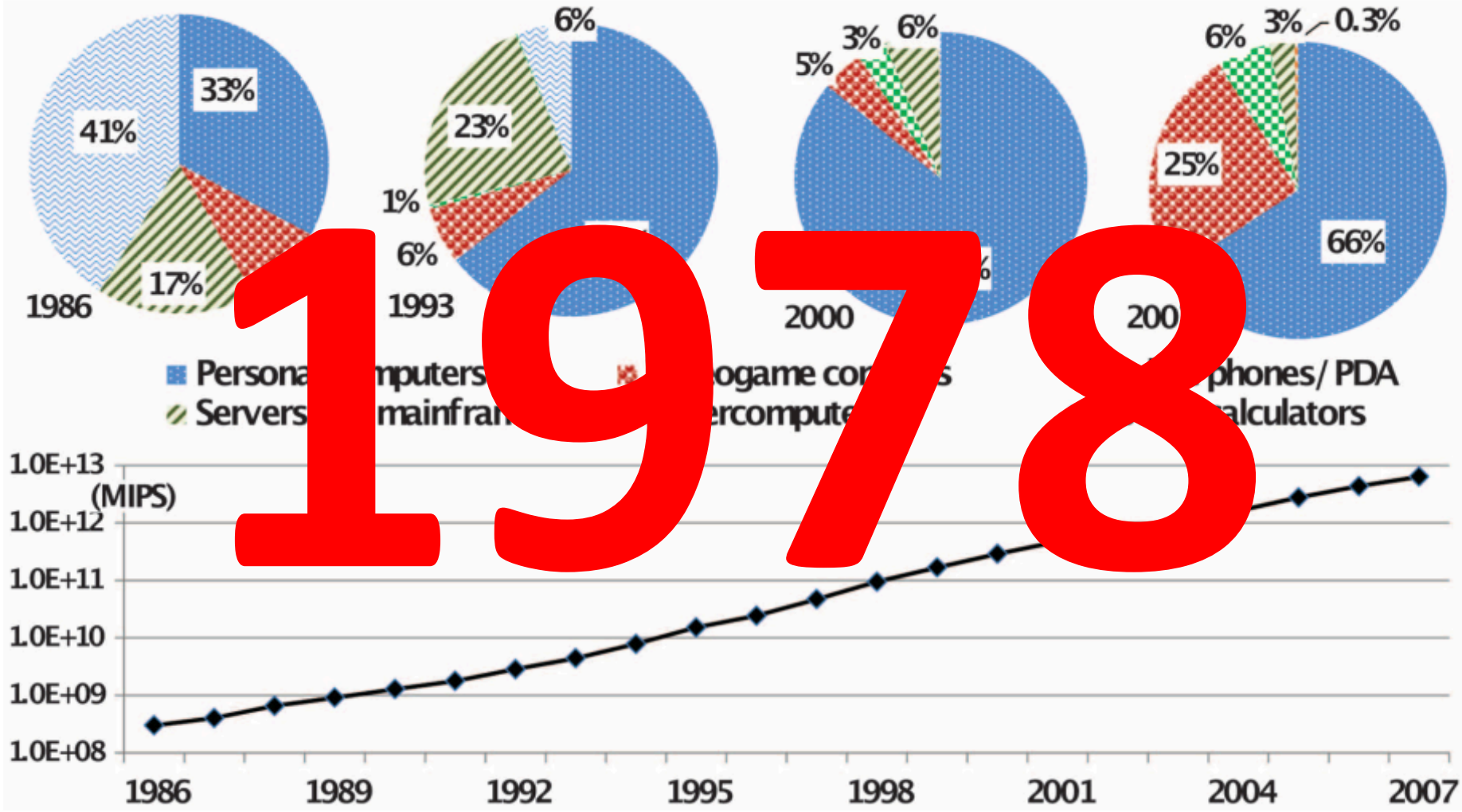
History: 3 – The current period

- Since early 2000s to now.
- Characterized by massive increases in capabilities to record from multiple electrodes at the same time as well as seeing the rise of optical imaging.
- All of this leads to truly massive datasets (compared to those of the past – megabytes to terabytes, several orders of magnitude).

And this is just the beginning

- There is a lot of data in there:
- "**the human brain produces in 30 seconds as much data as the Hubble Space Telescope has produced in its lifetime**." *–Konrad Körding*
- In time, we will come up with methods to harness this data.
- Then what?
- Would you be able to do anything with it, if you could get hold of all that data now?
- Can you handle the data?

Good news: Our ability to store and compute information has increased exponentially



Hilbert et al., *Science*, 2011

Times have changed

Cray 1 (1977)
$8.86 million
80 MFLOP/s

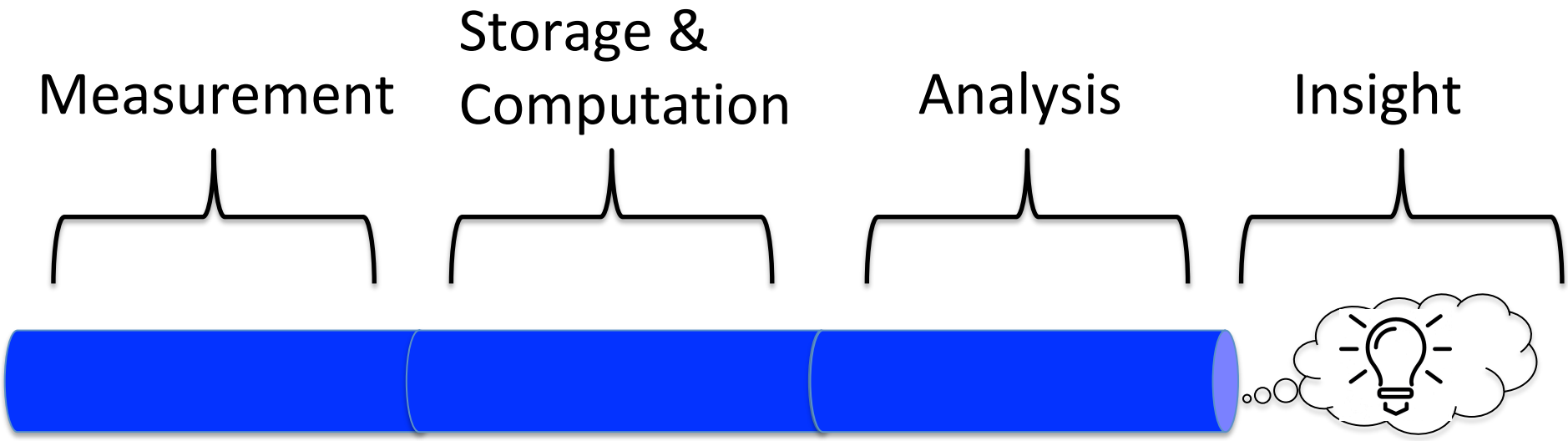
iPhone 7 (2016)
<$800
>160 GFLOP/s

This changes everything…

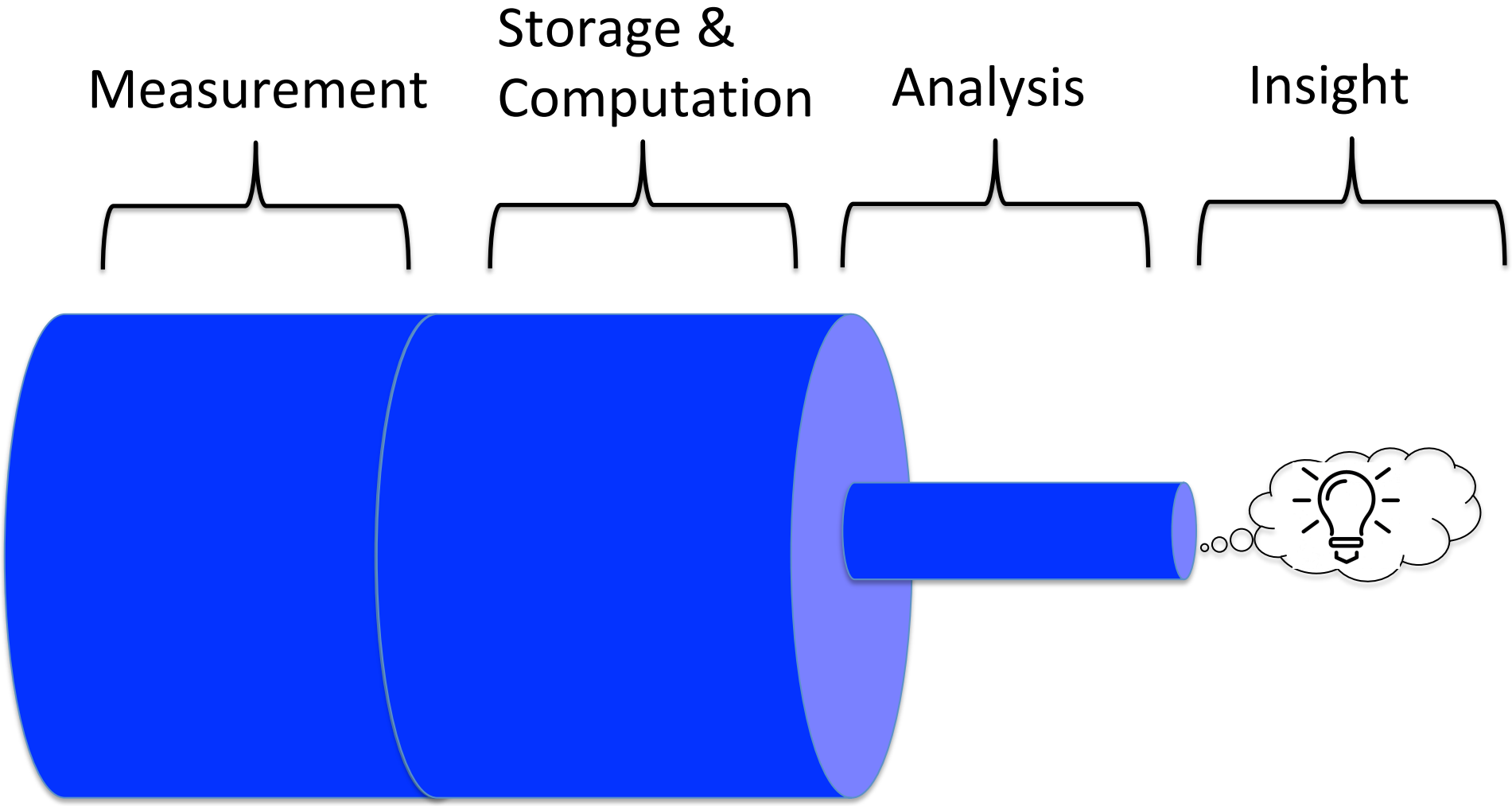
- Demonstrating the revolutionary computational power of computers by inverting a matrix.
- Inverse: $\mathbf{A} \times \mathbf{A}^{-1} = \mathbf{A}^{-1} \times \mathbf{A} = \mathbf{I}$ $\quad\quad$ $(\mathbf{A} \times \mathbf{I} = \mathbf{A})$
- Matrices are inverted in order to solve systems of linear equations (and other things)
- Say you know **A**, **B** and that **XA = B**. What's **X**?
- Can't divide!
- Can multiply with the inverse.
- Inverting a 24x24 matrix by hand takes hundreds of years of continuous work.
- How long does it take currently with a computer?

So where is the bottleneck?

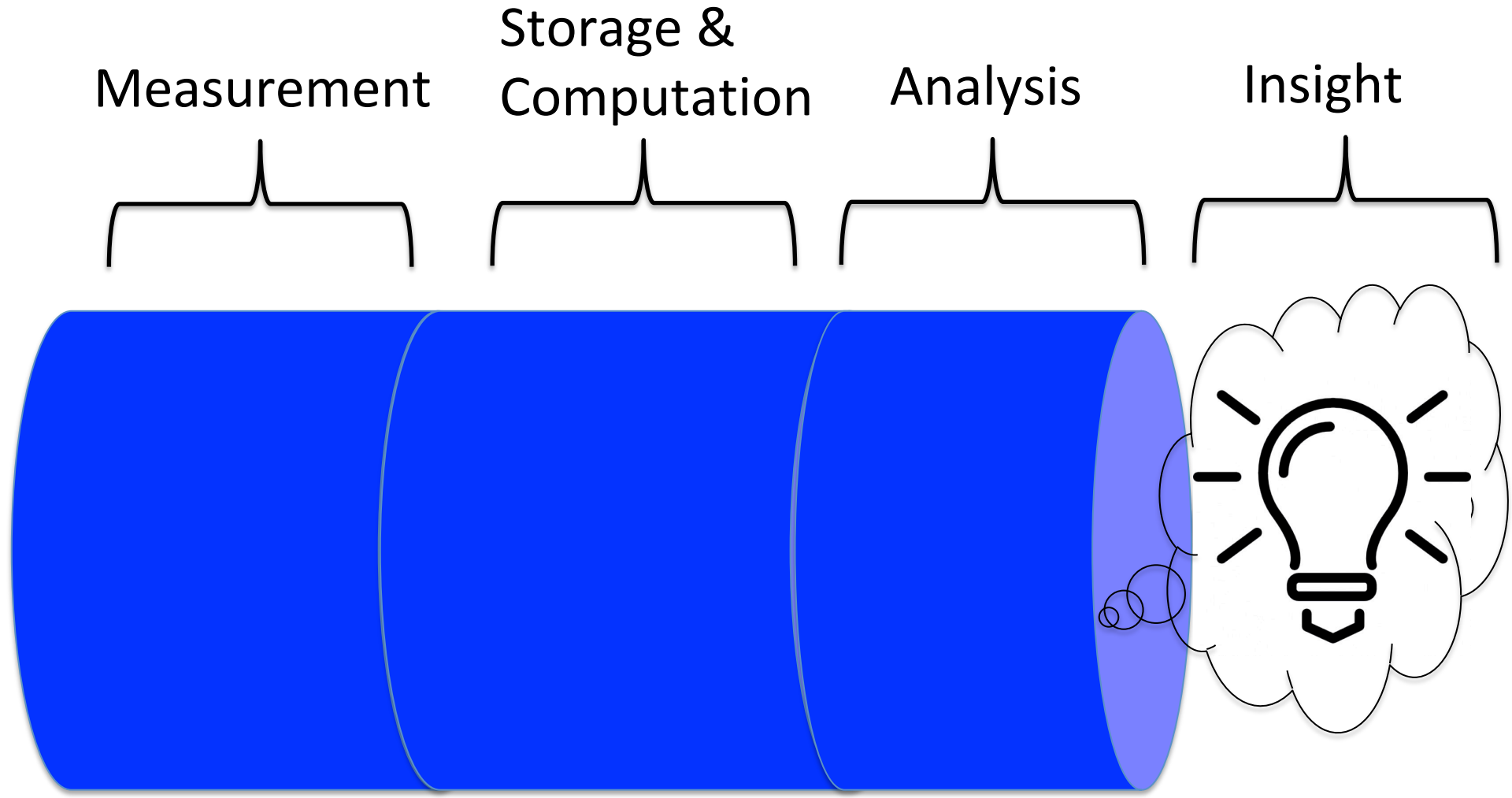
The data pipeline in the classical era



The data bottleneck in the current era



The point of this course



This also addresses Hubel & Wiesel's concerns

- In an environment of data scarcity, computational neuroscience might well unduly overweigh theory.
- But the days where one can stick an electrode into the brain, record from a single neuron and observe effects that are so dramatic that they are obvious on a single trial and win the Nobel prize are over.
- The point of Neural Data Science is to put the data front and center. Luckily, there is lots of it.
- But we need to find new conceptual ways to handle this data.

They had a point

- It supports the case for neural data science.
- If naïve approaches to understanding the brain are not working, it is probably **not** sufficient to borrow ill-fitting ones from mathematics.
- They are right to point this out.
- But this just reinforces the need for a new kind of science – new methods, methods that are specifically designed to tackle the peculiarities of the brain, which is quite unique physiologically – extremely diverse and heterogeneous ("polyphonic music"/"diplexity") and full of interesting spatiotemporal dynamics.
- It's time take this seriously.

But why bother?

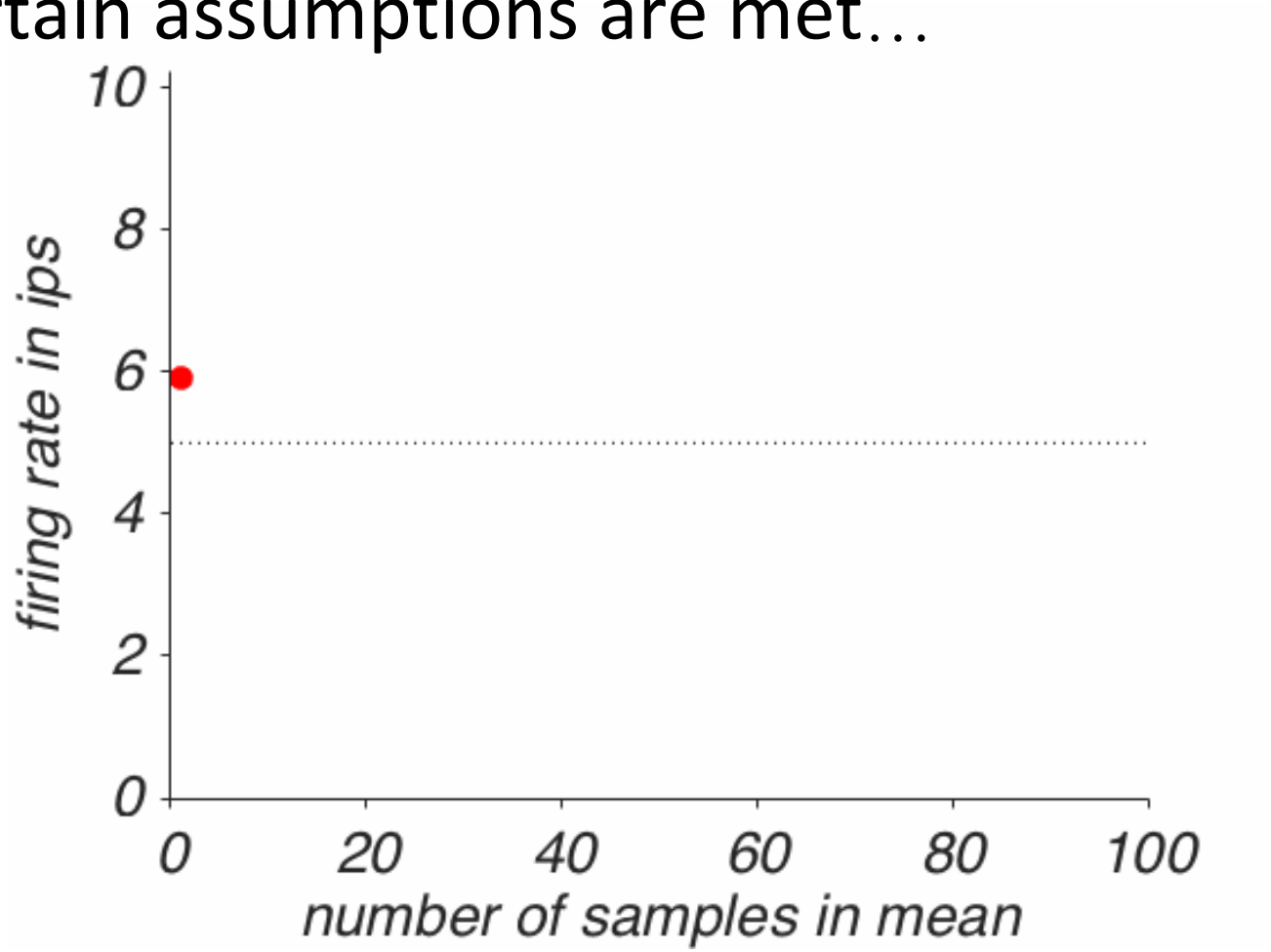
- A cynical view would be that more data and the capability to analyze it in more depth in a model-based way changes some of the details, but doesn't make a qualitative difference.
- So is there any evidence that this is worth doing?

Yes…

- For instance, by revisiting averaging and taking neural diversity seriously.

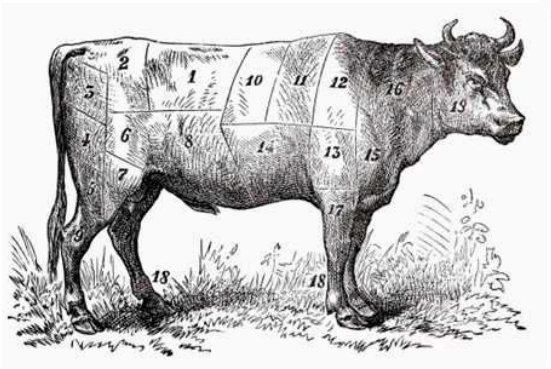
Why is averaging useful?

- It allows us to handle random “error” (variability)
- Convergence to the “true” value, as n increases.
- If certain assumptions are met…



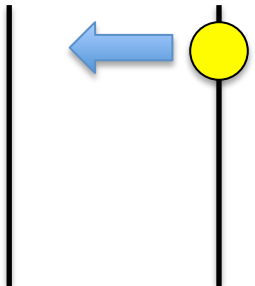
But even something as seemingly innocent as averaging is a radical processing step

- We have just gotten used to doing it.
- If the assumptions are met, it works very well:
- We do it routinely.
- Without thinking much.
- What does it presuppose?
- Where does it come from?



This comes to us from astronomy

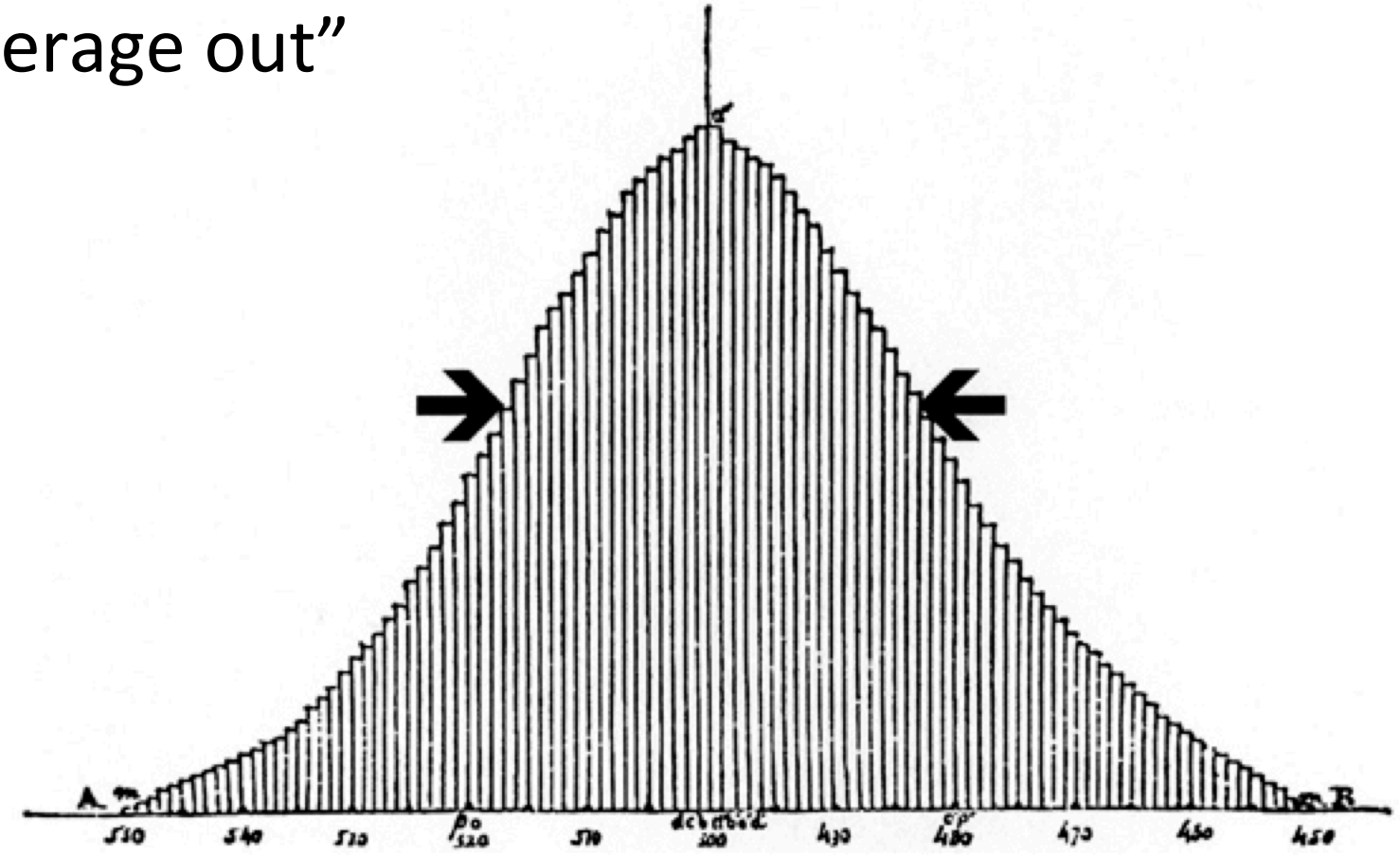
- Quetelet trained as an astronomer, brought statistical methods to other sciences. (e.g. “Social physics”)
- A chief task at the time: Determining the speed of heavenly objects by measuring the time it takes them to pass through two parallel lines etched on the glass of the telescope.



Adolphe Quetelet (1796-1874)

Statistical properties of this task

- Every measurement is not free of (random) error.
- But the object has an actual real speed.
- Combine enough measurements and the errors will "average out"



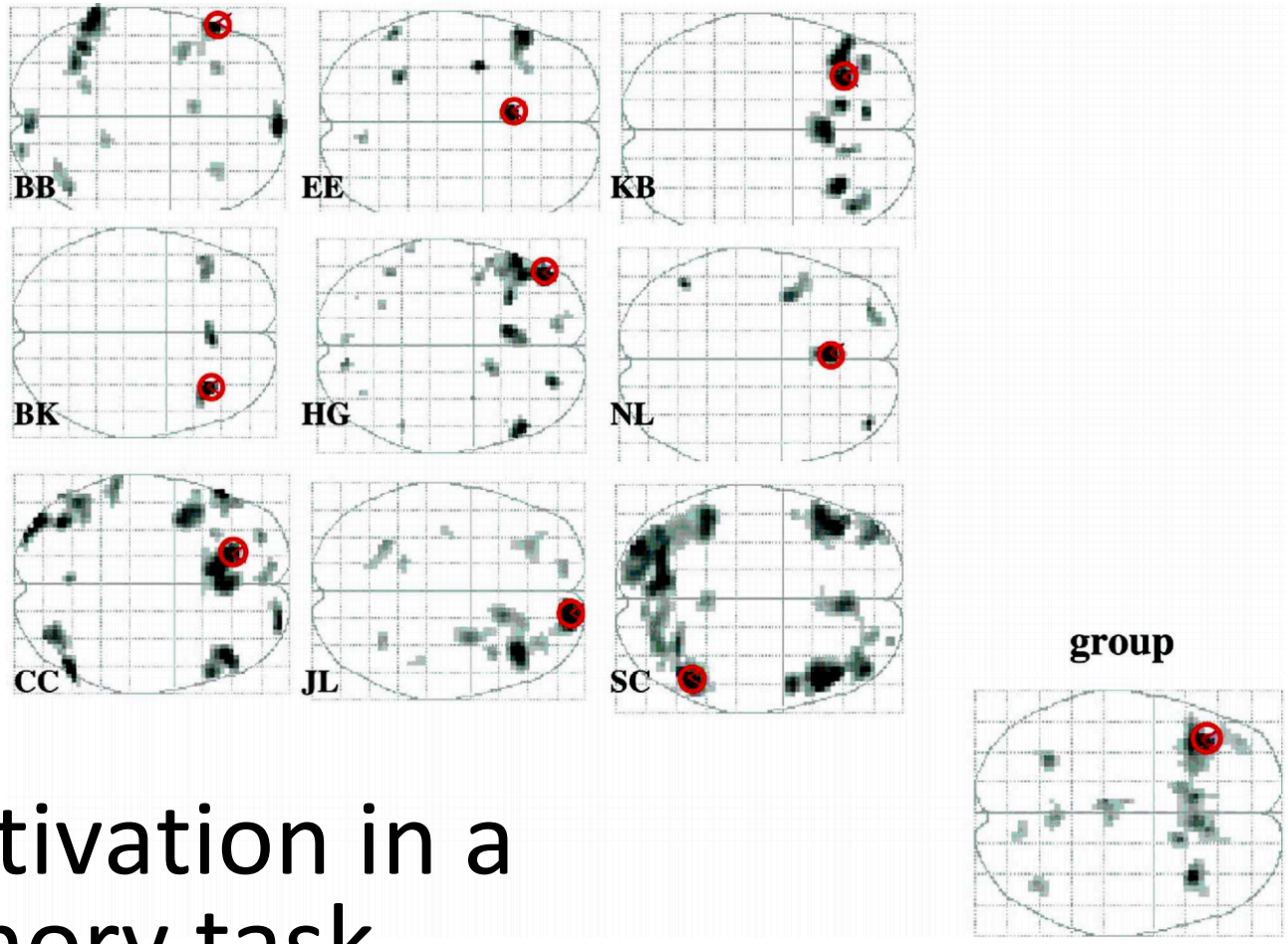
Ergodicity

- According to ergodic theory, one can use the group average to predict the average behavior of individuals if:
  - Every member of the group is identical
  - Every member of the group remains the same (doesn't change over time)

- This works reasonably well in statistical physics, e.g. predicting the behavior of ideal gases and gas molecules.
- How about neuroscience?

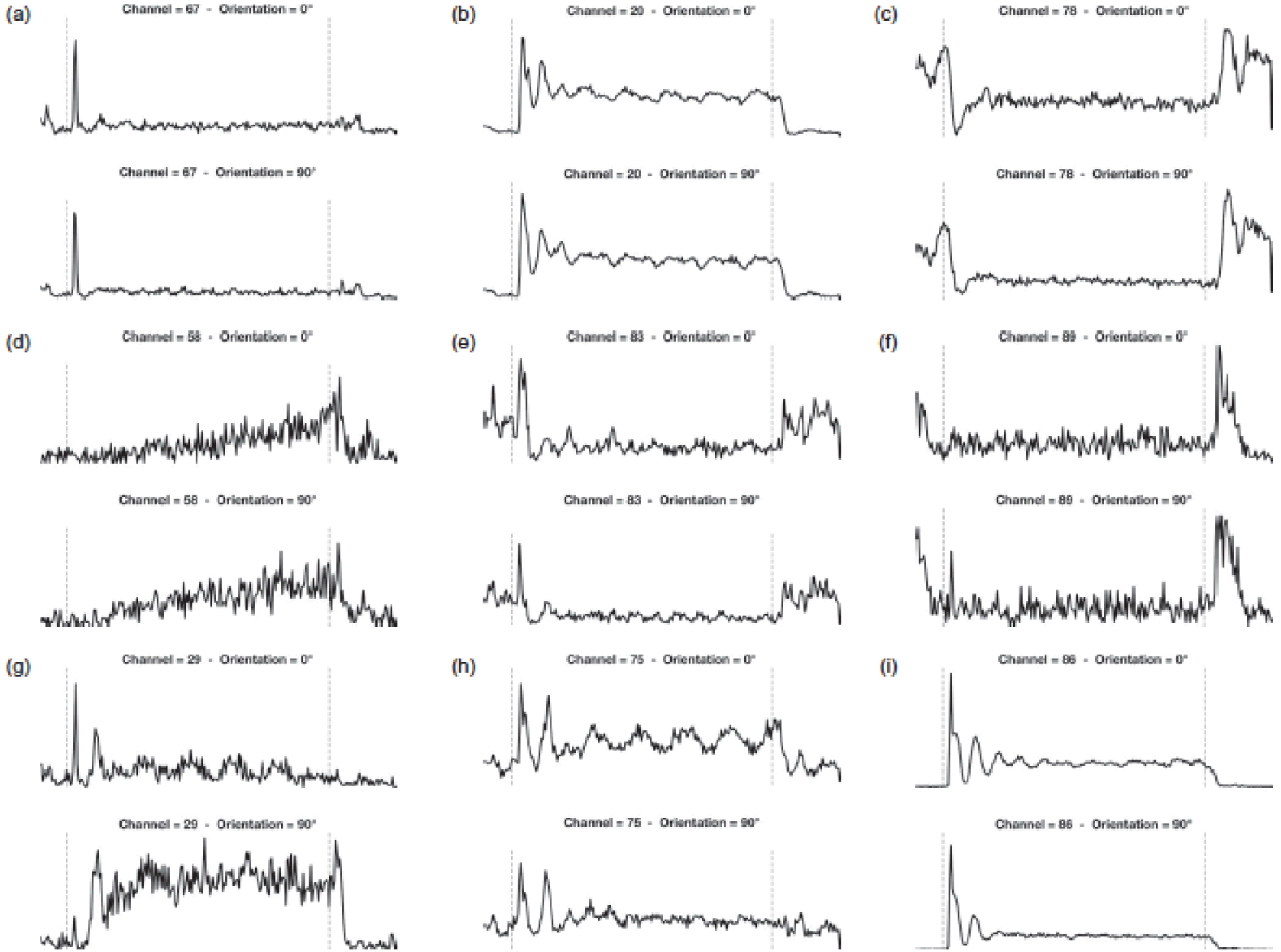
Or: Why is there more than one shoe size?

- Implied model: Deviations from average represent (measurement) error.



fMRI activation in a memory task

Neural diversity is real



Averaging is not an innocent operation

- How you do it can change the theoretical answer

Article

Neuron

Gamma and Beta Bursts Underlie Working Memory

Highlights

- Working memory information in neuronal spiking is linked to brief gamma bursts
- The narrow-band gamma bursts increase during encoding, decoding, and with WM load
- Beta bursting reflects a default network state interrupted by gamma
- Support for a model of WM is based on discrete dynamics and not sustained activity

Authors

Mikael Lundqvist, Jonas Rose, Pawel Herman, Scott L. Brincat, Timothy J. Buschman, Earl K. Miller

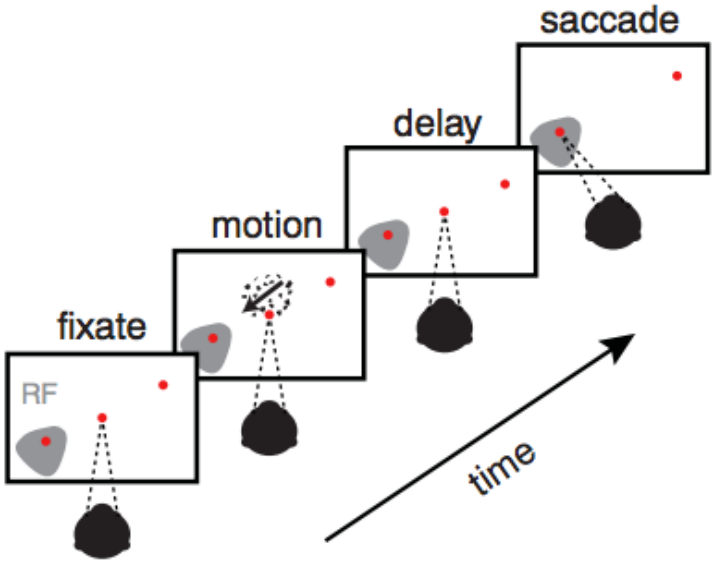
Correspondence

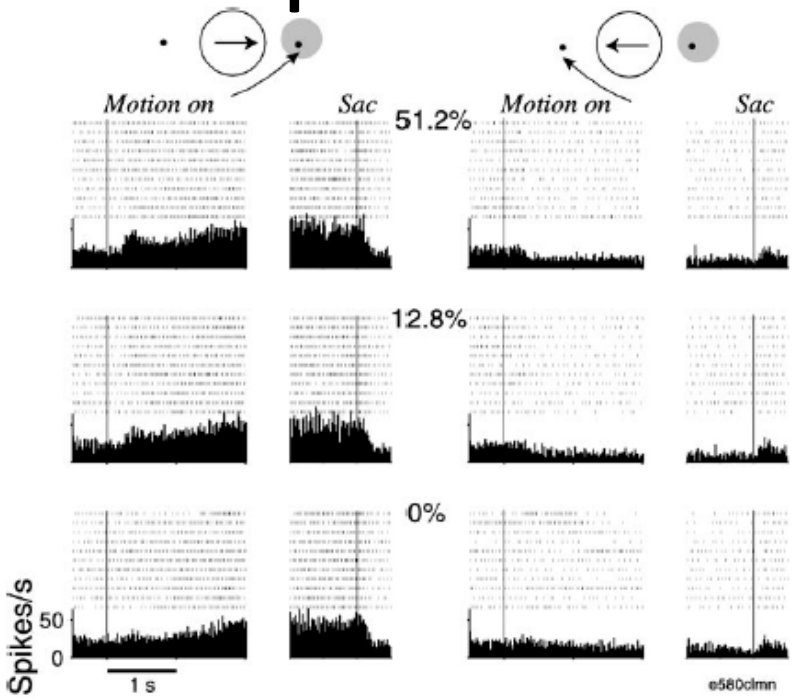
ekmiller@mit.edu

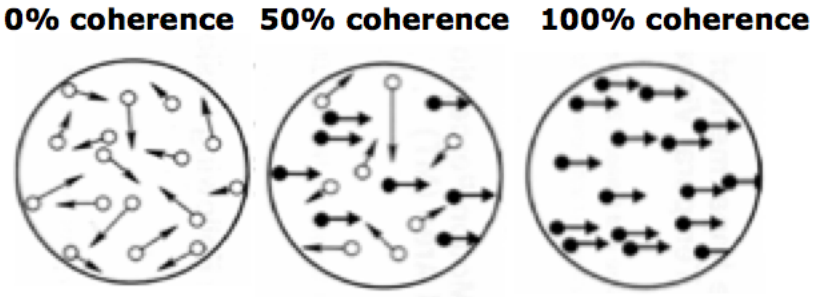
In Brief

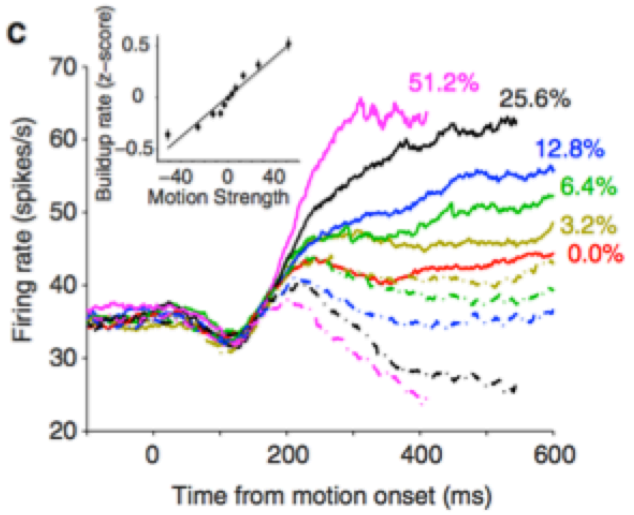
Lundqvist et al. confirm predictions of a working memory model. There is a tight link between gamma oscillations and neural information. Neural events are discrete and thus do not support the modal model of WM based on sustained activity.

The example of “ramps” in LIP

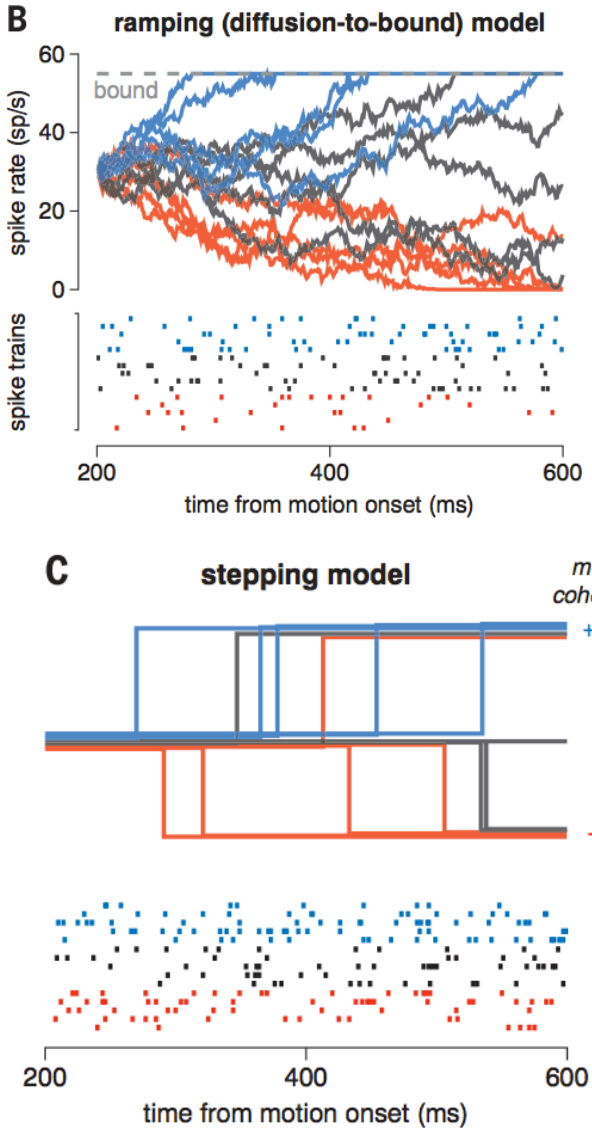


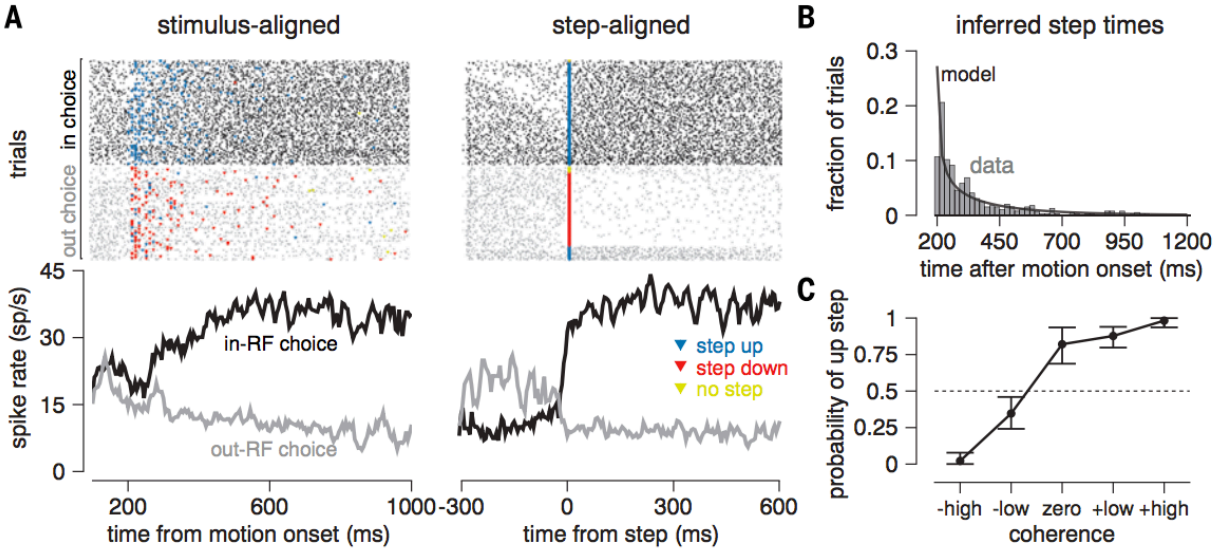






A more sophisticated analysis gives a dramatically different answer





Pillow et al., Science, 2015

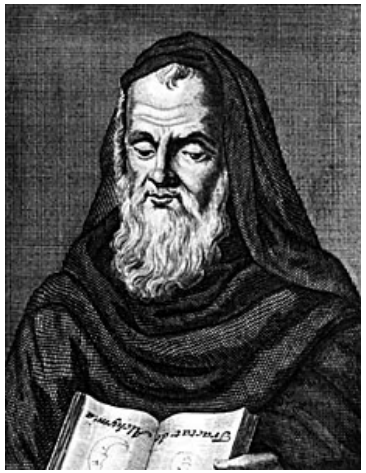
But isn't "data science" a silly name?

What is science?

- Derived from the latin word “scientia”, meaning knowledge.
- In this – broad – sense, any body of knowledge qualifies as a science.
- However, there is a much more narrow sense, in which we understand the term.
- In this sense, most bodies of knowledge are decidedly *not* a science.

Starting with 12th century scholastics

Robert Grosseteste: *Reasoning from the particular case to the general and back again (with respect to the empirical world).*



Roger Bacon: *"experiment and observation combined with mathematics when mathematics were available, and when they were not available, then experiment and observation pursued alone.*

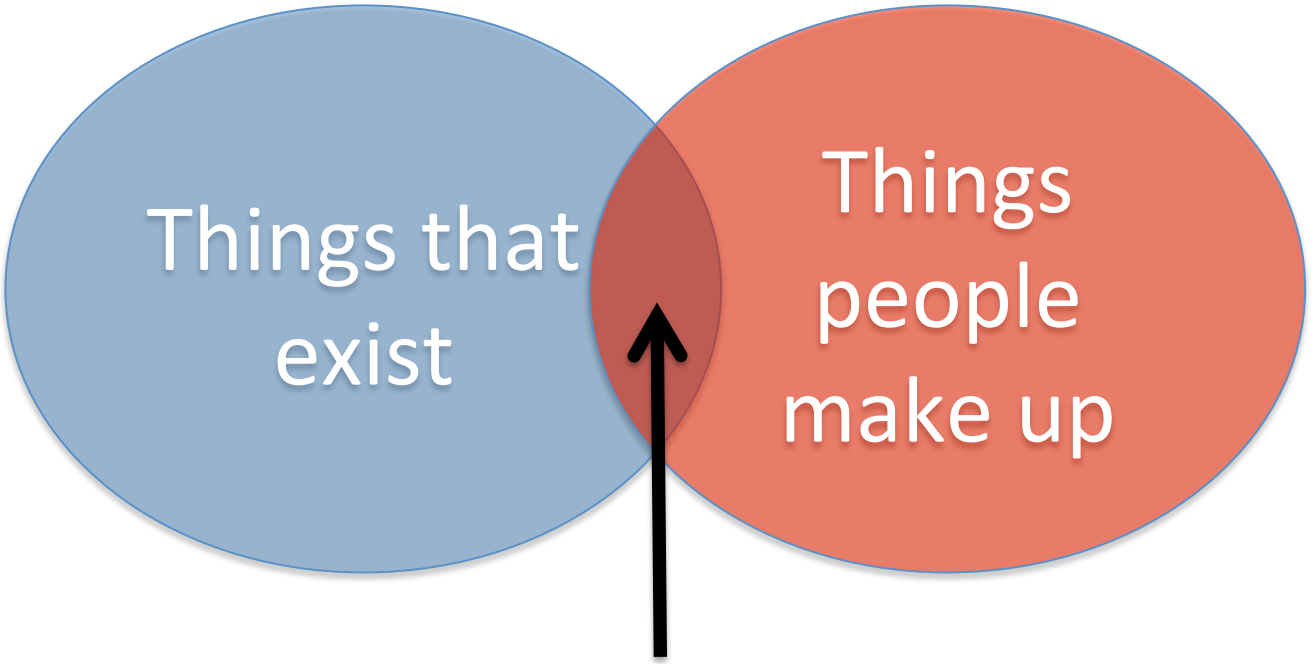
Characteristics of science

- A body of knowledge.
- Goal: A principled understanding of the natural world.
- **Inductive** by formulating principles from observations.
- **Deductive** insofar as principles are tested with new observations.
- The more precise the observations, the better.
- Can deal with simple or complex subjects.
- Experiments help with a certain kind of understanding (particularly for complex subjects) but are not strictly necessary.

What is data?

- Is data born or made - does it exist in the world or does it have to be generated?
- Is data singular or plural?
- Latin: “A given thing” (~1645) – no data in the ancient world. A relatively modern concept.
- However, it is *not* just given. Ironically named.
- Is born (AND made) as a result of a formal observation, a measurement process.

The birth/making of data



What is data science?

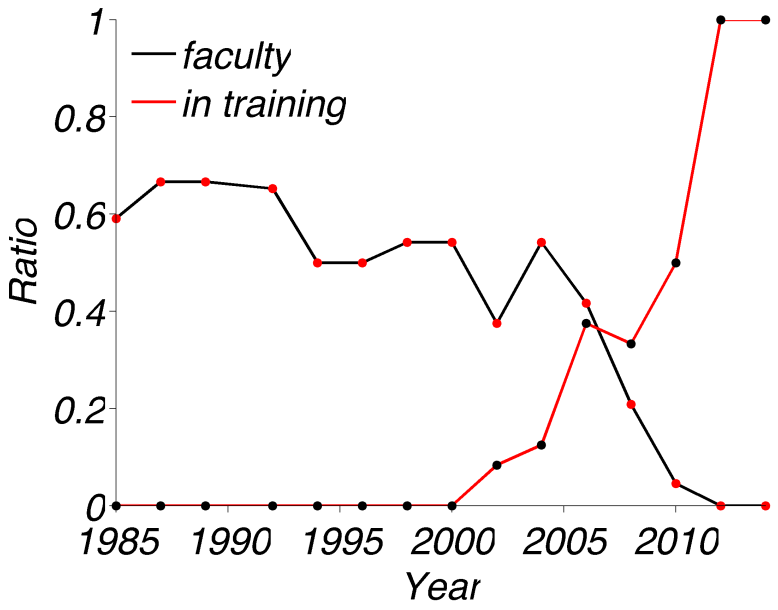
- As we saw, science is an inducto-deductive engine to understand the natural world.
- Data is the fuel that powers this engine.
- So data science is unfortunately named. Data is absolutely necessary for science of **any** kind.
- The idea is that “data science” per se looks for patterns in “big data” by any means necessary and across fields.
- Eliciting the structure of data.

What is neural data science?

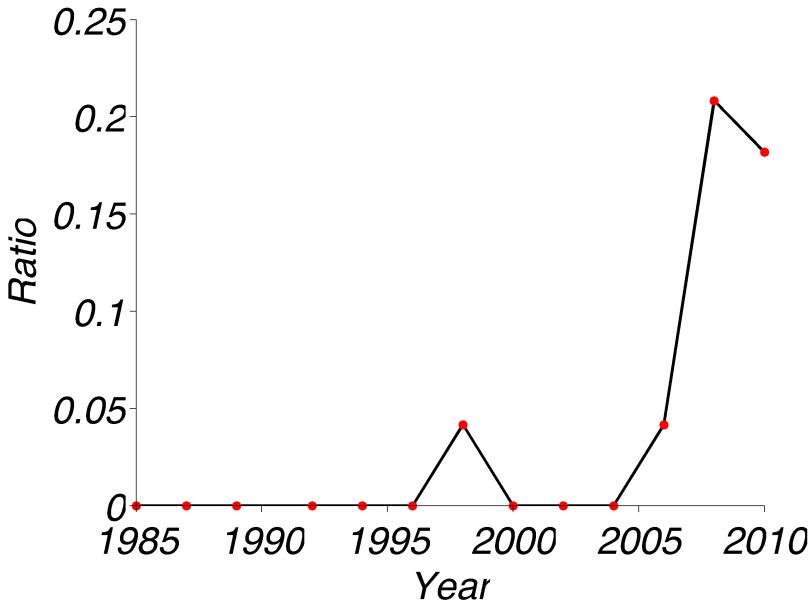
- The concept of neural data science implies that the "strong" program of data science is probably misguided.
- It is very likely that the field matters, as every field will create data with a particular structure (which will inform its analysis).
- So neural data science is the application of the principles of data science to neuroscience data.
- It is a good bet that this is a sensible endeavor, as the brain is complex enough to both necessitate this approach to understand it, as well as yield the data to do so.

A pragmatic consideration

p(faculty|taking course)



The rise of data science



Where are we going?