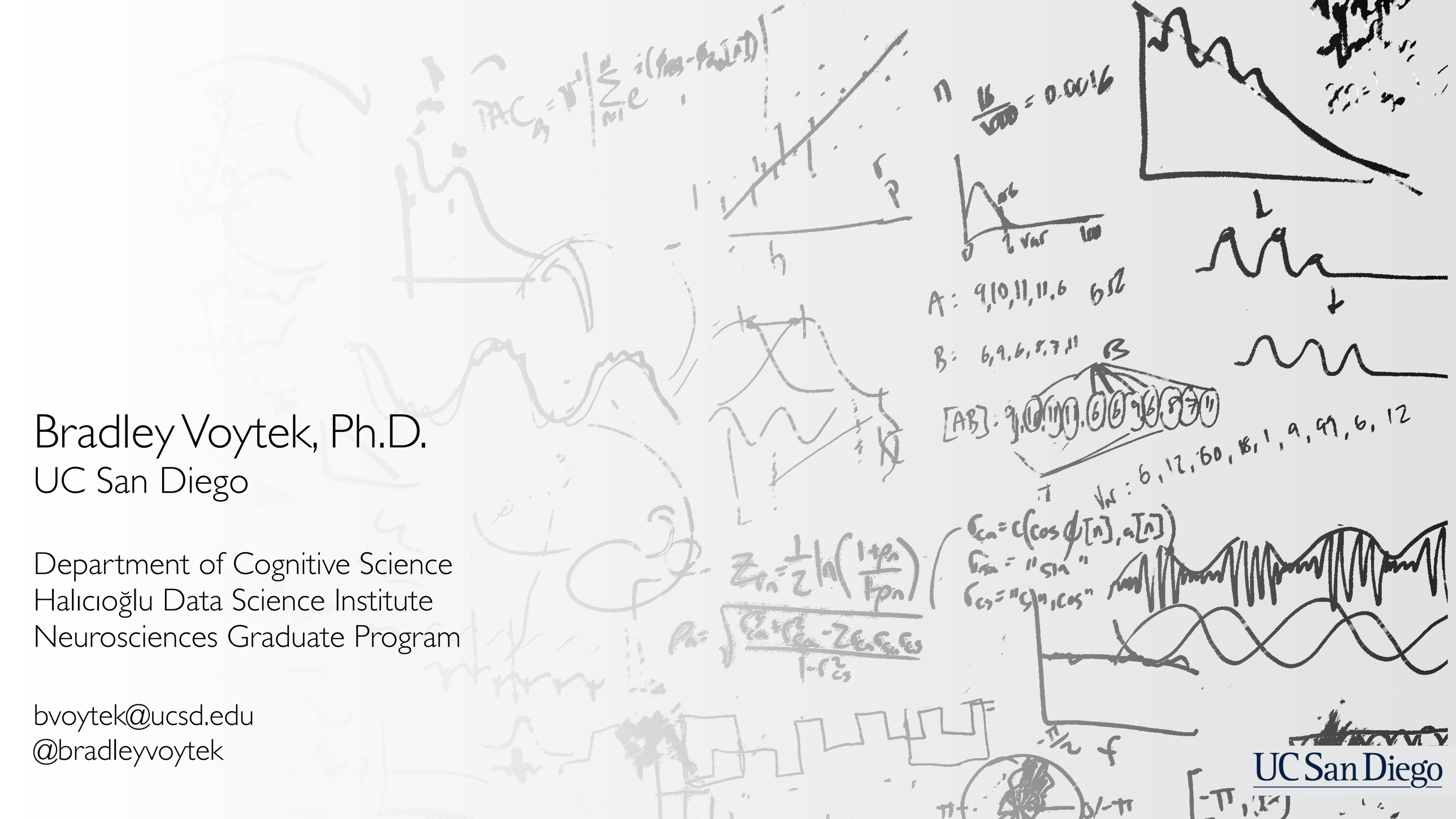
Bradley Voytek, Ph.D.

UC San Diego

Department of Cognitive Science
Halıcıoğlu Data Science Institute
Neurosciences Graduate Program

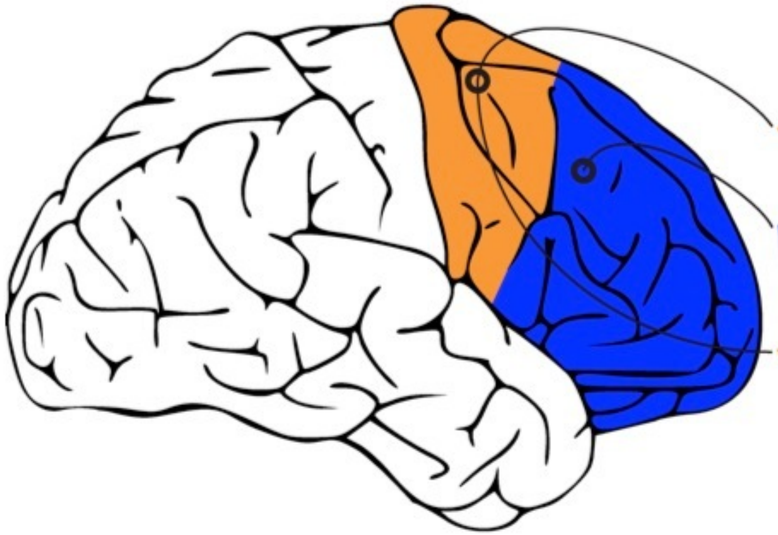
bvoytek@ucsd.edu
@bradleyvoytek



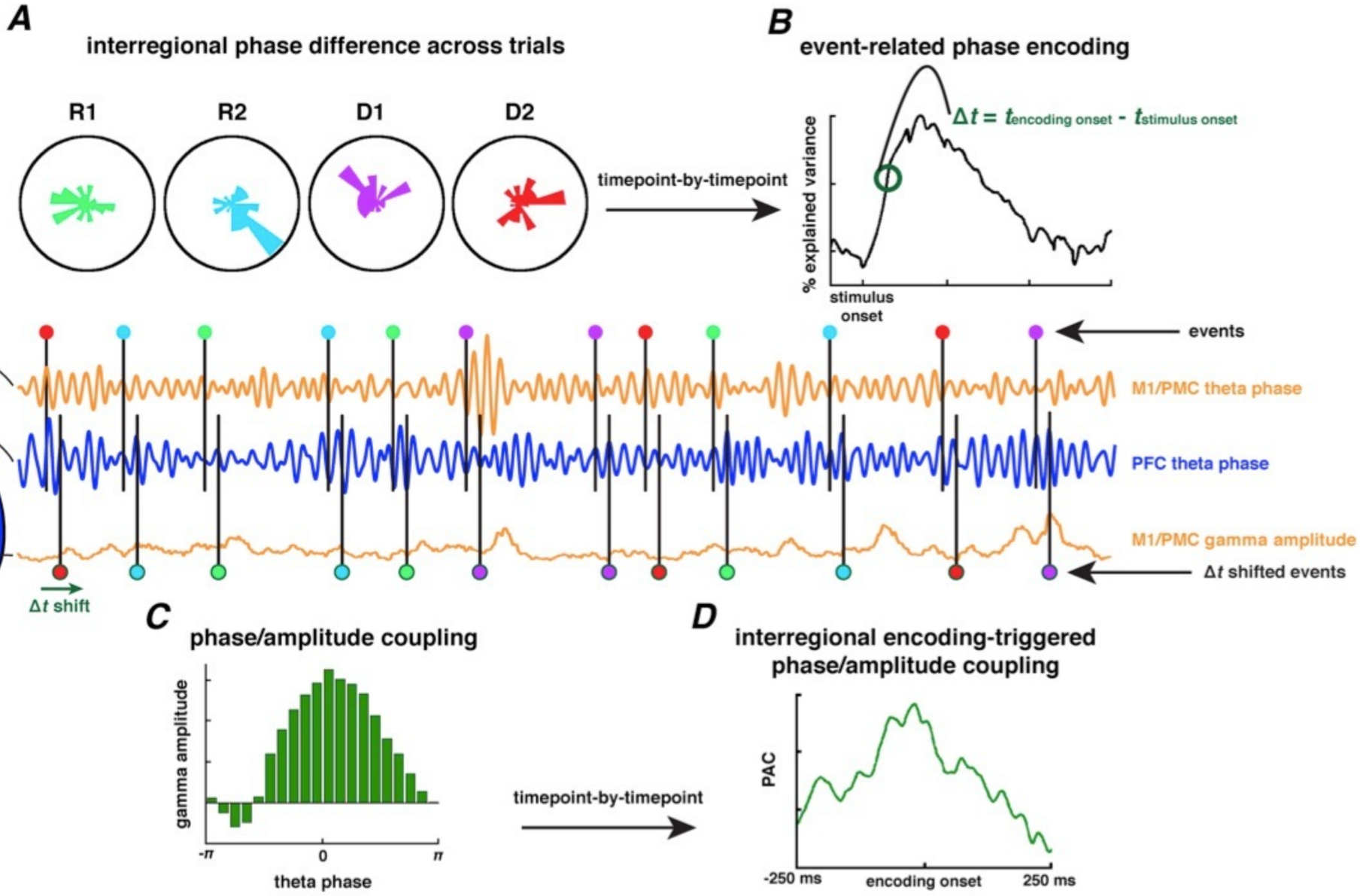
COGS 1
Introduction to Cognitive Science

UC San Diego

$$h_{a_{\gamma\theta}}[n] = a_{a_{\gamma\theta}}[n]\exp\left(i\phi_{a_{\gamma\theta}}[n]\right)$$



$$Coh_{i,j} = N^{-1}\left|\sum_{n=1}^{N} e^{i(\phi_{\theta i}[n]-\phi_{\theta j}[n])}\right|$$



Source: Voytek *et al.*, *Nature Neurosci* 2015

WIRED

THE SCIENCE OF SURVIVING THE ZOMBIE APOCALYPSE

THE ZOMBIE DISORDER

CONSCIOUSNESS DEFICIT HYPOACTIVITY DISORDER

Consciousness Deficit Hypoactivity Disorder (CDHD): The loss of rational, voluntary and conscious behavior replaced by delusional/impulsive aggression,stimulus-driven attention, the inability to coordinate motor linguistic behaviors and an insatiable appetite for human flesh.

SCANS OF THE ZOMBIE BRAIN

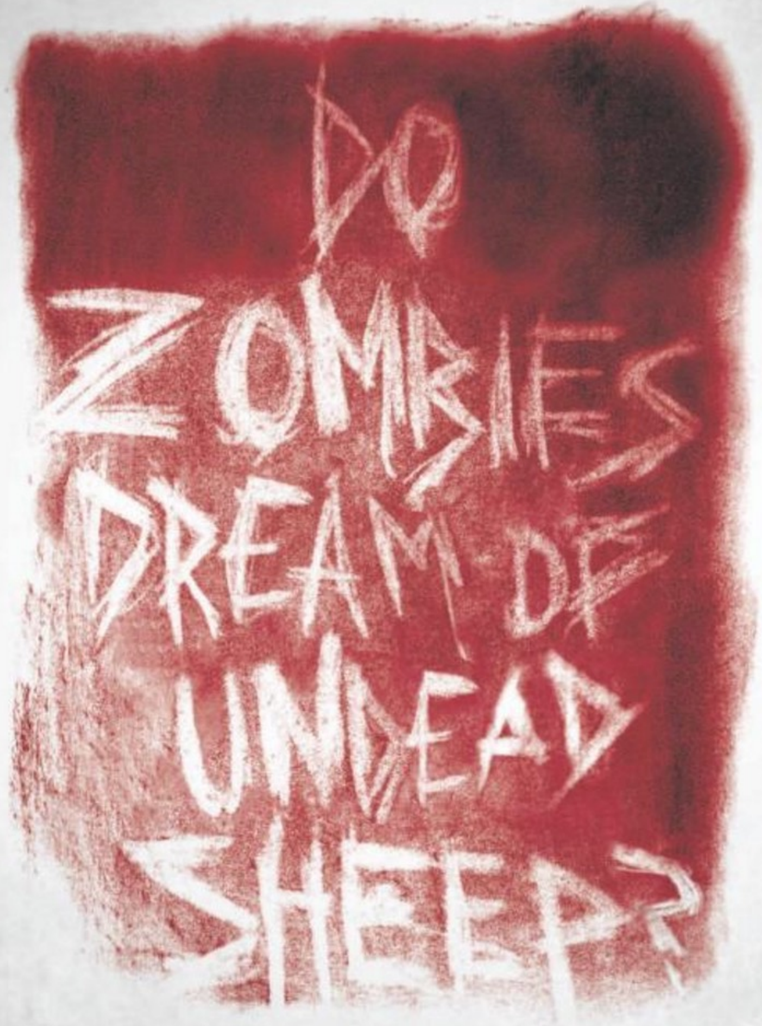
ZOMBIE

HUMAN

Through detailed scans, the exact brain areas that have been destroyed in the zombie can be reconstructed.

The scans show significant brain tissue loss in the zombie. The gray area shows what a human brain would look like. The profile of damage corroborates the behavioral observations of zombies.

TIMOTHY VERSTYNEN
BRADLEY VOYTEK



A NEUROSCIENTIFIC VIEW OF THE ZOMBIE BRAIN

Rave reviews

Five Stars

By rks1125 on February 6, 2015

Format: Hardcover | **Verified Purchase**

As described

Observations

Scientists: Advertise Your Failures!

They're a part of every career, and being upfront about them can help put things in perspective

By Elizabeth Landau on September 25, 2017

Most academics are not so vocal about the setbacks they've faced, but Voytek's story has gained traction. Online discussion boards about graduate admissions have invoked Voytek's name to show that it's possible to enter a Ph.D. program—he went to the University of California, Berkeley—with a subpar transcript. One even asks: "Is the neuroscientist Bradley Voytek a real or fictional person?"

Data Science!


HALICIOĞLU DATA SCIENCE INSTITUTE

UC San Diego Data Science

UC San Diego

Past guest lecturers

- Franziska Bell, PhD: Senior Data Science Manager, *Uber*
- Eli Bressert, PhD: Manager, Data Engineering & Analytics, *Netflix*
- Claire Dorman, PhD: Data Scientist, *Pandora*
- Mina Doroud, PhD: Data Scientist, *Twitter* (Senior Data Scientist, *LinkedIn*)
- Carlos Gomez-Uribe, PhD: Director, Core Data Science, *Facebook* (Statistician, *Google*; VP Product Innovation, *Netflix*)
- Hiroki Hiyama, PhD: Senior Data Scientist, *Uber*
- Emi Nomura, PhD: Senior Manager, Data Science, *Pandora*
- Kevin Novak: Head of Data Science, *Uber*
- DJ Patil, PhD: US Chief Data Scientist, *Obama White House* (Head of Data Science, *LinkedIn*; co-coined the term "data science")
- Maksim Pecherskiy: Chief Data Officer, *City of San Diego*
- Sarah Rich, PhD: Data Scientist, *Twitter*
- John Myles White, PhD: Research Scientist, *Facebook* (Author: *Machine Learning for Hackers*)
- Josh Wills, PhD: Head of Data Engineering, *Slack*

Why Data Science?

Harvard Business Review

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

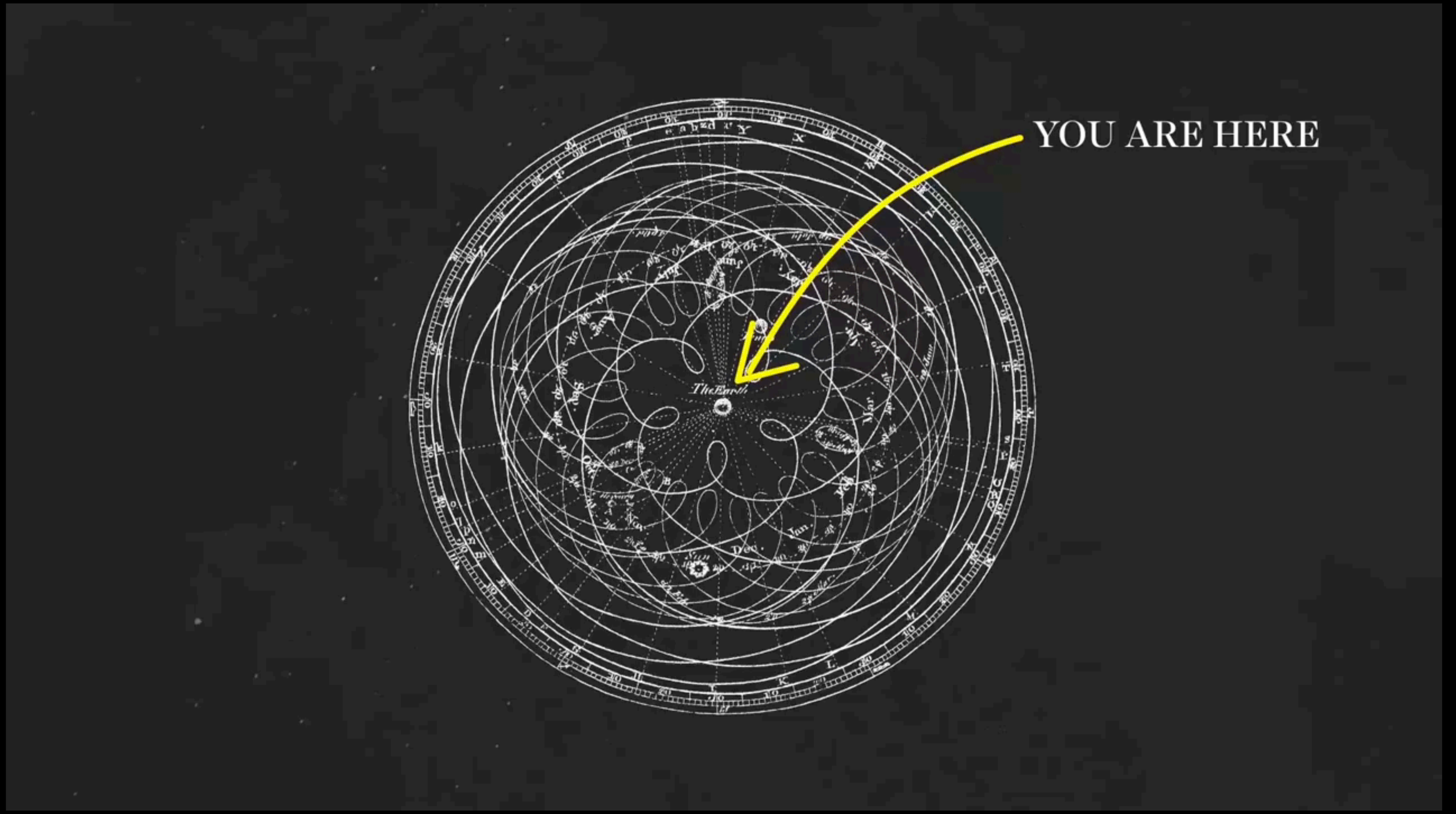
ntist: The Sexiest
e 21st Century

. Patil

Sexiest


NATIONAL
PARK
SERVICE

Source: NASA



Source: *Vox*

Ptolemaic model

- Ordering: Moon, Mercury, Venus, Sun, Mars, Jupiter, Saturn, Fixed Stars

Source: Wikipedia

Ptolemaic model

- Ordering: Moon, Mercury, Venus, Sun, Mars, Jupiter, Saturn, Fixed Stars
- Ptolemy did not invent or work out this order.

Source: Wikipedia

Ptolemaic model

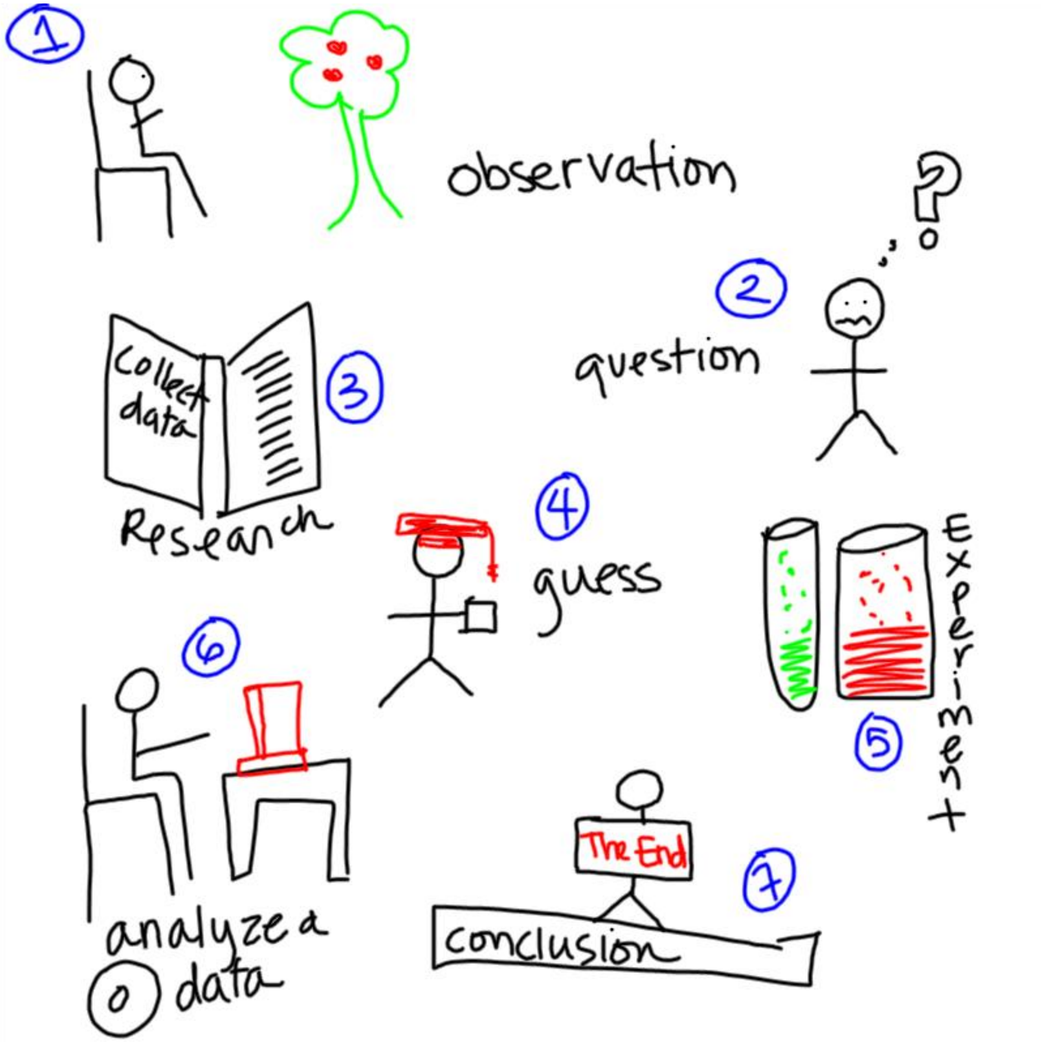
- Ordering: Moon, Mercury, Venus, Sun, Mars, Jupiter, Saturn, Fixed Stars
- Ptolemy did not invent or work out this order.
- It comes from the ancient "Seven Heavens" religious cosmology common to the major Eurasian religious traditions.

Source: Wikipedia

Ptolemaic model

- That is, Ptolemy begins with observations.

The Scientific Method



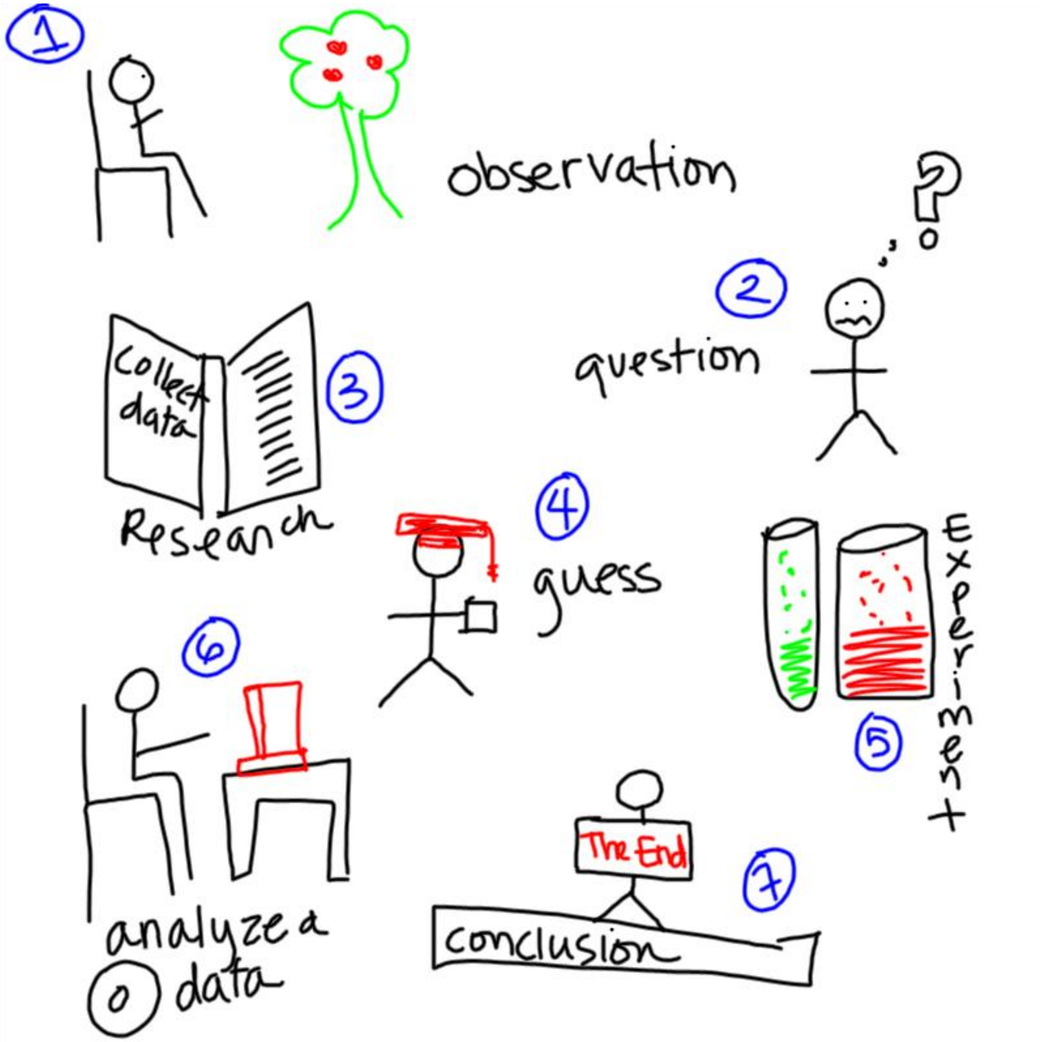
Ptolemaic model

- Accurate
- Predictive

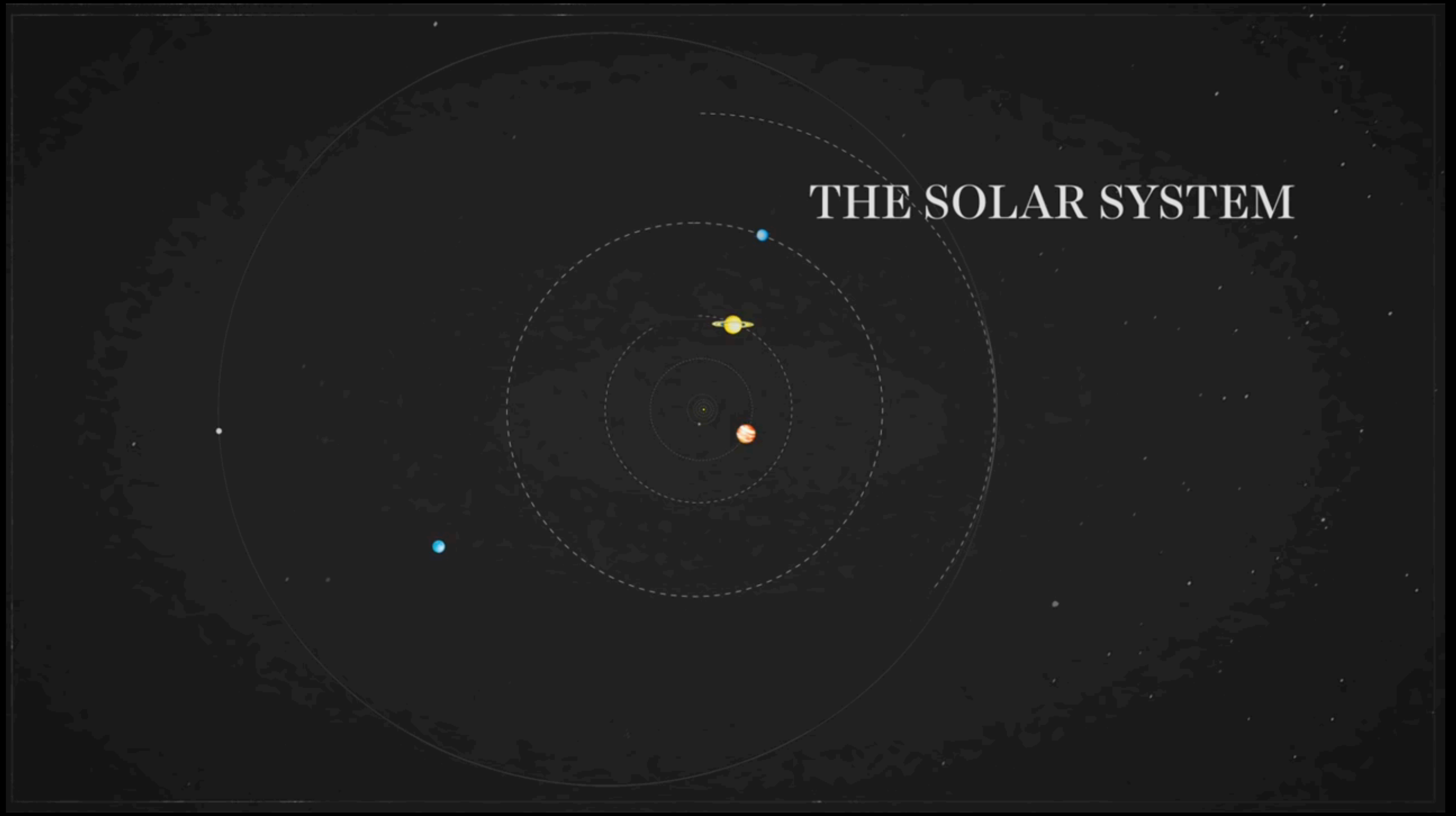
Ptolemaic model

- He builds a model given *social and political constraints*.

The Scientific Method



social and political constraints?!?



Source: *Vox*

So what is the point of modeling?

- Prediction

Modeling

2.3 Parsimony

Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity.

2.4 Worrying Selectively

Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.

Source: Box GEP, *J Am Stat Assoc*, 1976

Modeling

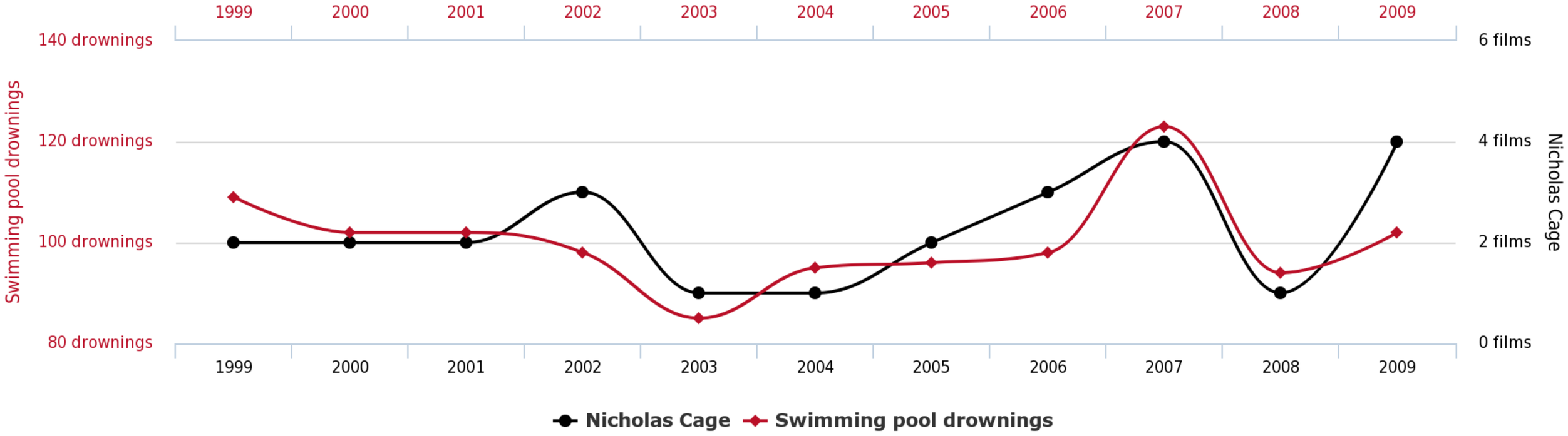
- All models are wrong.

Source: Box GEP, *J Am Stat Assoc*, 1976

Number of people who drowned by falling into a pool

correlates with

Films Nicolas Cage appeared in



tylervigen.com

Modeling

- All models are wrong.
- Some models are useful.

Source: Box GEP, *J Am Stat Assoc*, 1976

So what is the point of modeling?

- Prediction

So what is the point of modeling?

- Prediction
- Classification

So what is the point of modeling?

- Prediction
- Classification
- Knowledge discovery

So what is the point of modeling?

- Prediction
- Classification
- Knowledge discovery
- **DOING USEFUL SHIT**

What is Data Science?

Rigor and intuition

In today's pattern recognition class my professor talked about PCA, eigenvectors & eigenvalues.

I got the mathematics of it. If I'm asked to find eigenvalues etc. I'll do it correctly like a machine. But I didn't **understand** it. I didn't get the purpose of it. I didn't get the feel of it. I strongly believe in

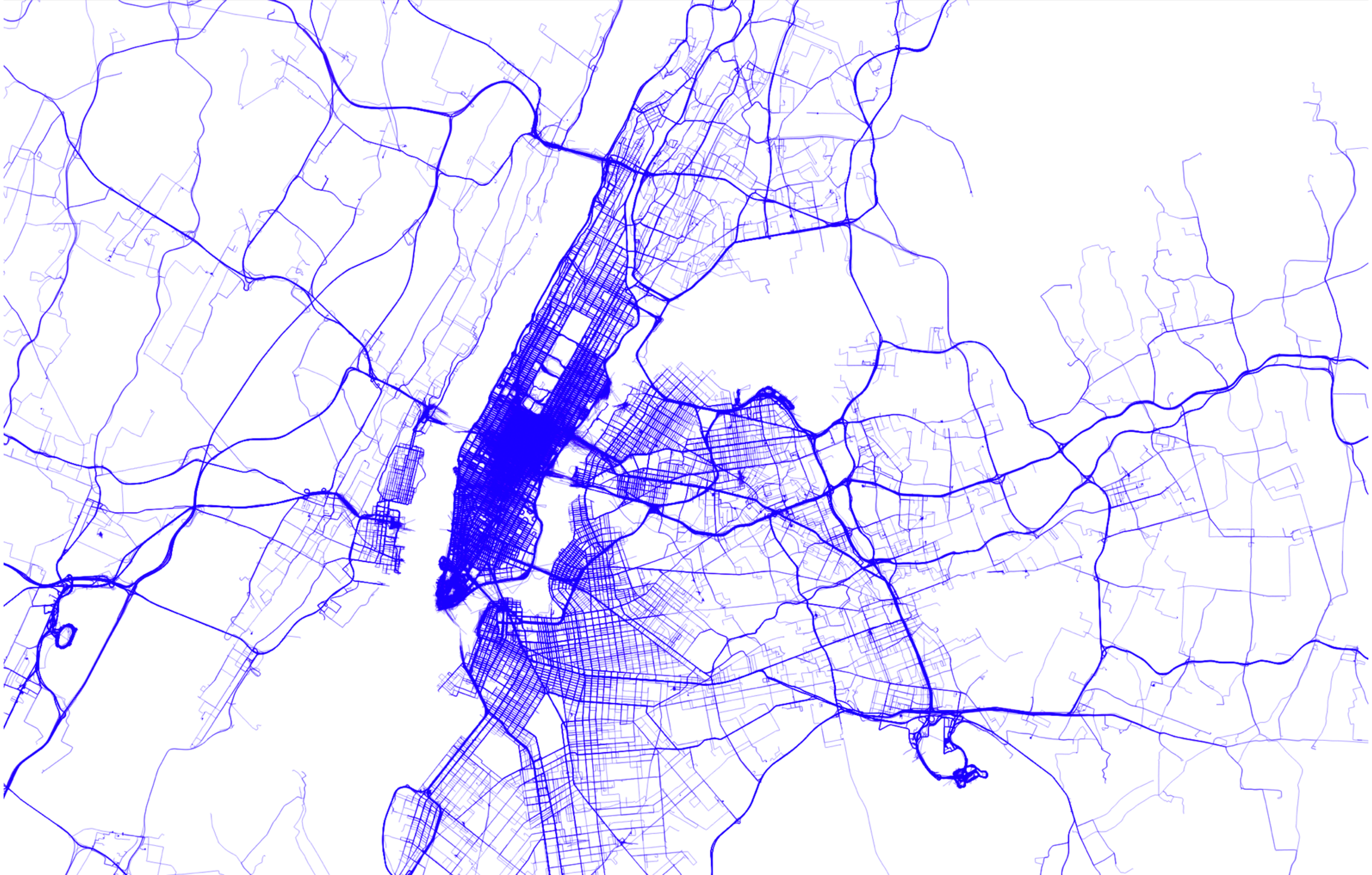
> you do not really understand something unless you can explain it to your grandmother -- Albert Einstein

Well, I can't explain these concepts to a layman or grandma.

1. Why PCA, eigenvectors & eigenvalues? What was the *need* for these concepts?
2. How would you explain these to a layman?

Uber


uber
Bradley Voytek, Ph.D.
voytek@uber.com
[redacted]
@bradleyvoytek
http://uber.com



Parametrization?

The previous plot has sparse data: just lat/long and time

Parametrization?

The previous plot has sparse data: just lat/long and time

From that you can get first-order calculated metrics—velocity, acceleration, etc.

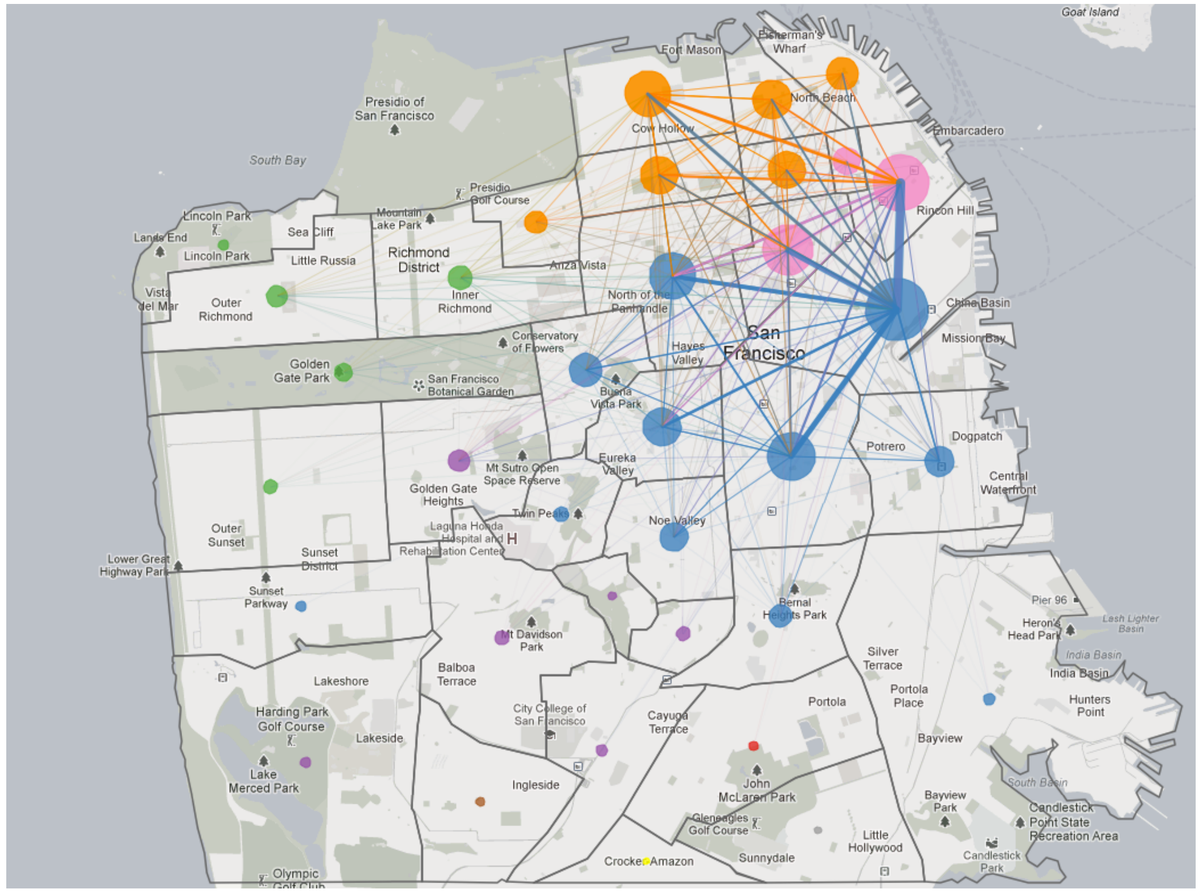
Parametrization?

The previous plot has sparse data: just lat/long and time

From that you can get first-order calculated metrics—velocity, acceleration, etc.

But how can you turn such relatively sparse data into actionable business decisions (KPIs)?

City dynamics



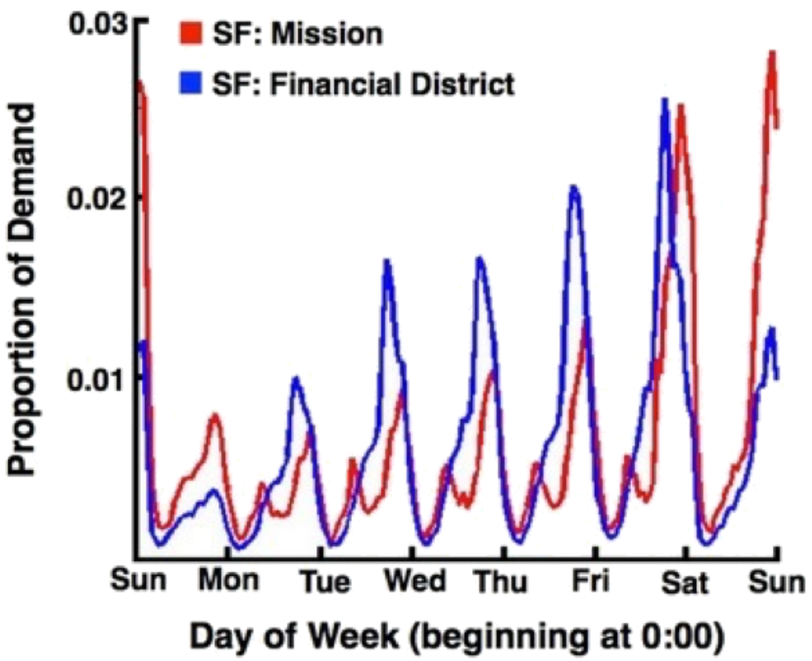
- Aggregating data into arbitrary geographical units: here neighborhoods, but under the hood you'd use e.g., hexagonal tiling.

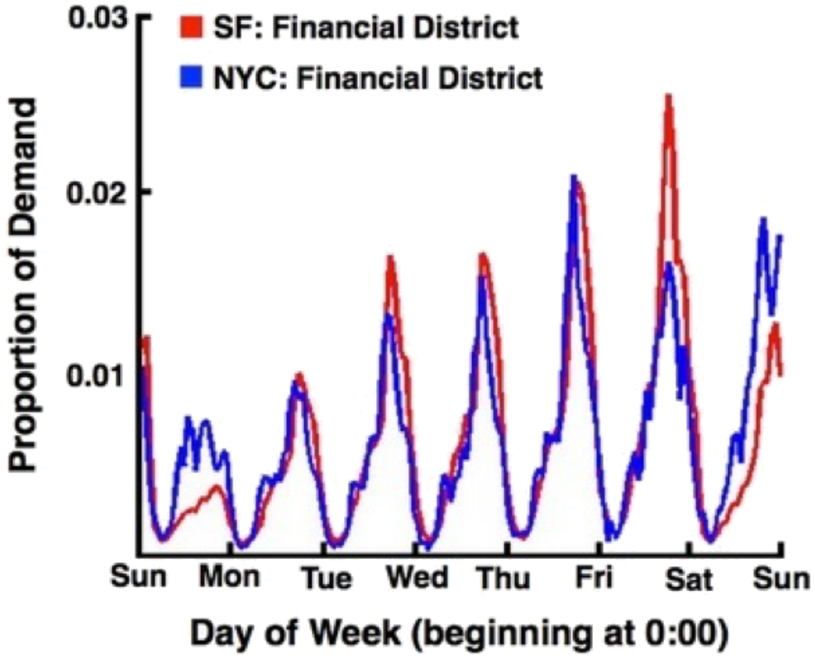
- Turns lat/lng pairs and time into temporal profiles of demand between parts of a city.

Source: Voytek, *Uber, Inc.*

UC San Diego

Spatiotemporal dynamics





Source: Voytek, *Uber, Inc.*

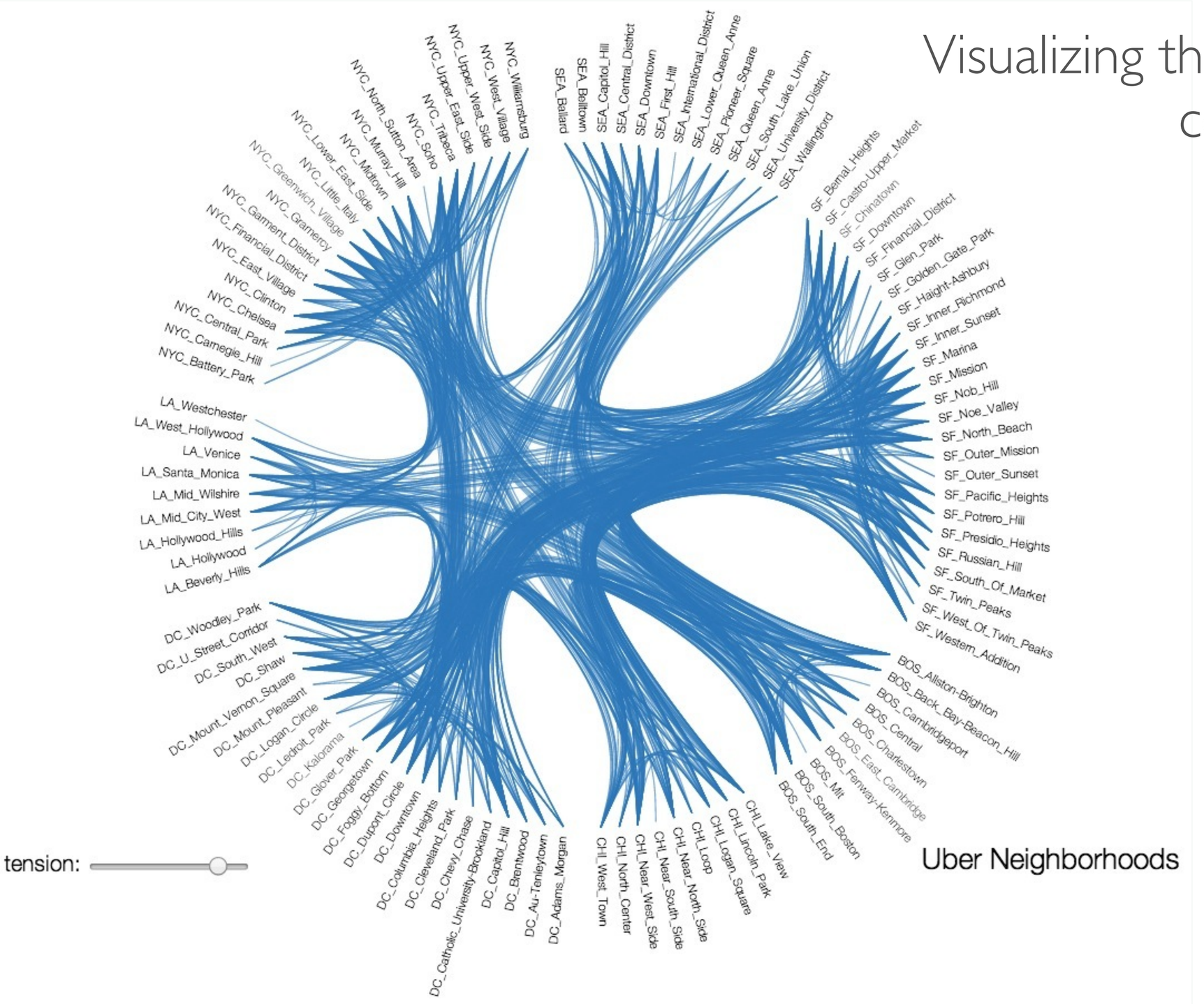
Spatiotemporal dynamics

- **Looking at dynamics over time allows you to correlate neighborhoods within and between cities.**

- **Can identify "types" of neighborhoods: those with peak weekend and late night demands are more "party"-like whereas M-F peaks are more "business" regions.**

Source: Voytek, *Uber, Inc.*

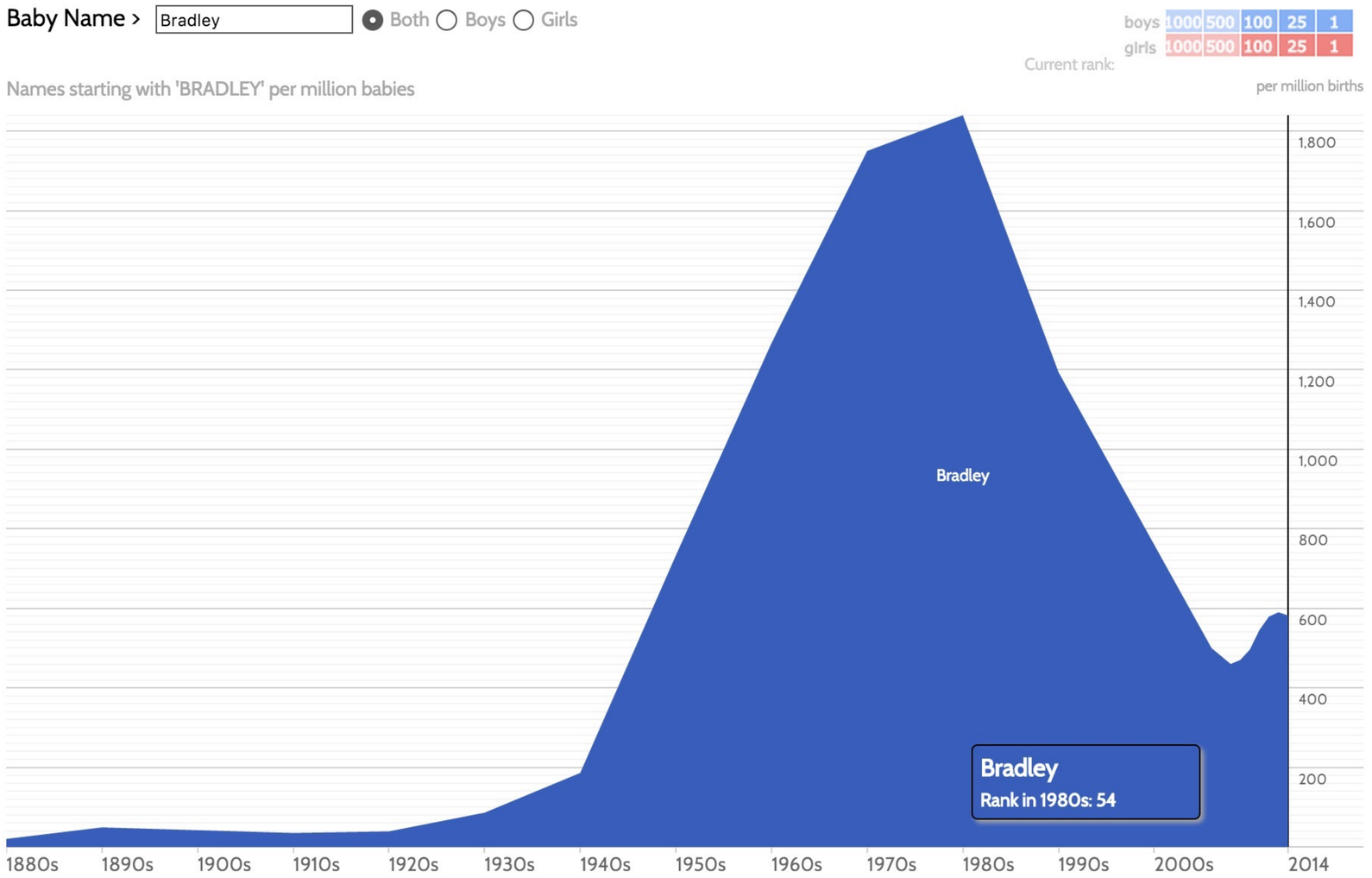
Visualizing the neighborhood correlation graph



The Human Side of Data Science

Names over time

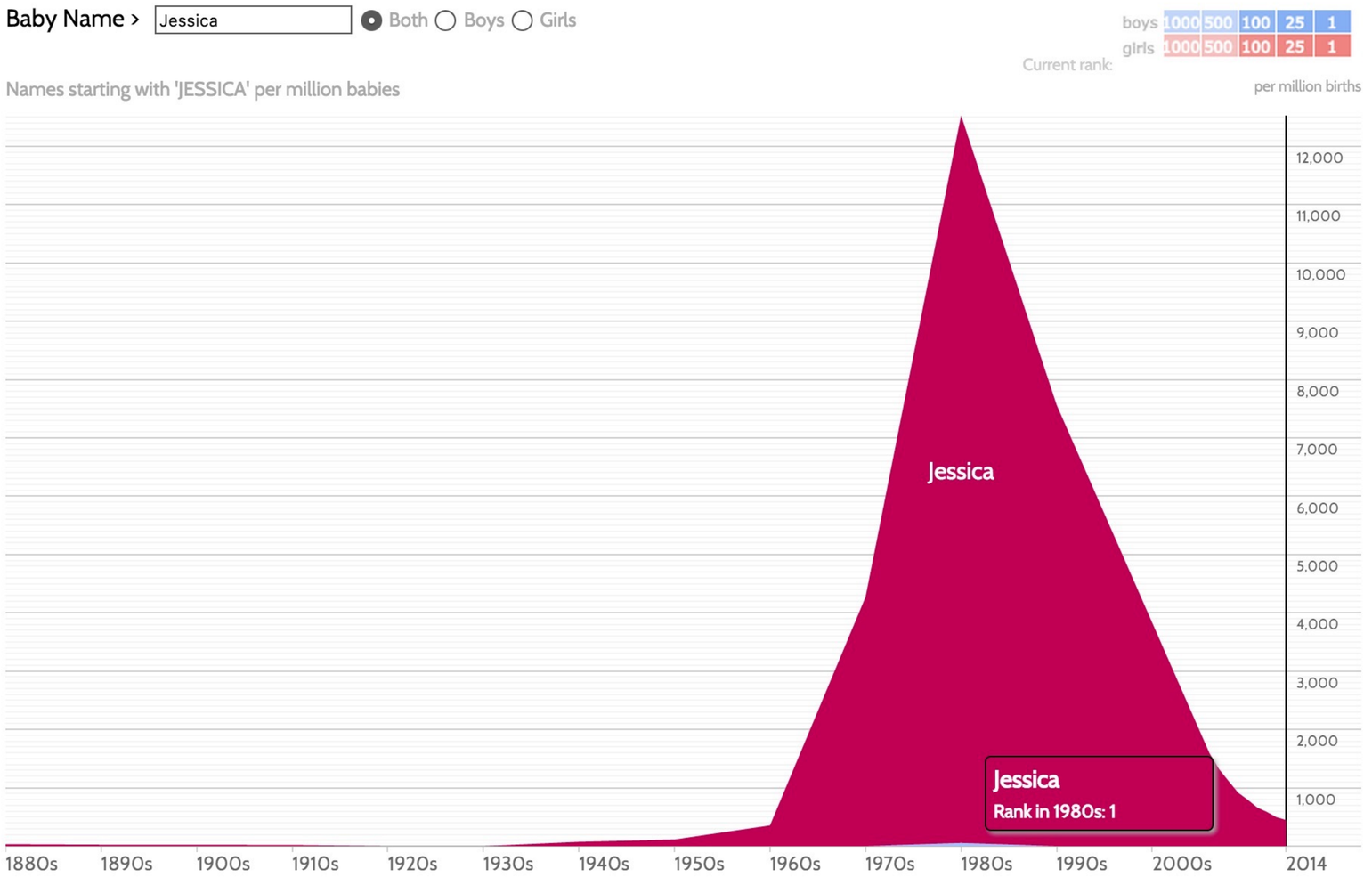
- Quantifying name popularity, trendiness, and "toxicity".



Source: babynamewizard.com

UC San Diego

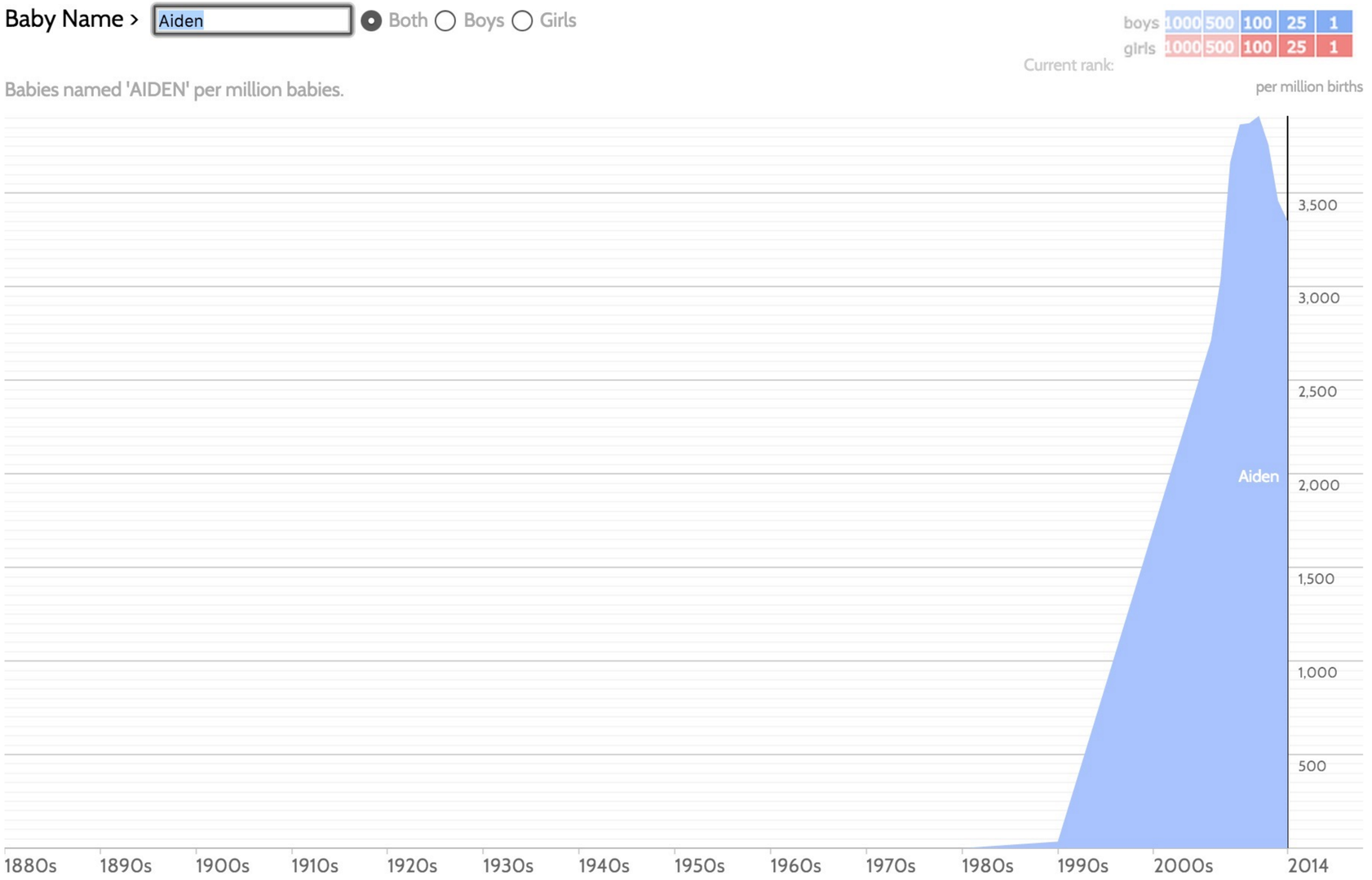
Names over time



Source: babynamewizard.com

UC San Diego

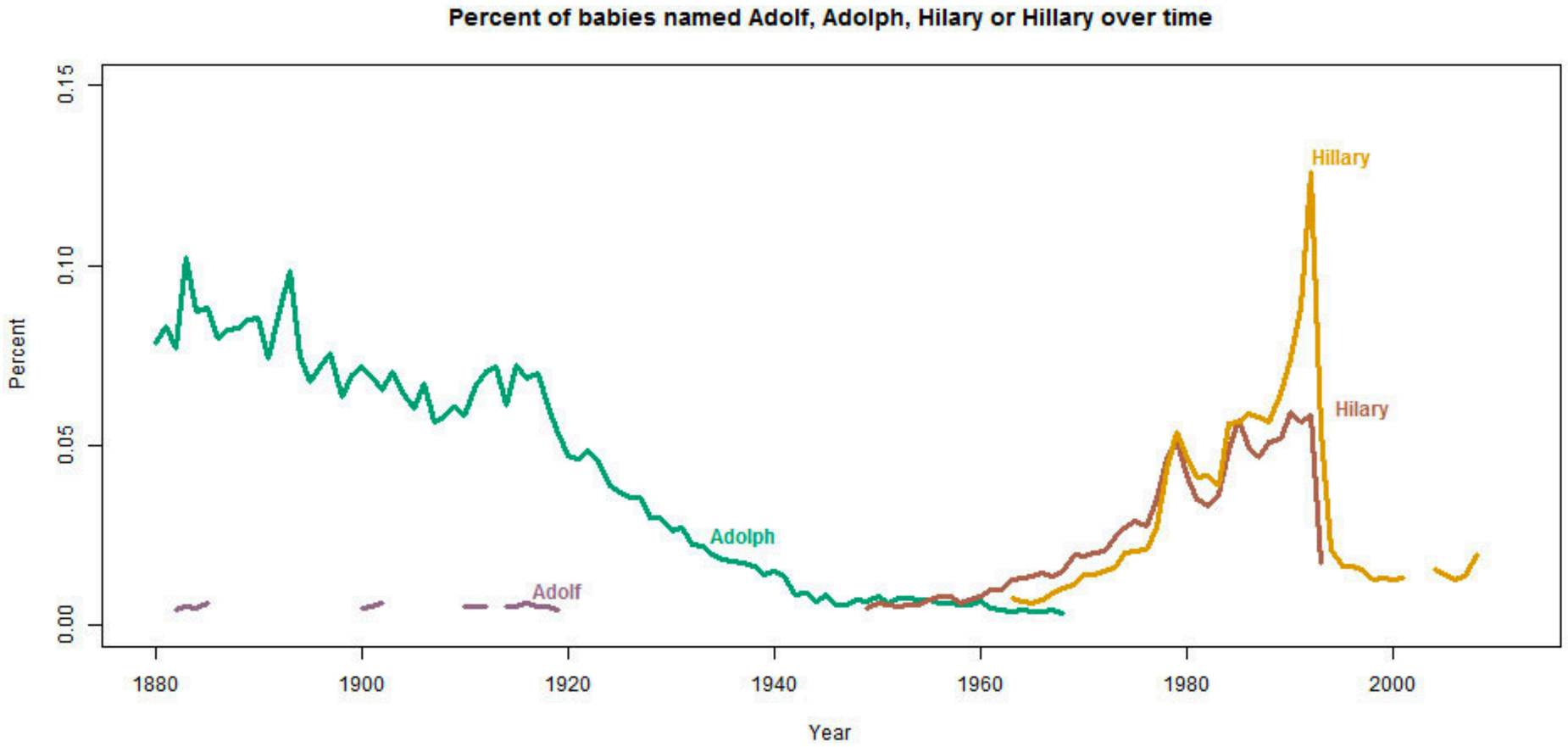
Names over time



Source: babynamewizard.com

UC San Diego

Name "toxicity"



Source: https://www.thecut.com/2013/01/hillary-most-poisoned-baby-name-in-us-history.html

Plurals

Source: McWhorter J, *TED-Ed* 2013

Verb regularization

Table 1 | The 177 irregular verbs studied

Frequency	Verbs	Regularization (%)	Half-life (yr)
10^{-1}–1	be, have	0	38,800
10^{-2}–10^{-1}	come, do, find, get, give, go, know, say, see, take, think	0	14,400
10^{-3}–10^{-2}	begin, break, bring, buy, choose, draw, drink, drive, eat, fall, fight, forget, grow, hang, help, hold, leave, let, lie, lose, reach, rise, run, seek, set, shake, sit, sleep, speak, stand, teach, throw, understand, walk, win, work, write	10	5,400
10^{-4}–10^{-3}	arise, bake, bear, beat, bind, bite, blow, bow, burn, burst, carve, chew, climb, cling, creep, dare, dig, drag, flee, float, flow, fly, fold, freeze, grind, leap, lend, lock, melt, reckon, ride, rush, shape, shine, shoot, shrink, sigh, sing, sink, slide, slip, smoke, spin, spring, starve, steal, step, stretch, strike, stroke, suck, swallow, swear, sweep, swim, swing, tear, wake, wash, weave, weep, weigh, wind, yell, yield	43	2,000
10^{-5}–10^{-4}	bark, bellow, bid, blend, braid, brew, cleave, cringe, crow, dive, drip, fare, fret, glide, gnaw, grip, heave, knead, low, milk, mourn, mow, prescribe, redden, reek, row, scrape, seethe, shear, shed, shove, slay, slit, smite, sow, span, spurn, sting, stink, strew, stride, swell, tread, uproot, wade, warp, wax, wield, wring, writhe	72	700
10^{-6}–10^{-5}	bide, chide, delve, flay, hew, rue, shrive, slink, snip, spew, sup, wreak	91	300

177 Old English irregular verbs were compiled for this study. These are arranged according to frequency bin, and in alphabetical order within each bin. Also shown is the percentage of verbs in each bin that have regularized. The half-life is shown in years. Verbs that have regularized are indicated in red. As we move down the list, an increasingly large fraction of the verbs are red; the frequency-dependent regularization of irregular verbs becomes immediately apparent.

Source: Lieberman *et al.*, *Nature* 2007

Culturomics

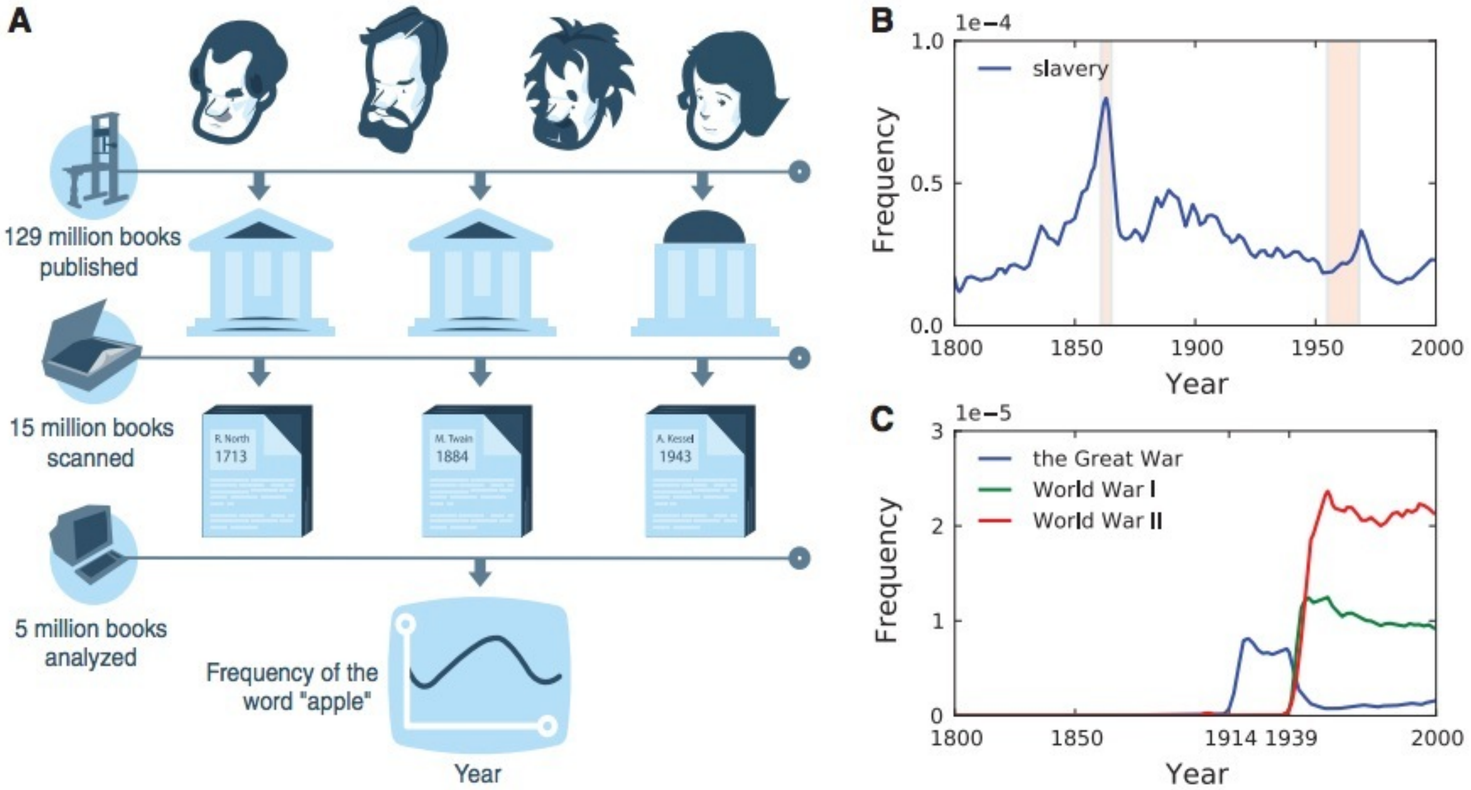
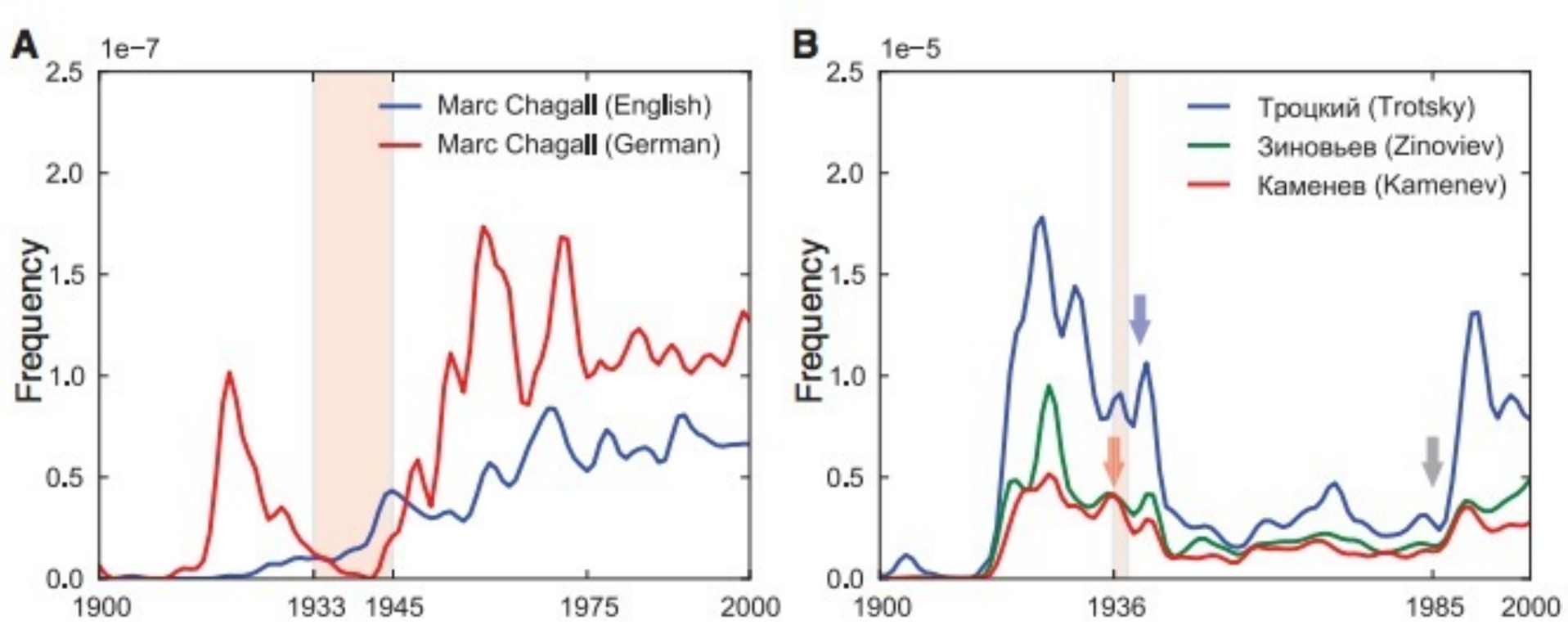


Fig. 1. Culturomic analyses study millions of books at once. **(A)** Top row: Authors have been writing for millennia; ~129 million book editions have been published since the advent of the printing press (upper left). Second row: Libraries and publishing houses provide books to Google for scanning (middle left). Over 15 million books have been digitized. Third row: Each book is associated with metadata. Five million books are chosen for computational analysis (bottom left). Bottom row: A culturomic time line shows the frequency of "apple" in English books over time (1800–2000). **(B)** Usage frequency of "slavery". The Civil War (1861–1865) and the civil rights movement (1955–1968) are highlighted in red. The number in the upper left (1e-4 = 10^{-4}) is the unit of frequency. **(C)** Usage frequency over time for "the Great War" (blue), "World War I" (green), and "World War II" (red).

Source: Michel *et al.*, *Science* 2011

UC San Diego

Culturomics



Source: Michel *et al.*, *Science* 2011

Cognitive Data Science!

So what is Cognitive Science?

What is Cognitive Science?

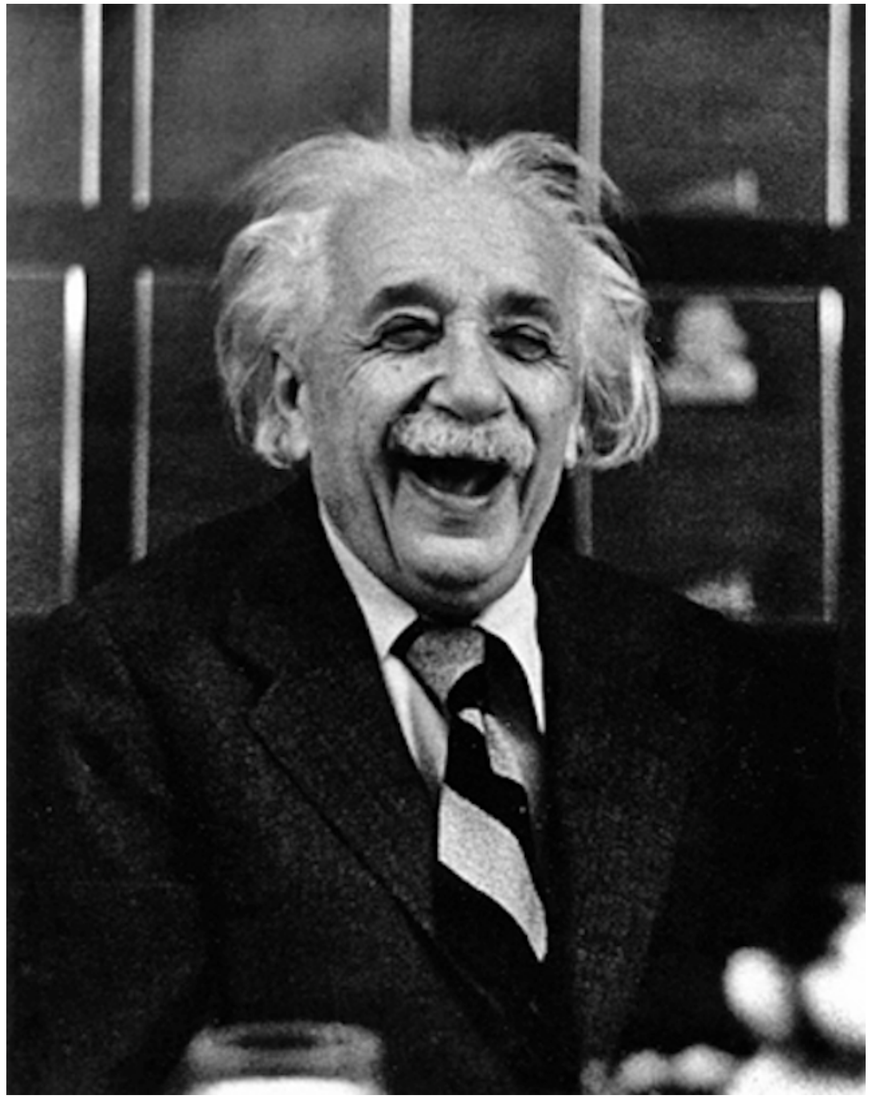
The study of intelligences.
Natural and artificial.

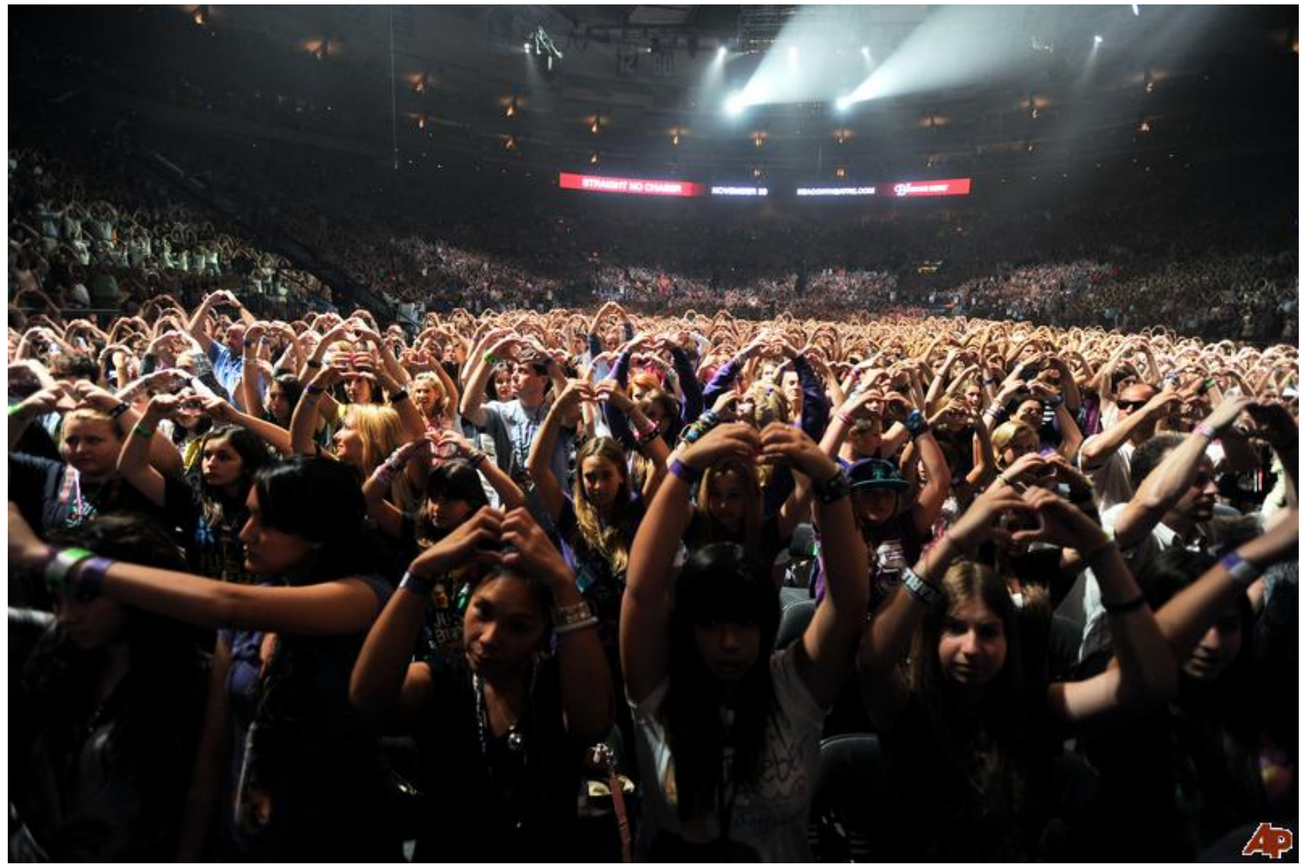
What is “cognition"?

Communication +
Computation +
Reasoning/inference +
Memory +
Planning/decision making +
…

What has cognition?

Humans?



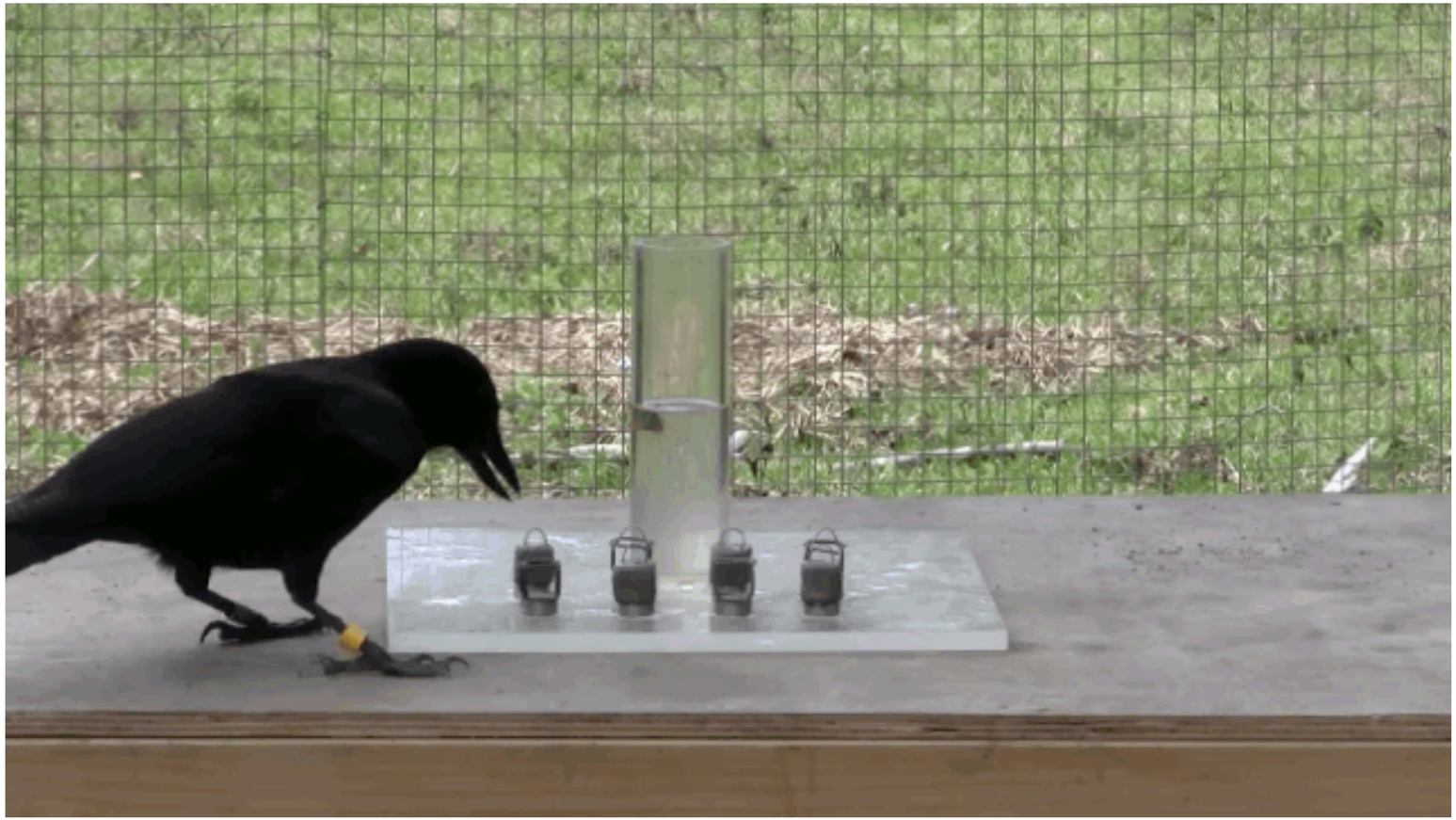


Babies?

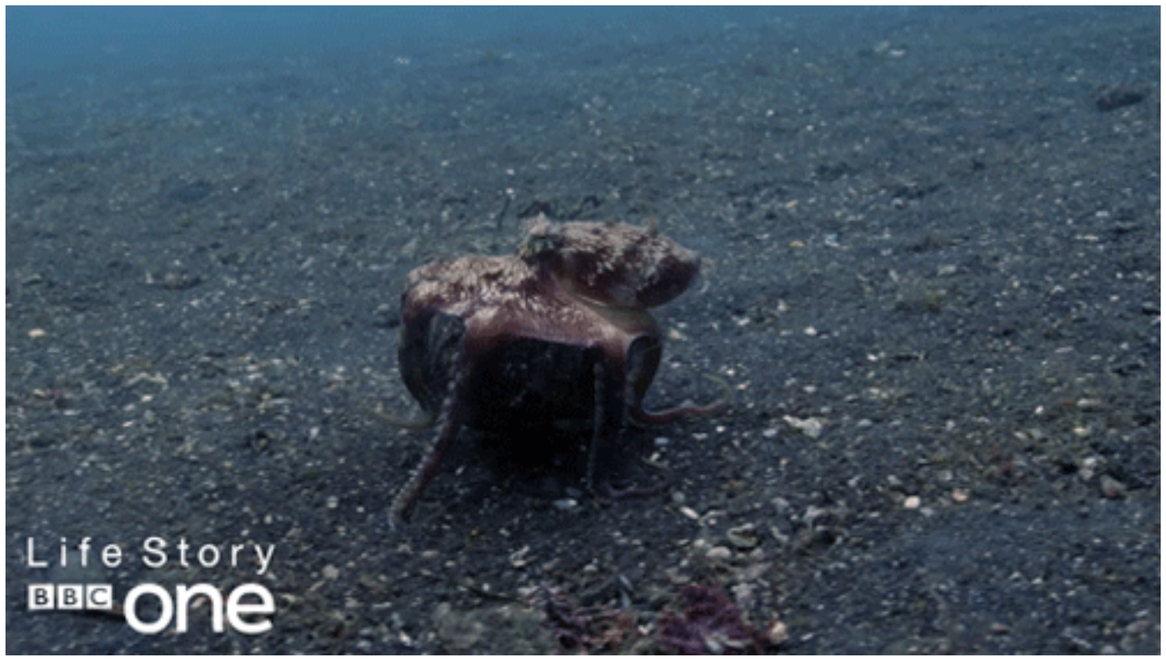
What about non-human primates?

Other mammals?

Crows?

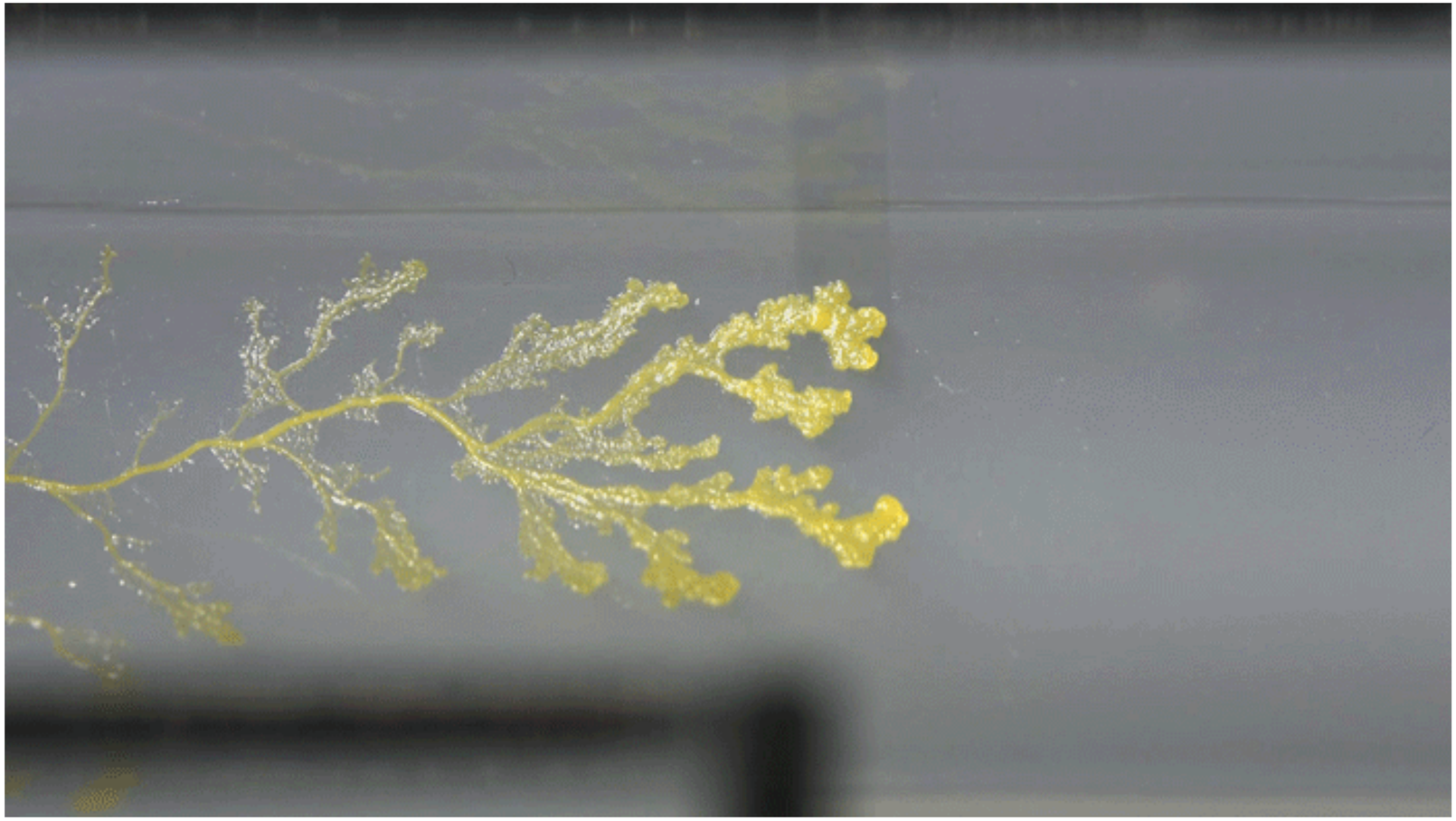


Octopuses?!?

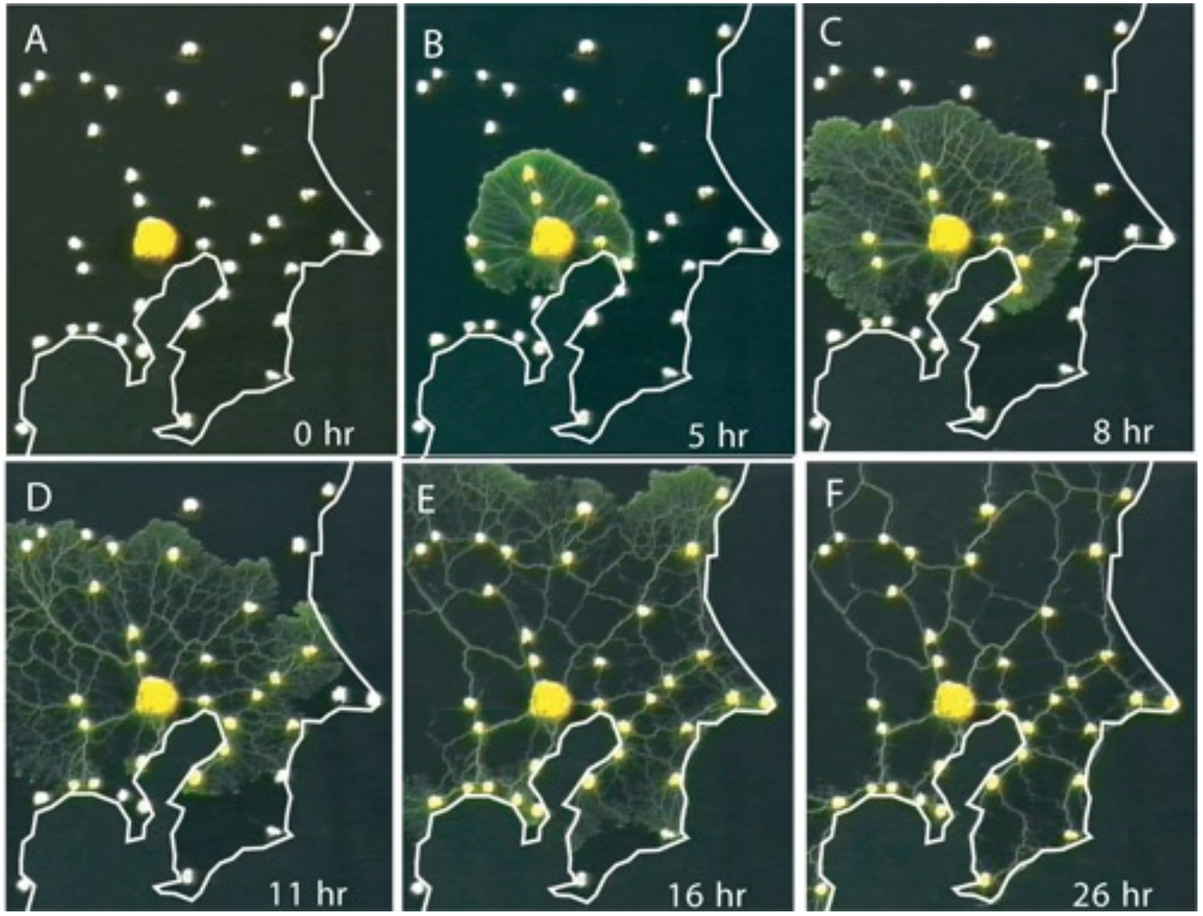


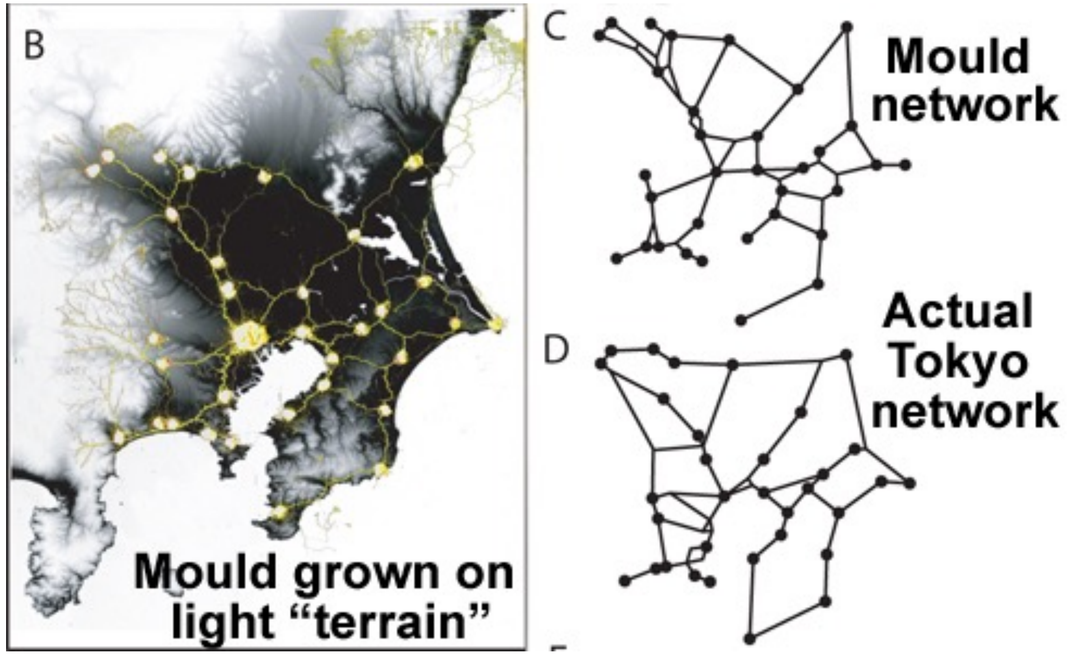
Source: Finn *et al.*, *Curr Biol* 2009

Slime molds?



Slime molds?





Source: Tero *et al.*, *Science* 2010

Linking Cognitive Science, Neuroscience, and Data Science

Oscillations!

Oscillations!

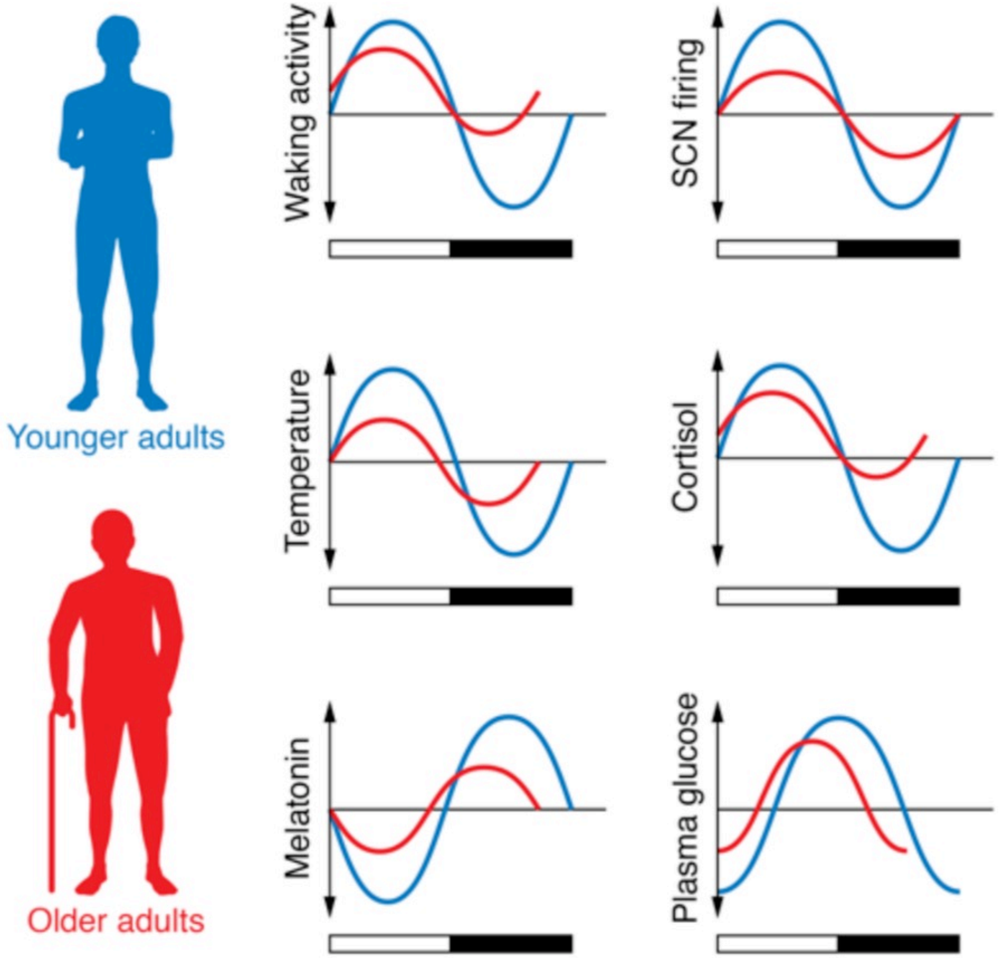
Brainwaves!

Oscillations!

Brainwaves!

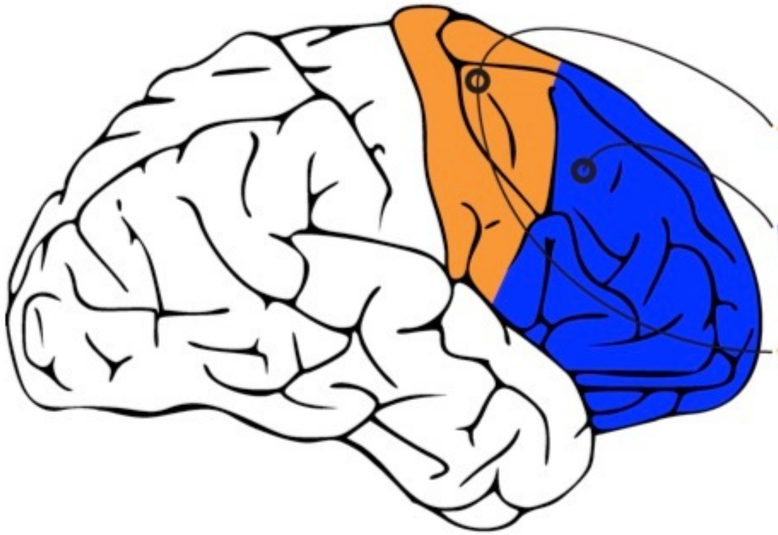
Rhythms of the Brain!

Circadian rhythms

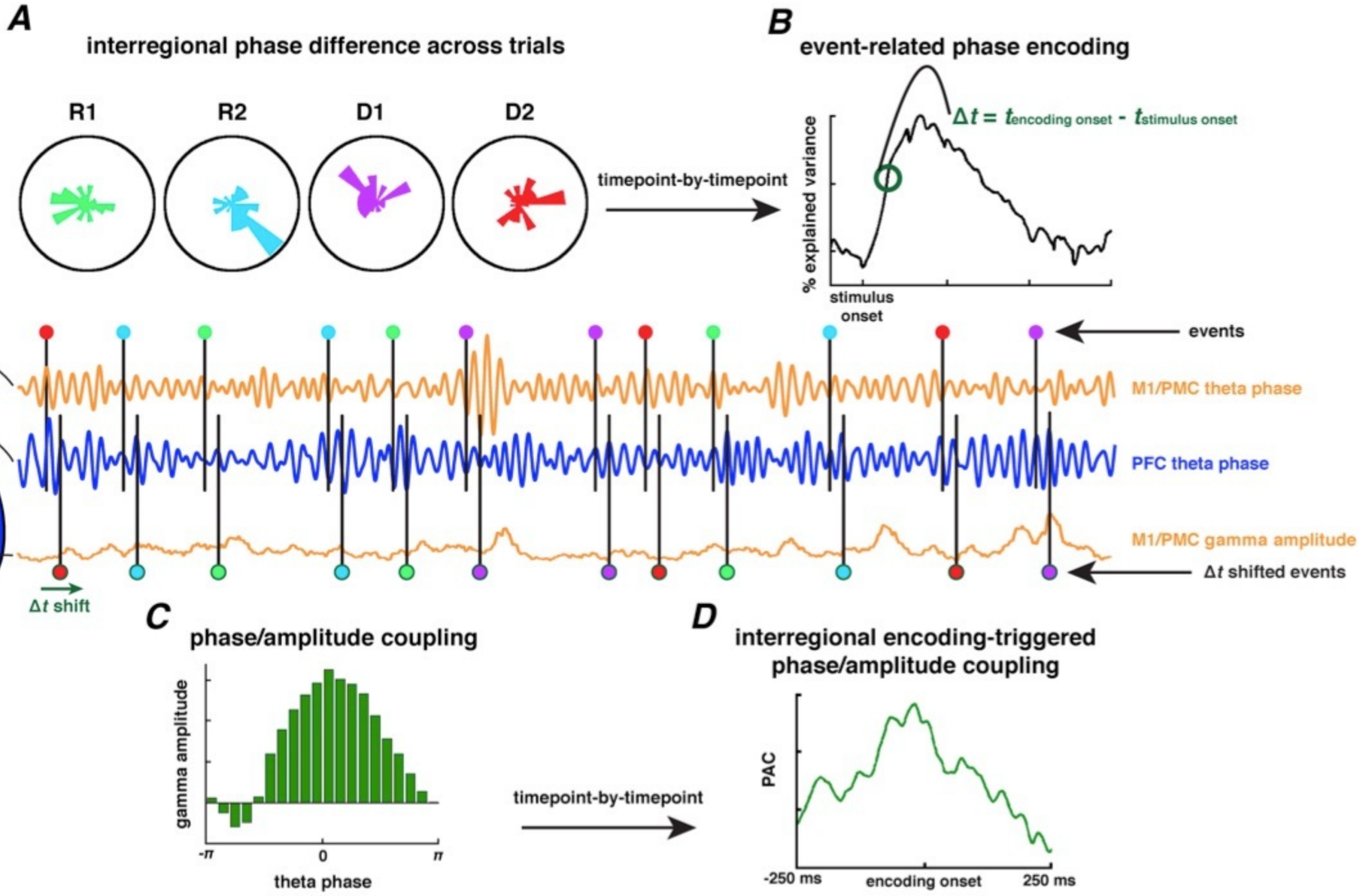


Source: Hood & Amir, *J Clin Invest* 2017

$$h_{a_{\gamma\theta}}[n] = a_{a_{\gamma\theta}}[n]\exp\left(i\phi_{a_{\gamma\theta}}[n]\right)$$



$$Coh_{i,j} = N^{-1}\left|\sum_{n=1}^{N} e^{i(\phi_{\theta i}[n]-\phi_{\theta j}[n])}\right|$$

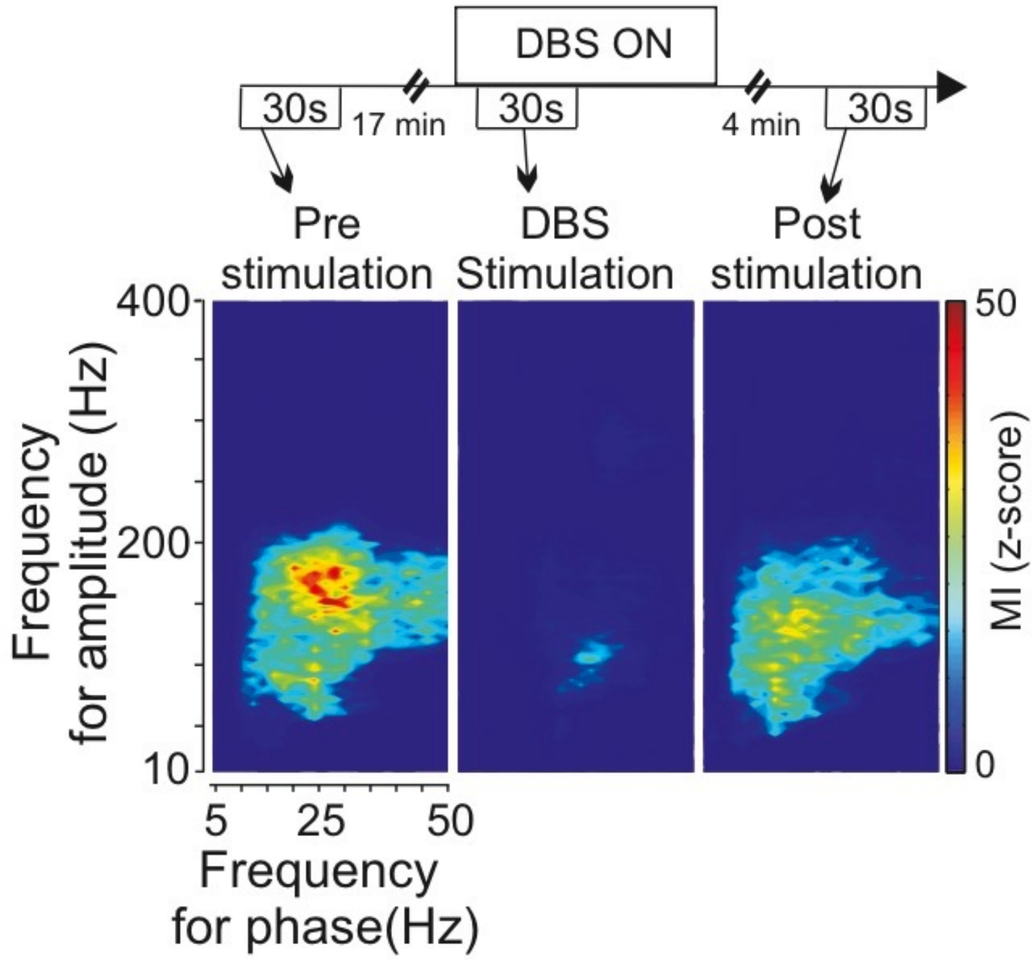


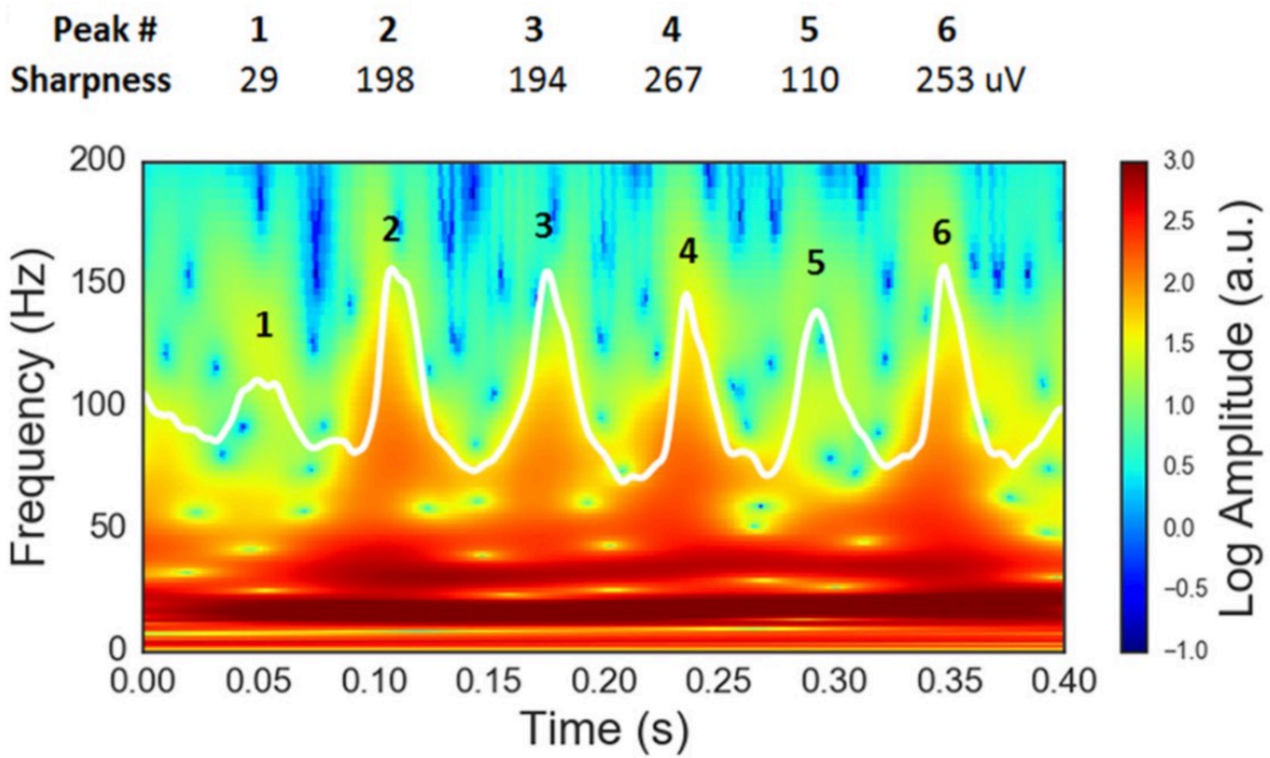
Source: Voytek *et al.*, *Nature Neurosci* 2015

WARNING!

Electrocorticography

Oscillations in Parkinson's disease





Source: Cole *et al.*, *J Neurosci* 2017

UC San Diego