UNDERSTANDING FEAR AND BEYOND IN NEURAL NETWORKS: AN INTERDISCIPLINARY END-TO-END DATA SCIENCE APPROACH

JORDAN STEINHAUSER

PHD STUDENT – PSYCHOLOGY DEPARTMENT

UNIVERSITY OF CALIFORNIA, RIVERSIDE

OUTLINE

How did I get here?

What do I study?

How do I use Data Science in my research?

What Data Mining Techniques do I use?

Why research scientists need Data Science?

MY PATH TO A PHD… SO FAR

UC Riverside

- B.S. Neuroscience
- Minor in Psychology
- *2 years research lab experience*
- *3.0 GPA by graduation*

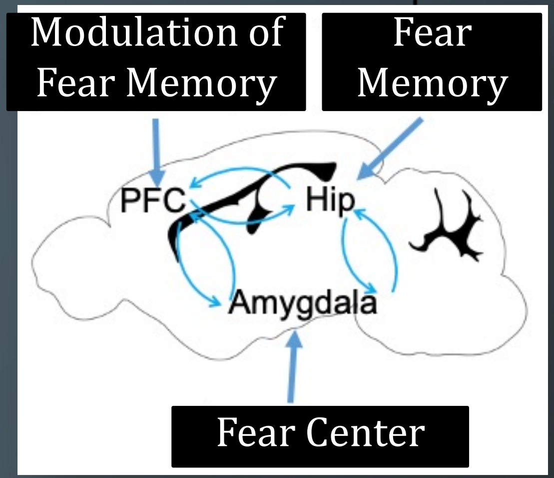
UC Law San Francisco

- Master of Legal Studies
- Specialization in Criminal Justice
- *Student Government Leadership experience*

UC Riverside

- PhD in Behavioral Neuroscience, Department of Psychology
- *M.A. Psychology*
- *Courses: Neuroscience, Statistics, Data Mining Techniques (CS235)*

FEAR

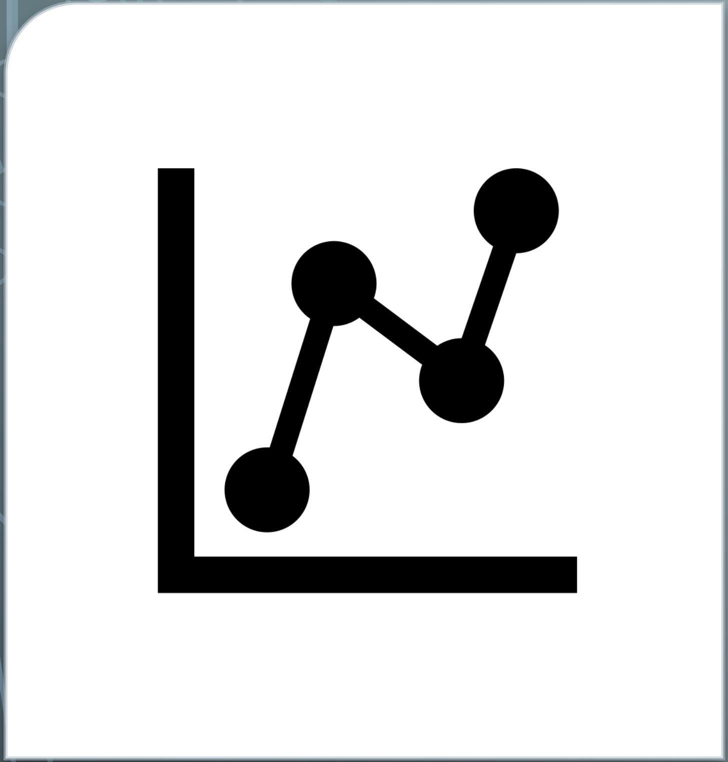


- Threat perception is critical for survival
  - But inappropriate threat responses are debilitating and hinder quality of life
- Overgeneralized fear = Not being able to discriminate between threating and non-threatening stimuli
  - Symptom of generalized anxiety disorder and post-traumatic stress disorder
    - In 2019, approximately 15% of adults in the U.S. experienced symptoms of anxiety (1)
    - In 2020, approximately 13 million Americans had PTSD (2)
- Modulation of fear circuit guides safety learning

1. Centers for Disease Control and Prevention – National Center for Health Statistics
2. U.S. Department of Veterans Affairs – National Center for PTSD

RESEARCH QUESTIONS TO PONDER…

- What regions of the brain are involved in fear interpretation, modulation, and generation?
- What is the underlying neuronal firing activity within these regions?
- Are we able to predict the type of behavior that will be exhibited if a certain neuronal network activity pattern is present?
- What will happen if we manipulate regions to be more or less active?

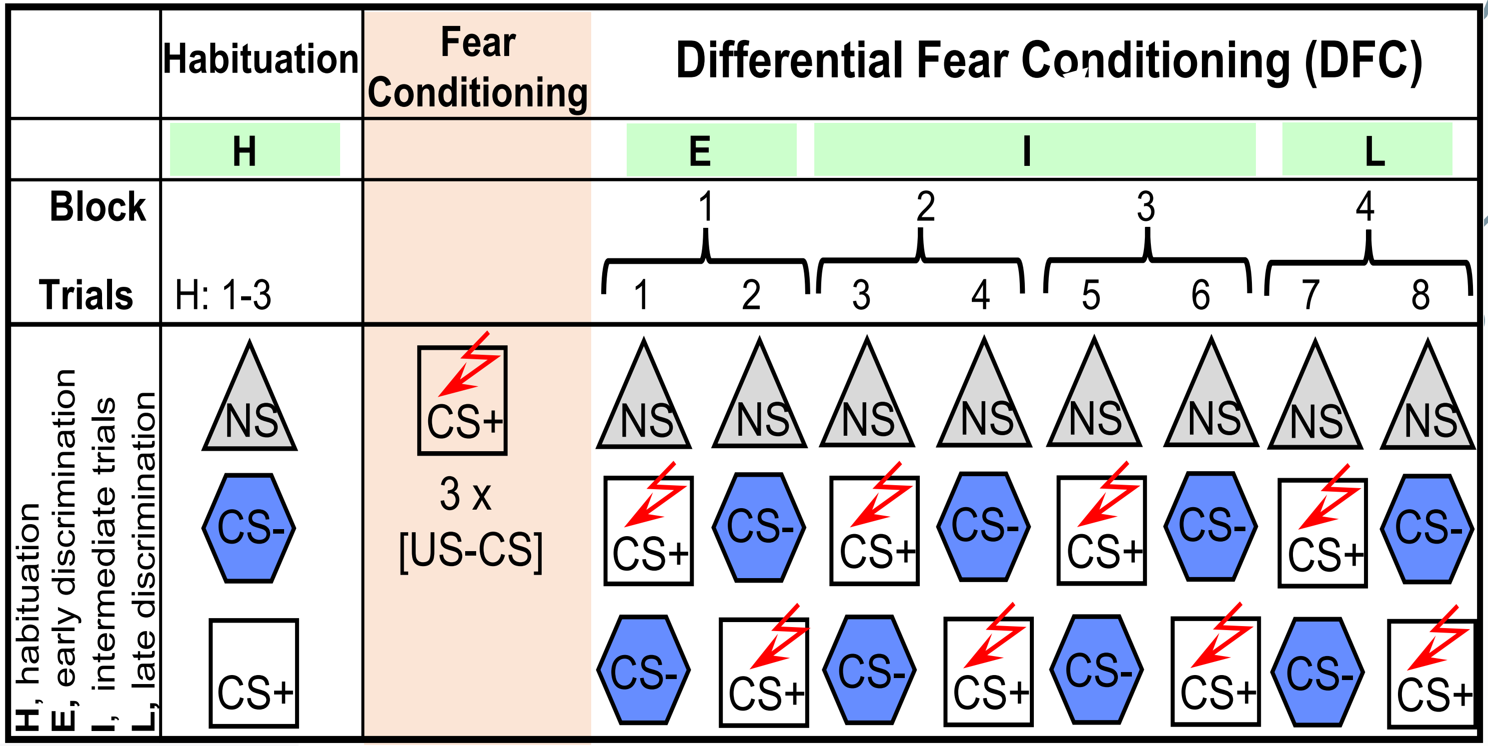


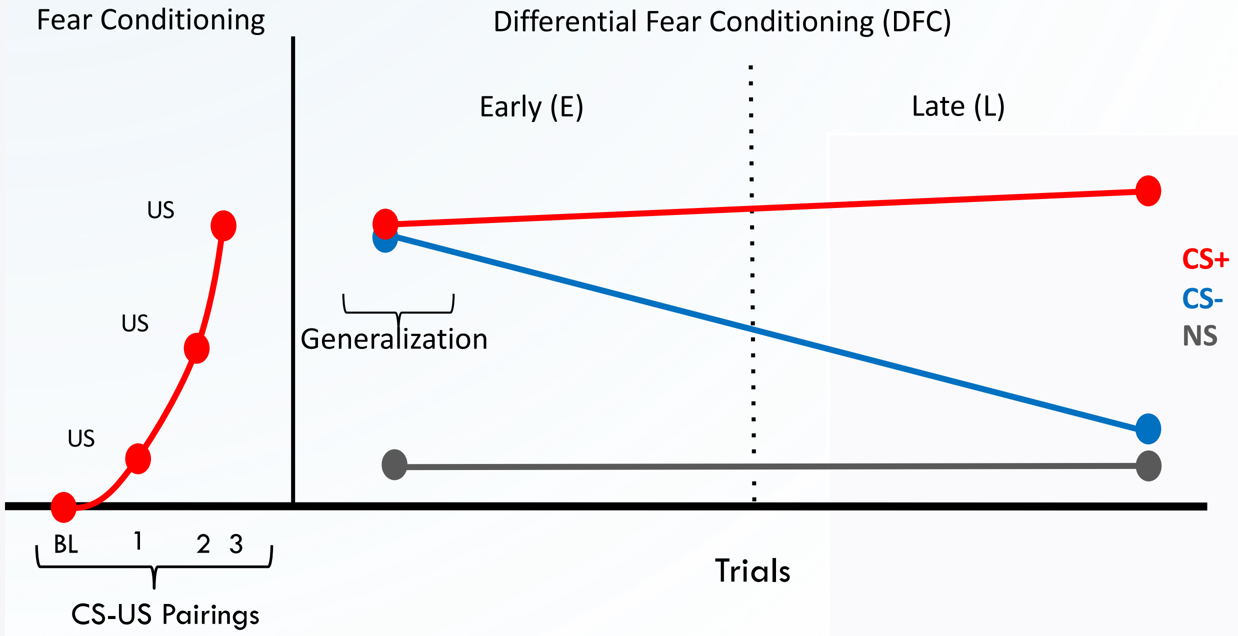
A BRIEF ORIENTATION TO MY DATA

EXPERIMENTAL DESIGN

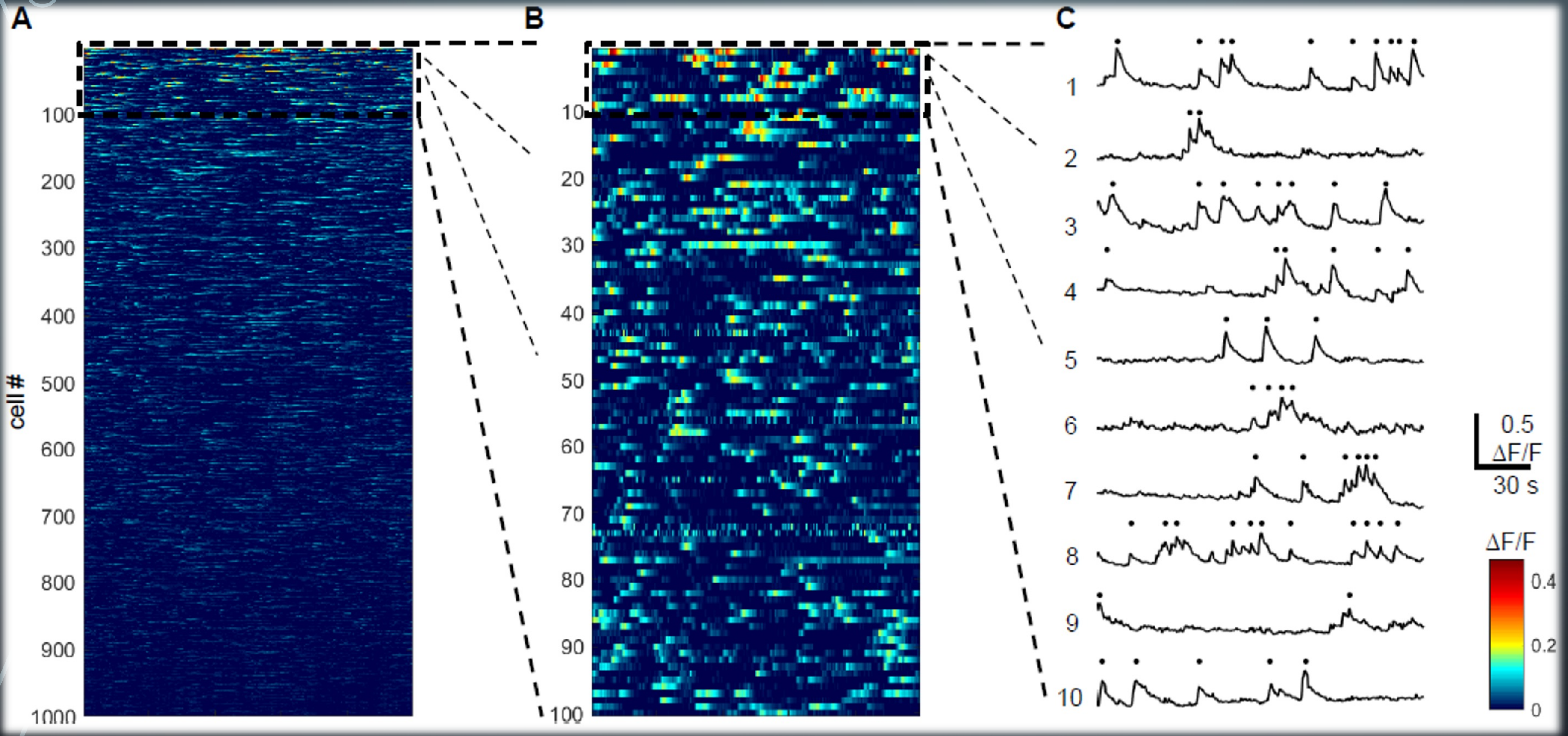
- **Habituation** → mice become familiar with environments
- **Fear Conditioning** → mice receive a foot shock at the same time as a conditioned stimulus (CS) presentation
- **DFC** → mice learn to discriminate between the threatening (CS+) and safe (CS-) environments

- **CS-** = safe environment (novel)
- **CS+** = threatening environment (novel)
- **NS** = neutral stimulus environment (familiar)

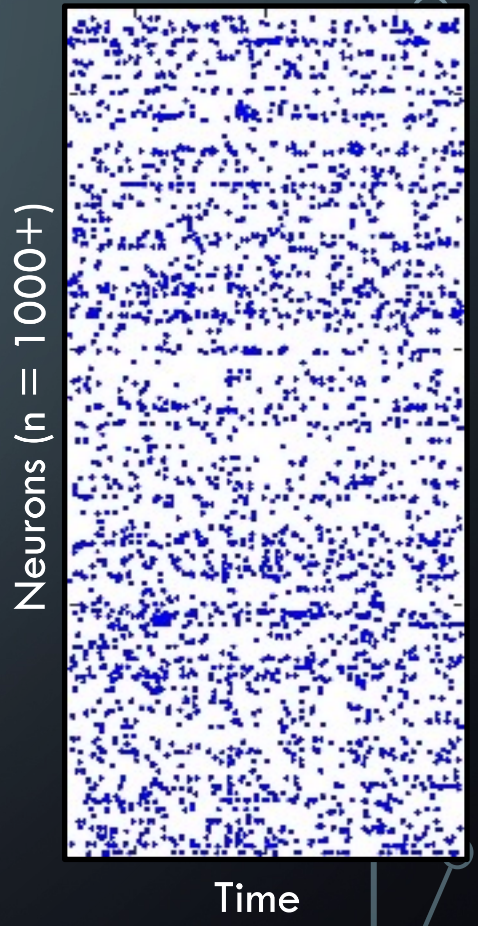




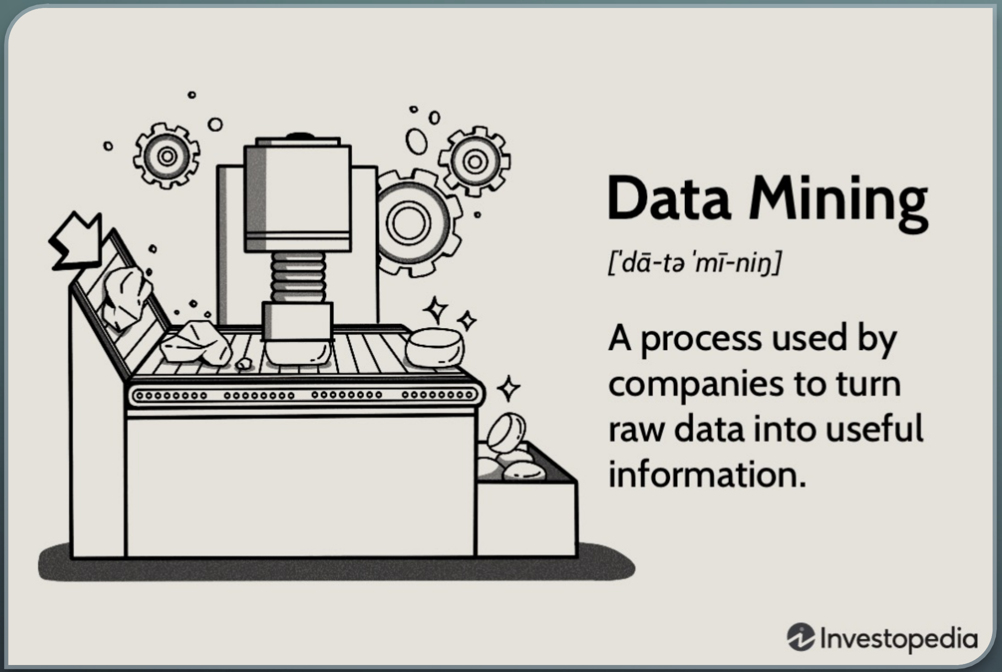
NEURON ACTIVITY



Spike Raster Plot



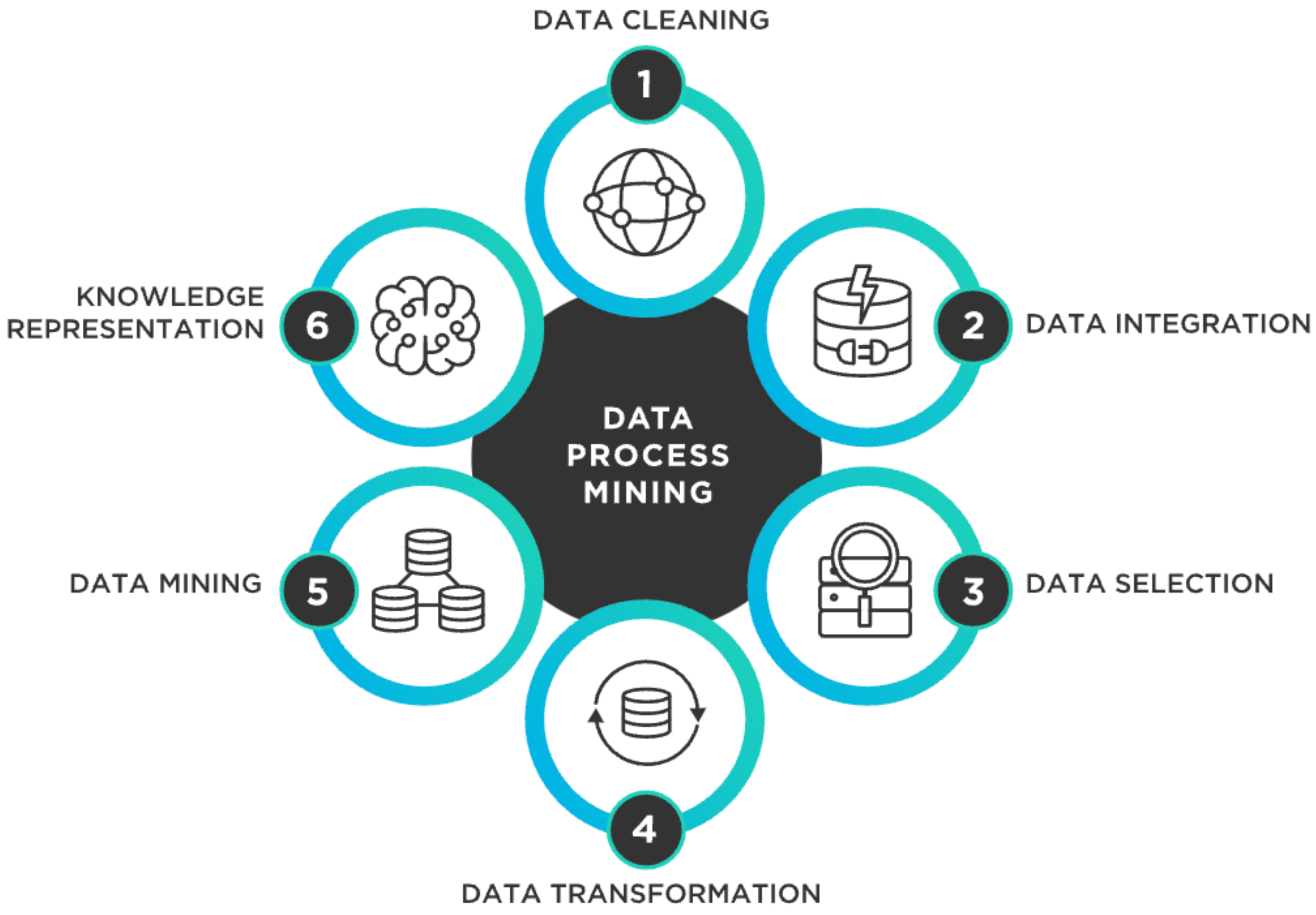
WHY DATA MINING?



- *Data Mining → Knowledge Discovery!*
  - *Find trends and patterns in large data sets*

- Goal of 10-12 subjects per group (control group + experimental group)
  - On average: 1000 neurons per subject (min = 500; max = 2100 neurons)
    - 11 days of calcium recording
      - 3 trials per day

= **528,000 data objects** (about 60GB alone of simple numerical data stored)

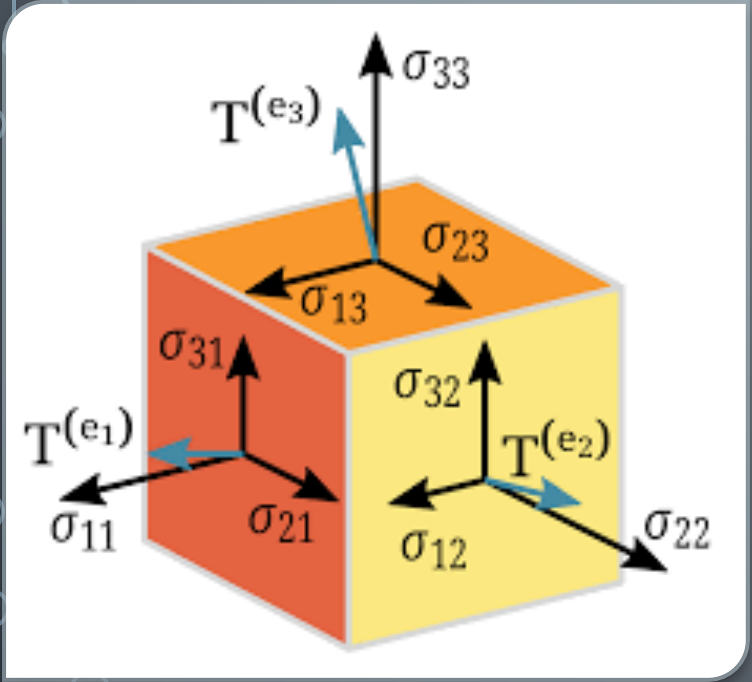


<u>Data Classification:</u>

- Binary → converted calcium spikes to 0 (neuron did not fire) / 1 (neuron did fire)
- Symmetric → both 0/1 are equally likely and valuable; carry same weight

<u>Data Cleaning:</u>

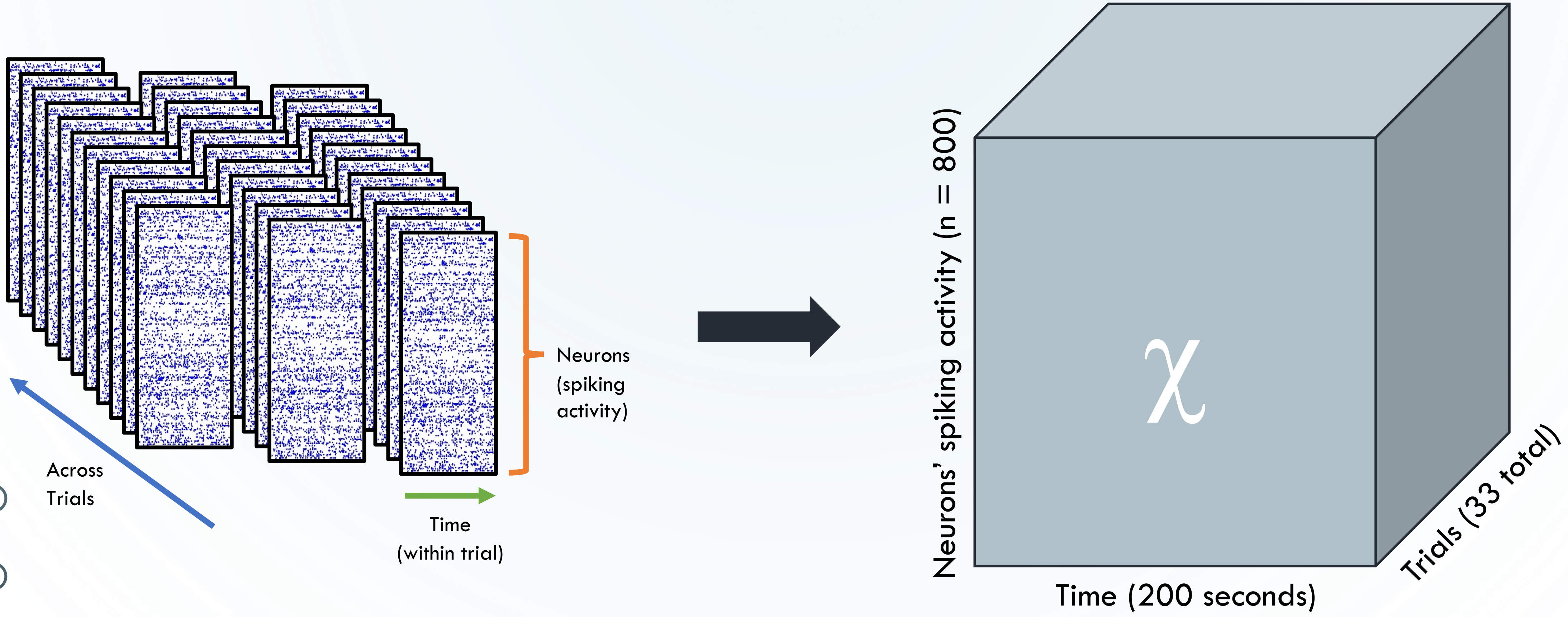
- Managed missing
- Binned data into interpretable and reasonable time periods



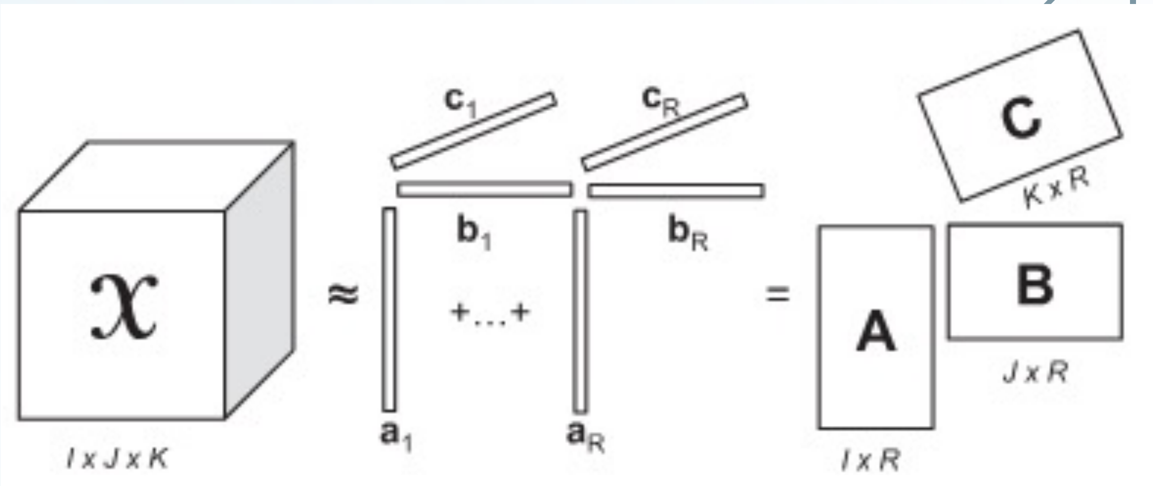
DATA MINING TECHNIQUE 1: TENSOR DECOMPOSITION

AN UNSUPERVISED TECHNIQUE TO DECOMPOSE TENSORS INTO NON-ORTHOGONAL LATENT FACTORS THAT CAN THEN BE INTERPRETED USING PREVIOUS DOMAIN KNOWLEDGE AND ADDITIONAL METHODS

DATA TRANSFORMATION AND REPRESENTATION



TENSOR DECOMPOSITIONS



Three-mode tensor: $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$

- Canonical polyadic (CP) decomposition
    - The sum of three-way outer products

$$\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r,$$

where $\mathbf{a}_r \in \mathbb{R}^I$, $\mathbf{b}_r \in \mathbb{R}^J$, $\mathbf{c}_r \in \mathbb{R}^K$, and their three-way outer product is given by

$$(\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r)(i, j, k) = \mathbf{a}_r(i)\mathbf{b}_r(j)\mathbf{c}_r(k) \quad \text{for all} \quad i, j, k.$$

Where R is the number of components

- **Components:** rank-one representations of the decomposition that serve as a latent "concept" or cluster in the data

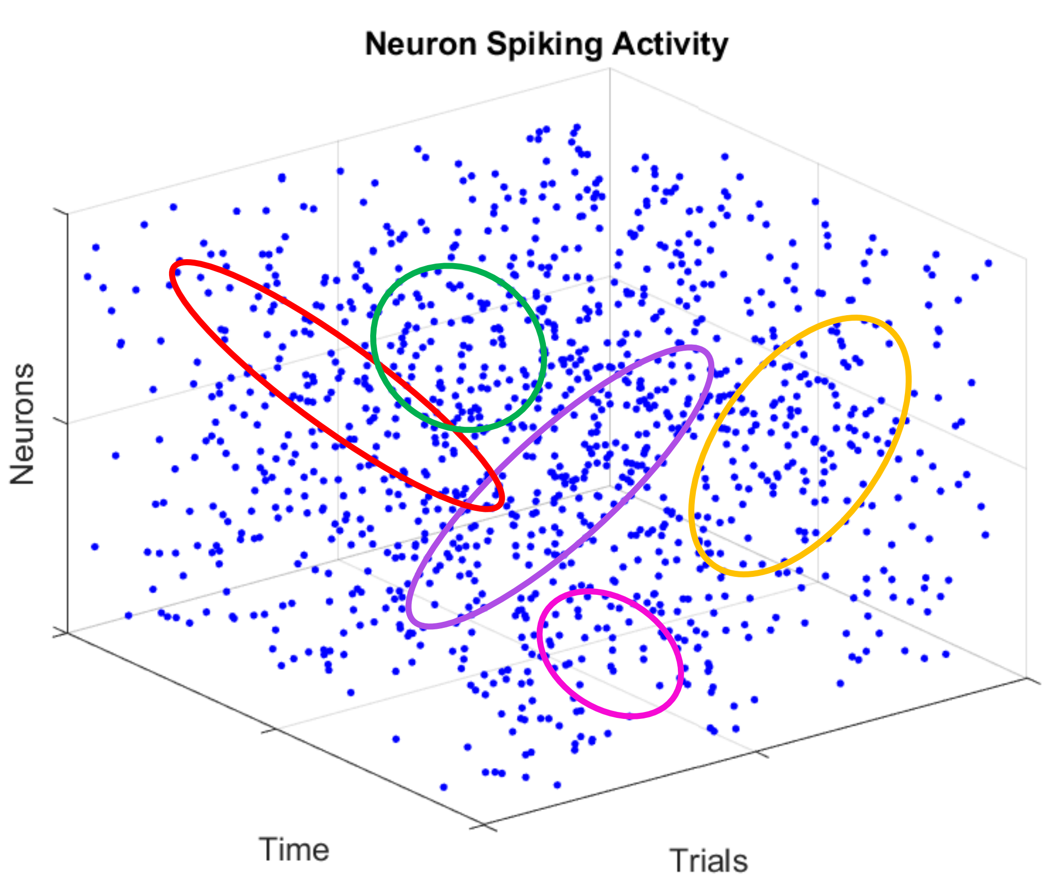
Papalexakis et al. (2016)

SO HOW DO I APPLY THIS IDEA OF TENSOR DECOMPOSITION TO MY DATA…

- *Tensor Decomposition: goal to find hidden co-clusters within data*

- 3 mode tensor (neurons x time x trials)
  - Factors: (1) neuron spiking activity, (2) time within a single trial, (3) total trials

- My research: find populations of neurons that are highly engaged during a particular epoch of the behavior paradigm



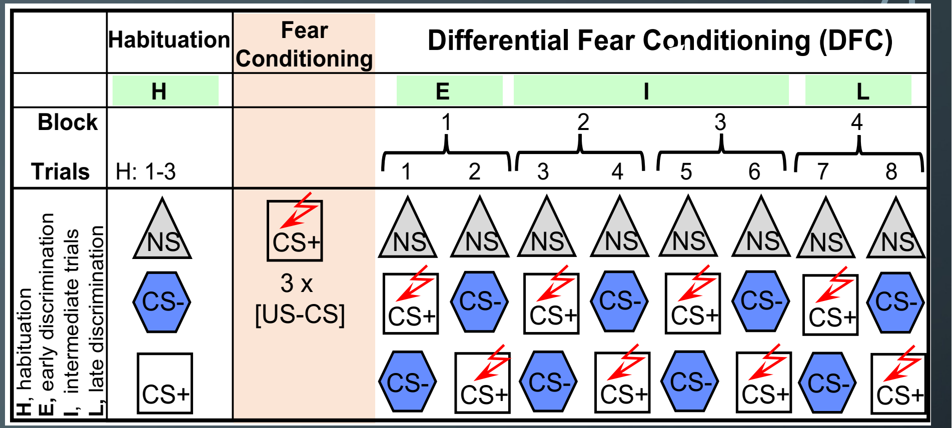
CHOOSING THE NUMBER OF COMPONENTS

- <u>Options:</u>

  1) Trial and Error → try running 100 components, 99 components, and so on
     - Computationally extensive; Takes a long time

  2) Choose an arbitrary number → stick with 10 components for all my data
     - Science hates arbitrary numbers you need a reason why you chose that number; Subjects are not all going to be the same

  3) Hypothesis driven → Think about what I expect to see and run that



HYPOTHESIS GENERATION

1 Component

	Habituation	Early Dis.	Late Dis.
NS			
CS-			
CS+			

2 Components

	Habituation	Early Dis.	Late Dis.
NS			
CS-			
CS+			

4 Components

	Habituation	Early Dis.	Late Dis.
NS			
CS-			
CS+			

9 Components

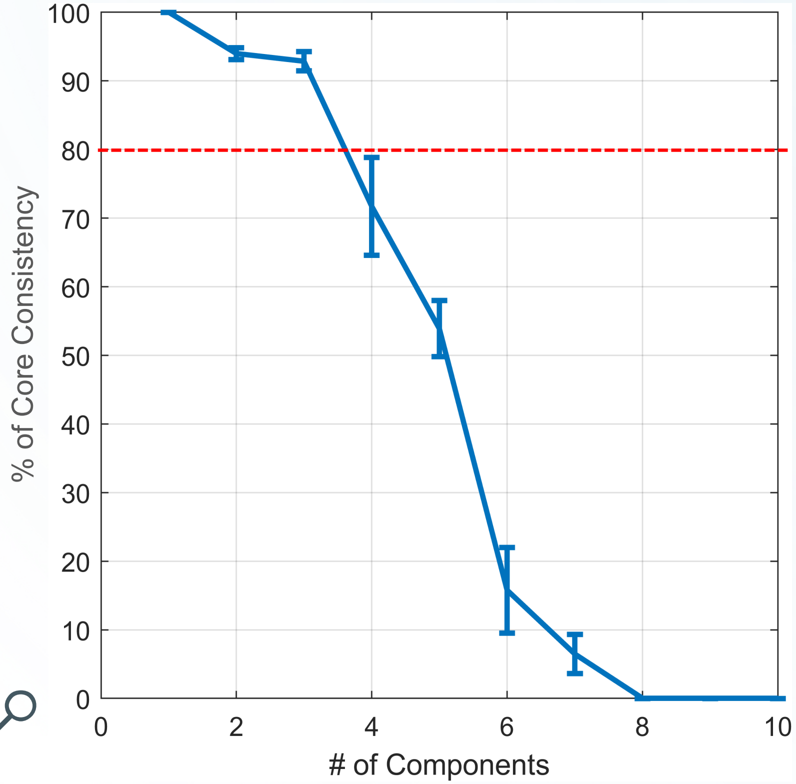
	Habituation	Early Dis.	Late Dis.
NS			
CS-			
CS+			

CHOOSING THE NUMBER OF COMPONENTS

- Options:
  1) Trial and Error → try running 100 components, 99 components, and so on
     - Computationally extensive; Takes a long time
  2) Choose an arbitrary number → stick with 10 components for all my data
     - Science hates arbitrary numbers you need a reason why you chose that number; Subjects are not all going to be the same
  3) Hypothesis driven → Think about what I expect to see and run that
     - Science also hates it when you include experimenter bias into your design/methods
  4) Look into tensor literature more, ask others in the field, and find an algorithm to determine the number of components I should choose

CORE CONSISTENCY DIAGNOSTIC (CORCONDIA)

- Addresses the problem: Determine the rank (number of components) of a tensor



- A model to guide the decision of how many components that should be specified in order to decompose the data
  - Takes into account magnitude of variation described by the components of the model

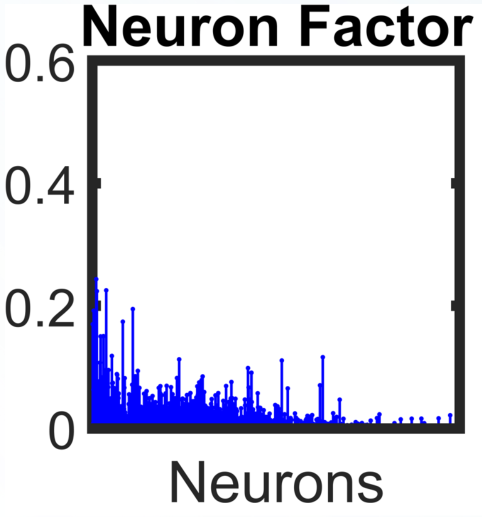
- % of Core Consistency:
  - 100% = appropriate model
  - > 90% = model fits very well
  - ~ 50% = model is problematic

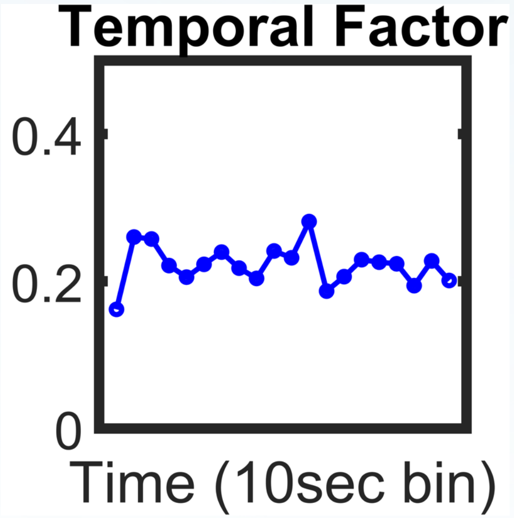
- Decided threshold rate of 80% of Core Consistency
  - > 80% = model represents the data well

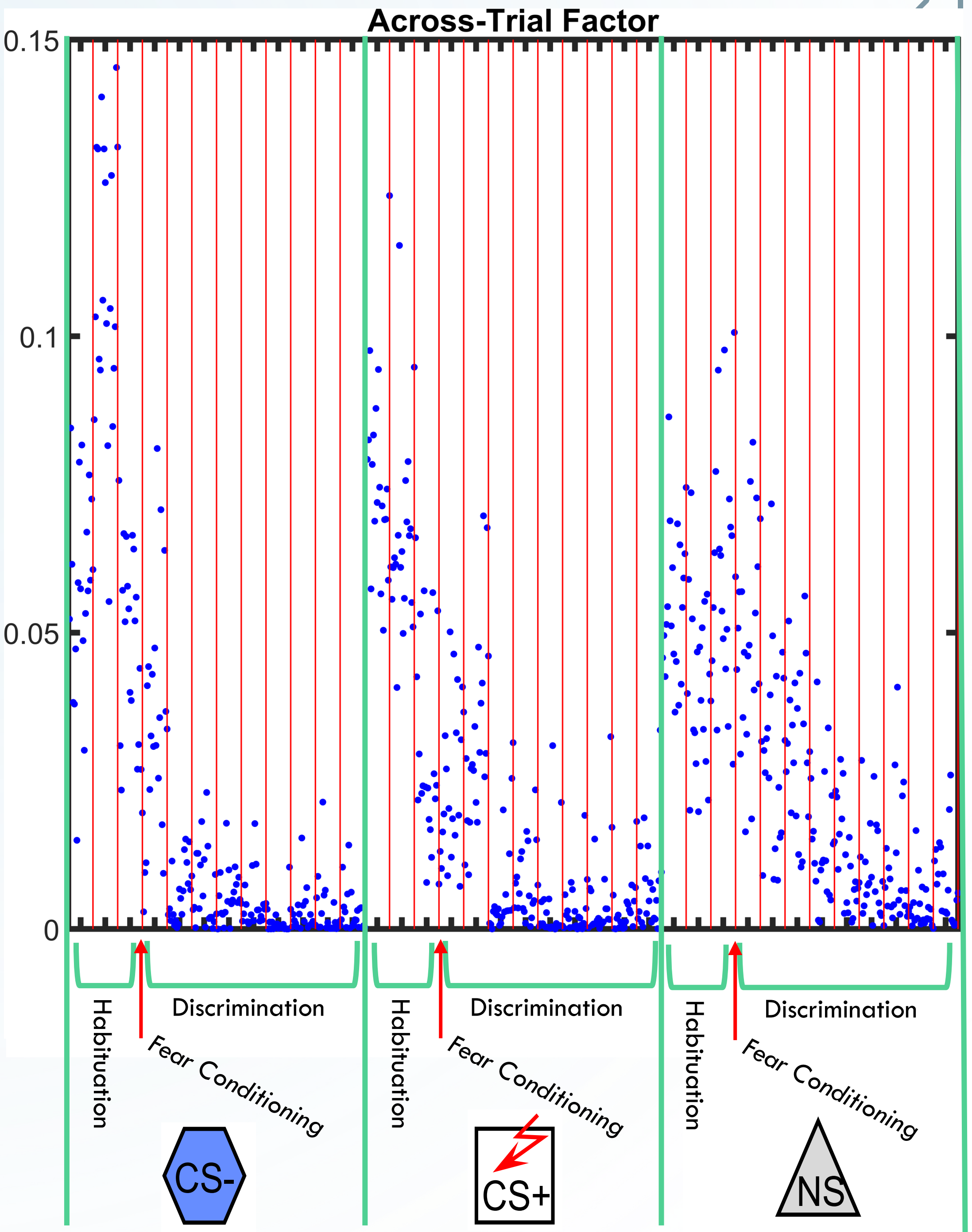
Papalexakis & Faloutsos (2015)

TCA

- **Neuron Factor:** how much each individual neuron contributes to that specific component
- **Temporal Factor:** how variable the activity of the grouped neurons attributed to that component are
- **Trial Factor:** how the neurons grouped within that component participated throughout the trials of the behavioral paradigm



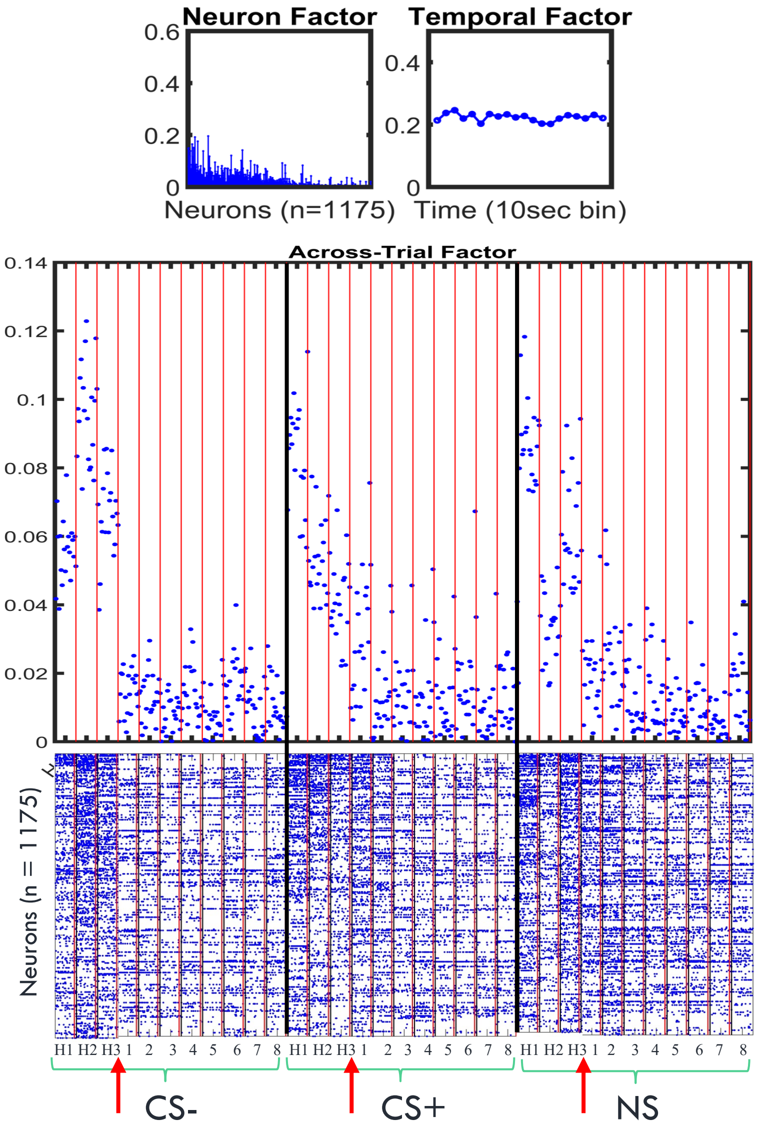




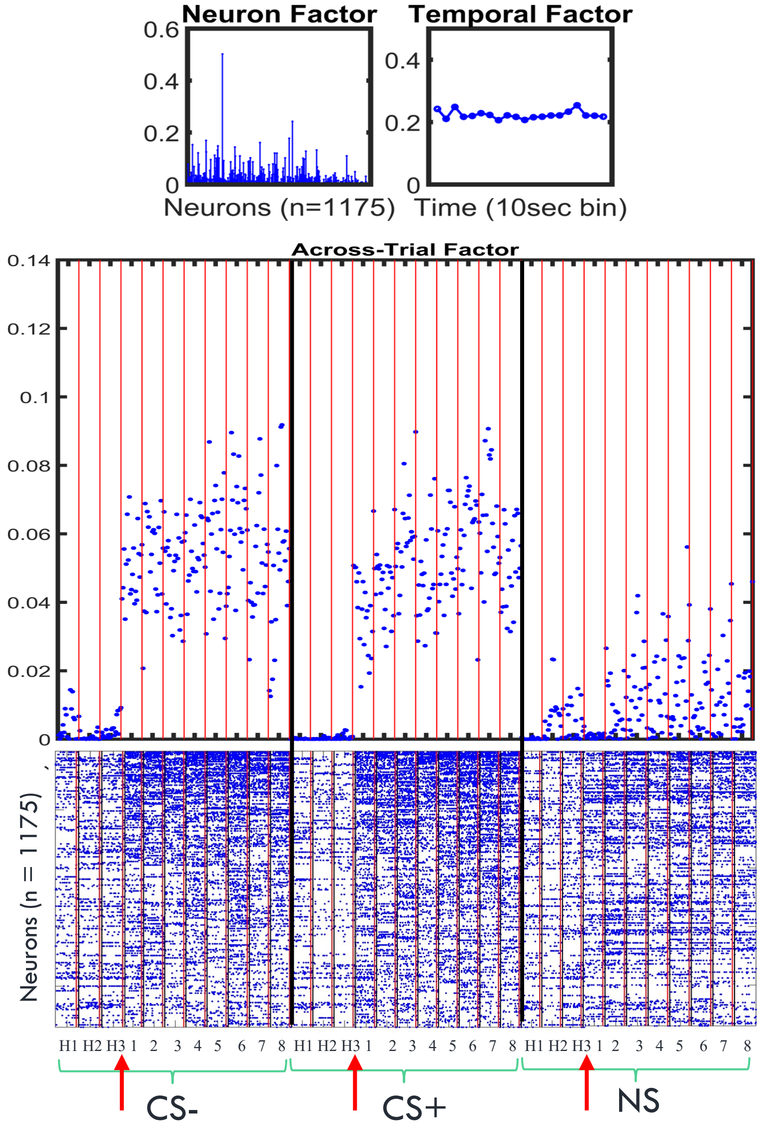
TCA – RESULTS FOR ONE SUBJECT

= Fear Conditioning

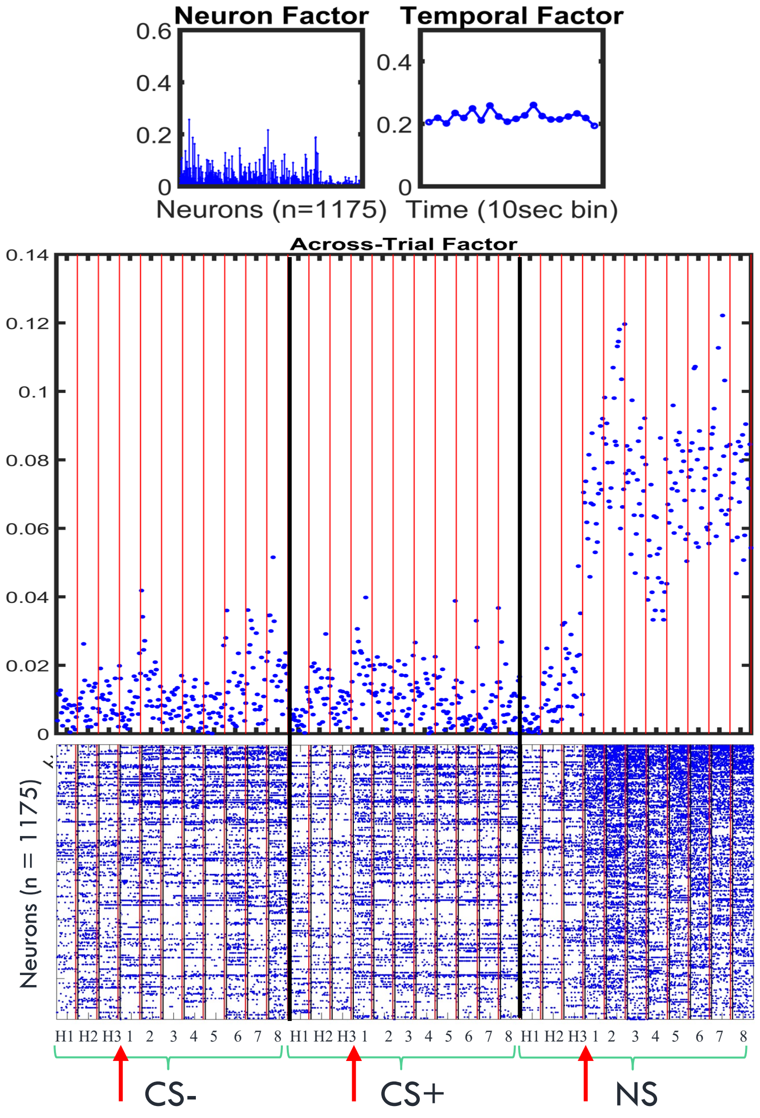
Component 1:

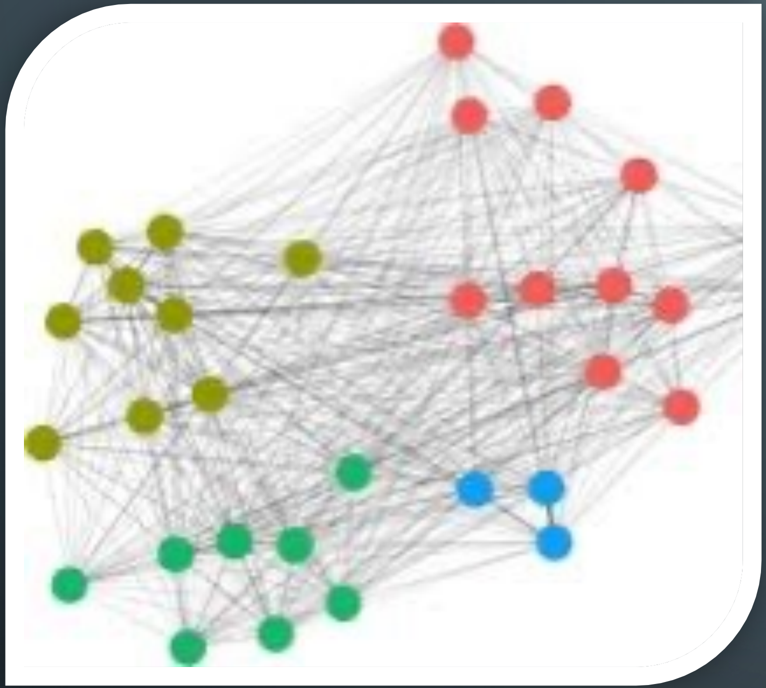


Component 2:



Component 3:





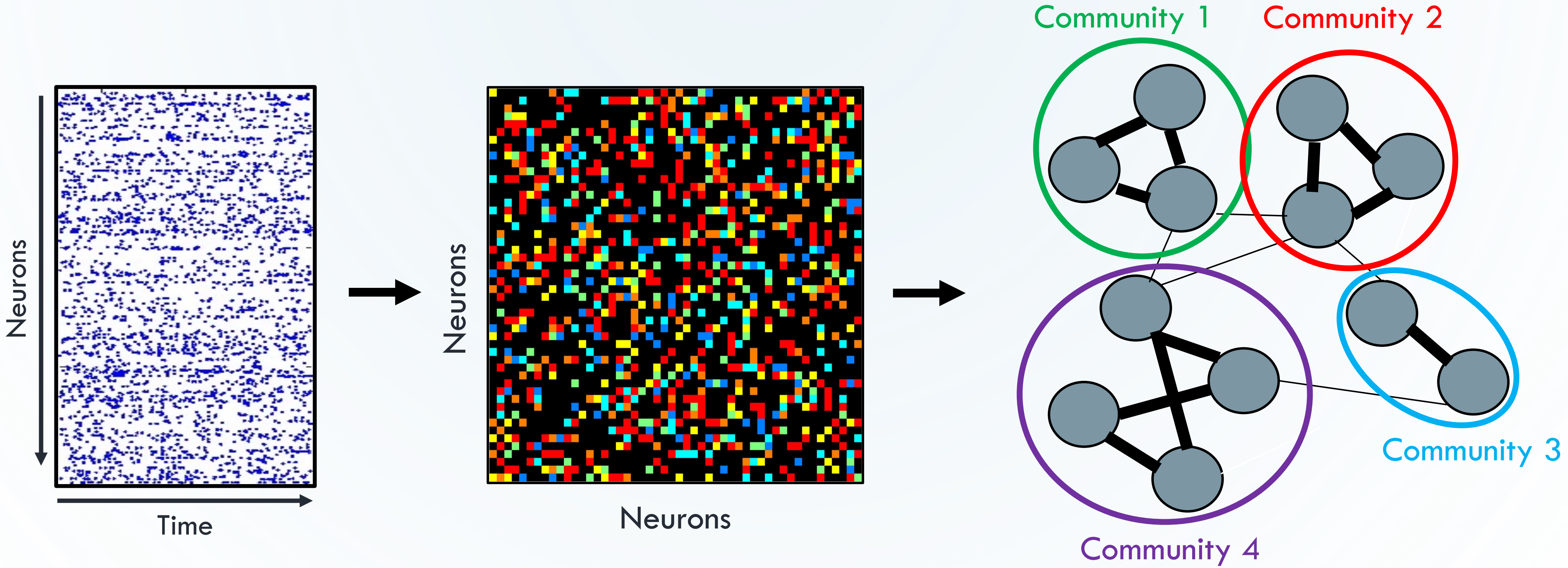
DATA MINING TECHNIQUE 2: COMMUNITY ANALYSIS

AN UNSUPERVISED TECHNIQUE TO FIND OPTIMIZED STRUCTURE OF NONOVERLAPPING GROUPS OF NODES

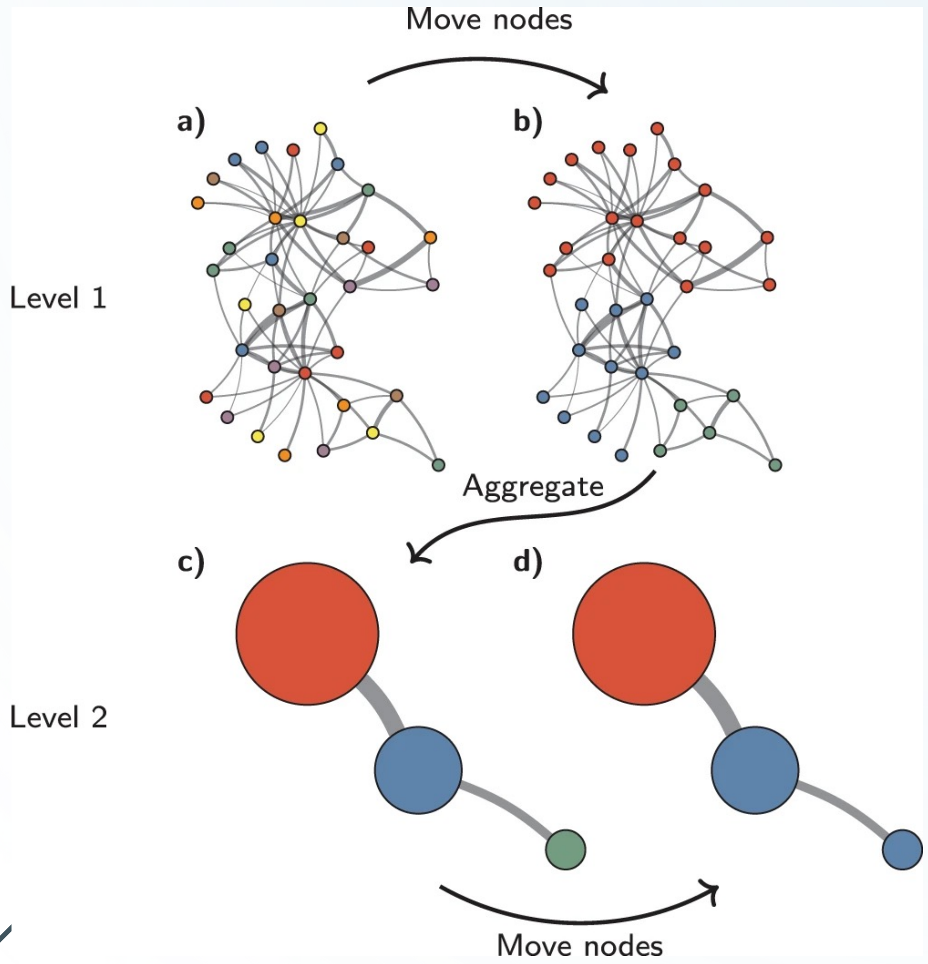
NETWORK SCIENCE

- Studies complex networks (i.e. biological and cognitive networks)
  - **Nodes:** distinct elements of the network (neurons)
  - **Links:** connections between the elements (structural, functional, or temporal connectivity)
- Find best algorithm for my data that answers my research question
    - *Goal: Find populations of neurons that are highly engaged during a particular epoch of the behavior paradigm*
  - Example Algorithms: centrality, similarity, path finding, node embeddings, community detection
  - Different production-quality Community Detection Algorithms:
    - K-core decomposition, K-means clustering, Louvain, Modularity optimization, Weakly connected components
  - **Louvain:** detects communities in large networks; maximizes modularity score (the quality of an assignment of nodes in a community) for each community

COMMUNITY ANALYSIS PIPELINE



LOUVAIN COMMUNITY ANALYSIS ALGORITHM



Traag et al. (2019)

- Optimizes quality function (i.e. modularity) in 2 phases:
  1) Local moving of nodes
  2) Aggregation of the network

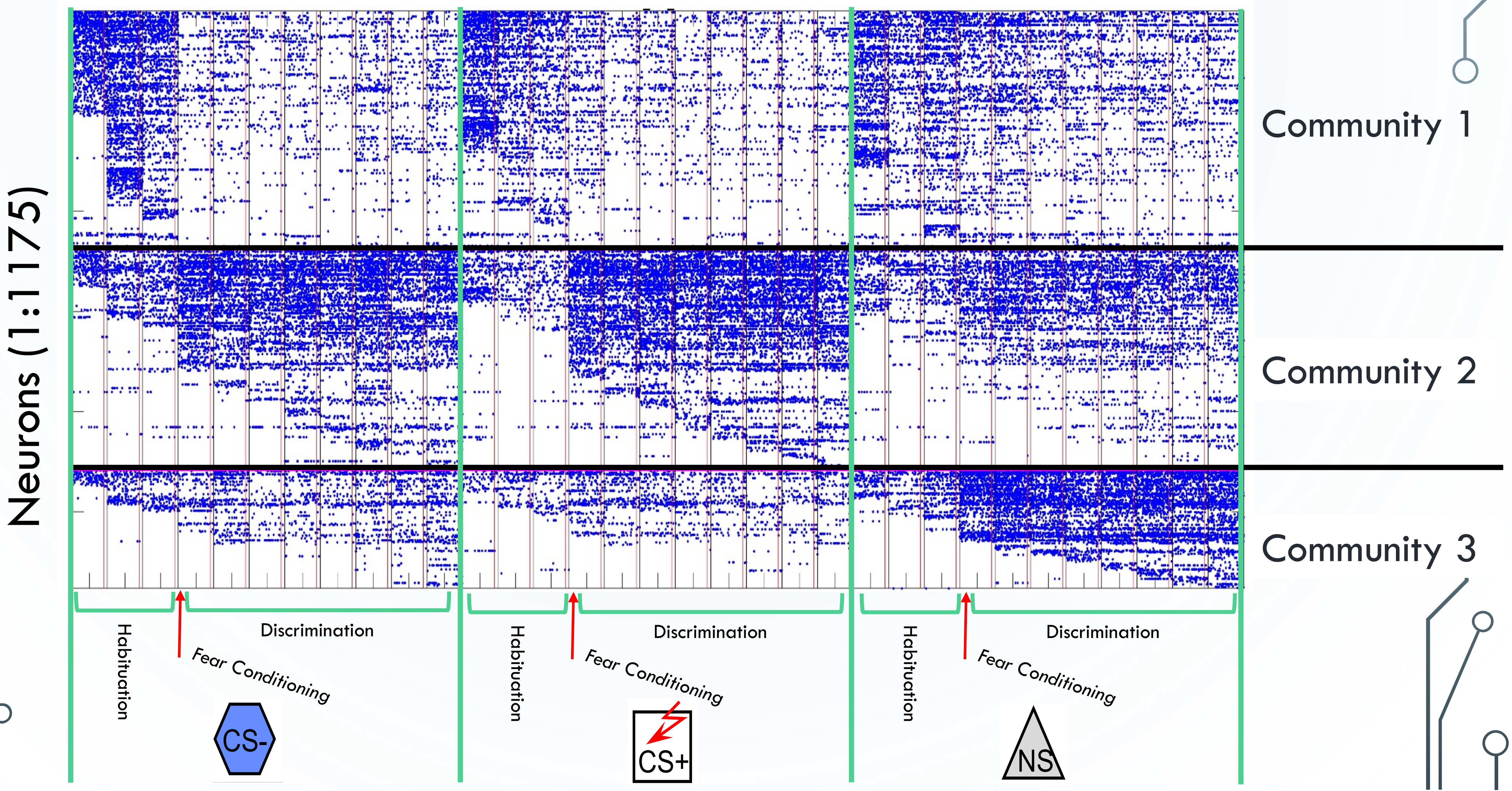
$$\mathcal{H} = \frac{1}{2m} \sum_c \left(e_c - \gamma \frac{K_c^2}{2m}\right),$$

where $\gamma > 0$ is a resolution parameter[4].

- Higher resolution = more communities
- Lower resolution = less communities

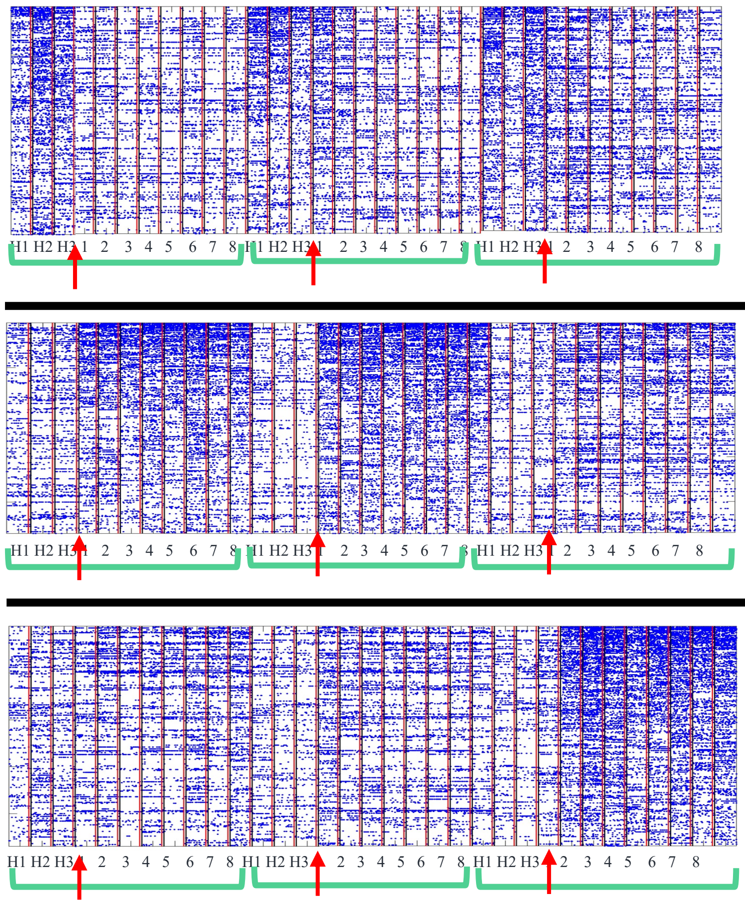
- Choosing the gamma value?

COMMUNITY ANALYSIS – RESULTS FOR ONE SUBJECT



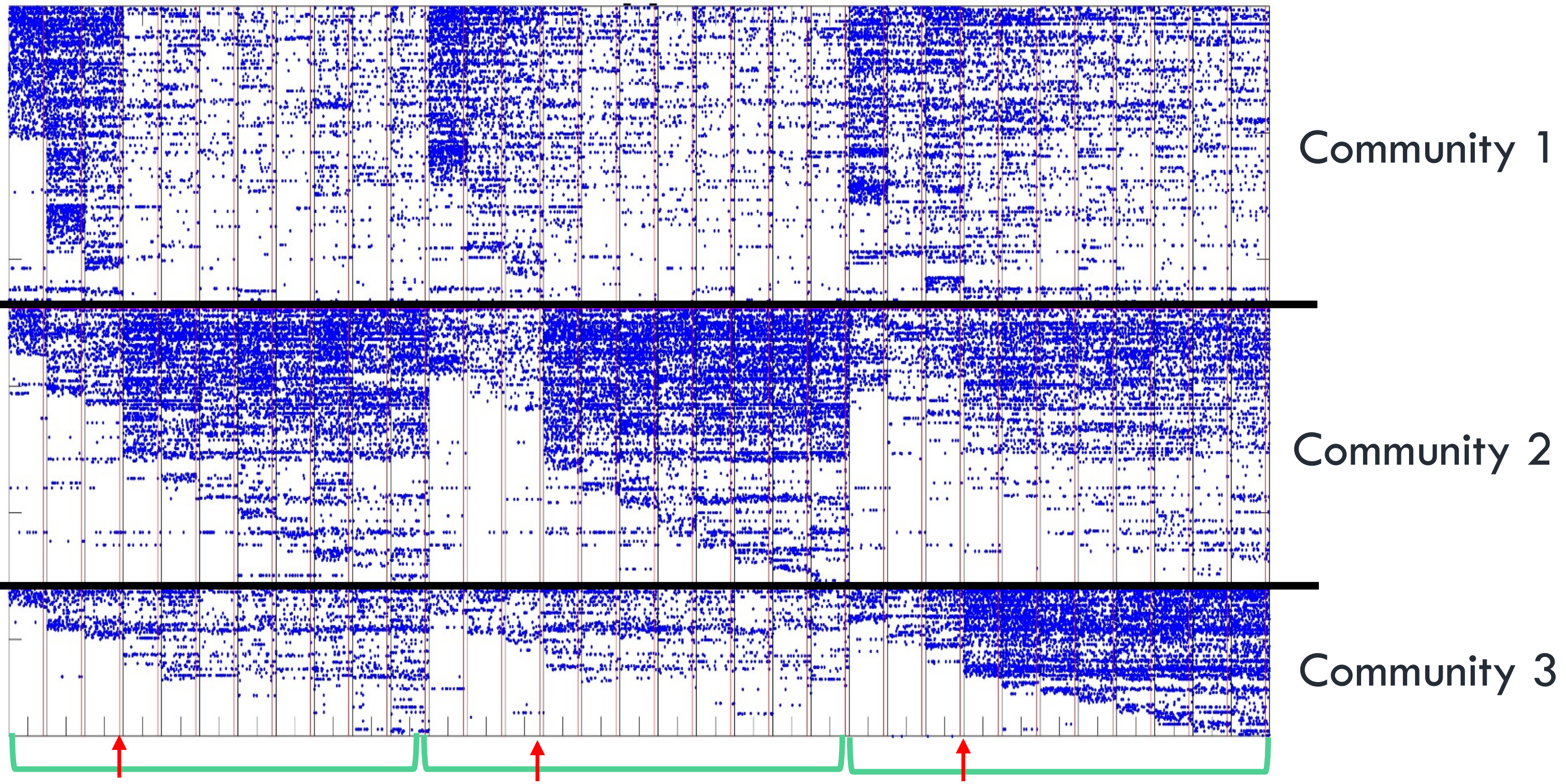
DATA MINING METHODS – ADVANCED TECHNIQUES

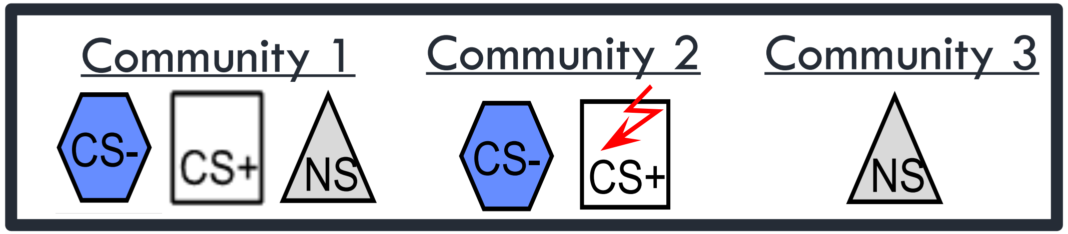
TENSOR DECOMPOSITION



= *Fear Conditioning*

COMMUNITY ANALYSIS





WHY SHOULD RESEARCH SCIENTISTS NEED DATA SCIENCE?

The Journal of Neuroscience, August 23, 2023 • 43(34):5989–5995 • 5989

TechSights

Neuroscience Needs Network Science

Dániel L. Barabási,[1,2] Ginestra Bianconi,[3,4] Ed Bullmore,[5] Mark Burgess,[6] S
Tina Eliassi-Rad,[9,10,11] Dileep George,[12] István A. Kovács,[13,14] Hernán Makse,[15] T
Christos Papadimitriou,[18] Olaf Sporns,[19] Kim Stachenfeld,[12,18] Zoltán Toroczkai
Anthony M. Zador,[25] Hongkui Zeng,[26] Albert-László Barabási,[9,27,28] Amy Bern
[1]Biophysics Program, Harvard University, Cambridge, 02138, Massachusetts, [2]Department of Molecular and C
Harvard University, Cambridge, 02138, Massachusetts, [3]School of Mathematical Sciences, Queen Mary Universit
Kingdom, [4]Alan Turing Institute, The British Library, London, NW1 2DB, United Kingdom, [5]Department of P
Centre, University of Cambridge, Cambridge, CB2 0QQ, United Kingdom, [6]ChiTek-i AS, Oslo, 0476, Norway,
University, New York, New York 10003, [8]Center for Computational Neuroscience, Flatiron Institute, Sim
10010, [9]Network Science Institute, Northeastern University, Boston, 02115, Massachusetts, [10]Khoury Coll
University, Boston, 02115, Massachusetts, [11]Santa Fe Institute, Santa Fe, New Mexico 87501, [12]DeepMind
[13]Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, [14]Northwest
Northwestern University, Evanston, Illinois 60208, [15]Levich Institute and Physics Department, City Colle
10031, [16]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Departme
Oxford, Oxford, OX3 7LF, United Kingdom, [17]Wellcome Centre for Integrative Neuroimaging, FMRIB, N
Neurosciences, University of Oxford, Oxford, OX3 9DU, United Kingdom, [18]Columbia University, New Y
Psychological and Brain Sciences, Indiana University, Bloomington, Indiana 47405, [20]Department of Phy
Dame, Indiana 46556, [21]Department of Computer Science, University of Calgary, Calgary, Alberta, AB T2
Institute, University of Calgary, Calgary, Alberta, AB T2N 1N4, Canada, [23]Department of Physics and Ast
Alberta, AB T2N 1N4, Canada, [24]Alberta Children's Hospital Research Institute, University of Calgary, Ca
[25]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, [26]Allen Institute for Brain Scien
[27]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachus
Data Science, Central European University, Budapest, H-1051, Hungary, [29]The Kavli Foundation, Los An
[30]Neuroscience Institute and Department of Neurology, NYU Grossman School of Medicine, New York U

The brain is a complex system comprising a myriad of interacting neurons, posing significant
ture, function, and dynamics. Network science has emerged as a powerful tool for studying such interconnected systems, offering a framework for integrating multiscale data and complexity. To date, network methods have significantly advanced functional imaging studies of the human brain and have facilitated the development of control theory-based applications for directing brain activity. Here, we discuss emerging frontiers for network neuroscience in the brain atlas era, addressing the challenges and opportunities in integrating multiple data streams for understanding the neural transitions from development to healthy function to disease. We underscore the importance of fostering interdisciplinary opportunities through workshops, conferences, and funding initiatives, such as supporting students and postdoctoral fellows with interests in both disciplines. By bringing together the network science and neuroscience communities, we can develop novel network-based methods tailored to neural circuits, paving the way toward a deeper understanding of the brain and its functions, as well as offering new challenges for network science.

Key words: Network Science; Connectomics; Neurodevelopment; Systems Neuroscience; Network Neuroscience; NeuroAI

Why Neuroscience Needs Data Scientists

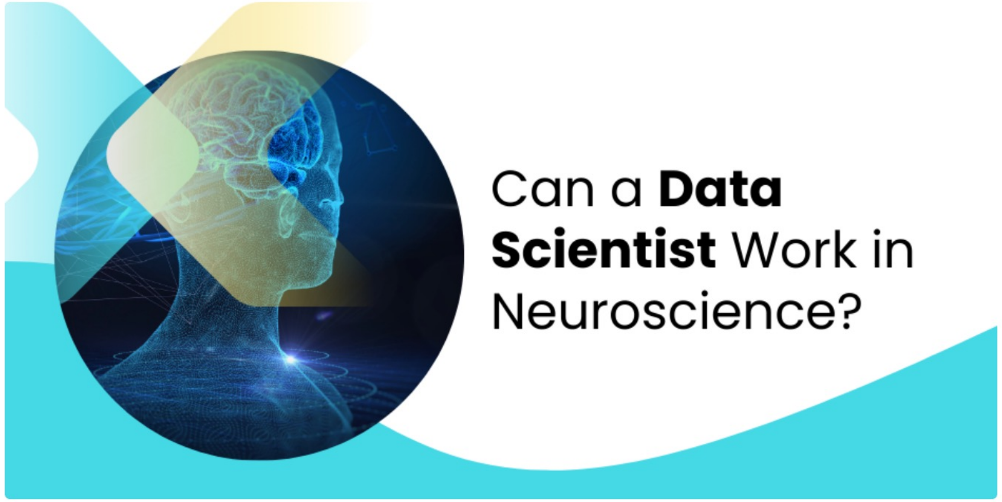
Liam Paninski describes his efforts to process calcium imaging and other neural data, an effort that he hopes will help drive a new era of data sharing.

By Emily Singer
November 19, 2018

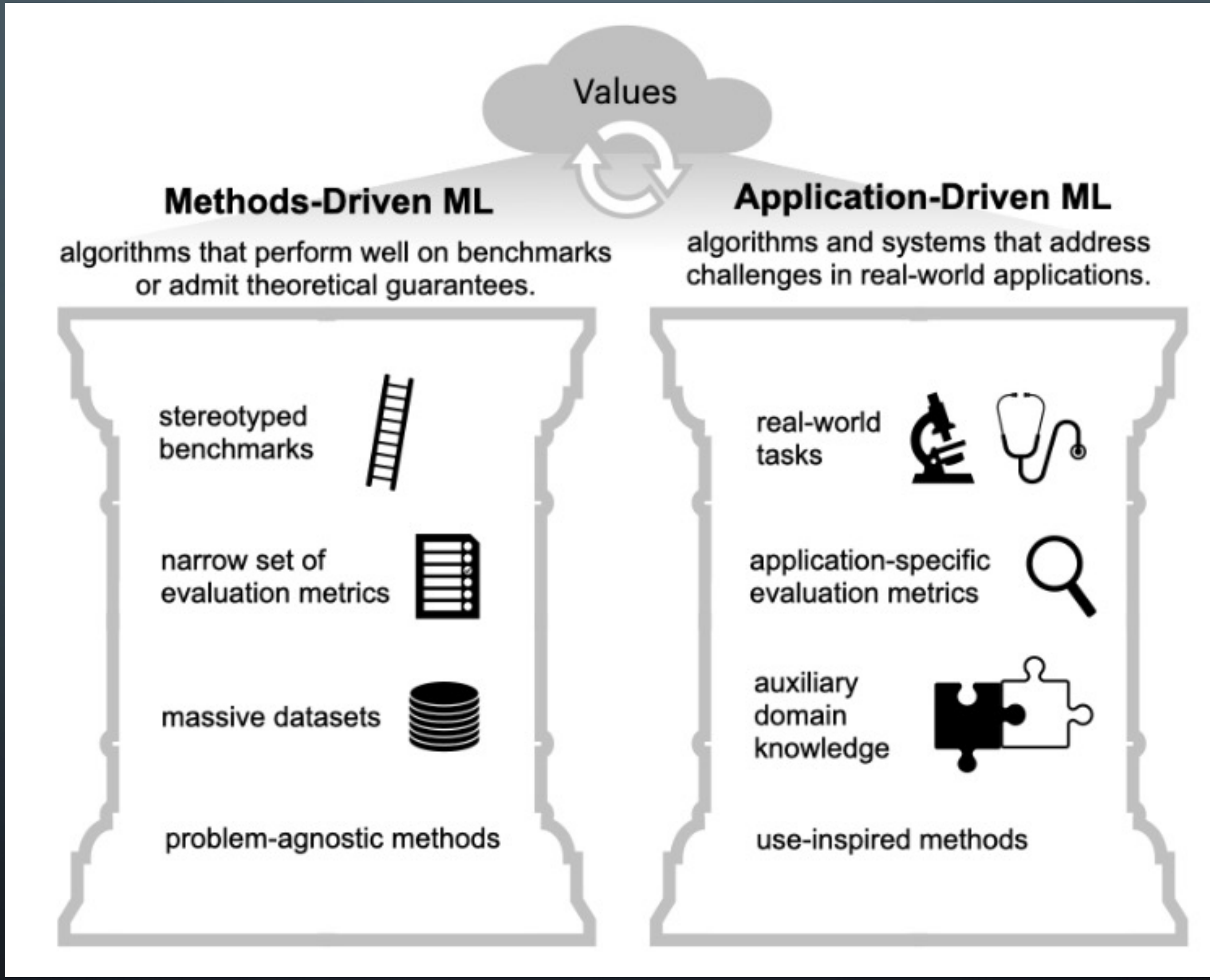
BIG DATA, DATA ANALYSIS, DATA SCIENCE

The Intersection of Neuroscience and Data Science: Exploring the Synergies

INSTITUTE OF DATA ON DECEMBER 13, 2023



APPLICATION-DRIVEN INNOVATION IN MACHINE LEARNING



Rolnick et al. (2024)

THANK YOU FOR YOUR ATTENTION! QUESTIONS?

JORDAN STEINHAUSER

4th year PhD student – Department of Psychology

Advisor – Dr. Edward Korzus

Email: jstei007@ucr.edu