Algorithms for Data Analysis

Statistics – Machine Learning

Matthieu Gilson

Chaire de Professeur Junior

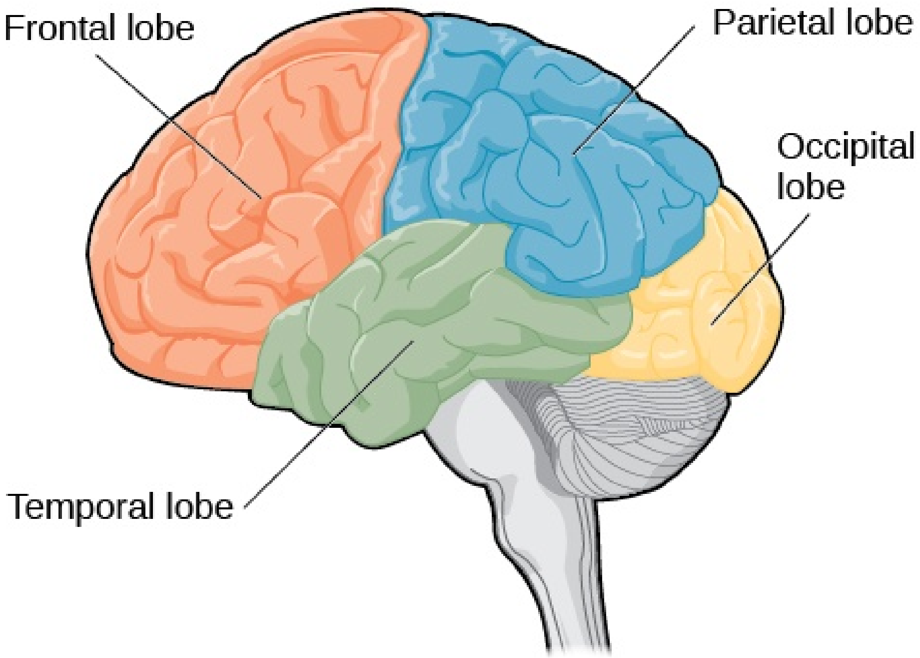
Introduction to Models in Neuroscience

- The brain as a distributed and complex network system
- Neuroimaging: quantifying the brain
- Statistical analysis versus classification
- Example 1: diagnosis / prognosis in stroke
- Example 2: whole-brain modeling, The Virtual Brain
- Example 3: characterize structure in multivariate data
- Scikit-learn: formatting data

Introduction to Models in Neuroscience

- **The brain as a distributed and complex network system**
- Neuroimaging: quantifying the brain
- Statistical analysis versus classification
- Example 1: diagnosis / prognosis in stroke
- Example 2: whole-brain modeling, The Virtual Brain
- Example 3: characterize structure in multivariate data
- Scikit-learn: formatting data

Localizing Behavioral and Cognitive Functions

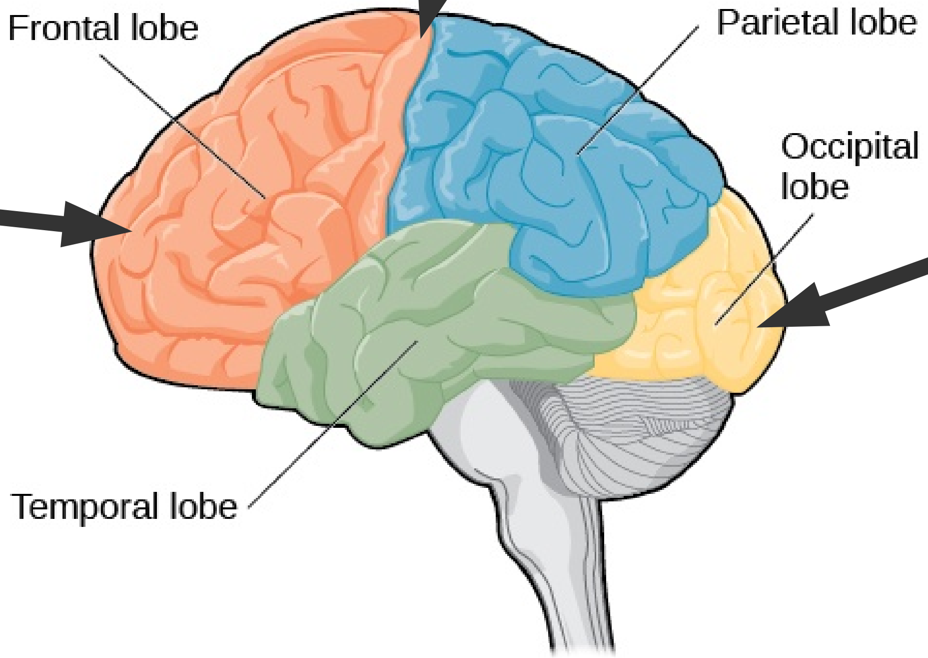


Localizing Behavioral and Cognitive Functions

voluntary muscle movements

long-term memory, planing, executive control

visual perception



Cognitive Neuroscience: Neuronal Implementation of Functions

- Example: matching of visual objects depending on color or shape, motor response

Context 1:
same color?

Cognitive Neuroscience: Neuronal Implementation of Functions

- Example: matching of visual objects depending on color or shape, motor response

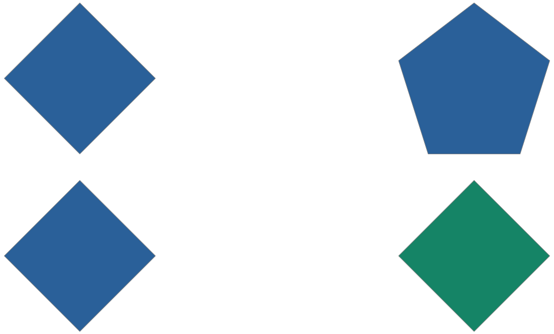
**Context 1:
same color?**

oui

Cognitive Neuroscience: Neuronal Implementation of Functions

- Example: matching of visual objects depending on color or shape, motor response

Context 1: same color?



oui

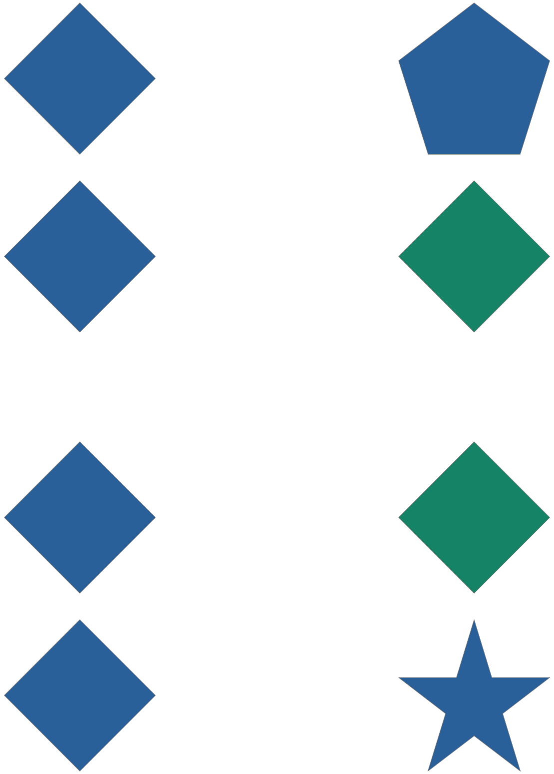
non

Cognitive Neuroscience: Neuronal Implementation of Functions

- Example: matching of visual objects depending on color or shape, motor response

Context 1: same color?

Context 2: same shape?



oui

non

oui

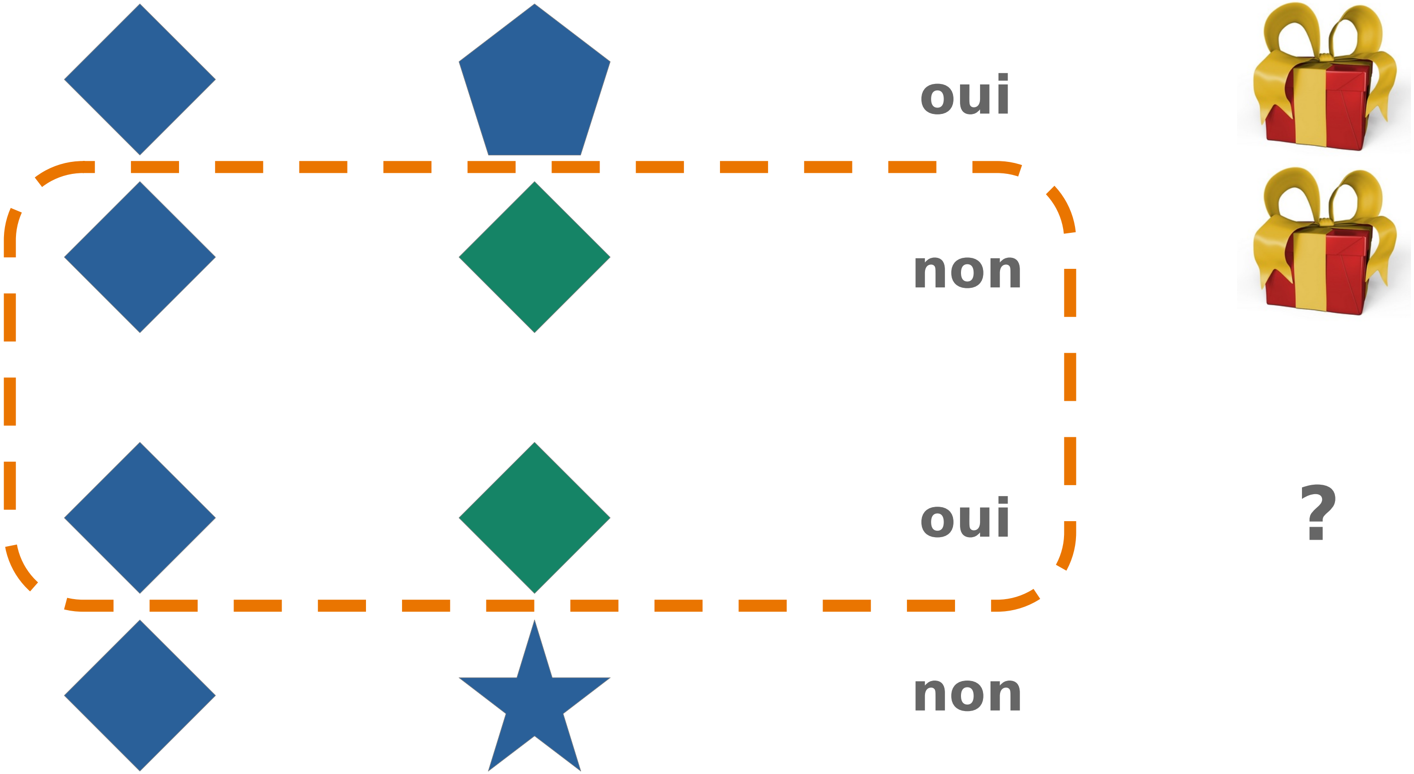
non

Cognitive Neuroscience: Neuronal Implementation of Functions

- Example: matching of visual objects depending on color or shape, motor response

Context 1:
same color?

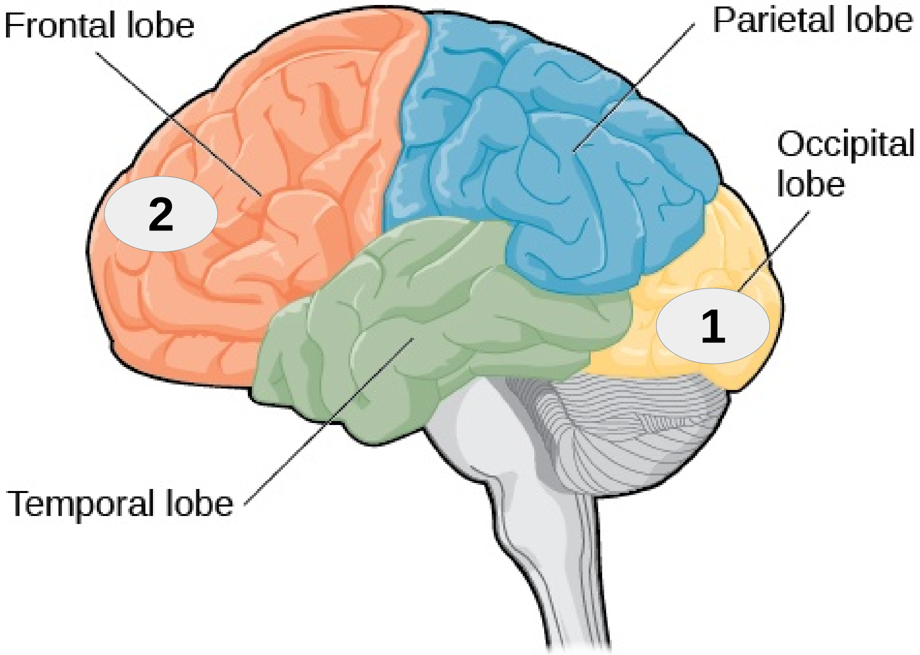
Context 2:
same shape?



Cognitive Neuroscience: Neuronal Implementation of Functions

- Example: matching of visual objects depending on color or shape, motor response
- Activation of brain regions and interactions (functional hierarchy)

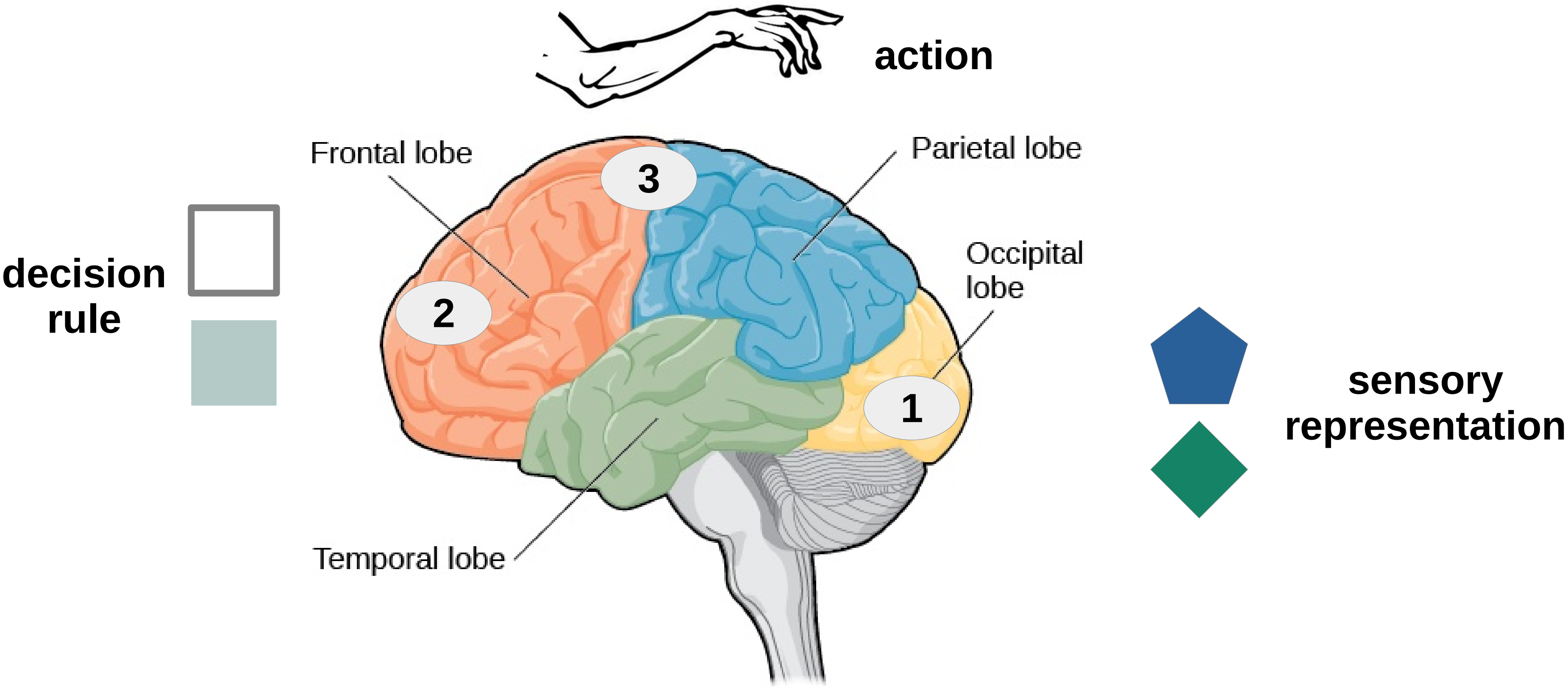
decision rule



sensory representation

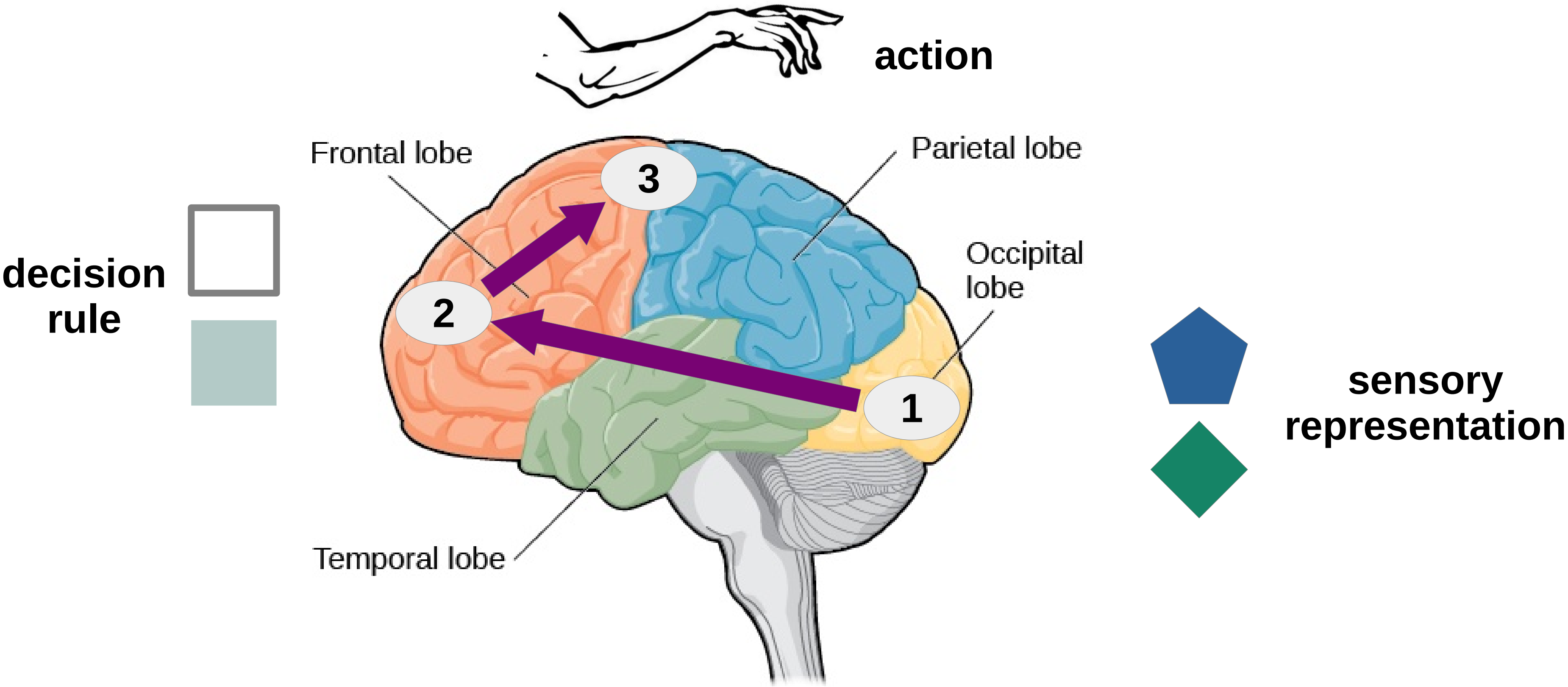
Cognitive Neuroscience: Neuronal Implementation of Functions

- Example: matching of visual objects depending on color or shape, motor response
- Activation of brain regions and interactions (functional hierarchy)



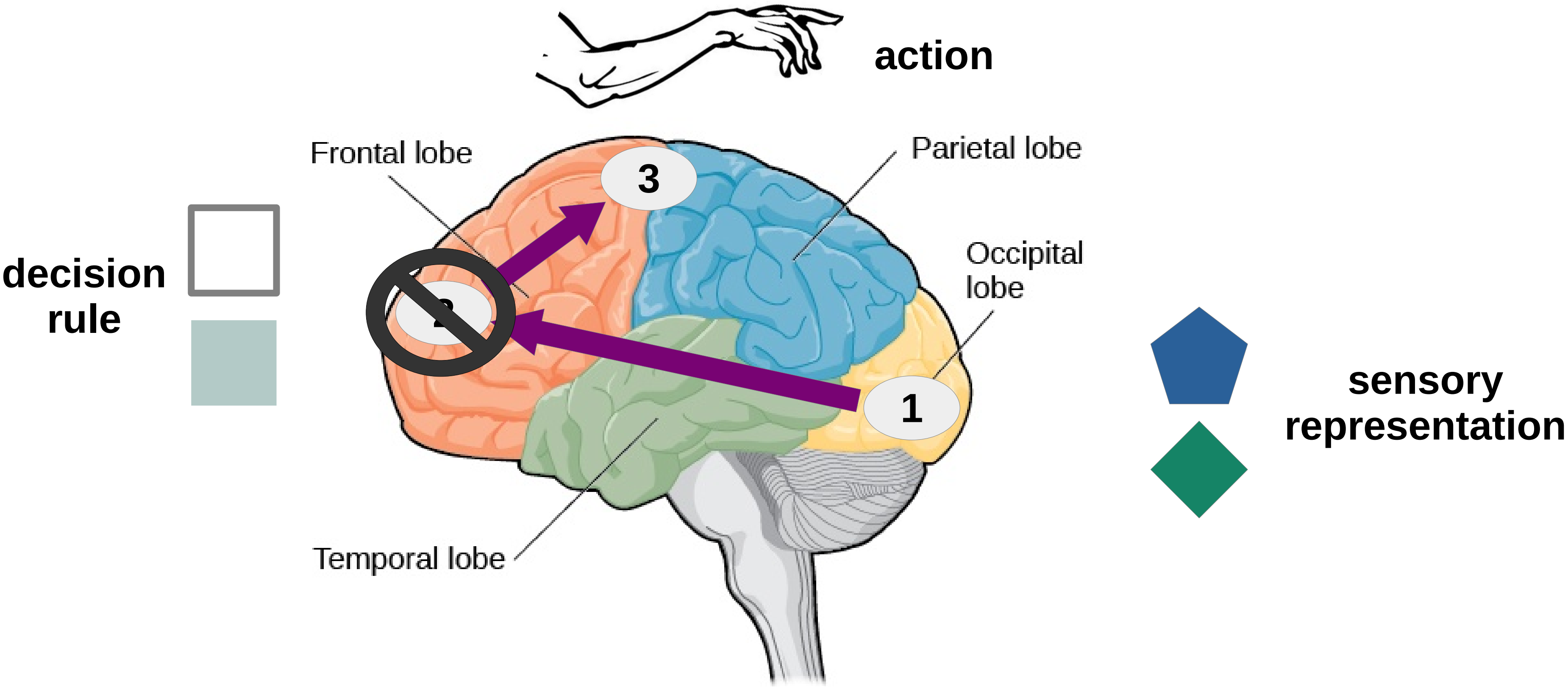
Cognitive Neuroscience: Neuronal Implementation of Functions

- Example: matching of visual objects depending on color or shape, motor response
- Activation of brain regions and interactions (functional hierarchy)



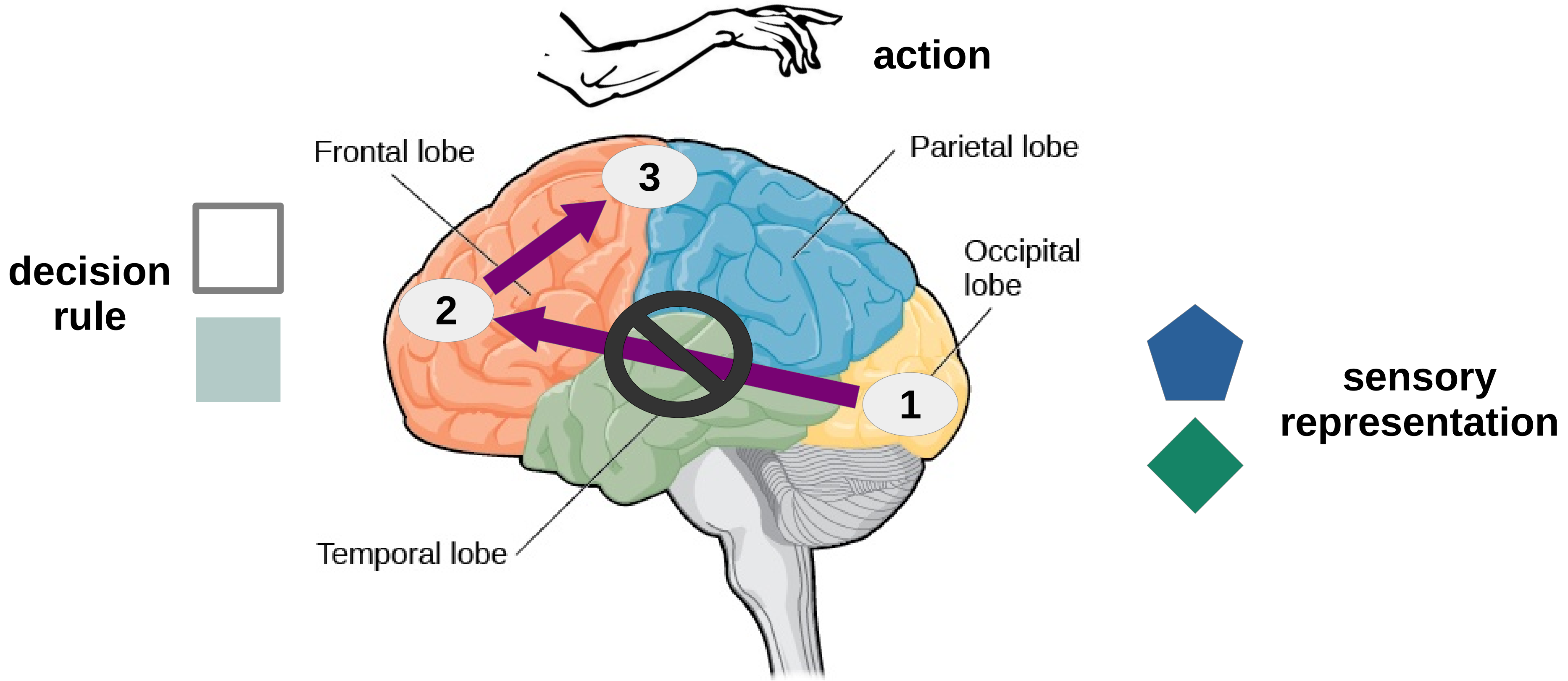
Characterization of Pathological Alterations

- Example: matching of visual objects depending on color or shape, motor response
- Activation of brain regions and interactions (functional hierarchy)



Characterization of Pathological Alterations

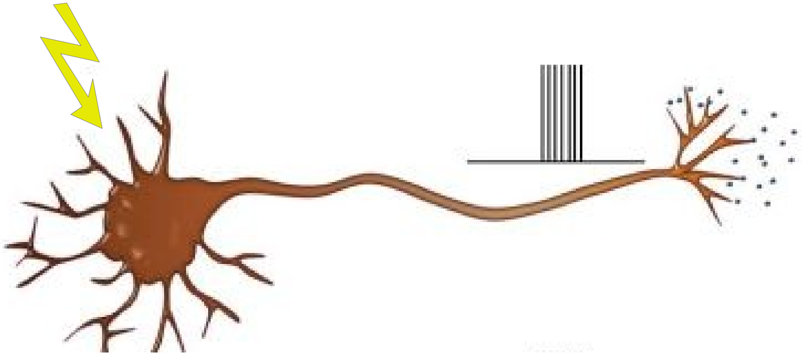
- Example: matching of visual objects depending on color or shape, motor response
- Activation of brain regions and interactions (functional hierarchy)



Multiscale Organization of the Brain

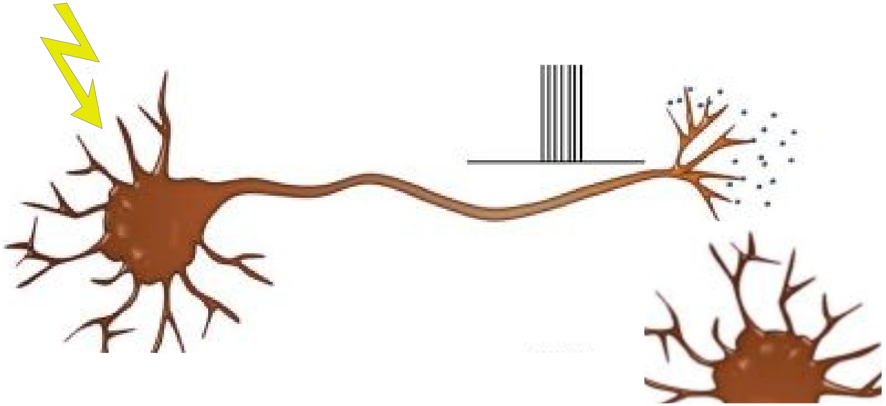
- 10^{11} neurons
  - 20% in cortex, 80% in cerebellum
- about 10000 synapses by neurons

Multiscale Organization of the Brain



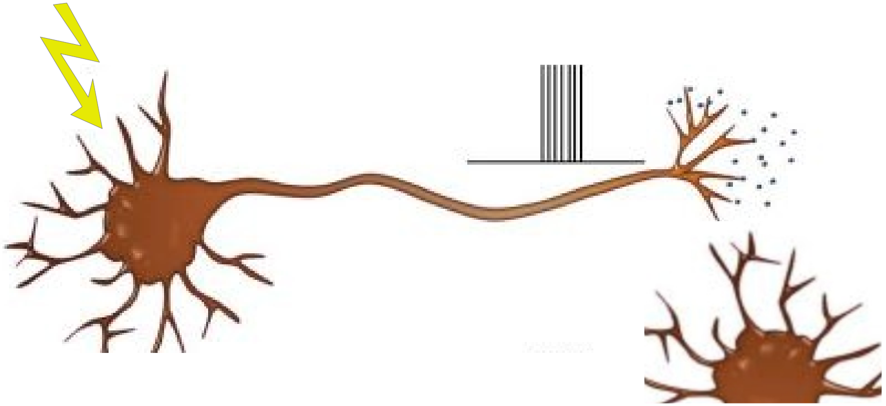
uclouvain.be, unisciel.fr

Multiscale Organization of the Brain



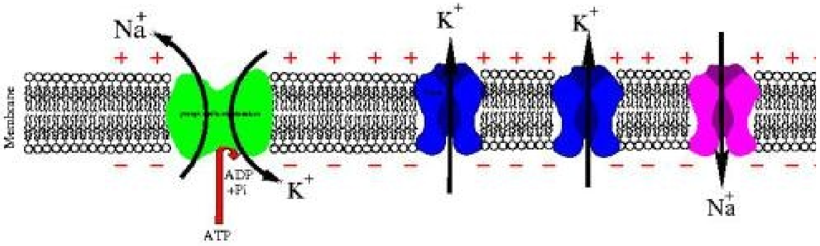
target
neuron

Multiscale Organization of the Brain

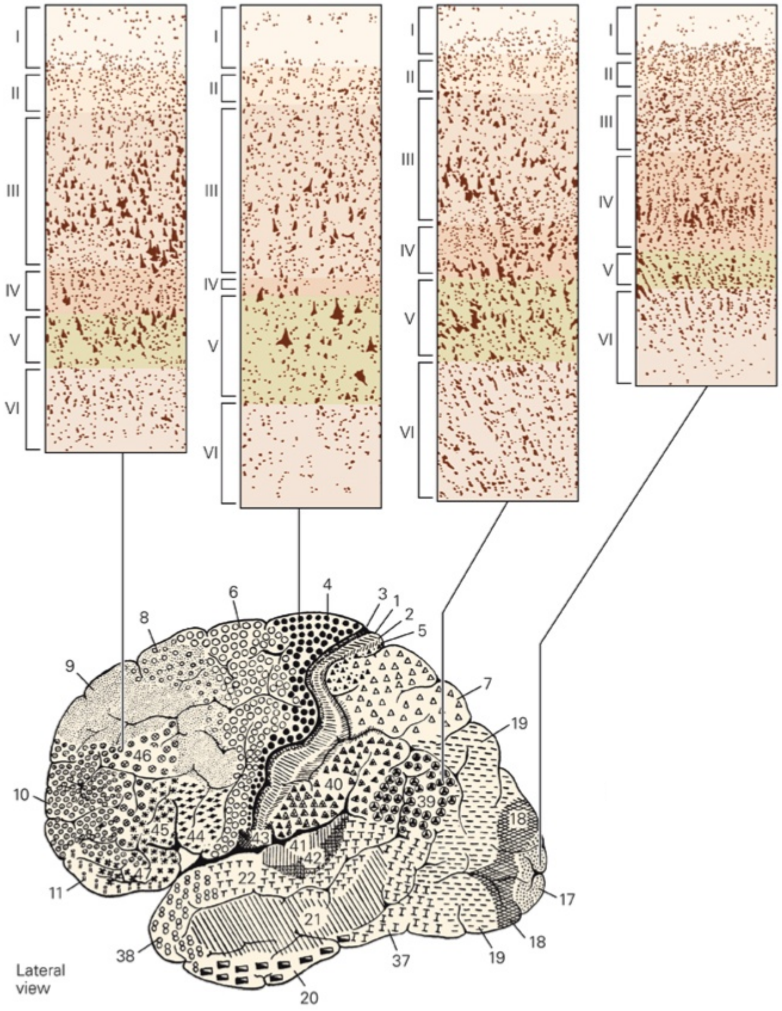


target
neuron

active ionic channels on neuronal cell membrane



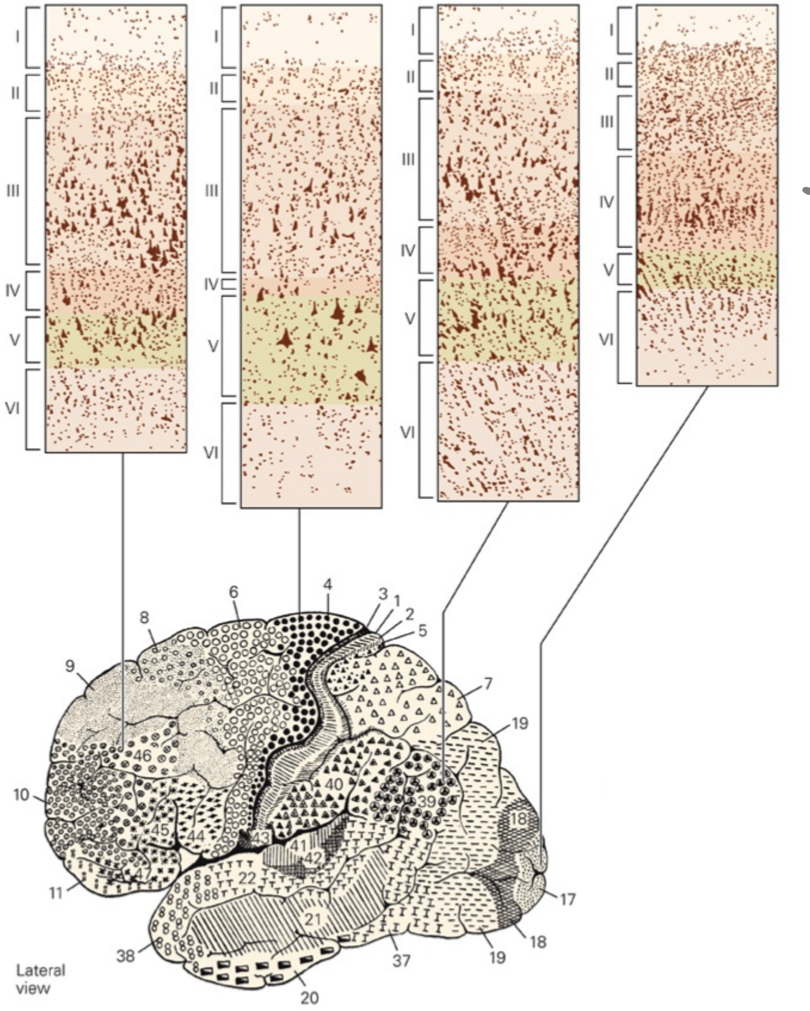
Multiscale Organization of the Brain



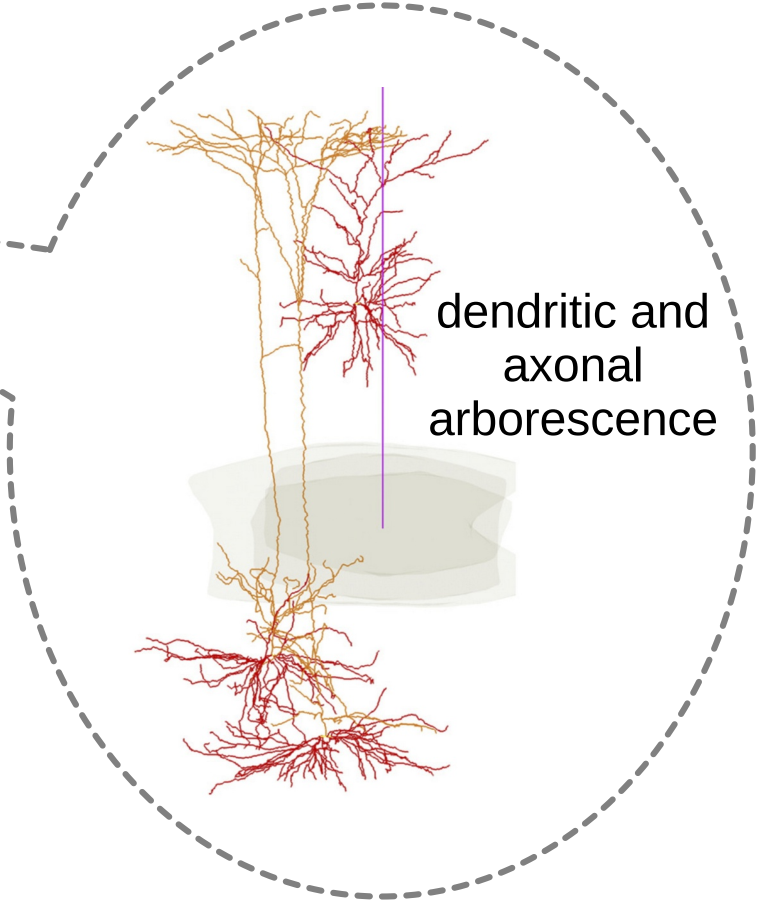
cortical layers

van Kan 2014; Helmstaedter BRR 2007

Multiscale Organization of the Brain



cortical layers



dendritic and axonal arborescence

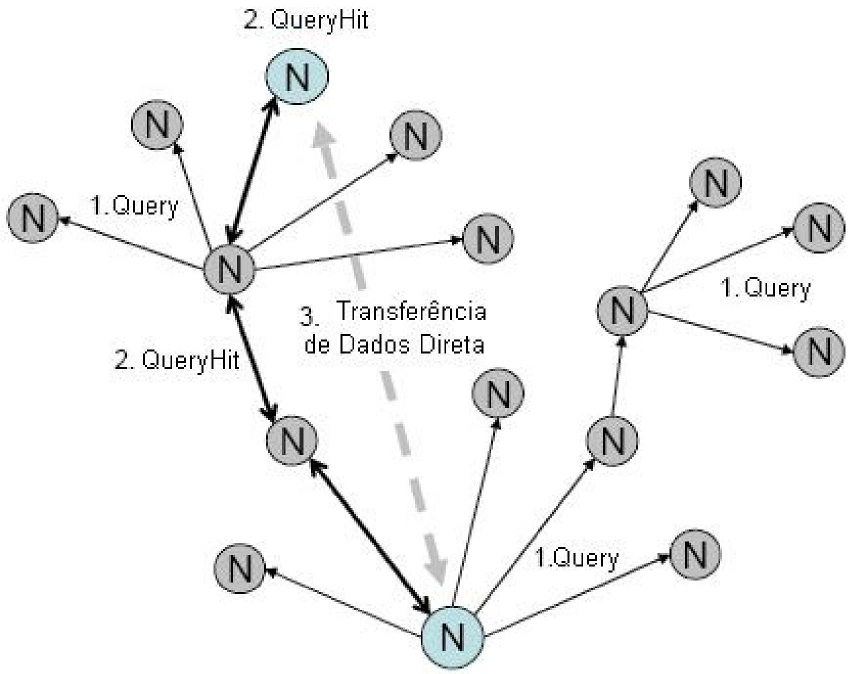
van Kan 2014; Helmstaedter BRR 2007

Analogy with Communication Network

- Complexity = vulnerability or resilience?
  - example in engineering: telecommunication, transport, logistic, financial, … networks
  - potentially fragile (internet cut, traffic jams, COVID, 2007 crisis)

Gnutella network

- peer-to-peer file sharing scheme
- efficiency: short paths

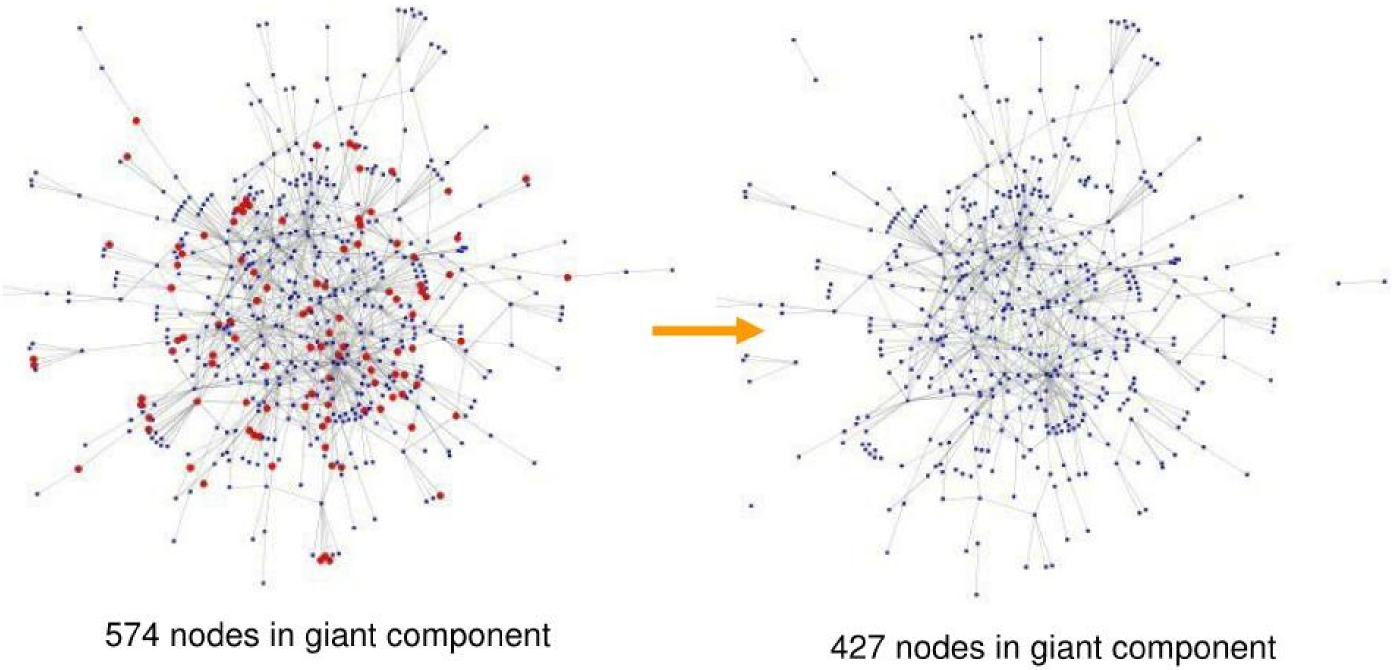


Analogy with Communication Network

- Complexity = vulnerability or resilience?
  - example in engineering: telecommunication, transport, logistic, financial, … networks
  - potentially fragile (internet cut, traffic jams, COVID, 2007 crisis)

Gnutella network

- peer-to-peer file sharing scheme
- efficiency: short paths
- resilience to node breakdown / attack
- scale-free topology



- Historically functions thought to be localized (cf. Phrenology)
- But modern viewpoint focuses on understanding how distributed neuronal implementations (high-level cognition)

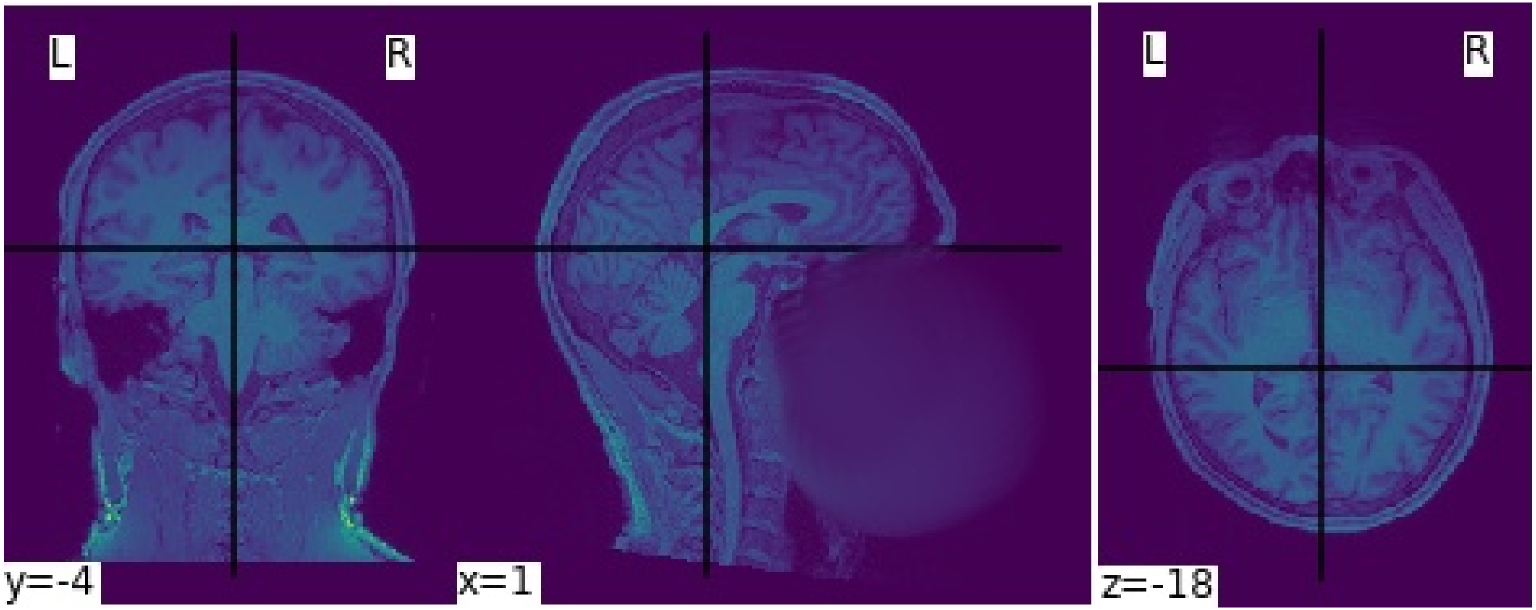
- Limits of analogy with telecommunication network
  - heterogeneous neuron types
  - multiple scales of organization
  - many molecular mechanisms in parallel: synaptic transmission + neuromodulation + metabolism + ...

- Still, some concepts remain important
  - redundant network structure (supporting robustness)
  - dynamic aspect of processes
- How are these properties that affected in neuropathologies?
  - systemic approach at network level

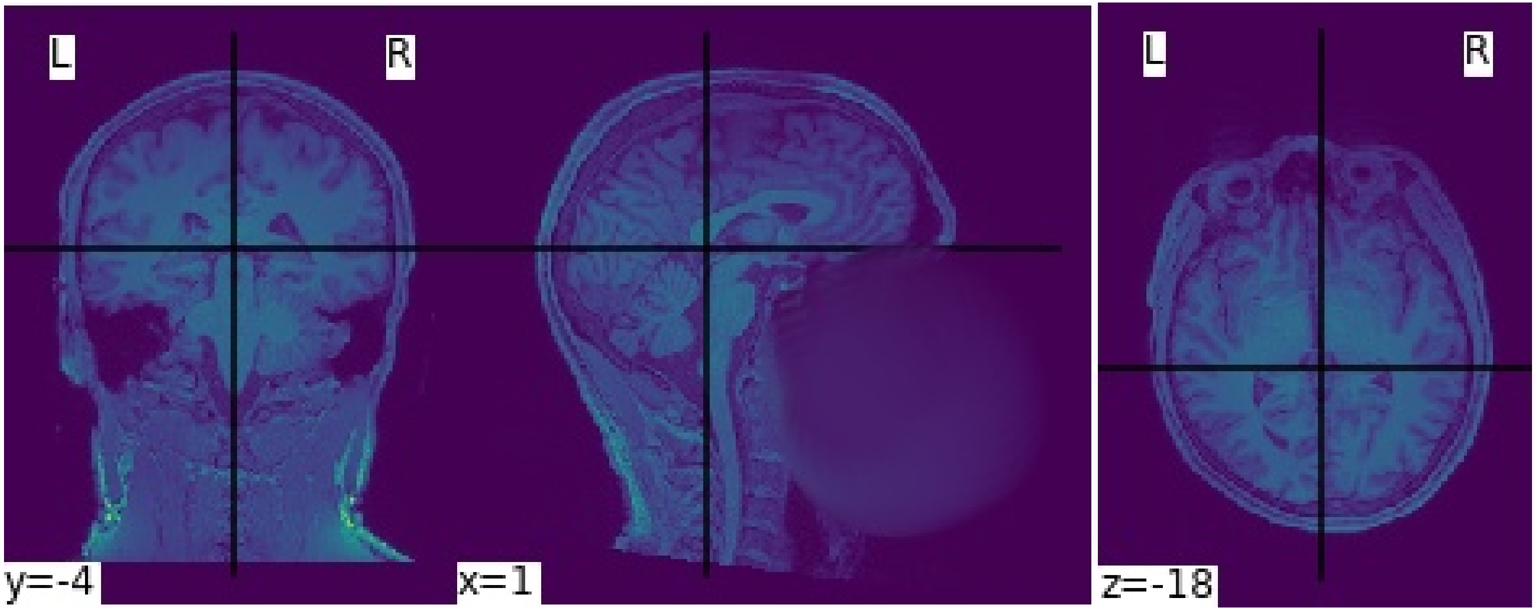
Introduction to Models in Neuroscience

- The brain as a distributed and complex network system
- **Neuroimaging: quantifying the brain**
- Statistical analysis versus classification
- Example 1: diagnosis / prognosis in stroke
- Example 2: whole-brain modeling, The Virtual Brain
- Example 3: characterize structure in multivariate data
- Scikit-learn: formatting data

Magnetic Resonance Imaging (MRI) to Reveal Brain Structure



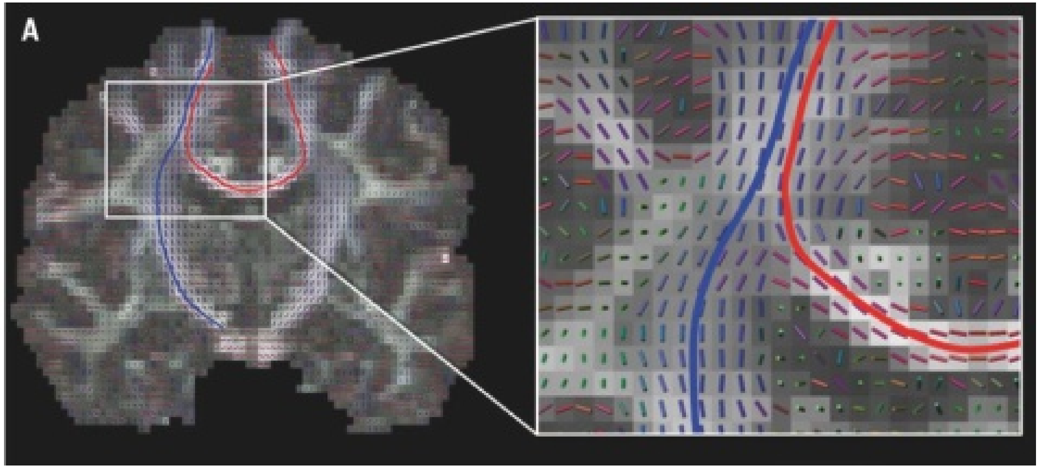
Magnetic Resonance Imaging (MRI) to Reveal Brain Structure



- Identification tissus neuronaux (matière grise versus blanche ; riches en eau, lipides, ...)
- Séquence du scanner : excitation certains atomes (hydrogène)

Tractography: Axonal Fibers = Connectome

diffusion IRM
(movement of
water molecules)

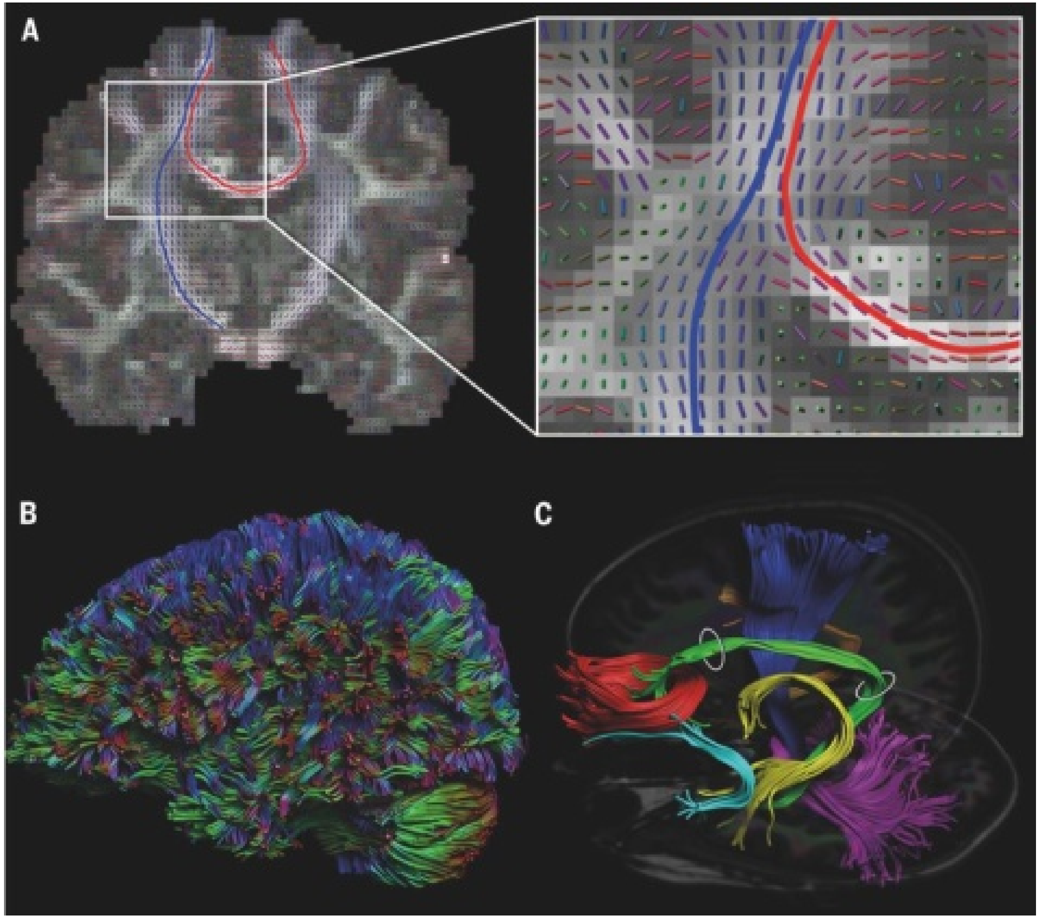


- corticospinal tract
- corpus callosum

Axer, Amunts (2022) *Science*

Tractography: Axonal Fibers = Connectome

diffusion IRM
(movement of
water molecules)



- corticospinal tract
- corpus callosum

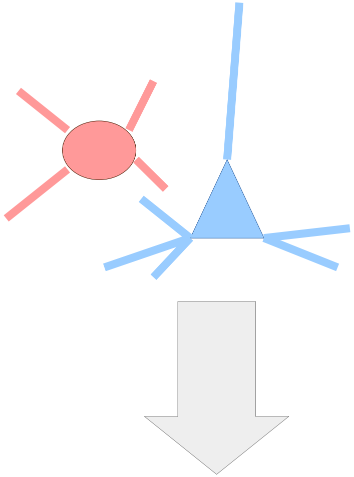
whole-brain
connectome

Example fiber
bundles (white
matter)

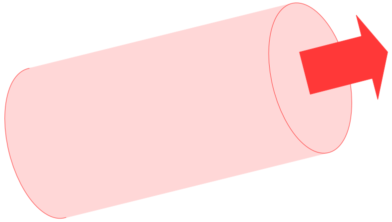
Axer, Amunts (2022) *Science*

Functional MRI: Proxy for Neuronal Activity

neuronal activity

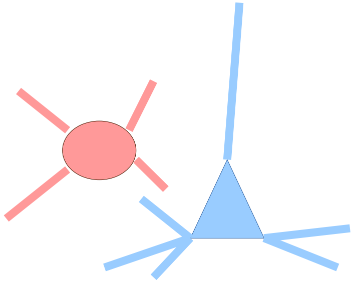


blood influx
(oxygenation)

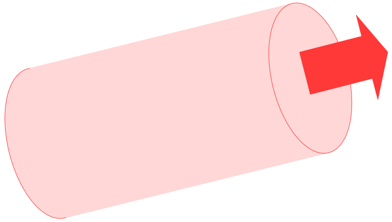


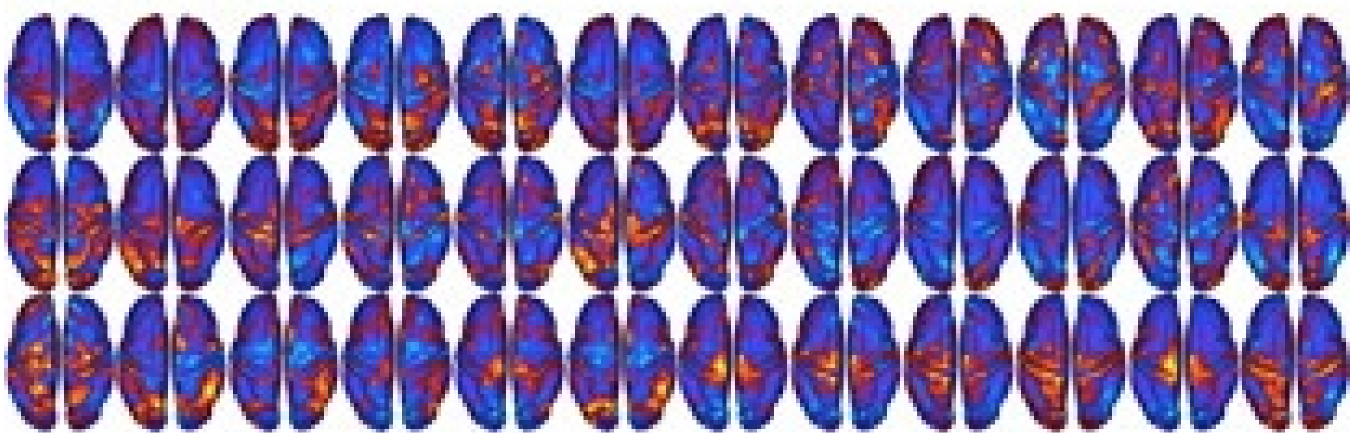
Functional MRI: Proxy for Neuronal Activity

neuronal activity



blood influx
(oxygenation)





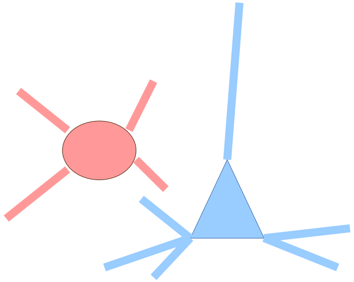
time

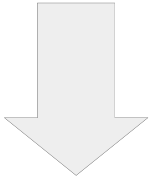
- blood-oxygen-level dependent (BOLD) signal
- indirect measure of neuronal activation
- spatio-temporal structure

Raichle (2010) *Trend Cog Sci*; Snyder (2012) *Neuroimage*; Biswal BB(2012) *Neuroimage*

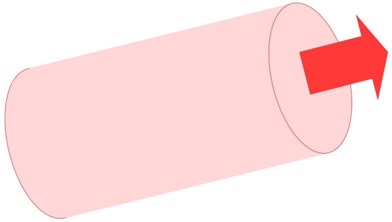
Functional MRI: Proxy for Neuronal Activity

neuronal activity

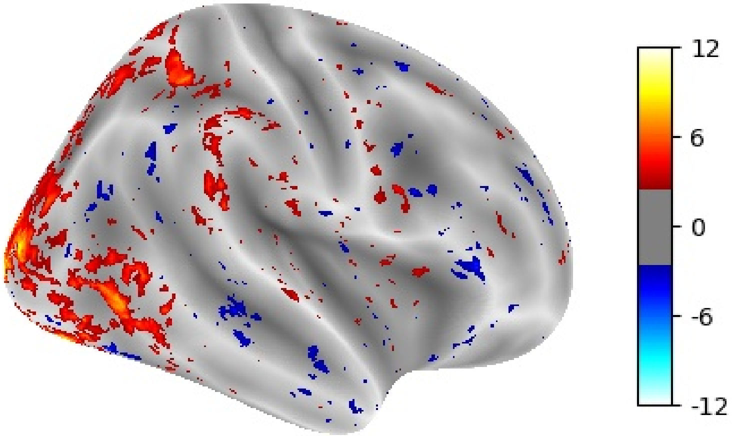




blood influx
(oxygenation)



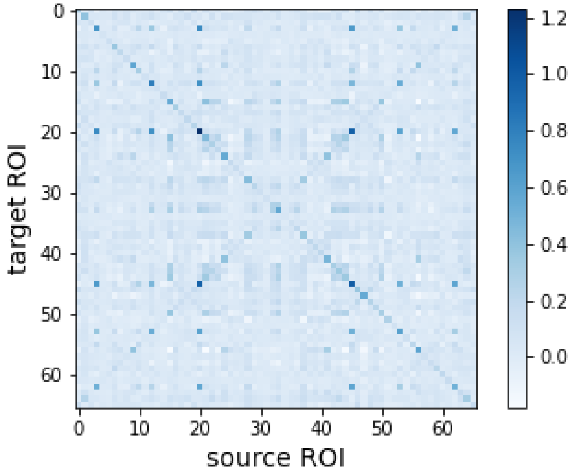
change in neuronal activity during emotional task



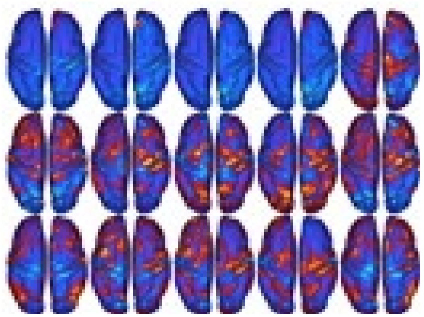
Introduction to Models in Neuroscience

- The brain as a distributed and complex network system
- Neuroimaging: quantifying the brain
- **Statistical analysis versus classification**
- Example 1: diagnosis / prognosis in stroke
- Example 2: whole-brain modeling, The Virtual Brain
- Example 3: characterize structure in multivariate data
- Scikit-learn: formatting data

Functional Connectivity: Signature of Brain Dynamic Activity

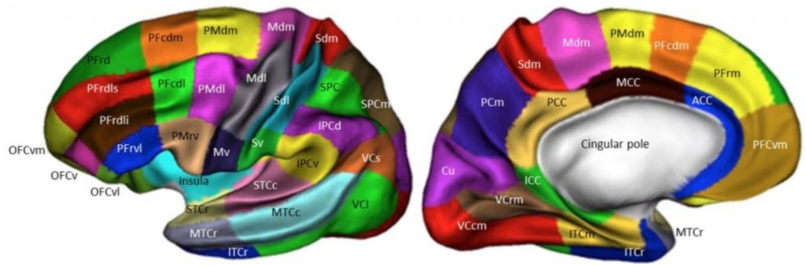


- 100 regions
- 5000 matrix elements



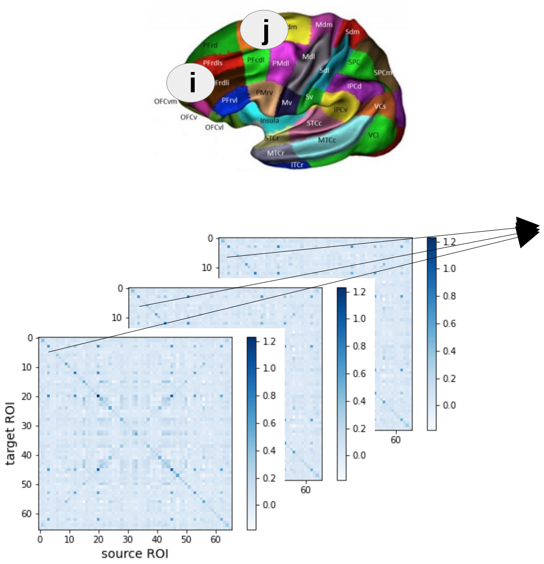
BOLD images
(voxel space)

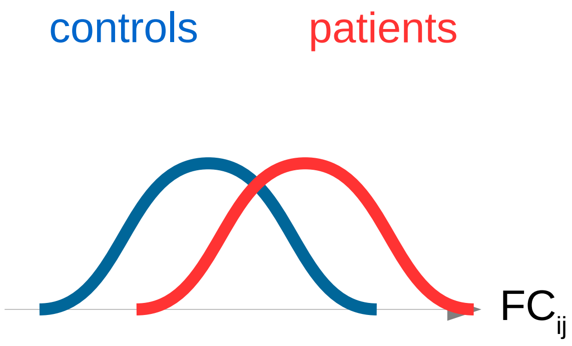
Brain Parcellation



MarsAtlas, Auzias, Brovelli, et al.

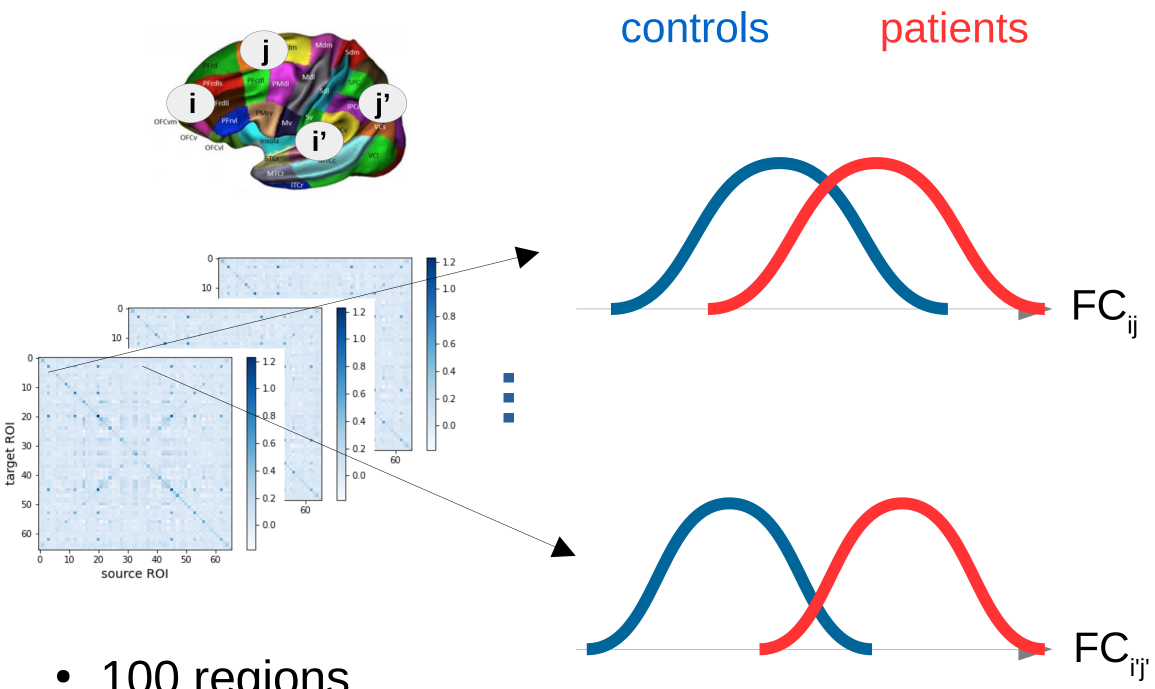
Descriptive Statistics





- 100 regions
- 5000 matrix elements
- patients et healthy controls

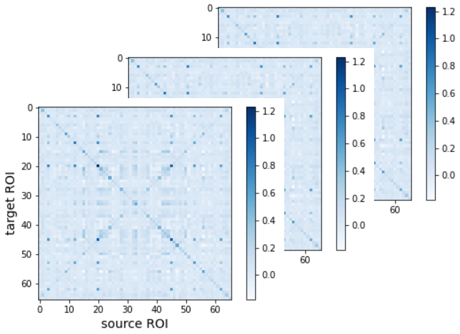
Descriptive Statistics



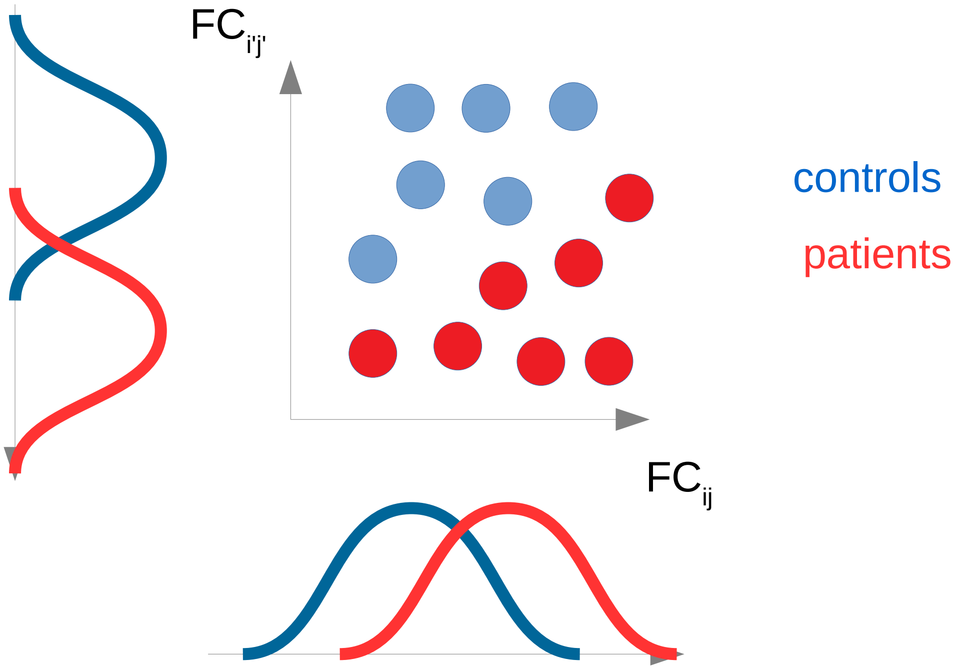
Which interactions change most across patients and controls?

- 100 regions
- 5000 matrix elements
- patients et healthy controls

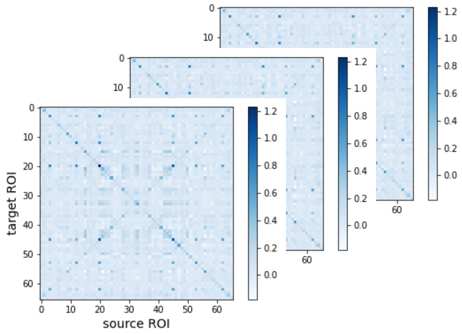
Descriptive Statistics



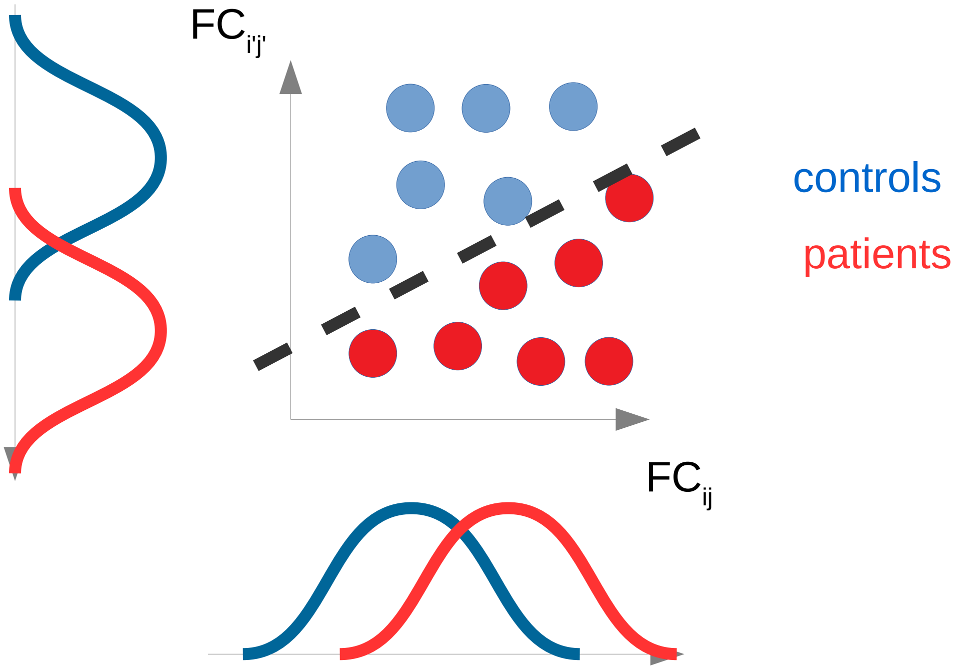
- 100 regions
- 5000 matrix elements
- patients et healthy controls



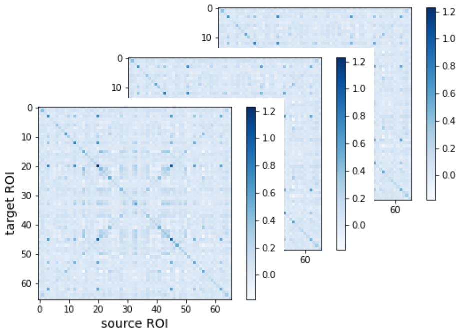
Descriptive Statistics



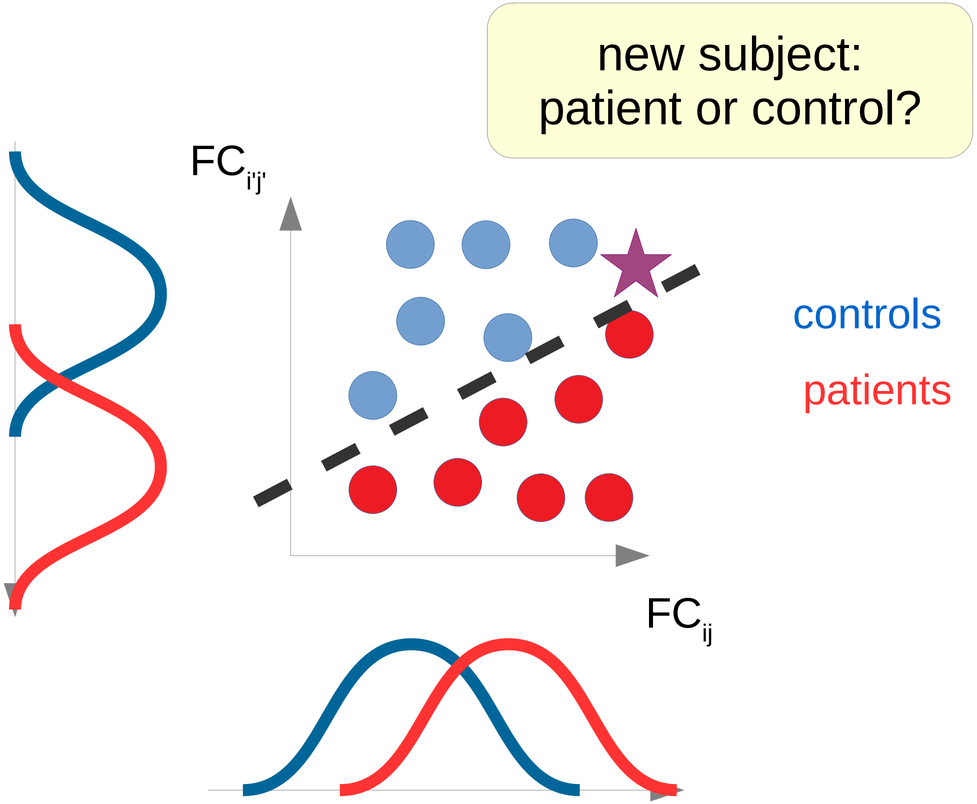
- 100 regions
- 5000 matrix elements
- patients et healthy controls



Predictive Statistics

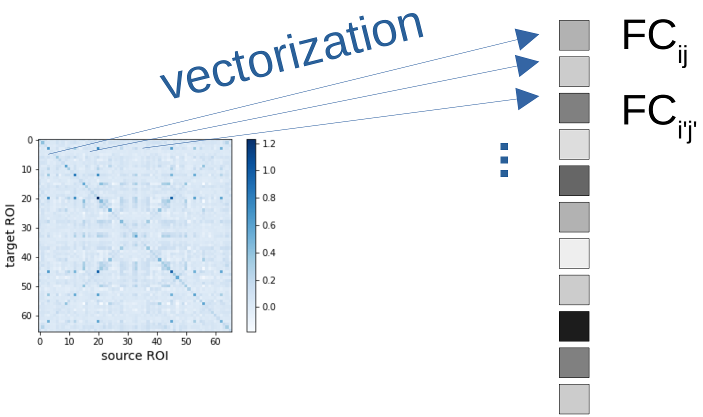


- 100 regions
- 5000 matrix elements
- patients et healthy controls

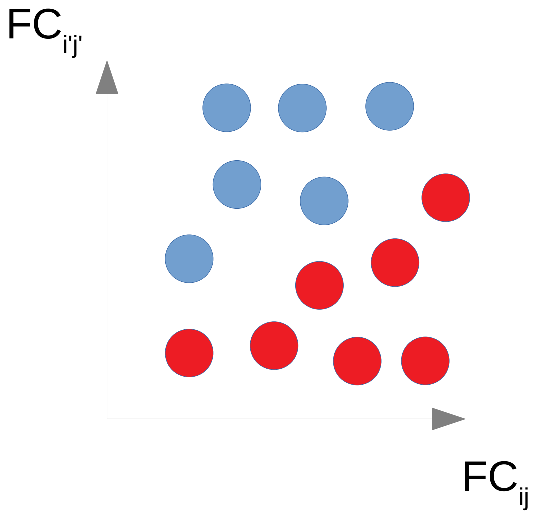


Predictive Statistics

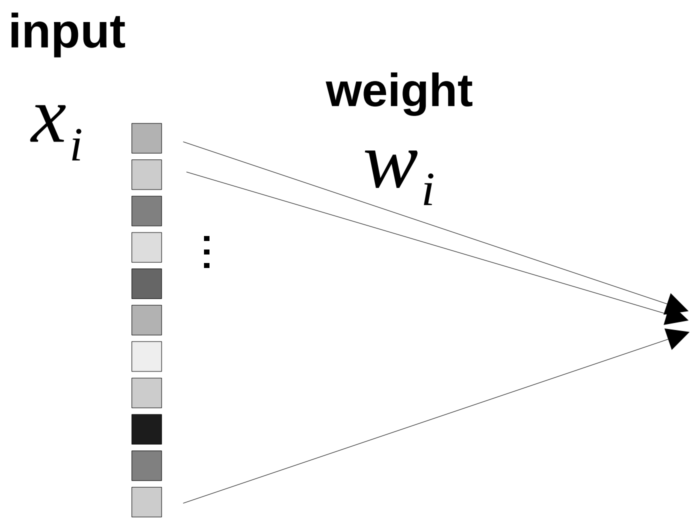
input



- 100 regions
- 5000 matrix elements

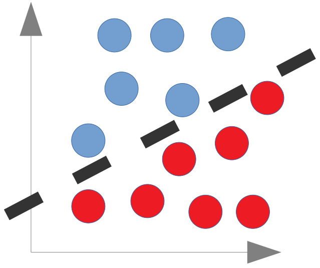


Predictive Statistics



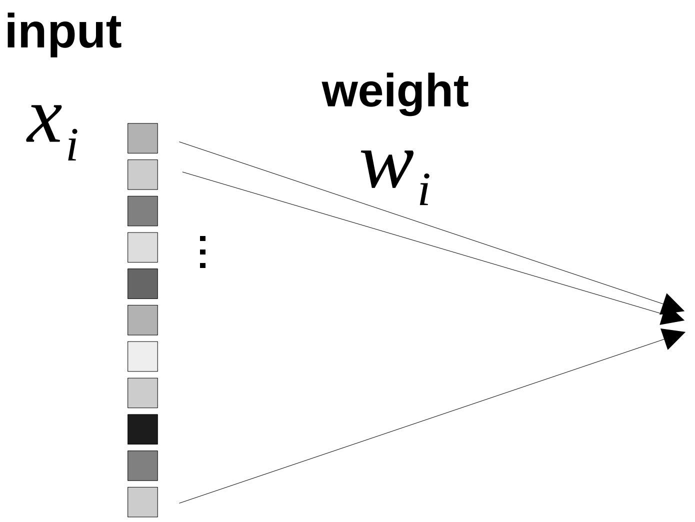
output

$$y = f\left(\sum_i w_i x_i\right)$$



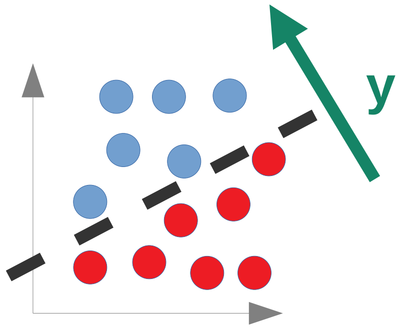
- Output indicates class/category for input (subject signature)

Predictive Statistics



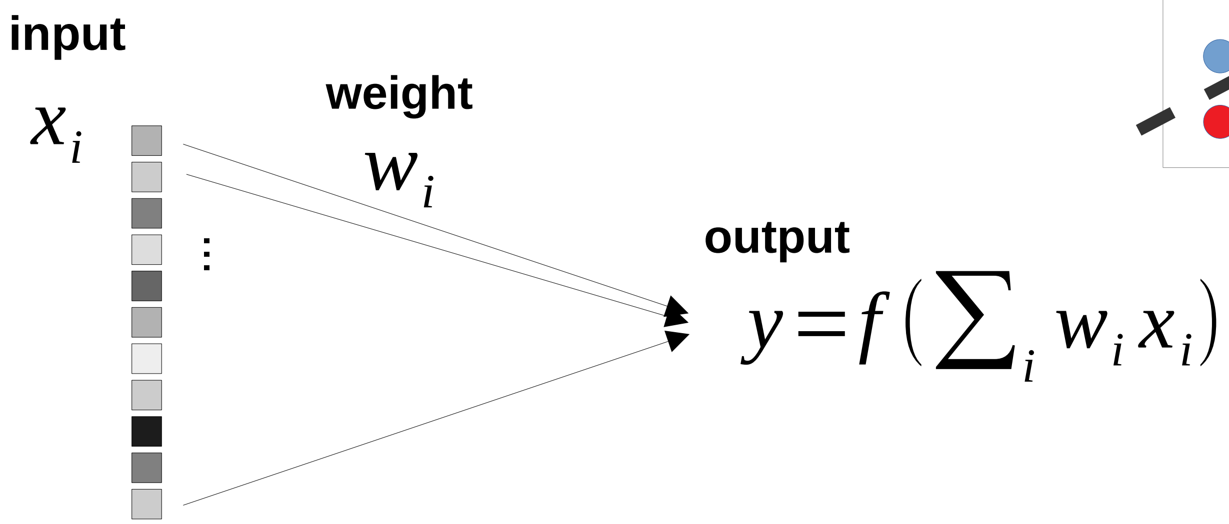
output

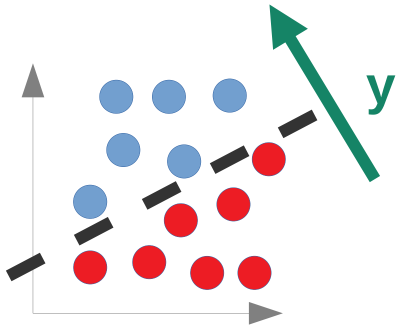
$$y = f\left(\sum_i w_i x_i\right) > 0$$



- Output indicates class/category for input (subject signature)

Predictive Statistics



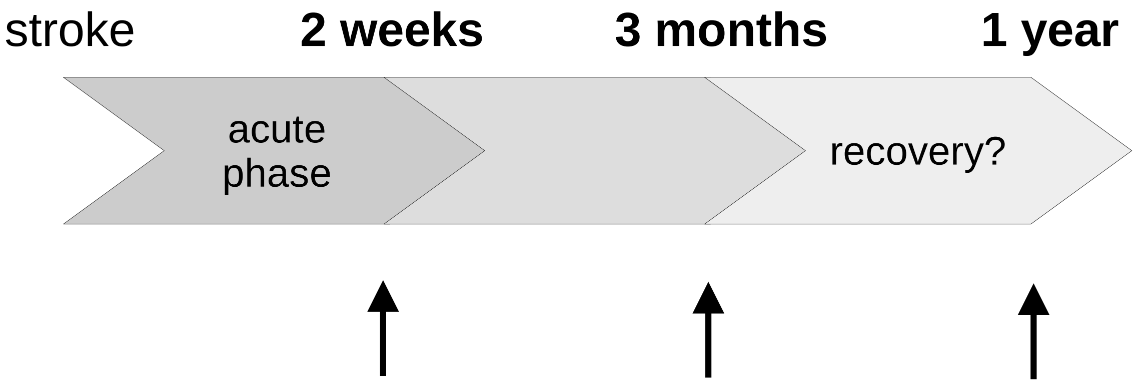


- Output indicates class/category for input (subject signature)
- Training weights for batch of inputs to obtain best possible prediction

Introduction to Models in Neuroscience

- The brain as a distributed and complex network system
- Neuroimaging: quantifying the brain
- Statistical analysis versus classification
- **Example 1: diagnosis / prognosis in stroke**
- Example 2: whole-brain modeling, The Virtual Brain
- Example 3: characterize structure in multivariate data
- Scikit-learn: formatting data

Monitoring and Predicting Recovery for Stroke Patients

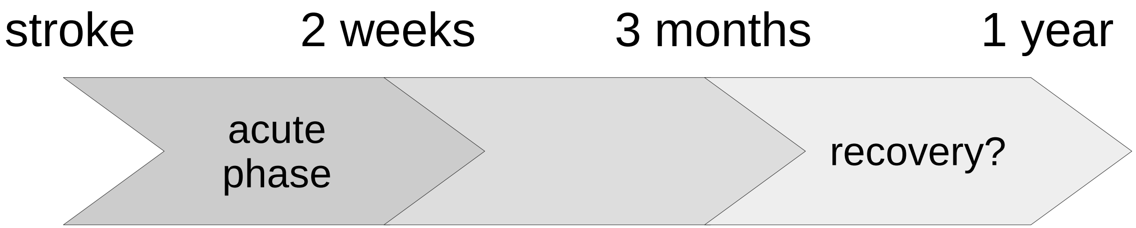


- Structural and functional MRI scans
- cognitive tests: memory, motor task, …

Adhikari, …, Gilson*, Corbetta* *Brain Commun* 2021

Human Brain Project

Monitoring and Predicting Recovery for Stroke Patients



- 132 patients, 25 controls
- 300 brain regions (ROIs)
- 80%-20% train-test (stratified split)
- classifier: logistic regression

Adhikari, …, Gilson*, Corbetta* *Brain Commun* 2021

Monitoring and Predicting Recovery for Stroke Patients

stroke **2 weeks** 3 months 1 year

fMRI scan

Diagnosis

patient or control?

Adhikari, …, Gilson*, Corbetta* *Brain Commun* 2021

Human Brain Project

Monitoring and Predicting Recovery for Stroke Patients

stroke **2 weeks** 3 months **1 year**

acute phase → → recovery?

fMRI scan

Diagnosis

patient or control?

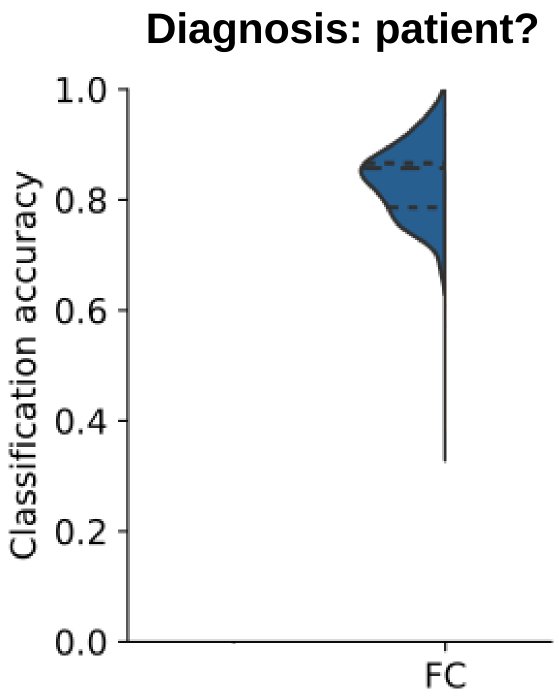
Prognosis

fMRI scan → **cognitive deficits**

Human Brain Project

Adhikari, …, Gilson*, Corbetta* *Brain Commun* 2021

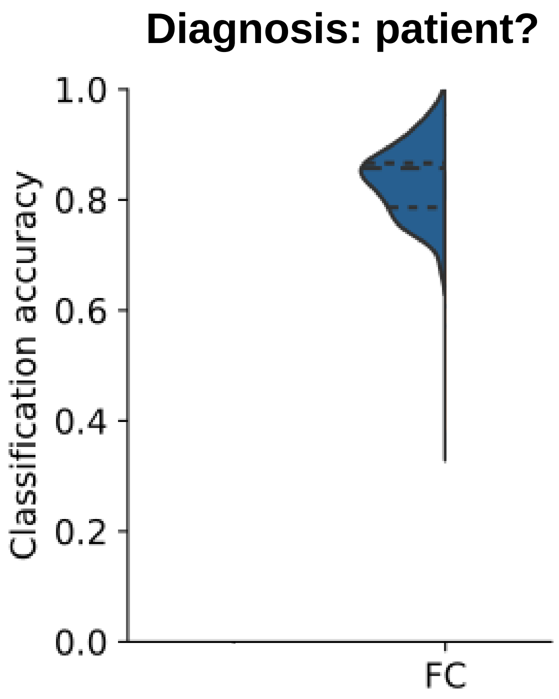
Prediction from FC



real data

Adhikari, …, Gilson*, Corbetta* *Brain Commun* 2021; Human Brain Project

Prediction from FC



100% correct

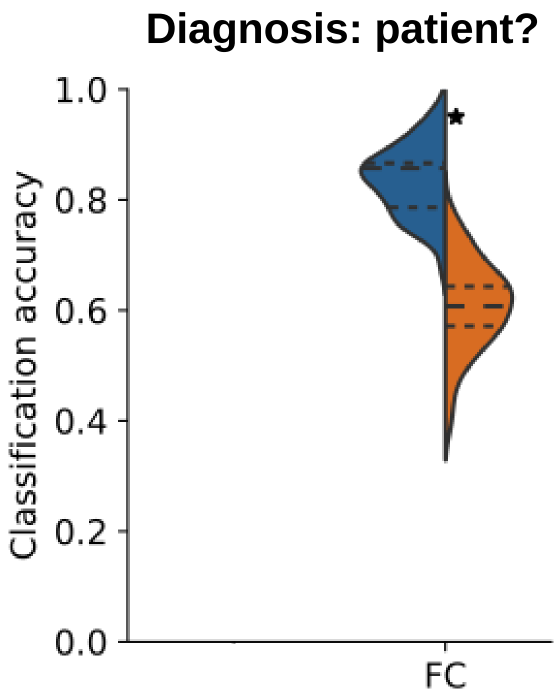
reference baseline???

100% wrong

real data

Adhikari, …, Gilson*, Corbetta* *Brain Commun* 2021; Human Brain Project

Prediction from FC

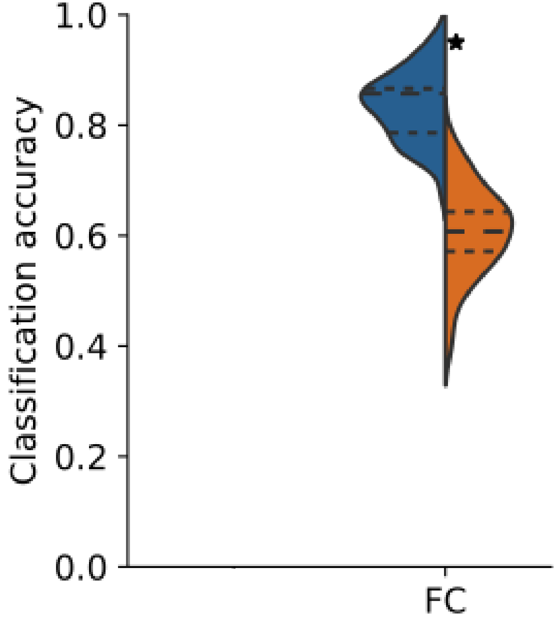


real data
chance level
(shuffling surrogates)

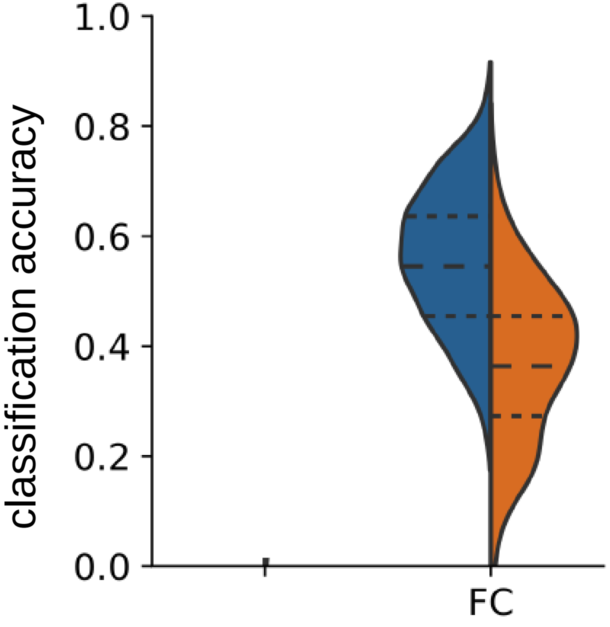
Adhikari, …, Gilson*, Corbetta* *Brain Commun* 2021; Human Brain Project

Prediction from FC

Diagnosis: patient?



Prognosis: cognitive deficits

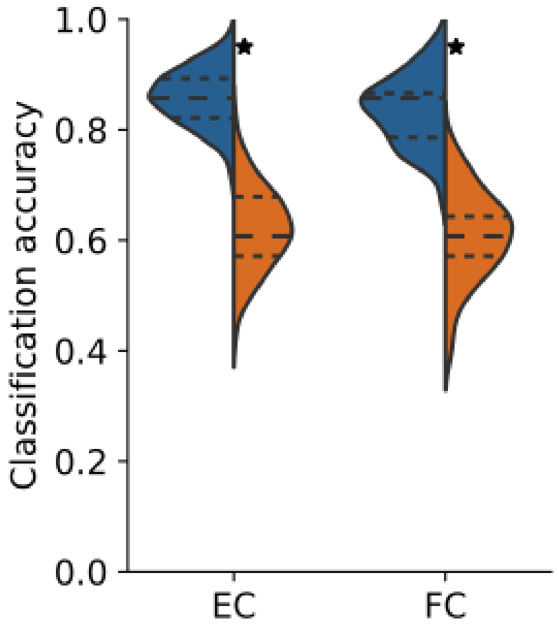


real data
chance level
(shuffling surrogates)

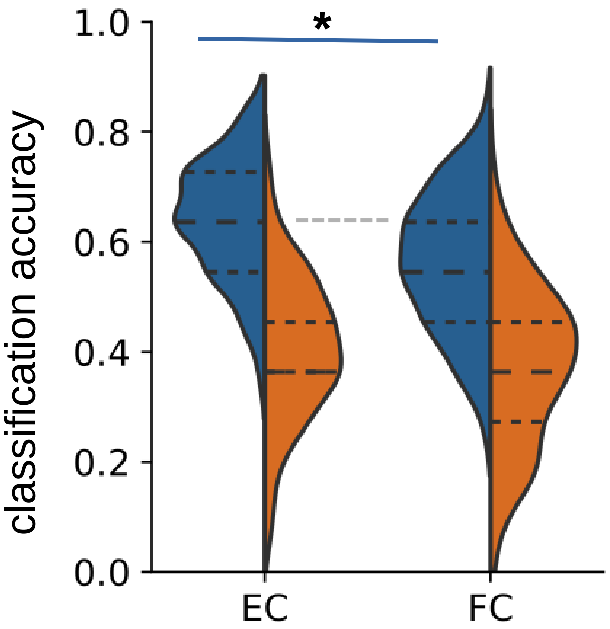
Adhikari, …, Gilson*, Corbetta* *Brain Commun* 2021; Human Brain Project

Model-based EC can Outperform FC

Diagnosis: patient?



Prognosis: cognitive deficits

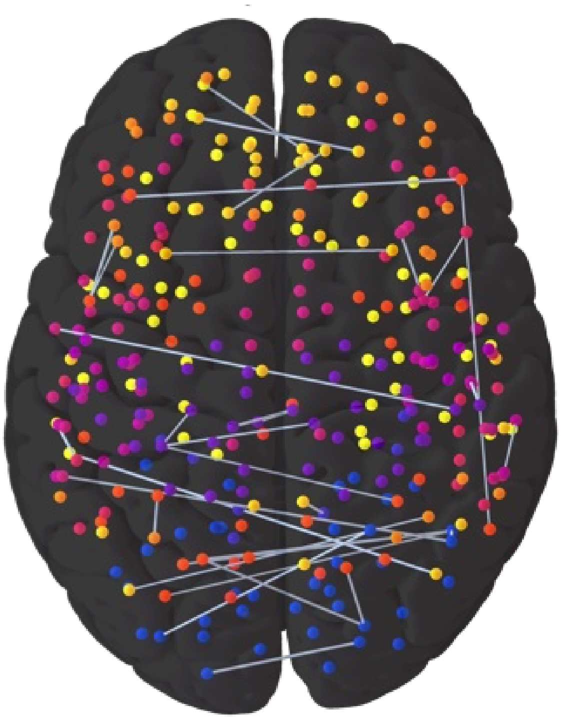


real data
chance level
(shuffling surrogates)

Adhikari, …, Gilson*, Corbetta* *Brain Commun* 2021; Human Brain Project

Interpretation as Cortico-Cortical Reconfiguration

Informative EC links



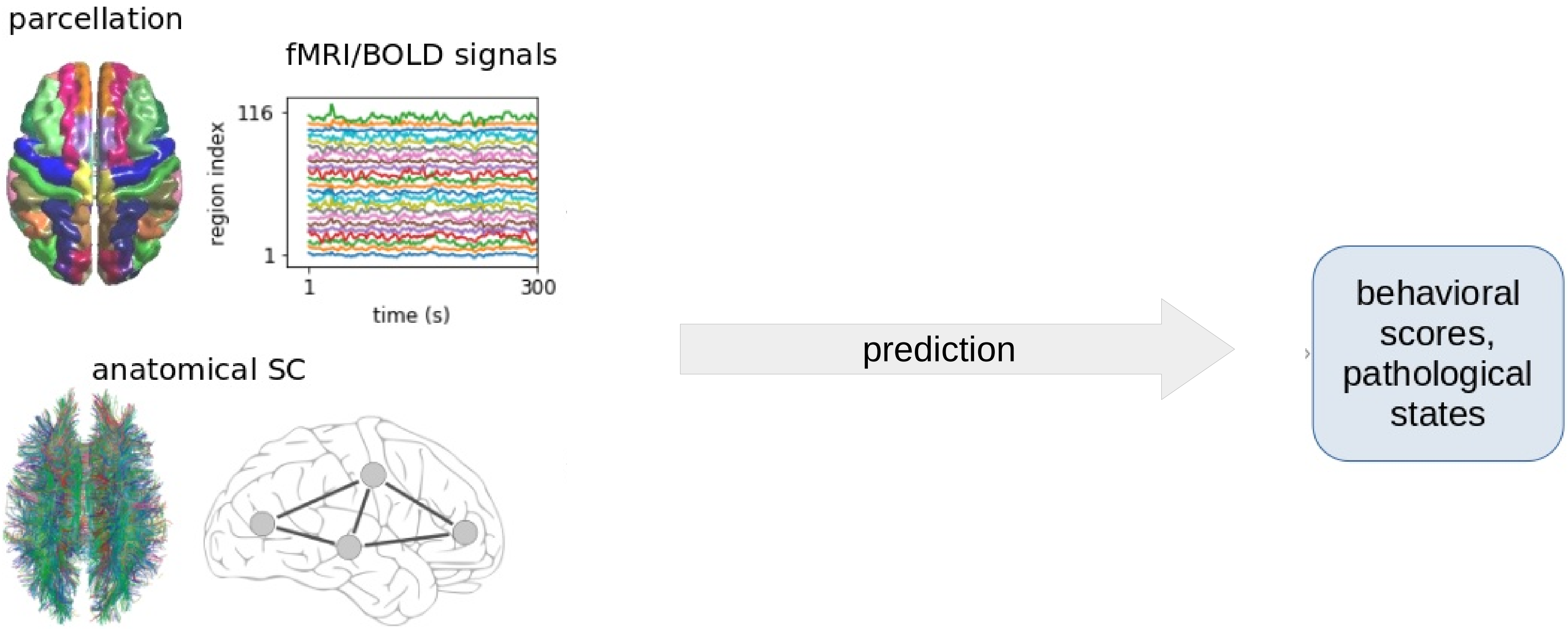
- What supports the recovery of cognitive deficits?
- Not only close to lesions
- Inter-hemispheric connections

Adhikari, …, Gilson*, Corbetta* *Brain Commun* 2021; Human Brain Project

Introduction to Models in Neuroscience

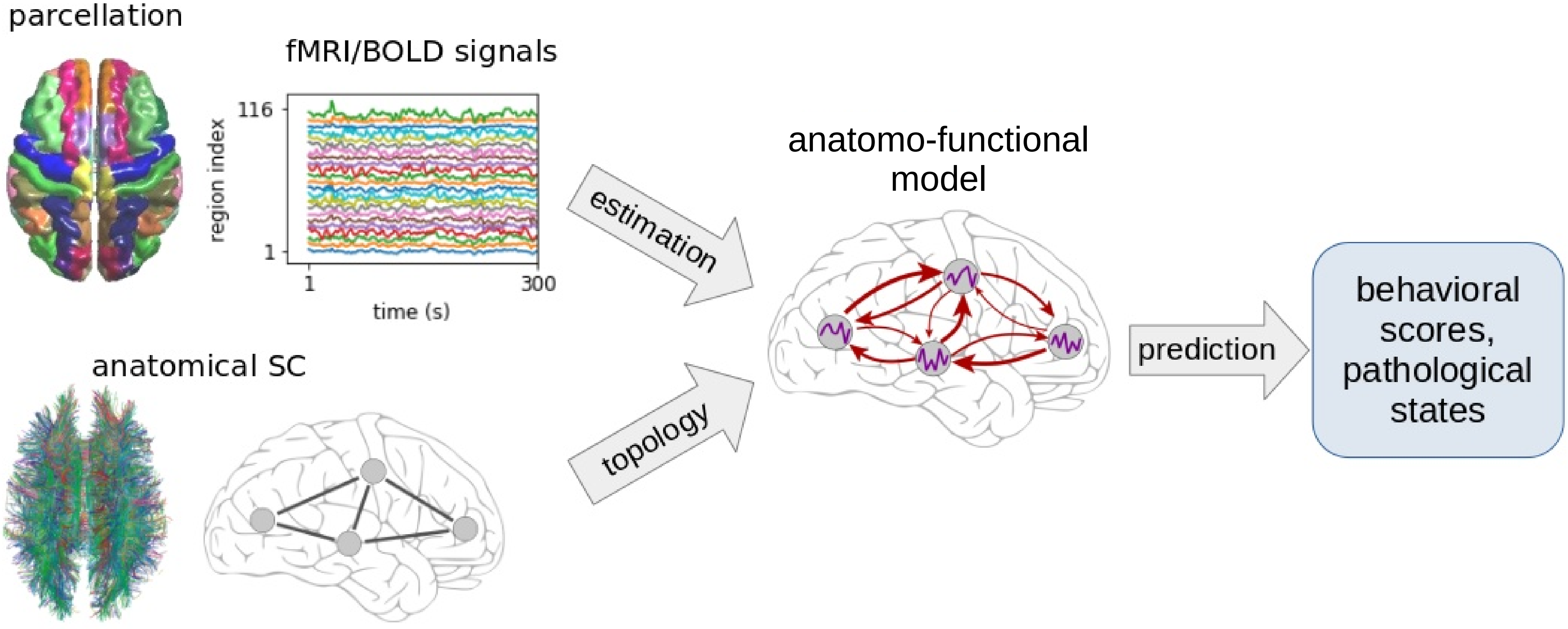
- The brain as a distributed and complex network system
- Neuroimaging: quantifying the brain
- Statistical analysis versus classification
- Example 1: diagnosis / prognosis in stroke
- **Example 2: whole-brain modeling, The Virtual Brain**
- Example 3: characterize structure in multivariate data
- Scikit-learn: formatting data

Model-Based Analysis of Neuroimaging Data



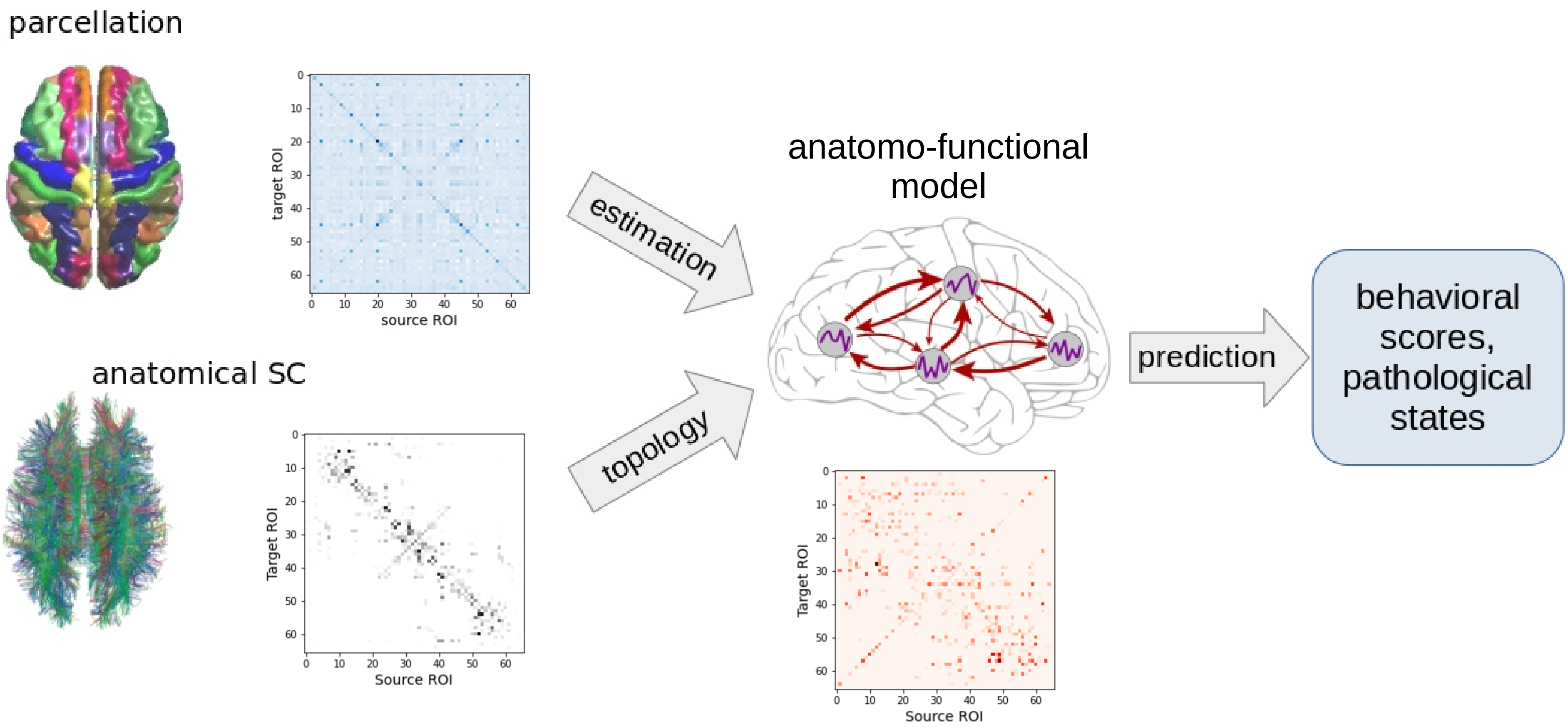
Gilson, ..., Deco *PLoS Comput Biol* 2016; Gilson et al. *Net Neurosci* 2020

Model-Based Analysis of Neuroimaging Data



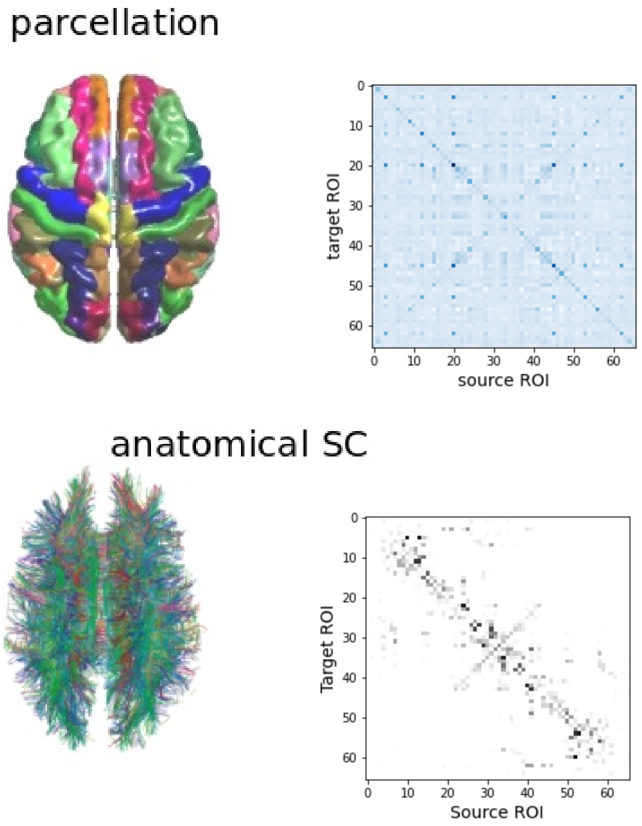
Gilson, ..., Deco *PLoS Comput Biol* 2016; Gilson et al. *Net Neurosci* 2020

Model-Based Analysis of Neuroimaging Data

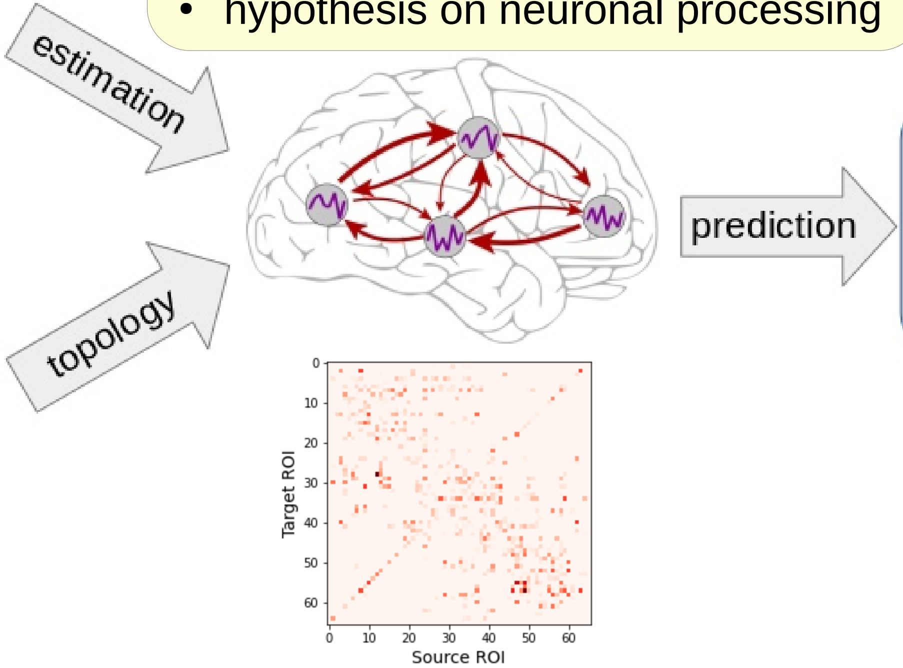


Gilson, ..., Deco *PLoS Comput Biol* 2016; Gilson et al. *Net Neurosci* 2020

Model-Based Analysis of Neuroimaging Data



- signature of brain dynamics (individual, task-specific)
- combine different types of data
- hypothesis on neuronal processing

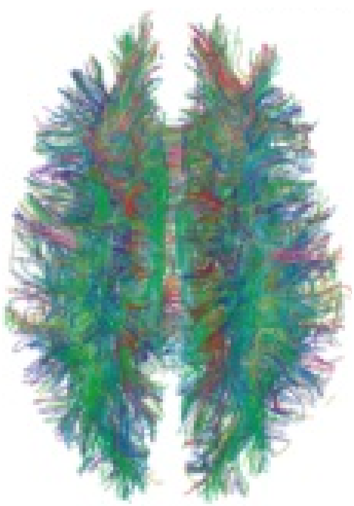


behavioral scores, pathological states

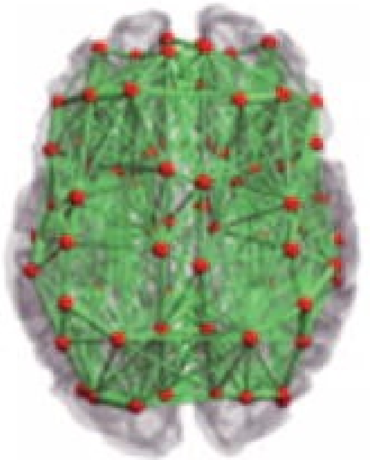
Gilson, ..., Deco *PLoS Comput Biol* 2016; Gilson et al. *Net Neurosci* 2020

Example of Biophysical Model: The Virtual Brain

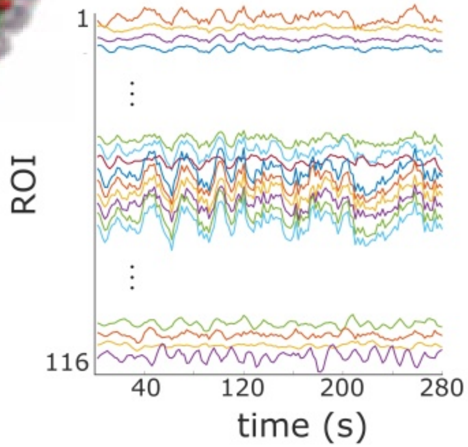
Anatomical connectivity



Dynamic model



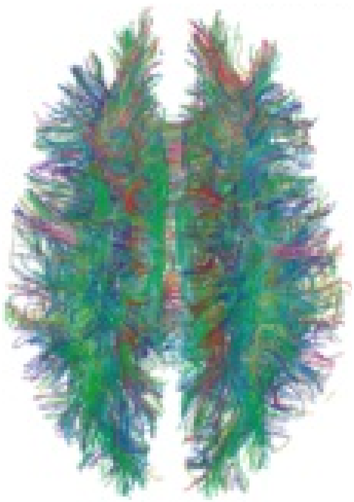
Measured activity
(e.g. fMRI, SEEG)



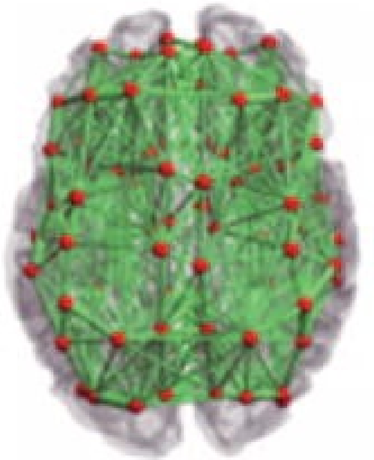
THEVIRTUALBRAIN.

Example of Biophysical Model: The Virtual Brain

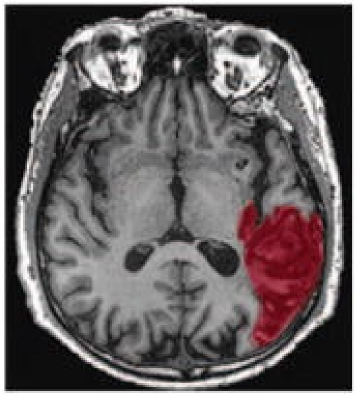
Anatomical connectivity



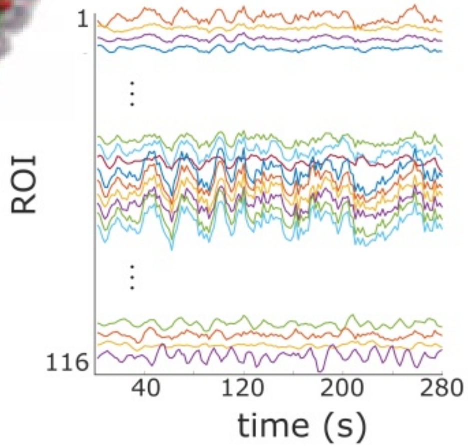
Dynamic model



brain lesions



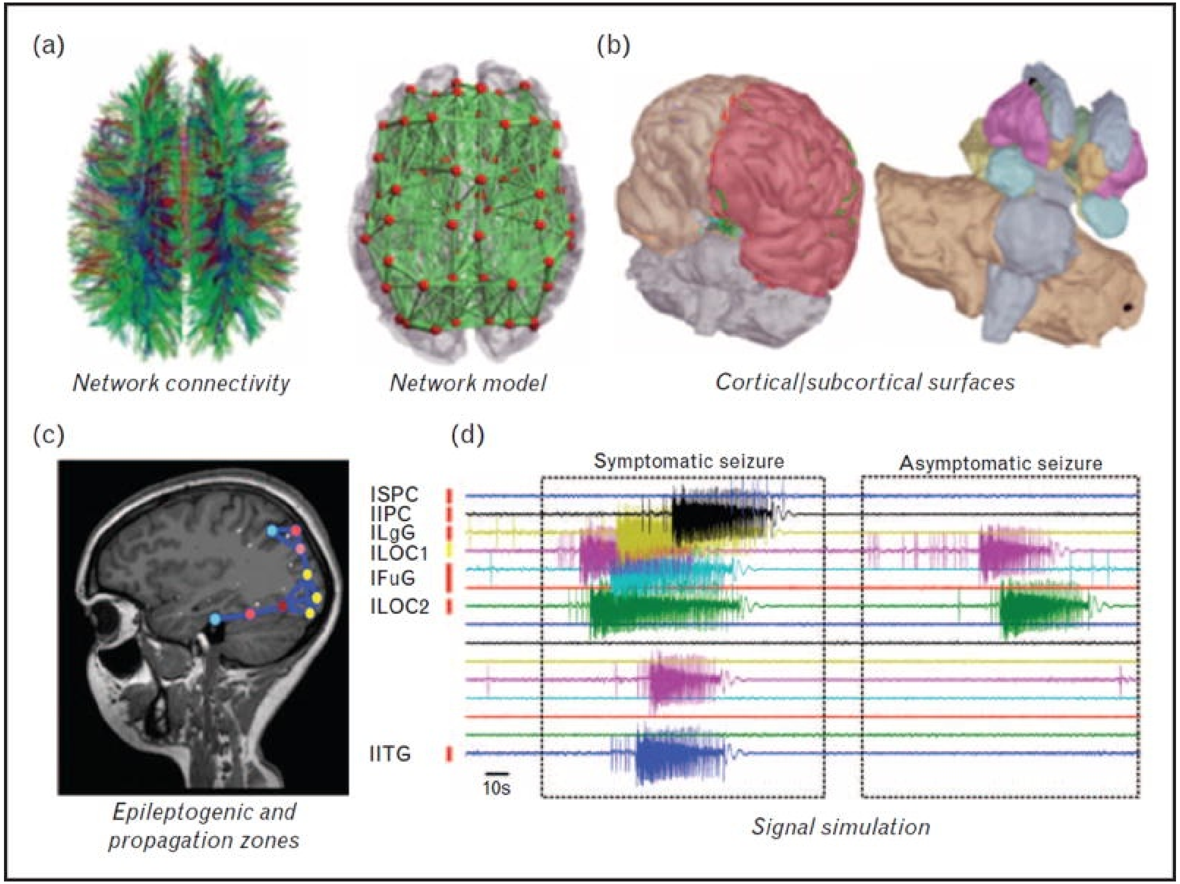
Measured activity
(e.g. fMRI, SEEG)



THEVIRTUALBRAIN.

Virtual Epileptic Patient

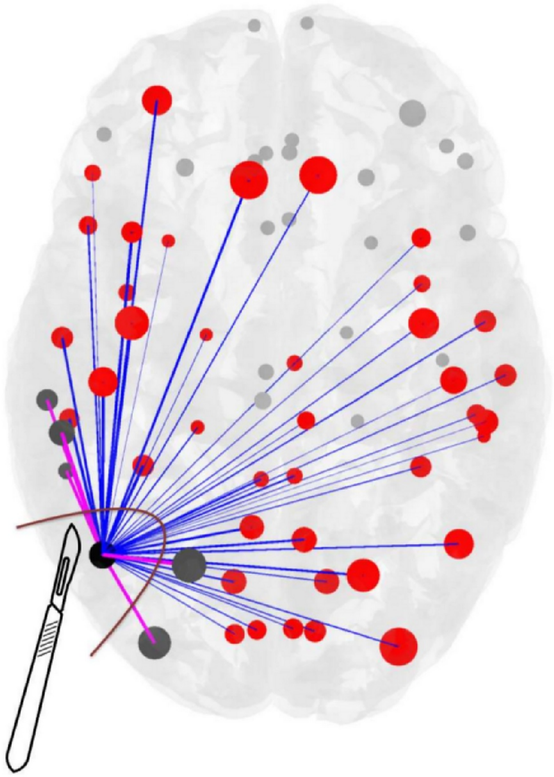
- Characterize propagation of pathological activity during seizure
- Influence of anatomical connectivity
- Inferred local excitability
- Choice of model for neuronal population for each region

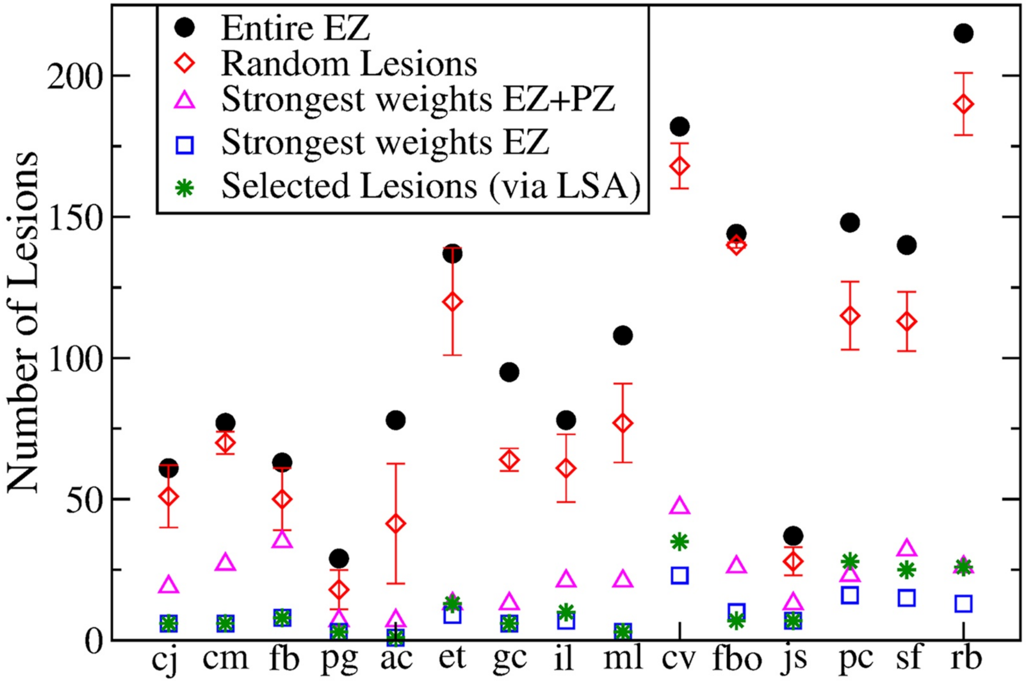


Falcon, *Curr Opin Biol* 2016; Proix, Bartolomei, Guye, Jirsa, *Brain* 2017

Virtual Epileptic Patient

- Choose best surgery strategy (resection) to free patients from seizures
- Inform resection scheme by modeling



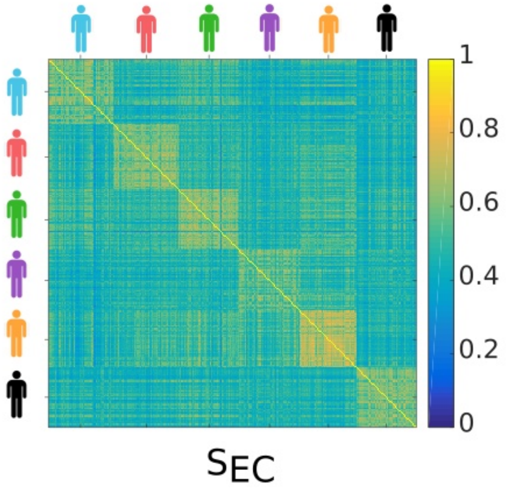


Introduction to Models in Neuroscience

- The brain as a distributed and complex network system
- Neuroimaging: quantifying the brain
- Statistical analysis versus classification
- Example 1: diagnosis / prognosis in stroke
- Example 2: whole-brain modeling, The Virtual Brain
- **Example 3: characterize structure in multivariate data**
- Scikit-learn: formatting data

Types of Variabilities in FC Data

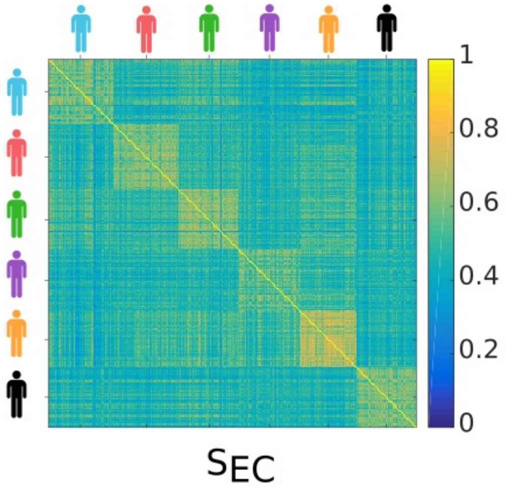
Similarity across
resting-state sessions



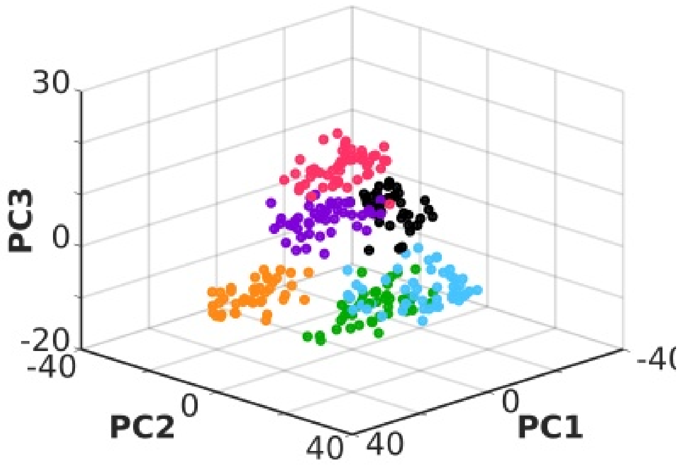
Pallarés, …, Gilson (2018) *Neuroimage*

Types of Variabilities in FC Data

Similarity across resting-state sessions



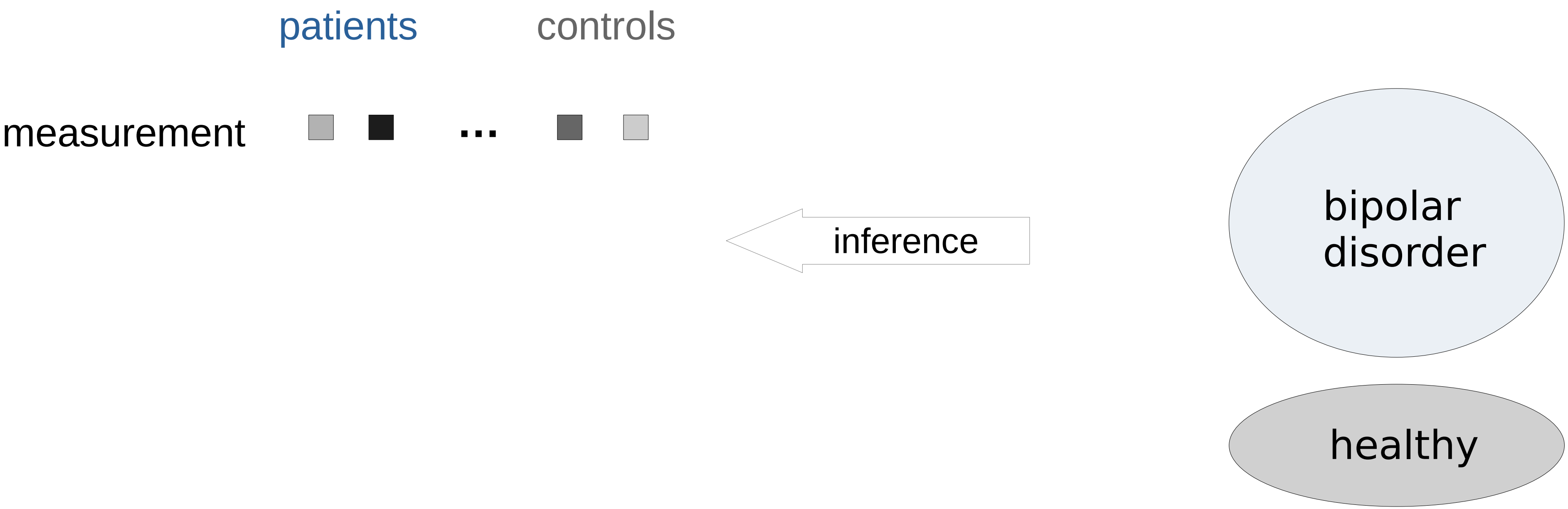
PCA representation (1 dot = 1 session)



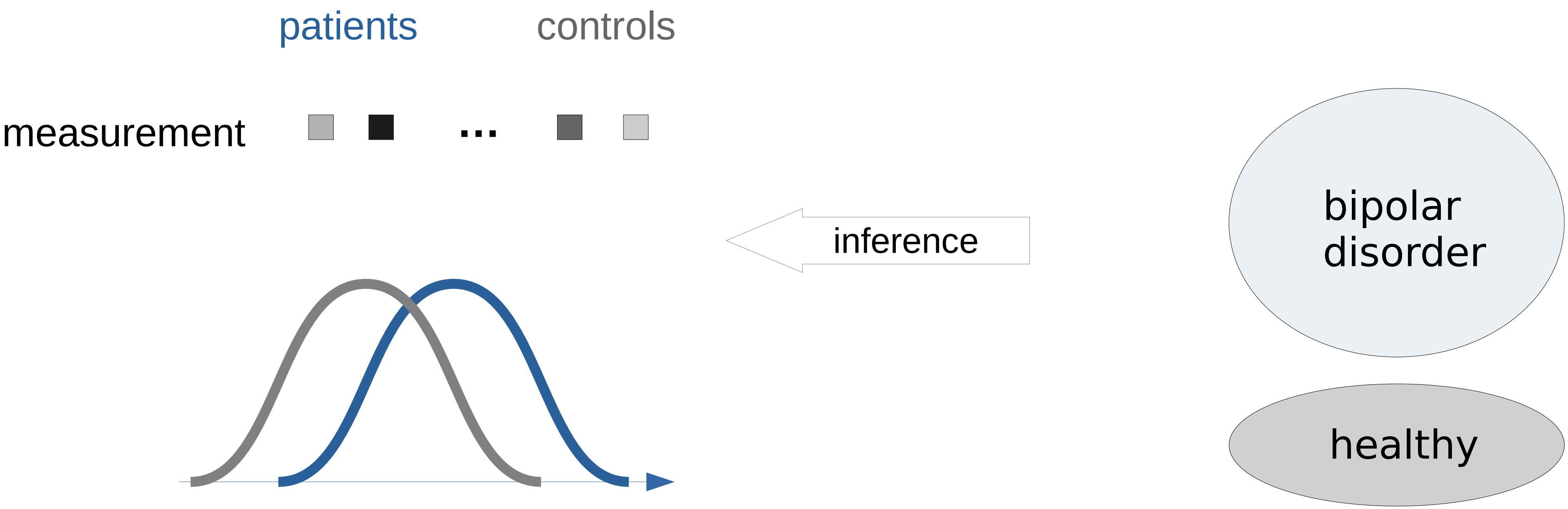
- Strong bias related to subject identity
- Signature for conditions needs to generalize across subjects
- Ignore session-to-session variability

Pallarés, …, Gilson (2018) *Neuroimage*

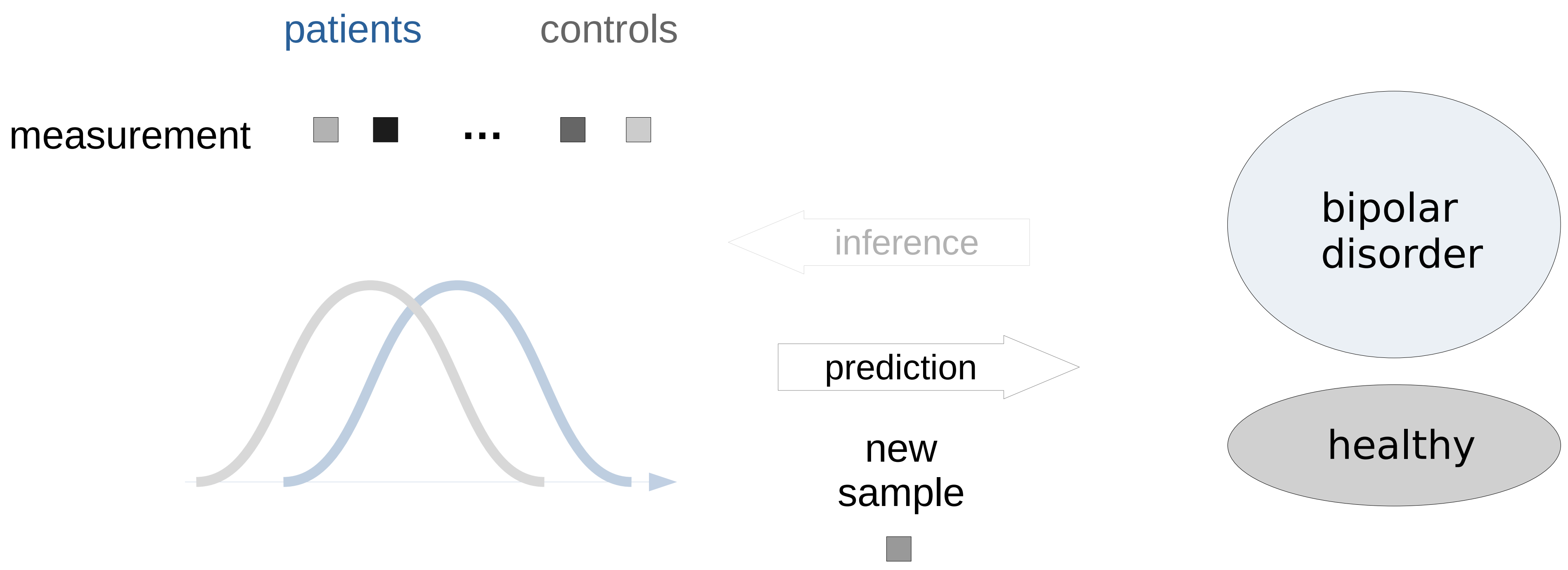
Statistical Analysis versus Machine Learning



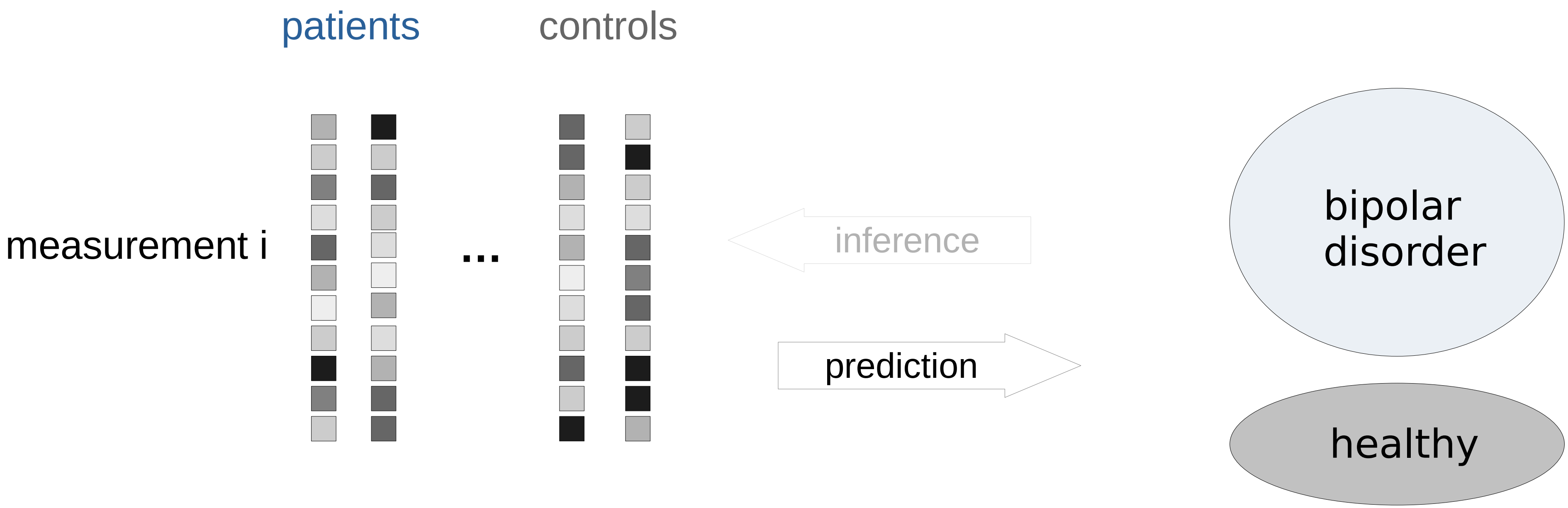
Statistical Analysis versus Machine Learning



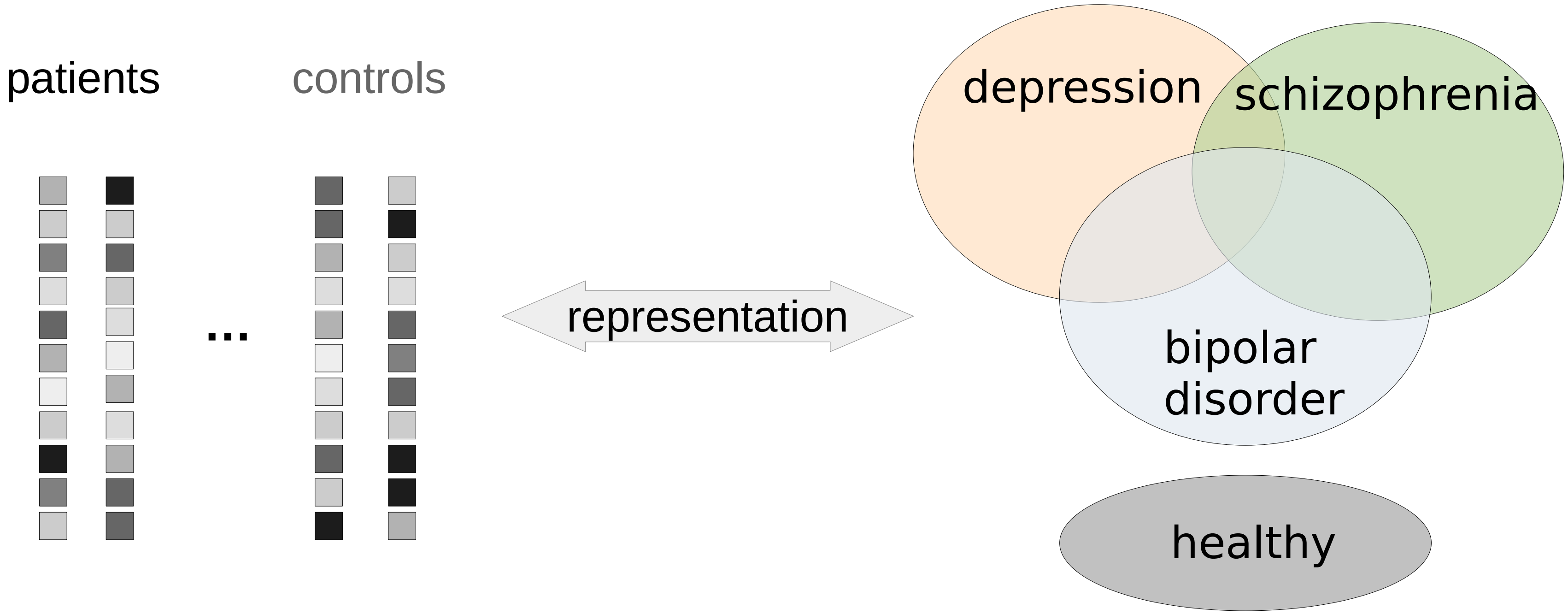
Statistical Analysis versus Machine Learning



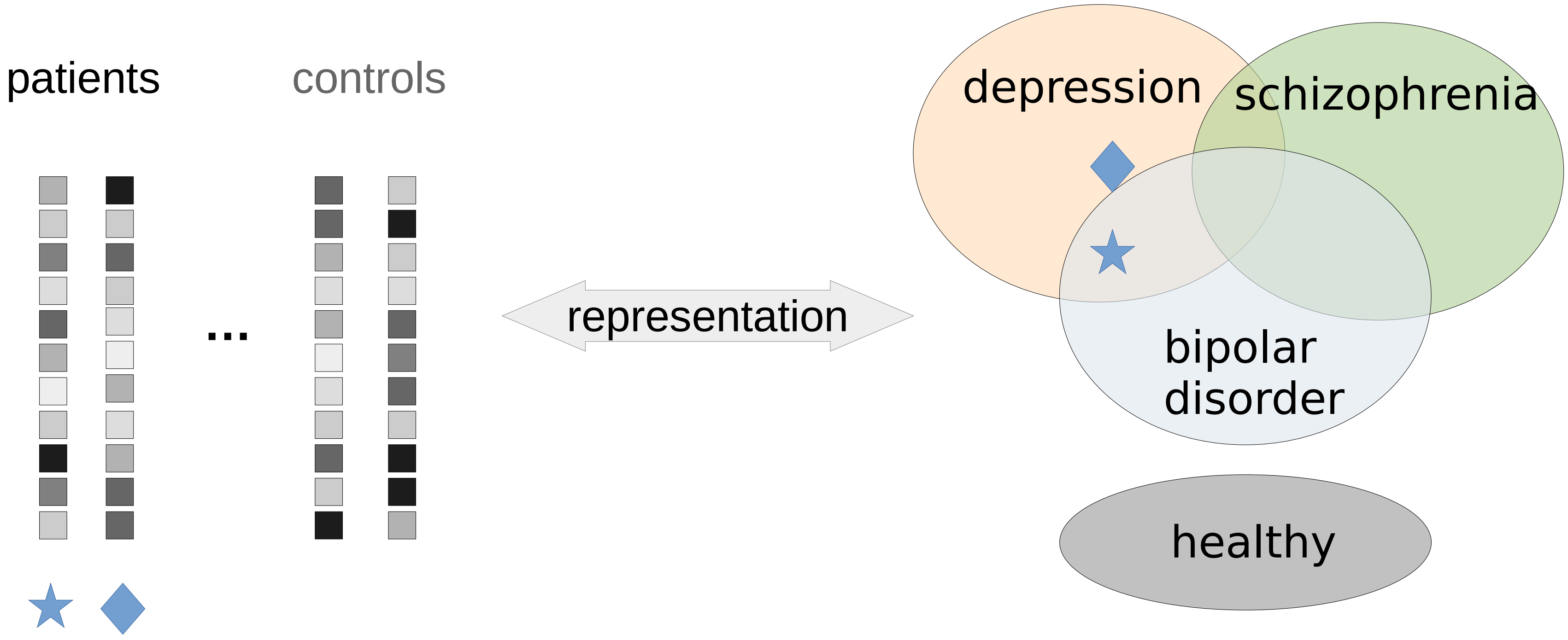
Statistical Analysis versus Machine Learning



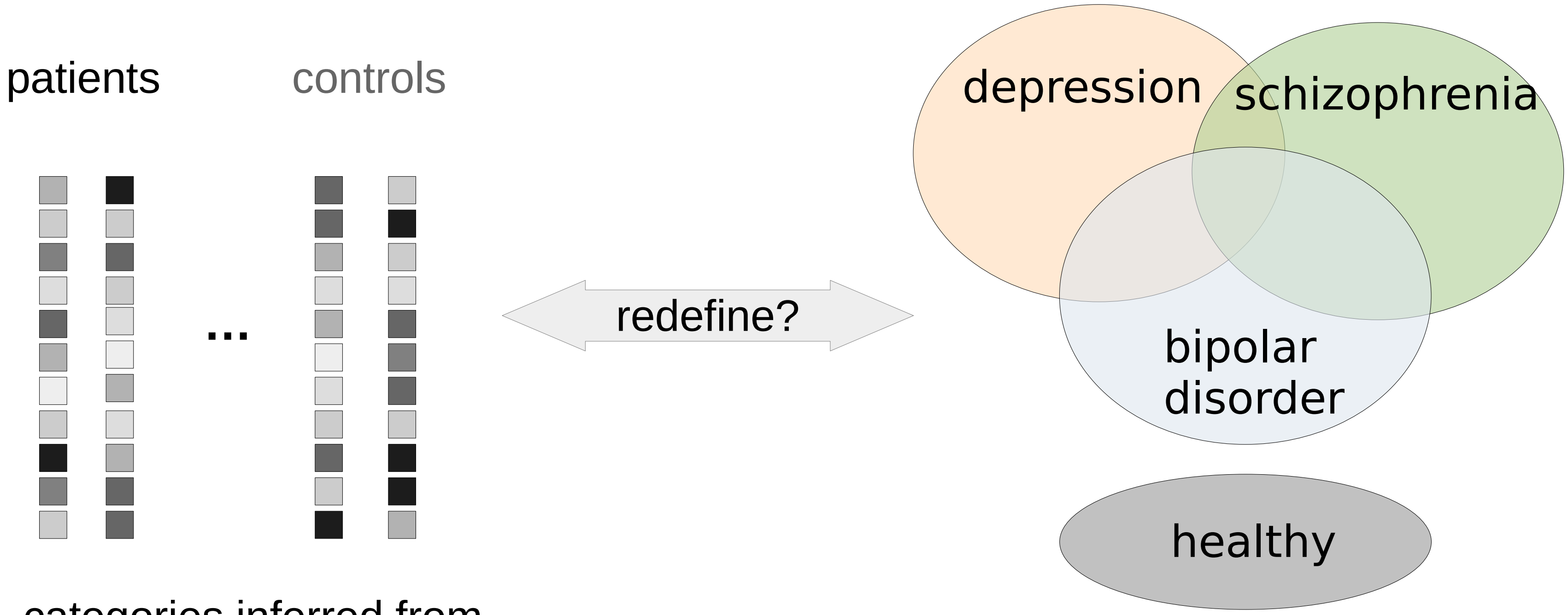
Statistical Analysis versus Machine Learning



Statistical Analysis versus Machine Learning



Statistical Analysis versus Machine Learning



- categories inferred from neuroimaging data
- multimodal data fusion

- How to defined normality in multivariate data?
  - not like blood samples...
  - blue / brown eyes

- Practical question for clinicians: pathology subtypes?
- Unsupervised learning (clustering), but less “powerful” or conclusive than supervised learning

Introduction to Models in Neuroscience

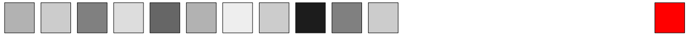
- The brain as a distributed and complex network system
- Neuroimaging: quantifying the brain
- Statistical analysis versus classification
- Example 1: diagnosis / prognosis in stroke
- Example 2: whole-brain modeling, The Virtual Brain
- Example 3: characterize structure in multivariate data
- **Scikit-learn: formatting data**

Organization of Data for Classification

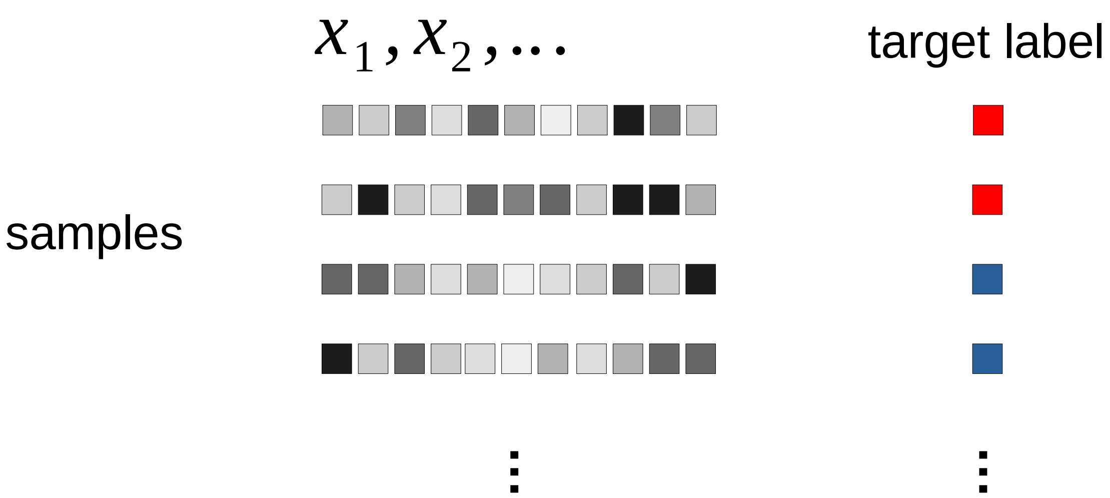
$x_1, x_2, \ldots$

target label

sample i

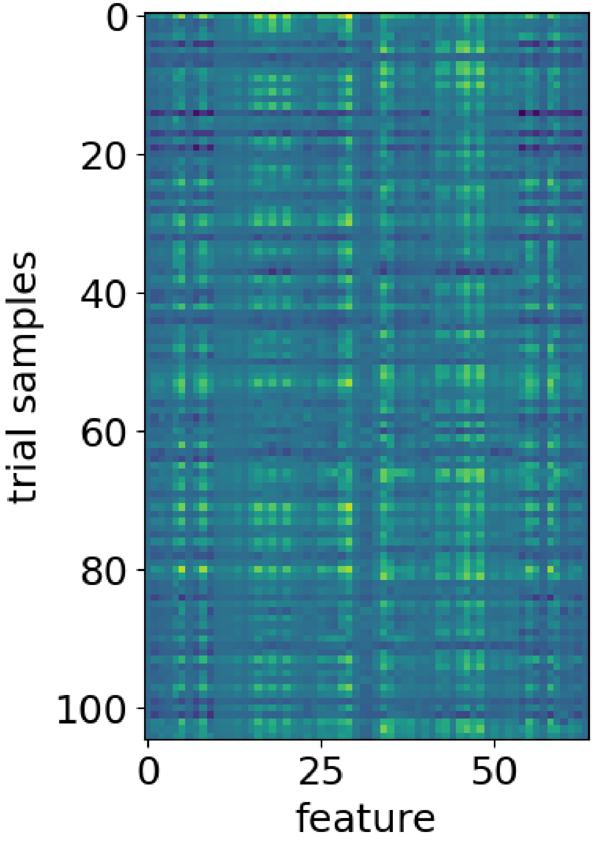


Organization of Data for Classification

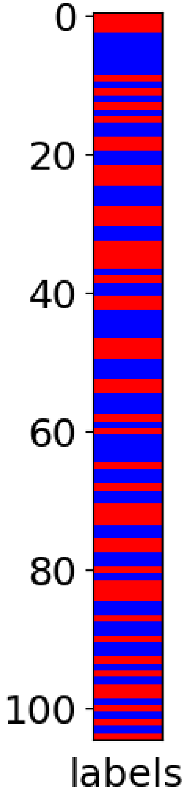


Organization of Data for Classification

matrix X



vector y



Classifier model:

$$y = f(X)$$

Supervised learning, unsupervised learning, etc.

Classification:	$f(X) = y$	labels
Regression:	$f(X) = y$	explained variables
Clustering:	$f(X) = y$	labels
Dimensionality reduction:	$f(X) = y$	new coordinates

Supervised learning, unsupervised learning, etc.

Classification: $f(X) = y$ labels

Regression: $f(X) = y$ explained variables

Clustering: $f(X) = y$ labels

Dimensionality reduction: $f(X) = y$ new coordinates

Optimize *f* for (X,y)

Supervised learning, unsupervised learning, etc.

fit	train the using data (X,y), sets the transformation (PCA with a number of components), etc.
transform	maps the X to y (prediction for classifier, new coordinates for PCA, etc.)
fit_transform	
score	evaluation (classification accuracy, clustering quality, etc.)

https://scikit-learn.org

- simple and efficient tools for predictive data analysis
- accessible to everybody, and reusable in various contexts
- built on NumPy, SciPy, and matplotlib
- open source, commercially usable - BSD license

classification MLR, SVM	**regression**	**clustering** k-means
dim reduction PCA, ICA	**model selection** CV, grid search	**preprocessing**

Practice

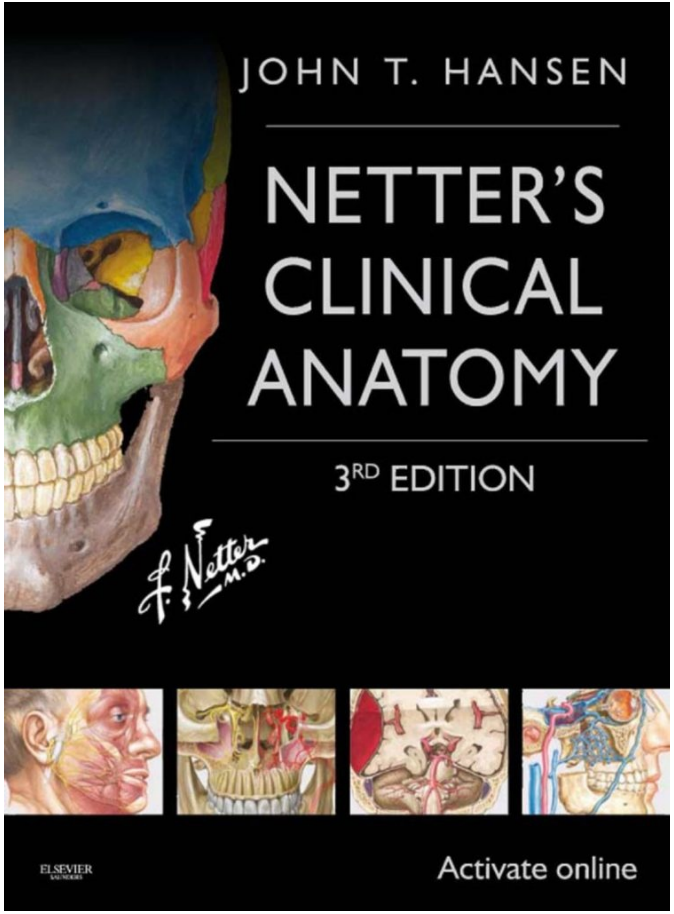
- Anaconda installation
- Environment yml file: required libraries
- Use jupyter-notebook (or jupyter-lab)

Introduction to Models in Neuroscience

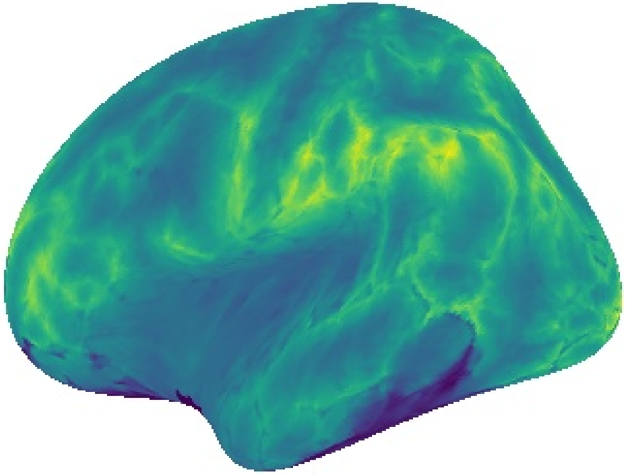
- Extra stuff...

Next-generation Database in Neuroimaging

electronic version (pdf) = 500 Mo

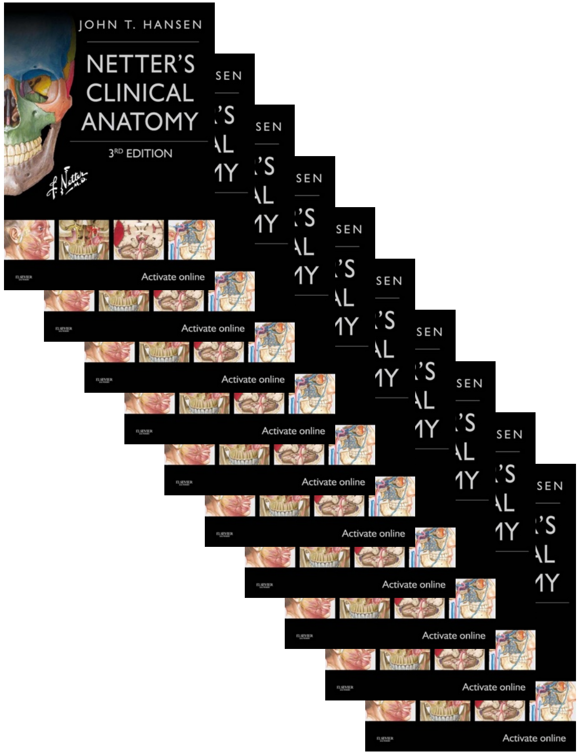


1h fMRI = 5 Go

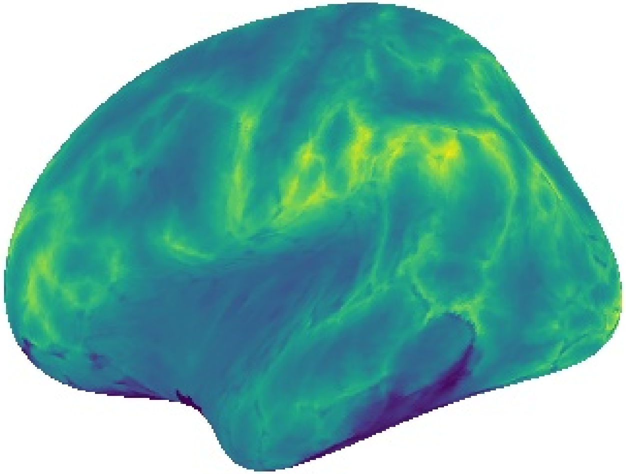


Next-generation Database in Neuroimaging

1h fMRI = 5 Go

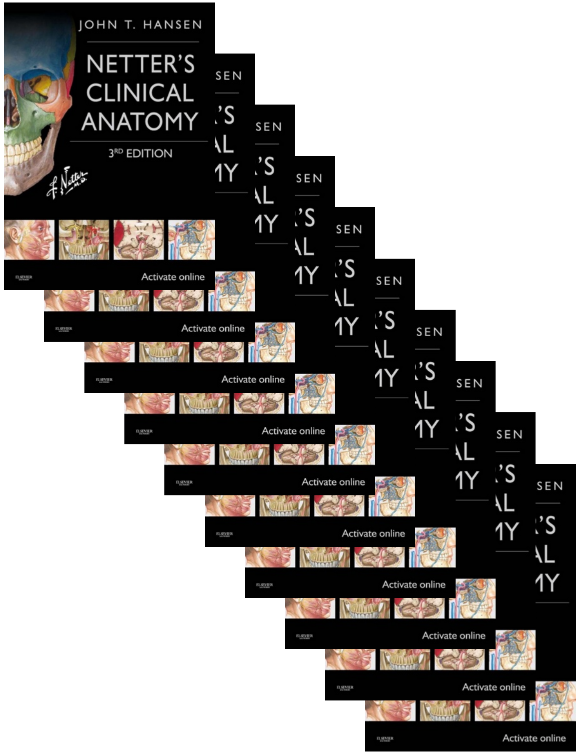


=

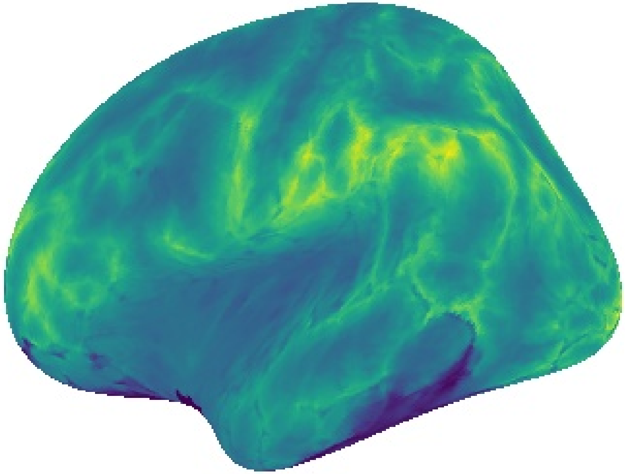


Next-generation Database in Neuroimaging

1h fMRI = 5 Go



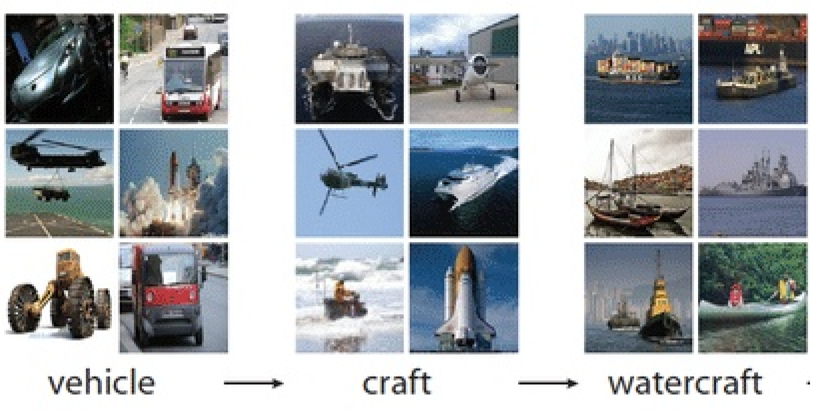
=



- Human Connectome Project : 1000 healthy controls, 2-3h MRI
- UK Biobank : 20000 subjects and patients

Next-generation Database in Neuroimaging

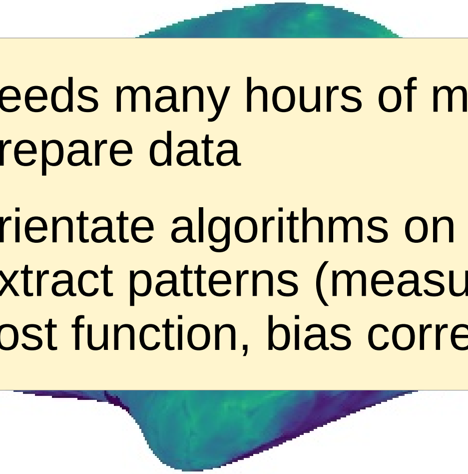
1 image = 5 Mo



- ImageNet : 14 000 000 annotated images
- ResNet : 200 000 000 weights trained

>>
<<

1h fMRI = 5 Go



- needs many hours of manual work to prepare data
- orientate algorithms on how to extract patterns (measure on data, cost function, bias correction)

- Human Connectome Project : 1000 healthy controls, 2-3h MRI
- UK Biobank : 20000 subjects and patients