Big Data for Small Brain

Prof. Yike Guo
Department of Computing
Imperial College London

Science was observational

- Galileo, Newton and the birth of modern science: c. 1600
- Problem: single "particle" (apple) in gravitational field (General two-body problem already too hard)
- Methods
  - Data: notebooks (Kbytes)
  - Theory: driven by data
  - Computation: calculus by hand (1 Flop/s)
- Collaboration
  - 1 brilliant scientist, 1-2 students

Science is still observational …but with different scale

Genome Sequencing: Understanding the life functions at the system level through molecular profiling and relating the molecular information with phenotypic data. Data generated by high throughput device (e.g. NGS machine) : 1TB/day for one machine , 1 Lab: 20-100 machines, global collaboration

Connectome: Mapping the connectome at the micrometer resolution means building a complete map of the neural systems, neuron-by-neuron. The human cerebral cortex alone contains on the order of 10^{10} neurons linked by 10^{14}synaptic connections. By comparison, the number of base-pairs in a human genome is 3×10^{9}. In 2012, a Citizen science project called EyeWire began attempting to crowdsourcing the mapping of the connectome through an interactive game

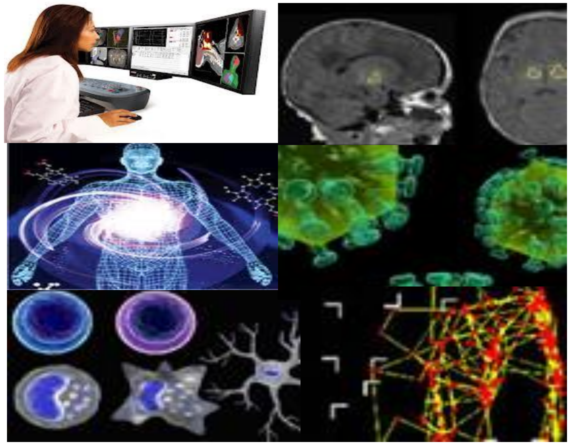
Datafication: Data Science as the Glue for Multidisciplinary Research

Medical

System oncology

real-time metabolic profiling

Cardio-vascular Science



Imaging

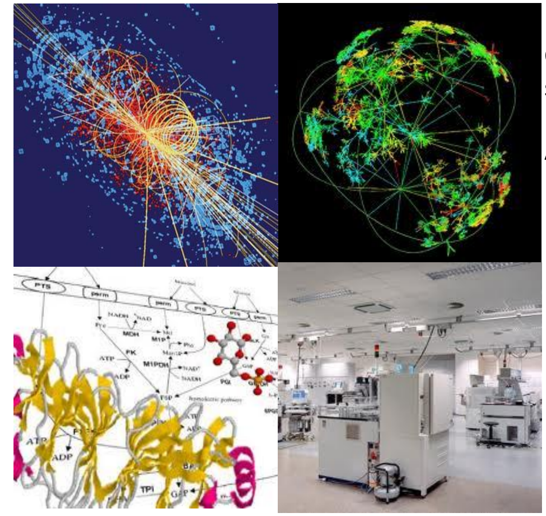
Infection/Epidemiology

Neurosciene

Natural Science

Particle Physics

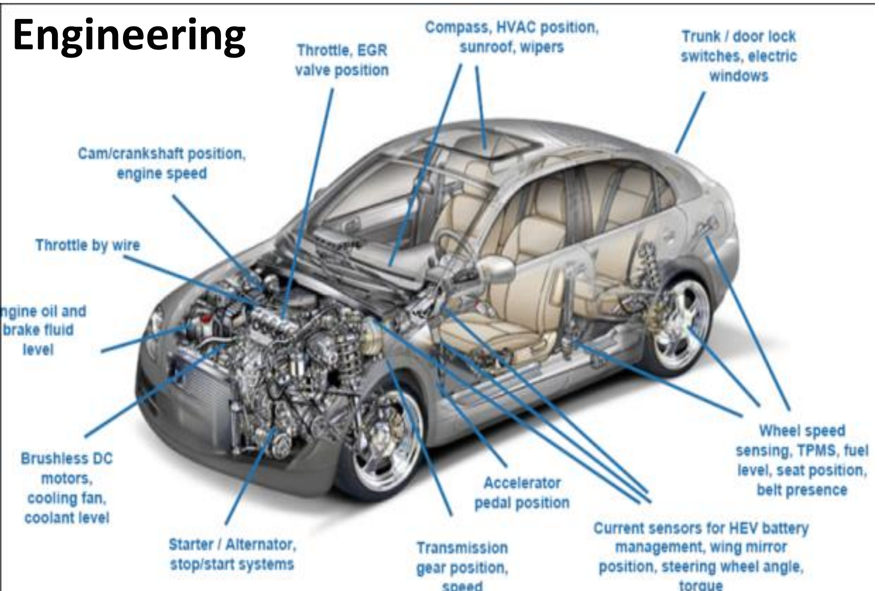
System Biology



Complex system & Network Data Analytics

High throughput screening

Engineering

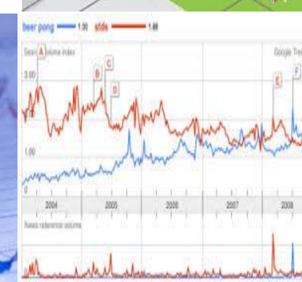


Business

Social Media and New Data Business

Digital City and Urban Life

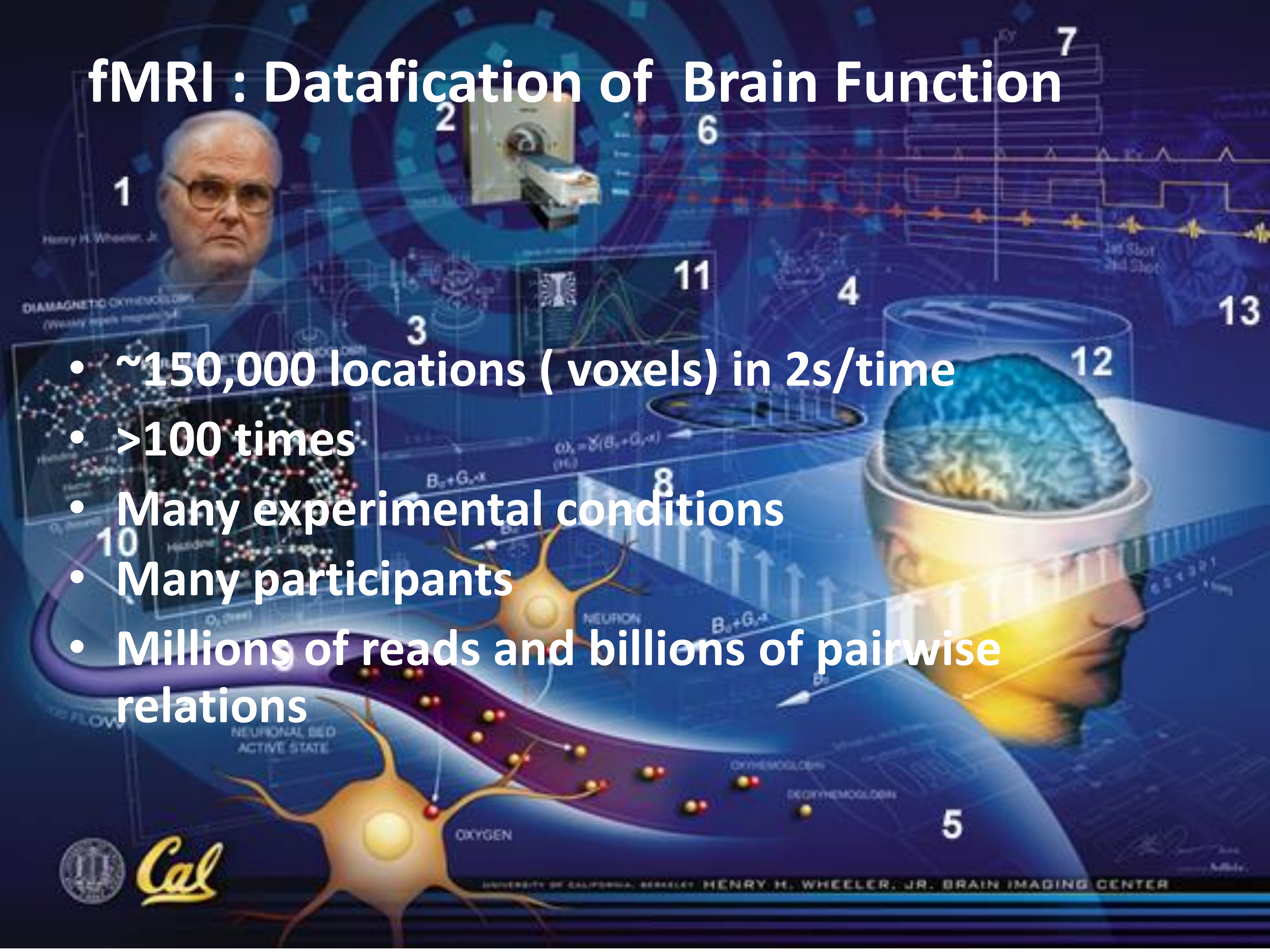
Algorithmic Trading



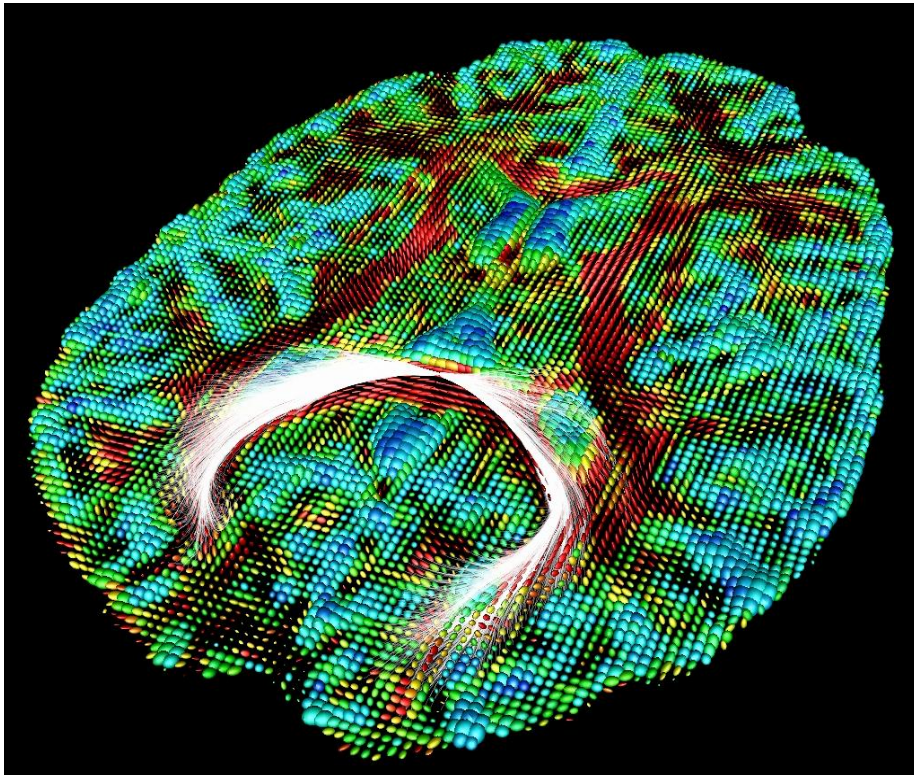
Public Health management

fMRI : Datafication of Brain Function

- ~150,000 locations (voxels) in 2s/time
- >100 times
- Many experimental conditions
- Many participants
- Millions of reads and billions of pairwise relations



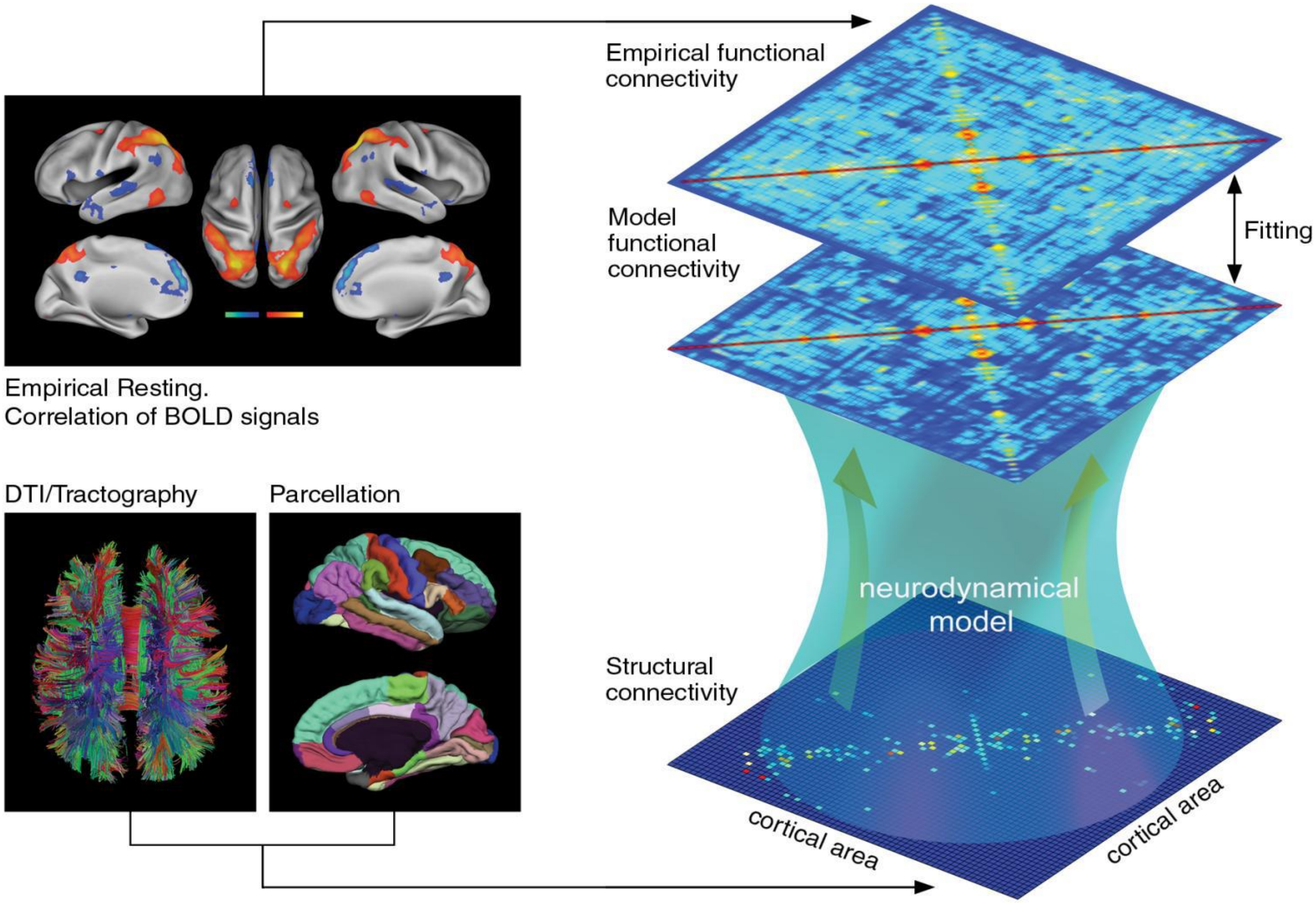
Brain and Web : Small and Big



Neuron: 100 billions
Synapse: 300 trillions – 1000 trillions
Change : 700—1000 new synapses/s

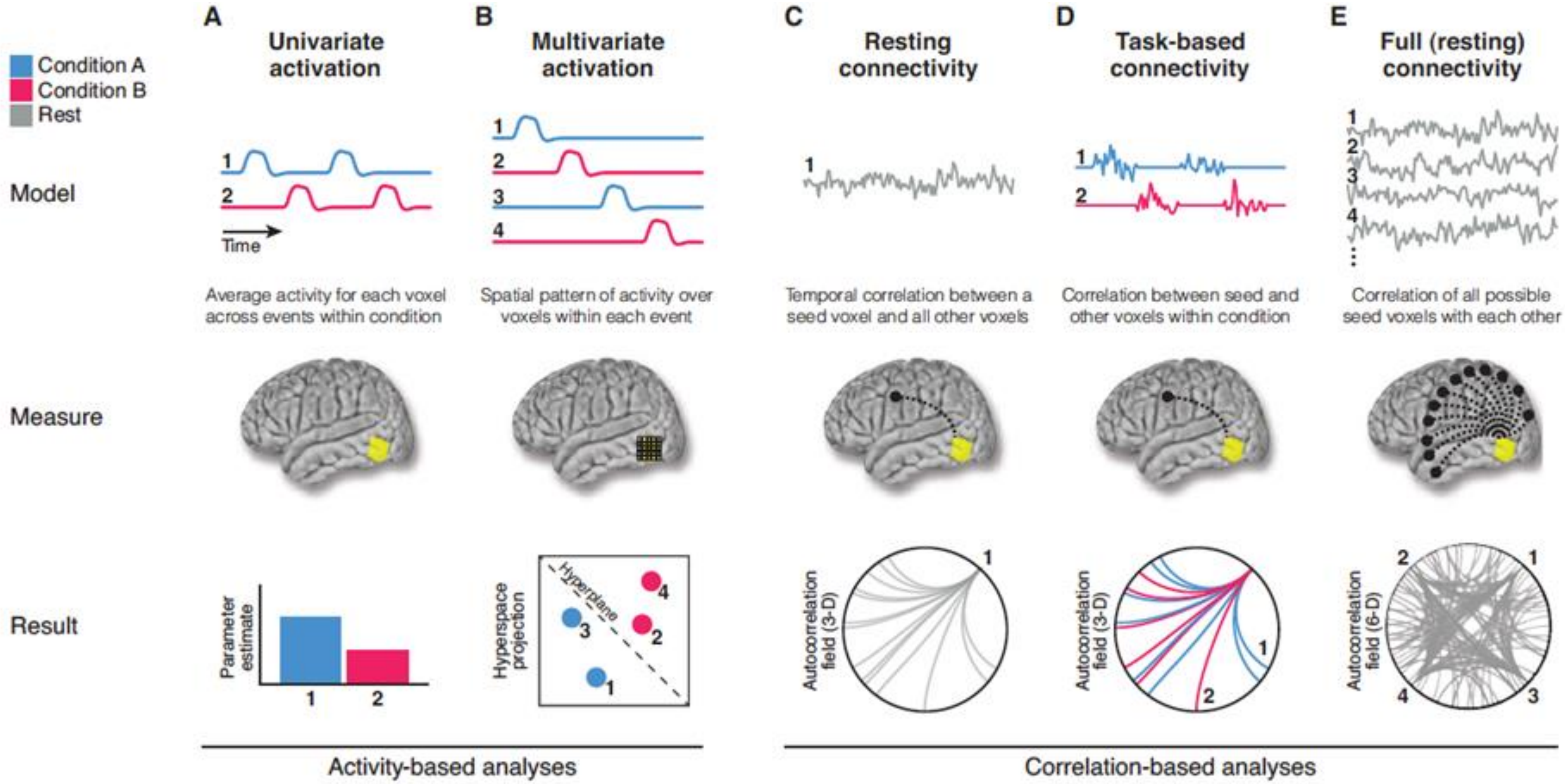
Web : 1 trillions (14 billon pages)
Link: 100 trillions
Change : 8 new website /s

Data Driven Brain Research



Deco, Gustavo, et al. "Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations." *The Journal of Neuroscience* 33.27 (2013): 11239-11252.

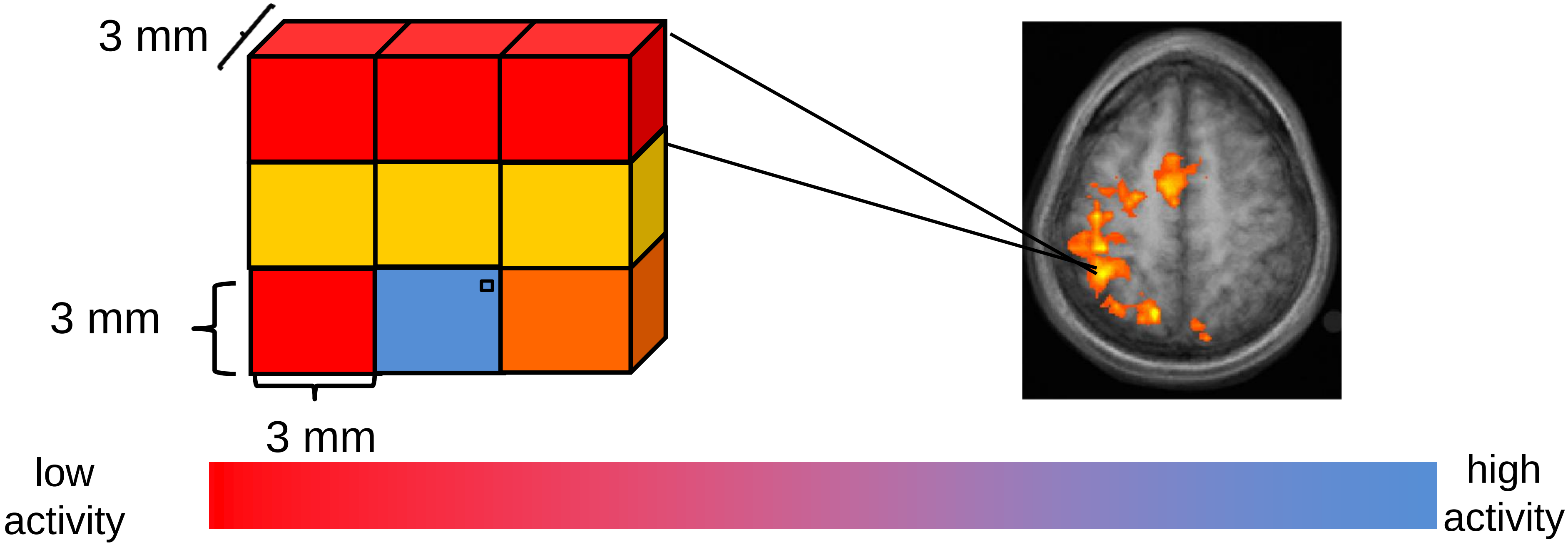
fMRI Analysis

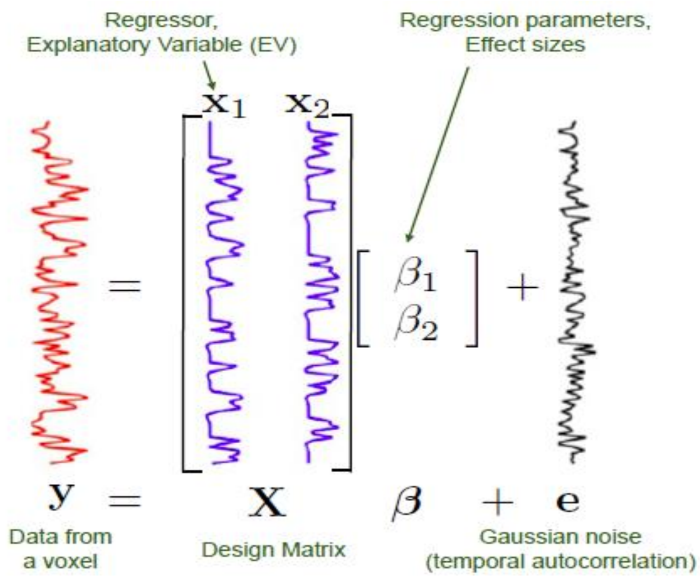


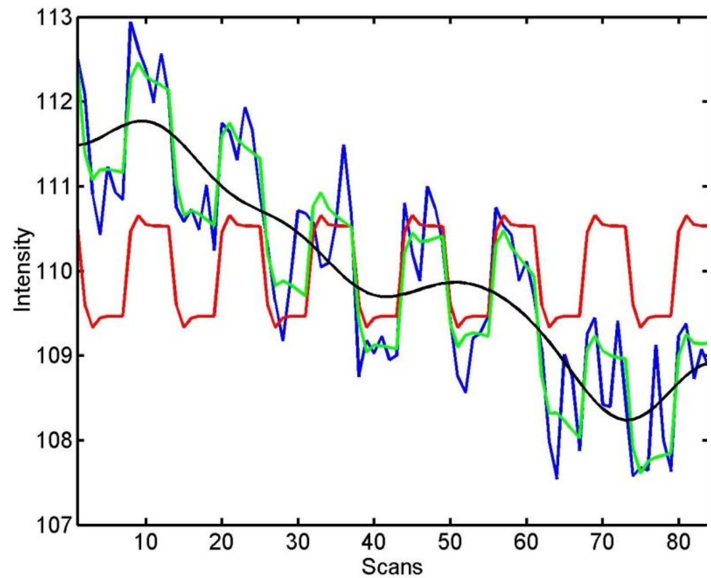
Small Data → **Big Data**

Turk-Browne, Nicholas B. "Functional Interactions as Big Data in the Human Brain." *science* 342.6158 (2013): 580-584.

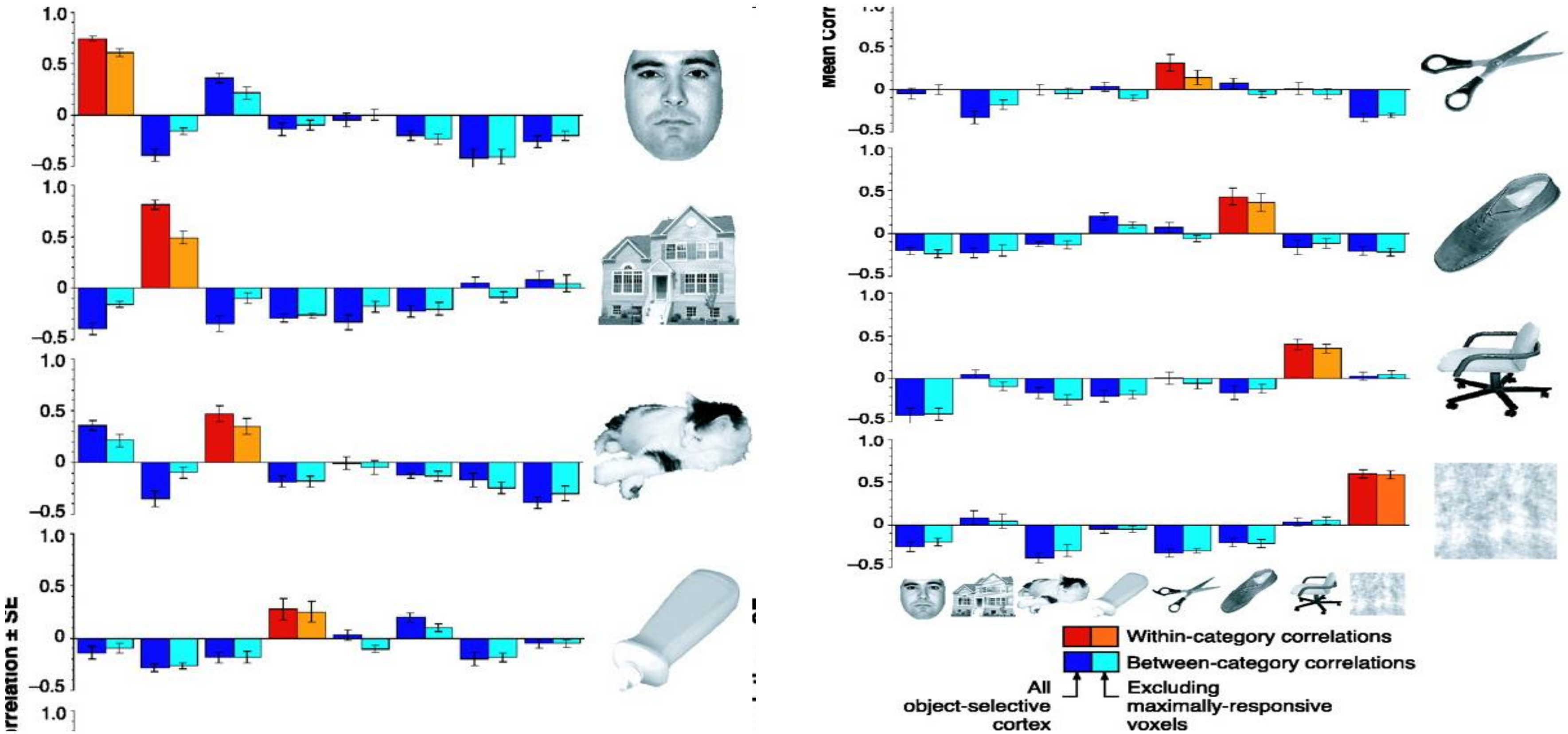
Single Voxel : Small Data Analysis



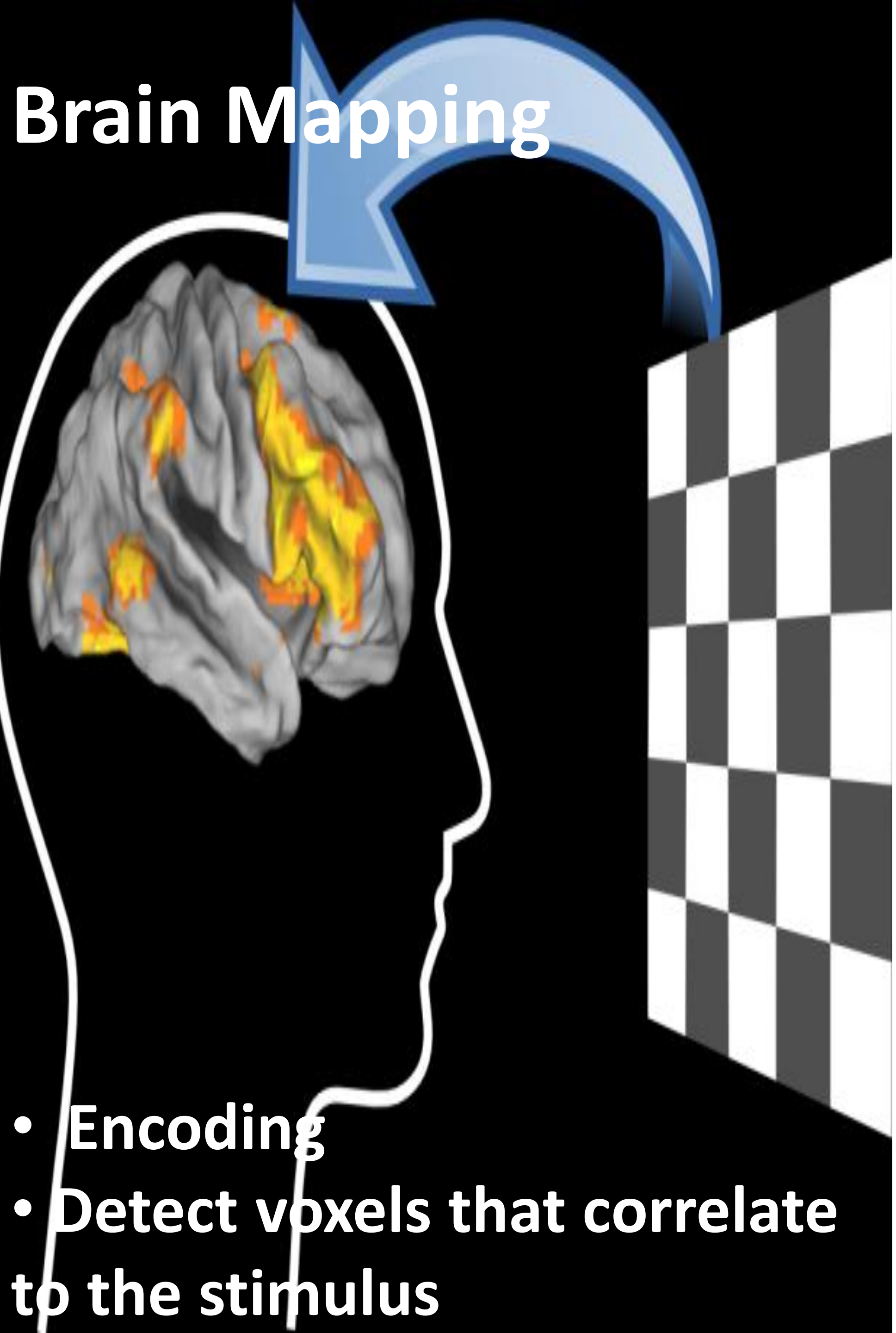




MVPA: Big data approach for brain analysis

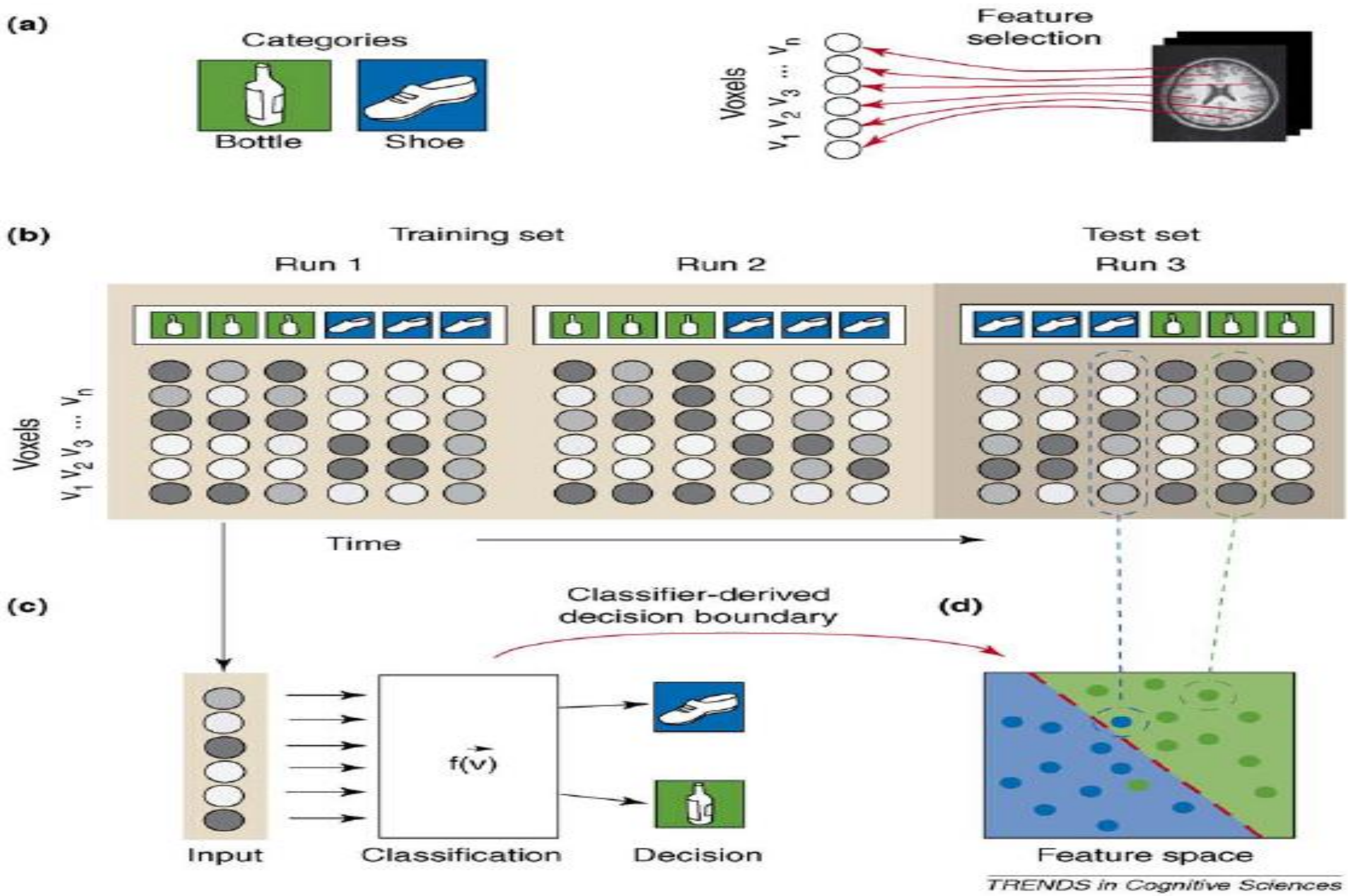


Haxby, James V., et al. "Distributed and overlapping representations of faces and objects in ventral temporal cortex." *Science* 293.5539 (2001): 2425-2430.


Brain Mapping
• Encoding
• Detect voxels that correlate to the stimulus


Brain Reading
•Decoding
•Find multiple voxels to decode(predict) the stimulus

MVPA Algorithm

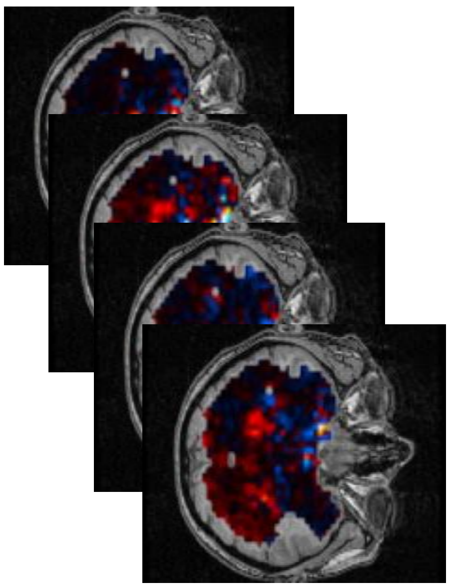


Norman, Kenneth A., et al. "Beyond mind-reading: multi-voxel pattern analysis of fMRI data." *Trends in cognitive sciences* 10.9 (2006): 424-430.

Linear Sparse Model for MVPA

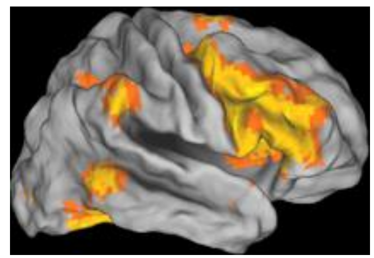
Design Matrix x

N samples



×

Coefficients w
(weights of voxels)



=

Stimuli(Observations) y

Relevant patterns = K non-zeros coefficients $\ll$ the number of brain features M

Why use Linear Sparse Model?

The number of patterns related to the stimuli is always far less than the number of brain features

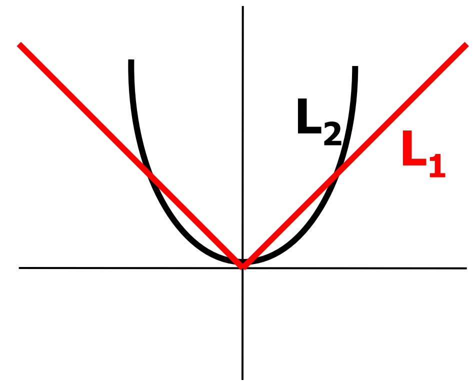
The prediction model constructed by the sparse coefficients does not easy to overfit the training data

Lasso for Feature Selection

Class label | Input features | Feature strength

=

x

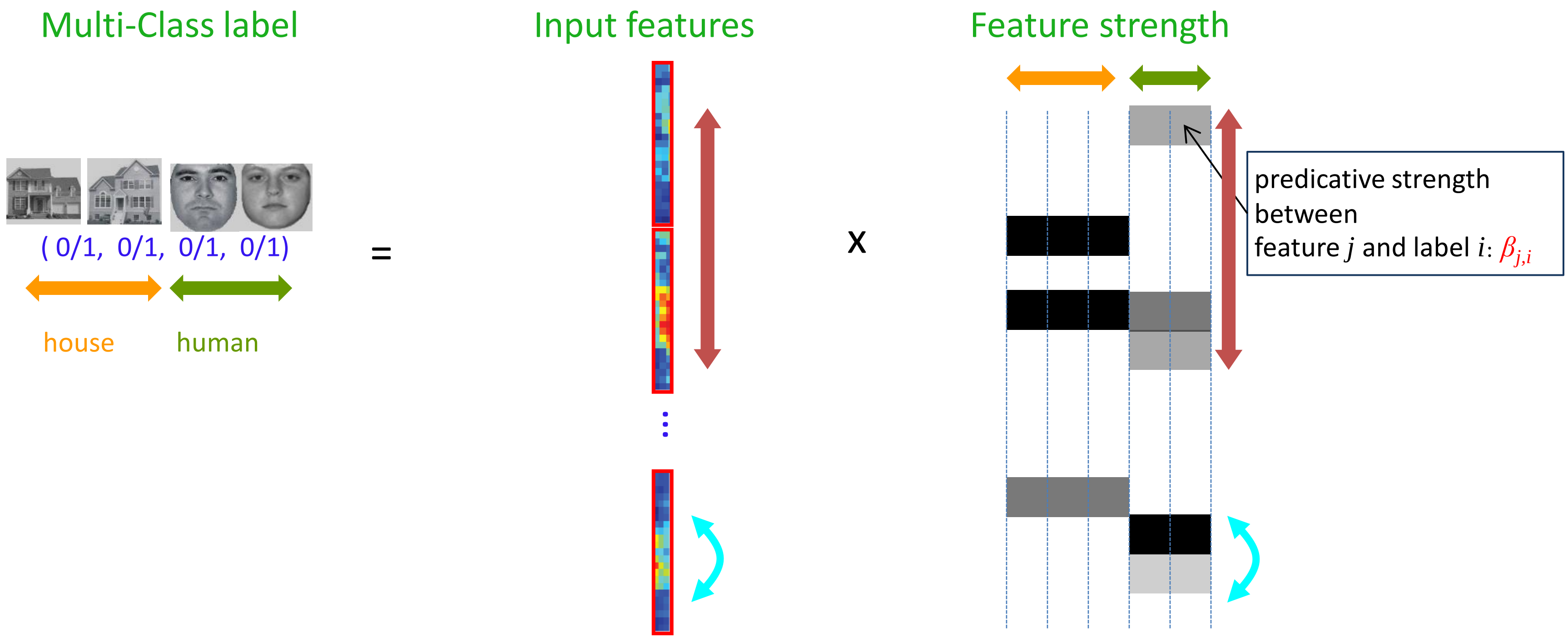


Lasso Penalty for sparsity

$$\beta^* = \arg\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) + \lambda \sum_{j=1}^{J} |\beta_j|$$

Many zero strengths (sparse results), but what if the features are correlated?

Multi-task Feature Selection

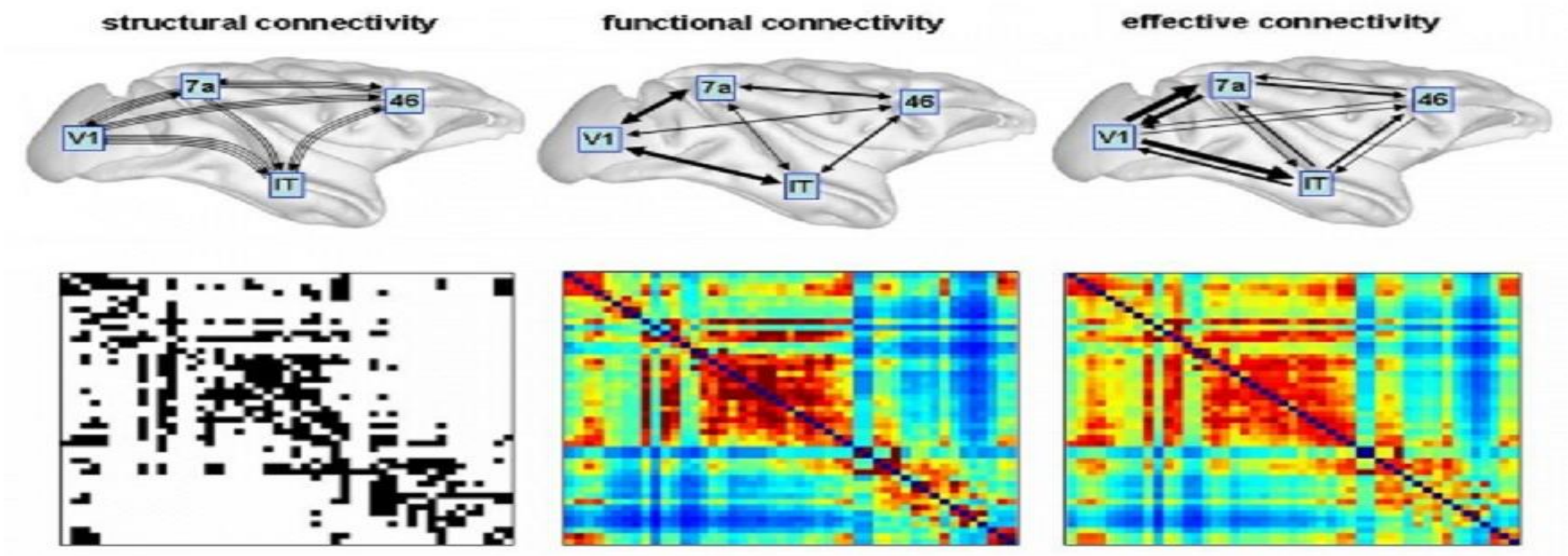


$$\beta^* = \arg\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) + \lambda \sum_{j=1}^{J} |\beta_j|$$

\+ We introduce **Structured L1/L2 norm**

$$\sum_k \sqrt{\sum_{(j,i)\in\mathcal{G}_k}^{J} (\beta_{i,j})^2}$$

Connectivity



Anatomical/structural connectivity presence of axonal connections

Functional connectivity statistical dependencies between regional time series

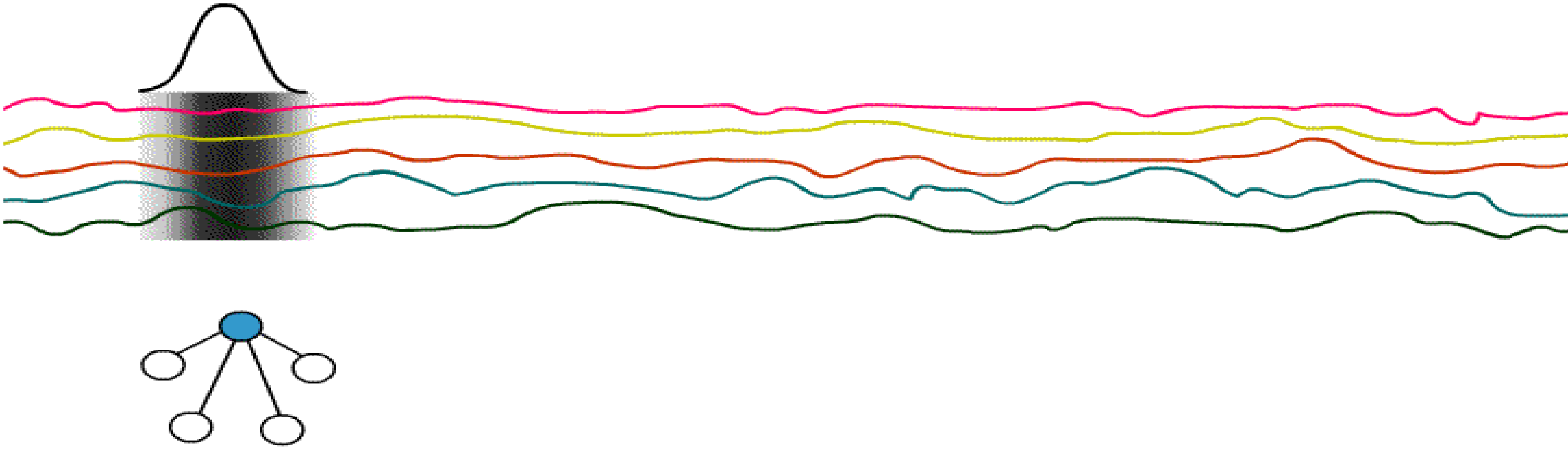
(Descriptive in nature; establishing whether correlation between areas is significant

Effective connectivity causal/directed influences between neurons or populations

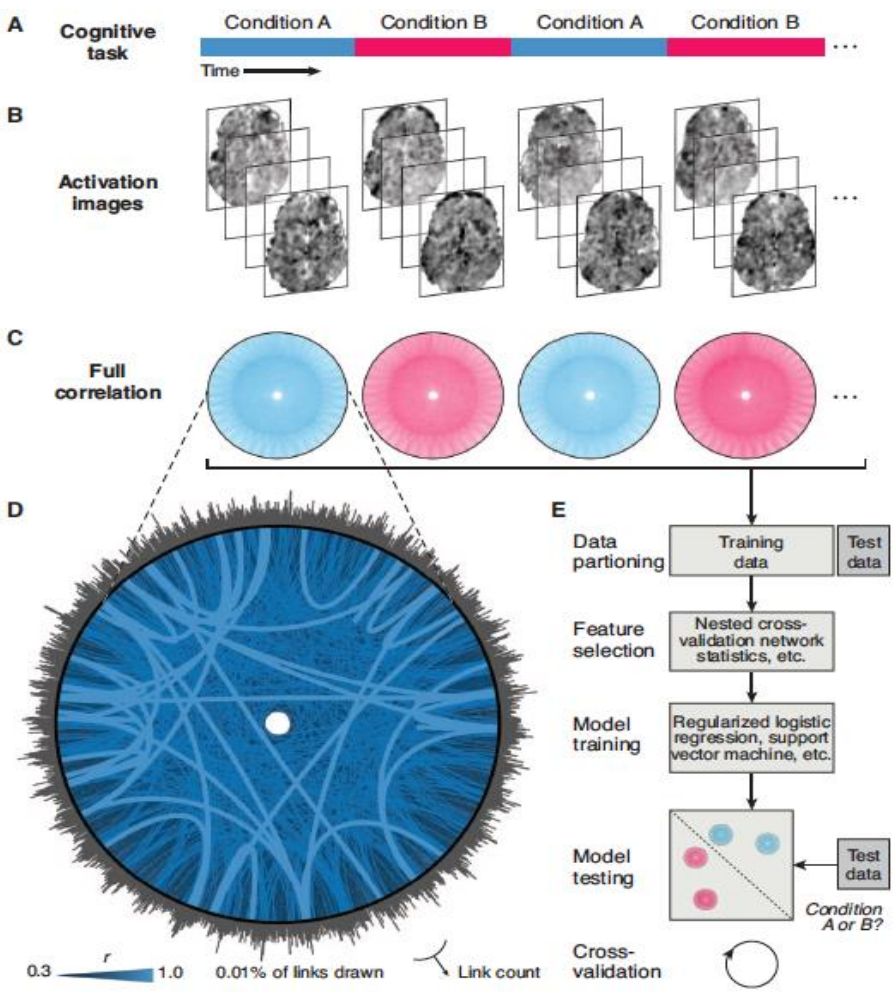
(Model-based; analysed through model comparison or optimisation)

Sporns, Olaf. "Brain connectivity." *Scholarpedia* 2.10 (2007): 4695.

Learning Functional Connectivity



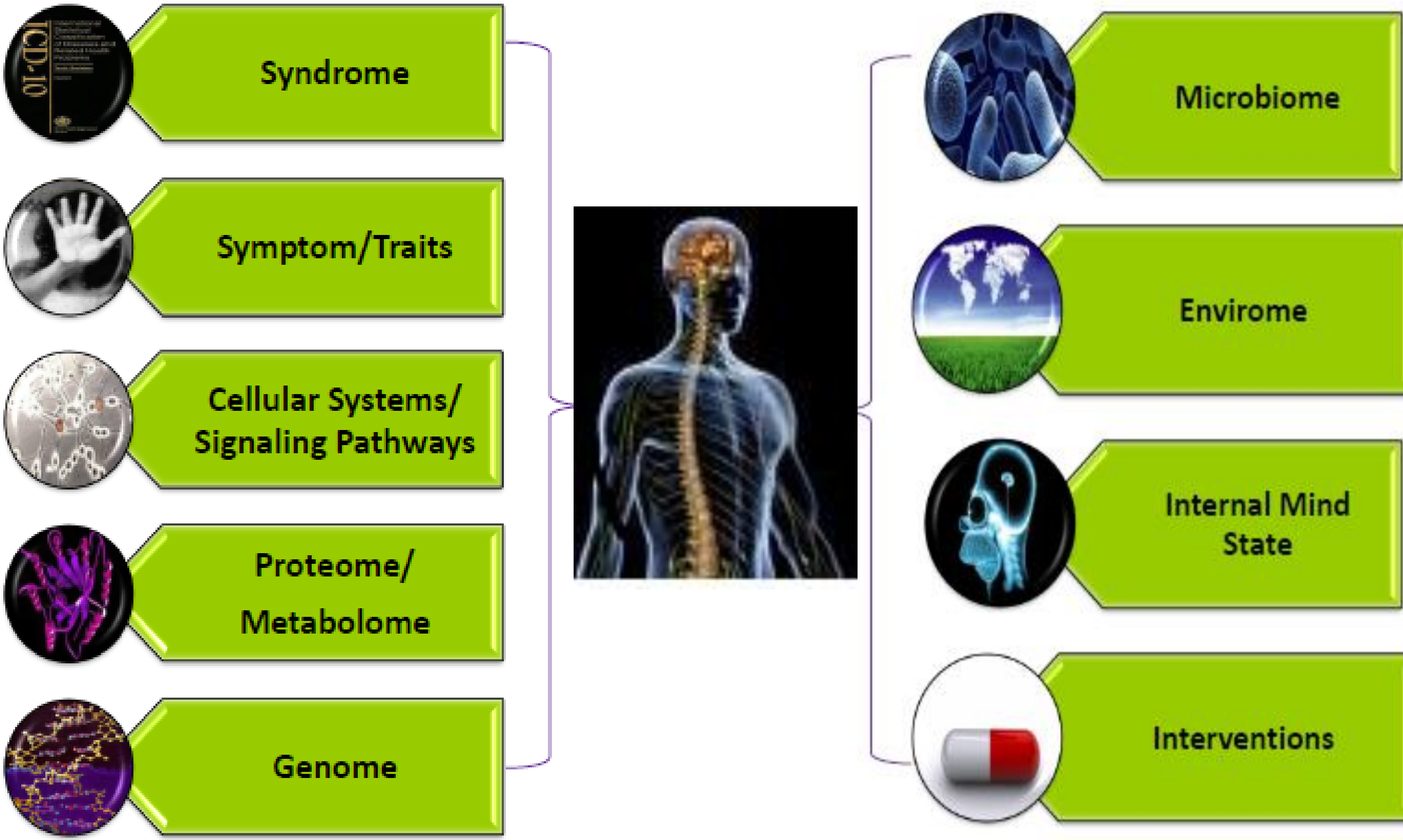
Full correlation matrix analysis: voxel level (active tasks)



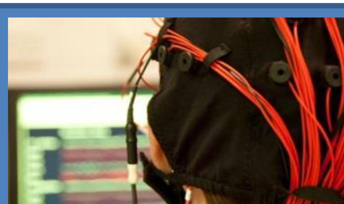
(A) An fMRI data set is divided into time windows, which are labeled with an experimental condition.
(B) Each window contains multiple time points, and each time point corresponds to a 3-D brain image.
(C) The time course of BOLD activity in every voxel is correlated with every other voxel to produce a full correlation matrix for each window.
(D) An example matrix from a 36-s block of fMRI data is depicted with 39,038 voxels arranged in a circle and 0.01% of correlations of >0.3 plotted as links.
(E) These matrices can be submitted as examples to MVPA, with each voxel pair as an input dimension.

Turk-Browne, Nicholas B. "Functional Interactions as Big Data in the Human Brain." *science* 342.6158 (2013): 580-584.

A translational view of research in brain disease



Combining knowledge of neuroscience and big data facilitates understanding of human behaviour



Unconstraint conditions

Non-invasive measuring system

Highly Constraint conditions

Simple & Mobile Neuro- & daily-life activitie measuring system

Link between lab and daily life

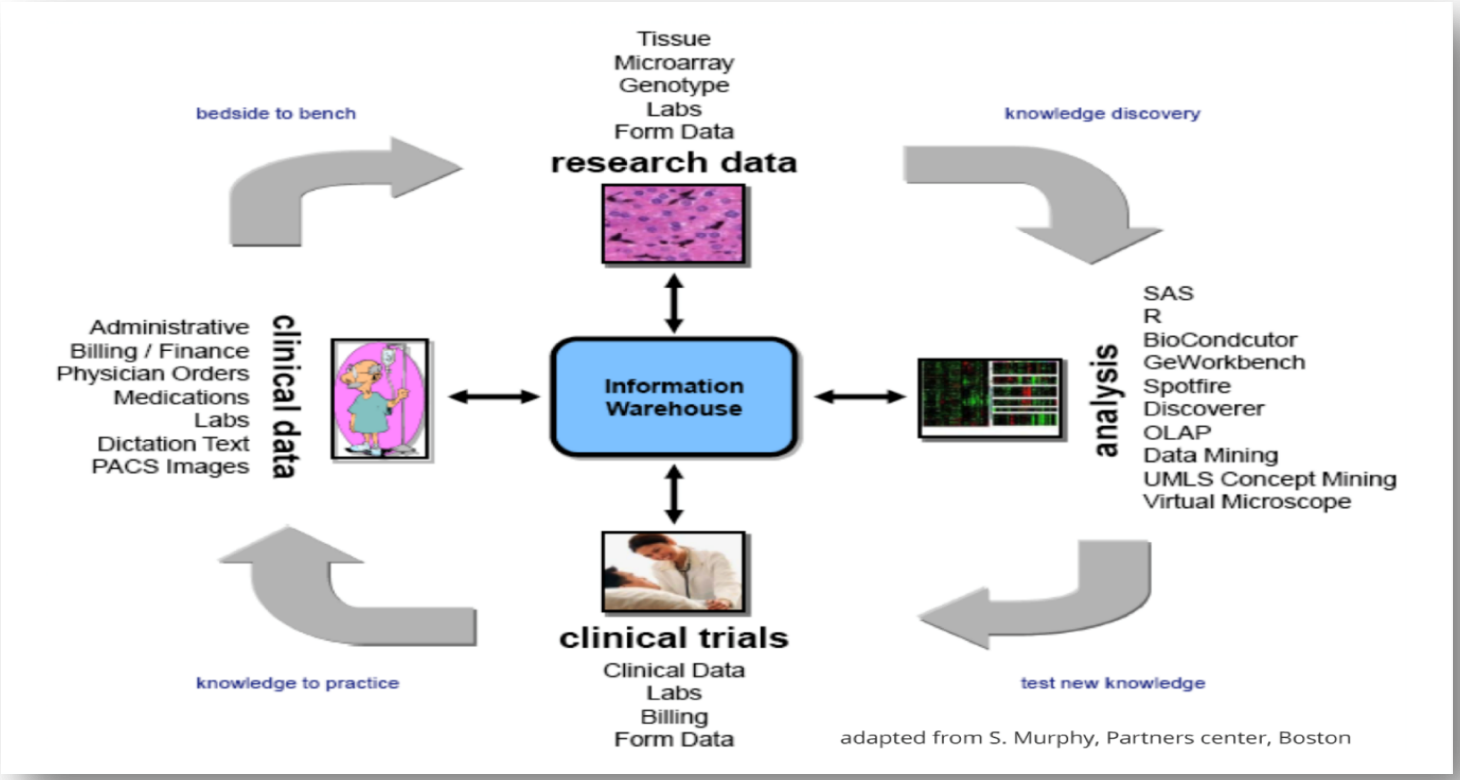
Collection of activity in daily life

Activity data

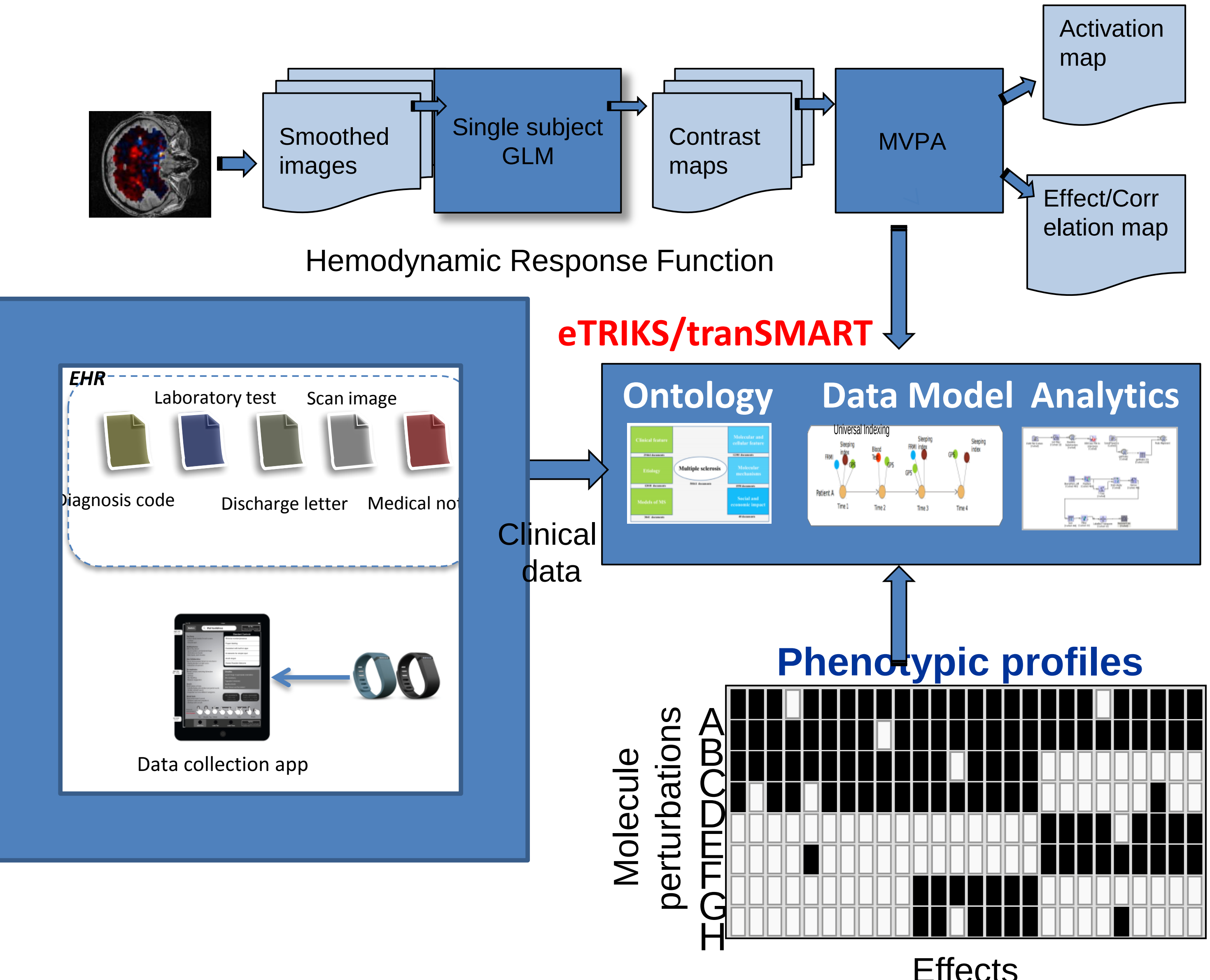
Unbiased /biased stimuli

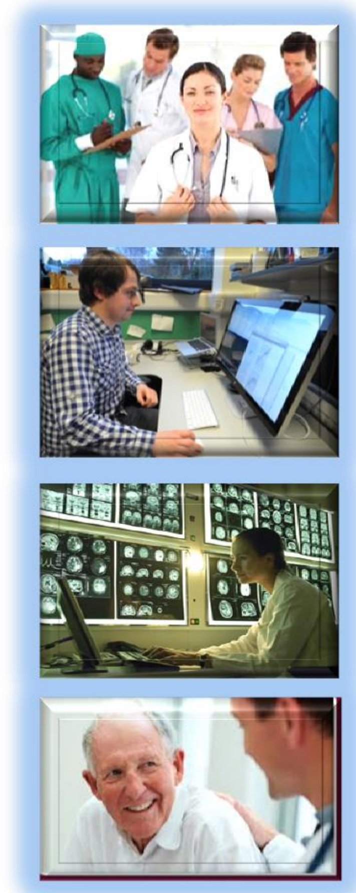
eTRIKS: European translational informatics platform and service

- *A €23.79m for 5 years (Oct 2012---Sept 2017) project for building a platform to support translational research*
- *Support €2 billion IMI projects in translation medicine study*
- *Imperial College leads with 3 major partners and 10 pharmaceutical companies*

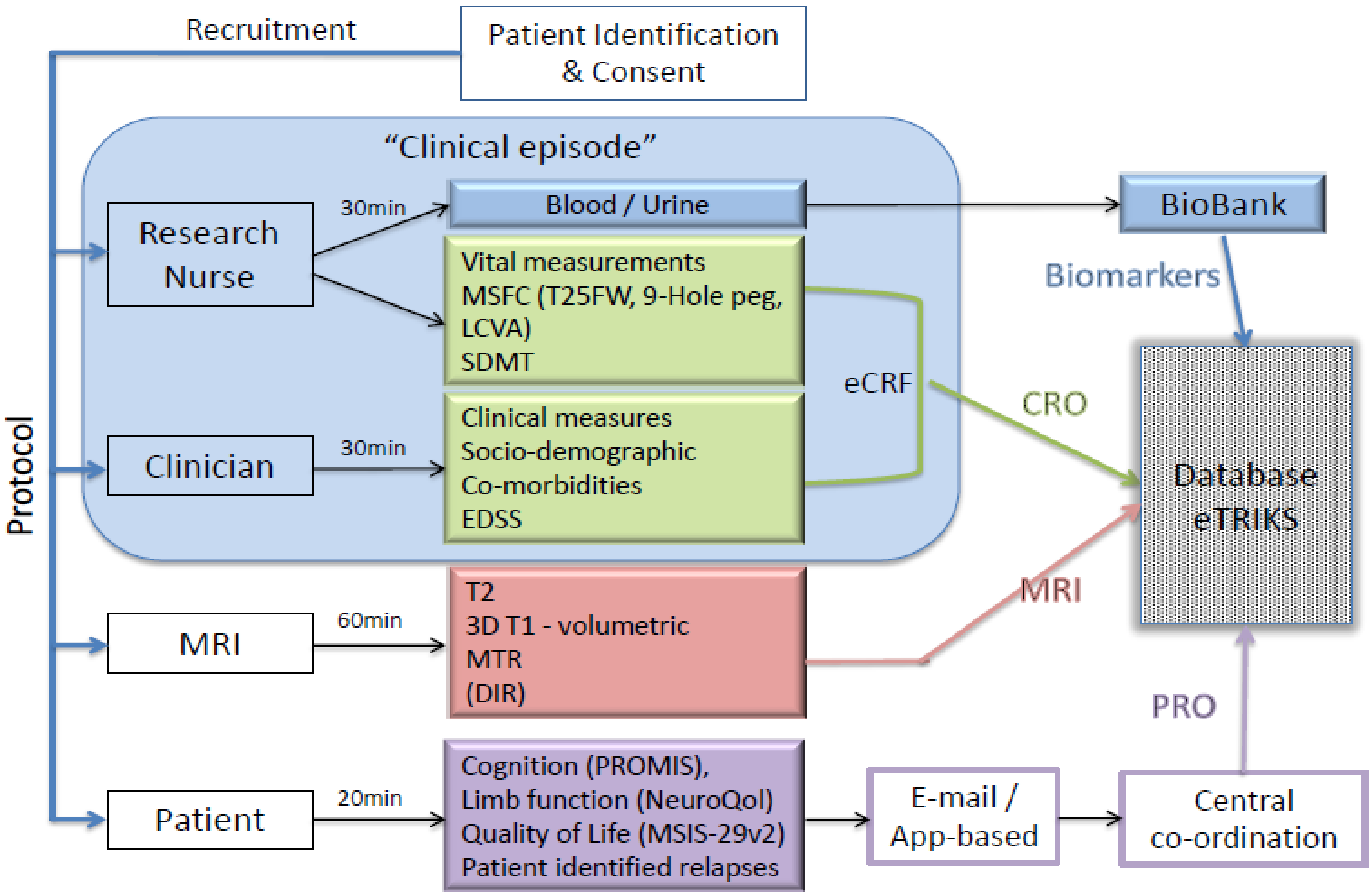


eTRIKS platform for brain disease translational research





OPTIMISE: stratified therapy in multiple sclerosis



Impact of neuroscience to data science

Brain function measurement

Non-invasive measuring systems

Generates big data

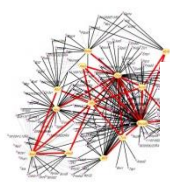
Provides stimulus set

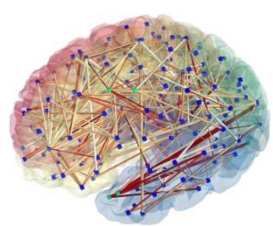
Interpretation of big data

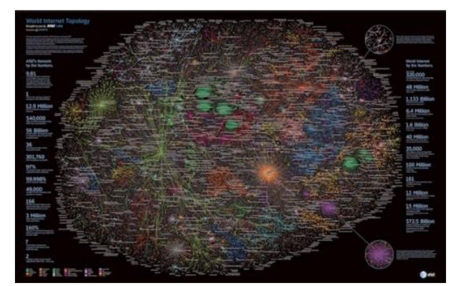
Generates human behaviour data

Information Network
Sensor Network

Complex network science





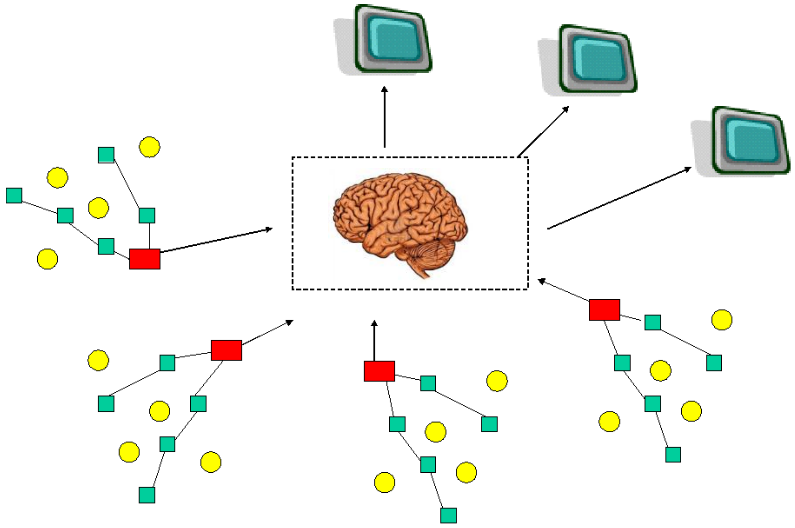


brain network

Analysis of network topologies

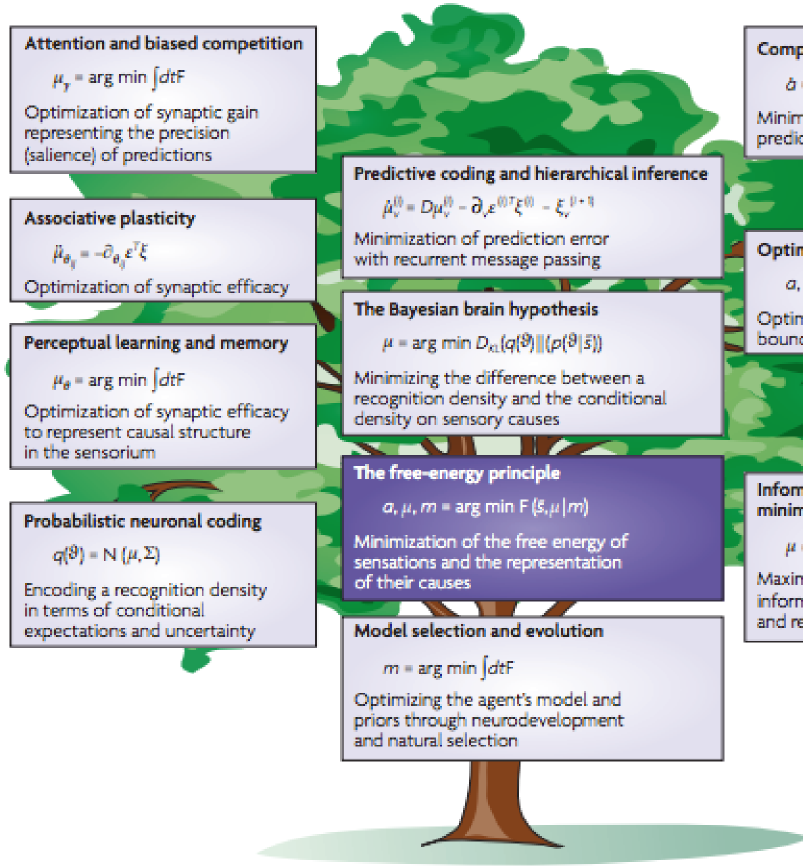
Cognitive sensing

- Applying cognitive science to computer sensing system (Brain as prediction machine -> Intelligent Sensing)

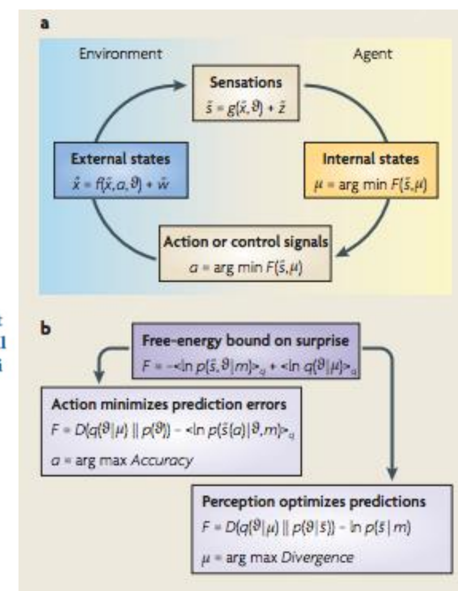


- Enabling the sensing system certain state of ‘consciousness’.
  - Make the system more adaptive (to resist a natural tendency to disorder).
  - Make the system more intelligent (to gain wanted knowledge from multifarious target).
  - Make the system more resource optimised (to balance the approximation and local accuracy).

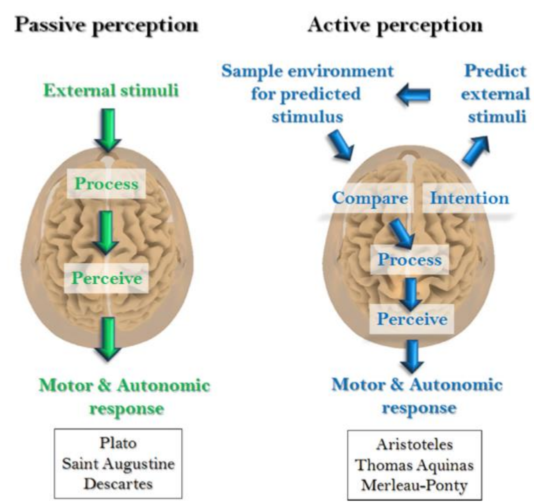
Bayesian brain and surprise



The **free-energy** principle: *an information theory measure that bounds or limits the surprise on sampling some data, given a generative model.*

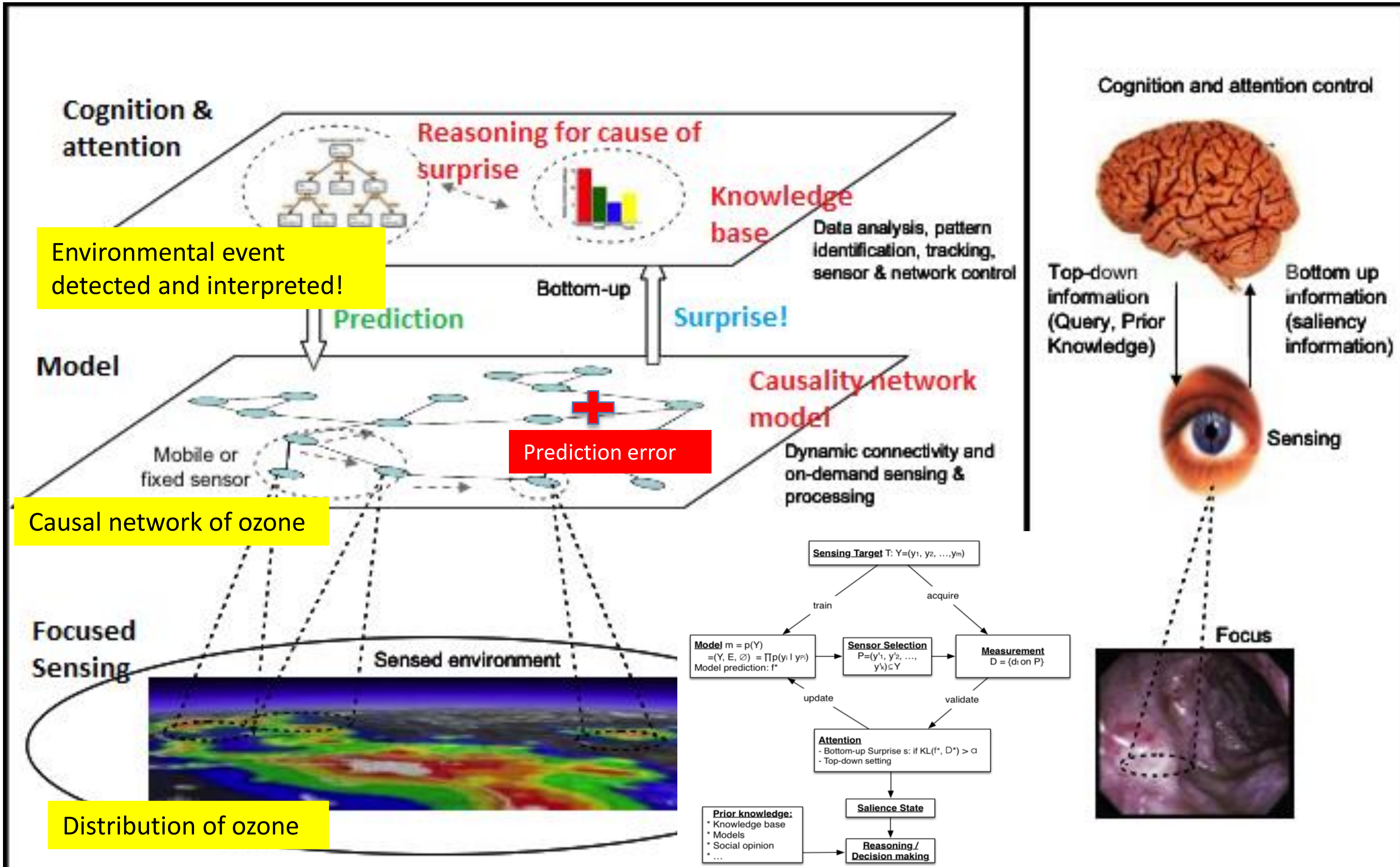


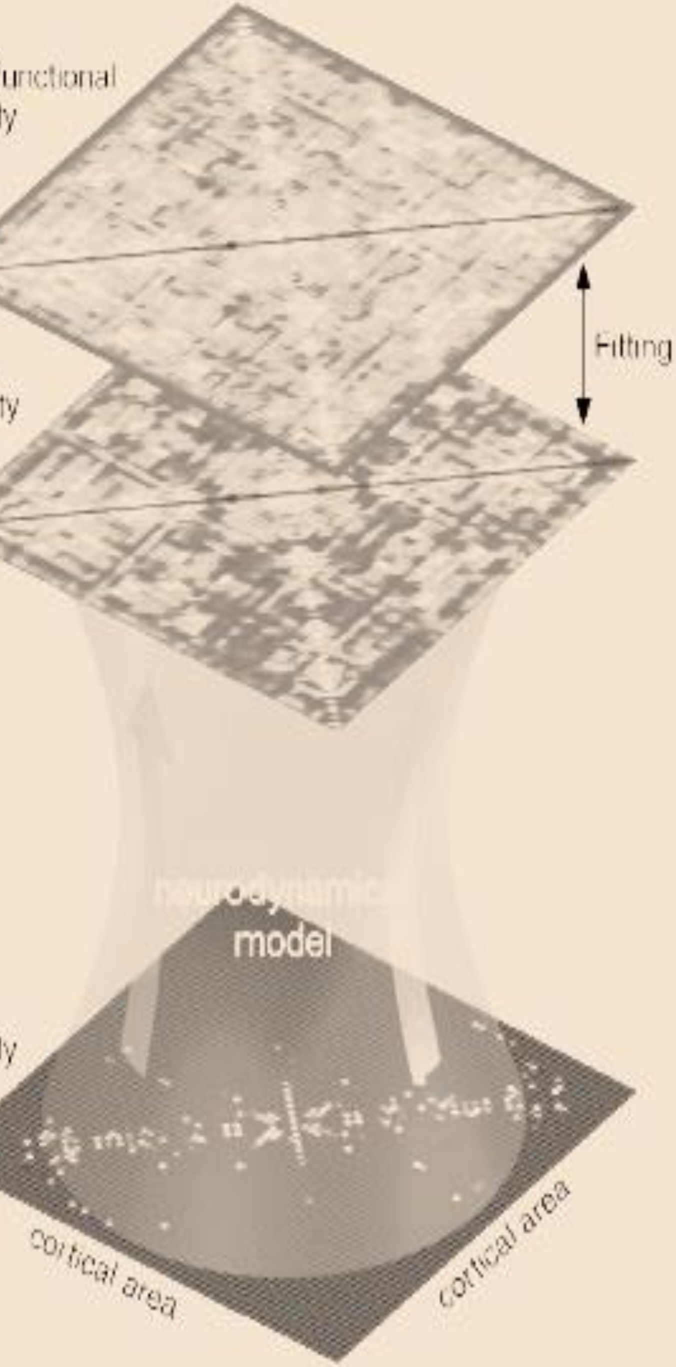
Any self-organising system that is at equilibrium with its environment must minimise its free energy -> minimise the long-term average of surprise to ensure that their sensory entropy remains low.



Friston, Karl. "The free-energy principle: a unified brain theory?." *Nature Reviews Neuroscience* 11.2 (2010): 127-138.

Cognitive sensing design





Conclusion

- Small brain is the best place for big data research
- Big data research is the key to demystify small brains
- We are in the beginning of innovations for brain research and neuro-technology
- Data science plays the key role in these endeavours