RSA (1): Representation

Choong-Wan Woo
Director of the Cocoan Lab

Key references

frontiers in SYSTEMS NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 24 November 2008
doi: 10.3389/neuro.06.004.2008

Representational similarity analysis – connecting the branches of systems neuroscience

Nikolaus Kriegeskorte[1,], Marieke Mur[1,2] and Peter Bandettini[1]*

[1] *Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA*
[2] *Department of Cognitive Neuroscience, Faculty of Psychology, Maastricht University, Maastricht, The Netherlands*

- the original RSA paper
- Cited by 1996 (2021. 4. 23)

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

A Toolbox for Representational Similarity Analysis

Hamed Nili[1]*, Cai Wingfield[2], Alexander Walther[1], Li Su[1,3], William Marslen-Wilson[3], Nikolaus Kriegeskorte[1]*

1 MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom, **2** Department of Computer Science, University of Bath, Bath, United Kingdom, **3** Department of Experimental Psychology, University of Cambridge, Cambridge, United Kingdom

Decoding Neural Representational Spaces Using Multivariate Pattern Analysis

James V. Haxby,[1,2] Andrew C. Connolly,[1] and J. Swaroop Guntupalli[1]

[1]Department of Psychological and Brain Sciences, Center for Cognitive Neuroscience, Dartmouth College, Hanover, New Hampshire 03755;
email: james.v.haxby@dartmouth.edu, andrew.c.connolly@dartmouth.edu, swaroopgj@gmail.com

[2]Center for Mind/Brain Sciences (CIMeC), University of Trento, Rovereto, Trentino 38068, Italy

Annu. Rev. Neurosci. 2014. 37:435–56

Review

Cell PRESS

Representational geometry: integrating cognition, computation, and the brain

Nikolaus Kriegeskorte[1] and Rogier A. Kievit[1,2]

[1] Medical Research Council, Cognition and Brain Sciences Unit, Cambridge, UK
[2] Department of Psychological Methods, University of Amsterdam, Amsterdam, The Netherlands

Representational similarity analysis

Let’s start with the following three questions:

What is…

Representation?

Representational space?

Representational geometry?

Key references

The physics of representation

Russell A. Poldrack

Abstract The concept of "representation" is used broadly and uncontroversially throughout neuroscience, in contrast to its highly controversial status within the philosophy of mind and cognitive science. In this paper I first discuss the way that the term is used within neuroscience, in particular describing the strategies by which representations are characterized empirically. I then relate the concept of representation within neuroscience to one that has developed within the field of machine learning (in particular through recent work in deep learning or "representation learning"). I argue that the recent success of artificial neural networks on certain tasks such as visual object recognition reflects the degree to which those systems (like biological brains) exhibit inherent inductive biases that reflect on the structure of the physical world. I further argue that any system that is going to behave intelligently in the world must contain representations that reflect the structure of the world; otherwise, the system must perform unconstrained function approximation which is destined to fail due to the curse of dimensionality, in which the number of possible states of the world grows exponentially with the number of dimensions in the space of possible inputs. An analysis of these concepts in light of philosophical debates regarding the ontological status of representations suggests that the representations identified within both biological and artificial neural networks qualify as first-class representations.

http://philsci-archive.pitt.edu/17455/

Representation?

Russ Poldrack:

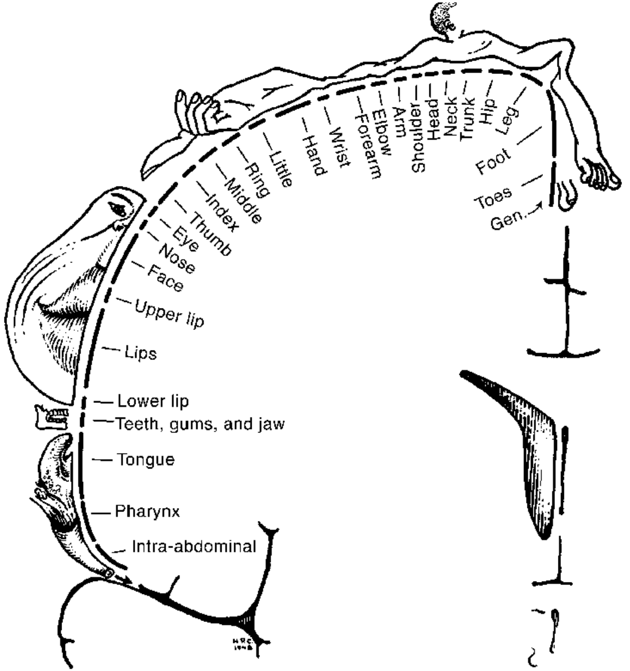
Representation refers to… *“patterns of activity that bear a systematic relationship to the structure of the external world and play a causal role in behavior, is fundamentally necessary for any intelligent organism”*

Relevant terms:

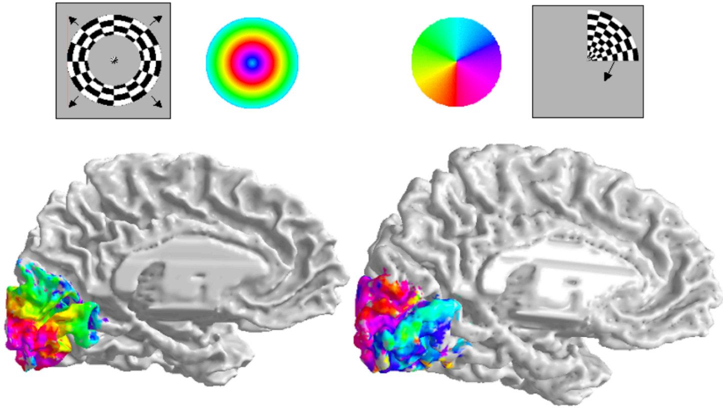
- Functional isomorphism
- Functional approximation
- Dimensionality reduction

Representation?

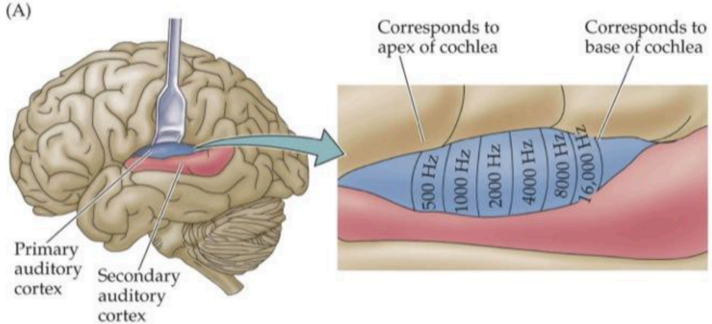
Simple forms of brain representation



Penfield's **sensory homunculus** (Penfield & Rasmussen, 1950)



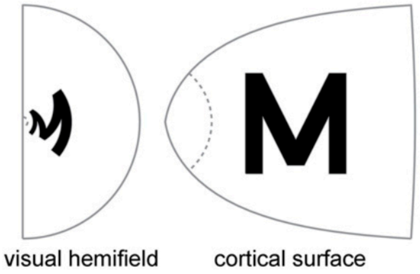
Retinotopy (Dougherty et al., 2003)

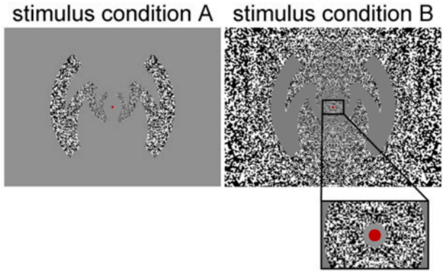


Tonotopy
from https://medium.com/@mosaicofminds/maps-in-the-brain-f236998d544f

Representation?

structurally isomorphic (between the neural and external worlds)





NeuroImage 52 (2010) 1334–1346

Contents lists available at ScienceDirect

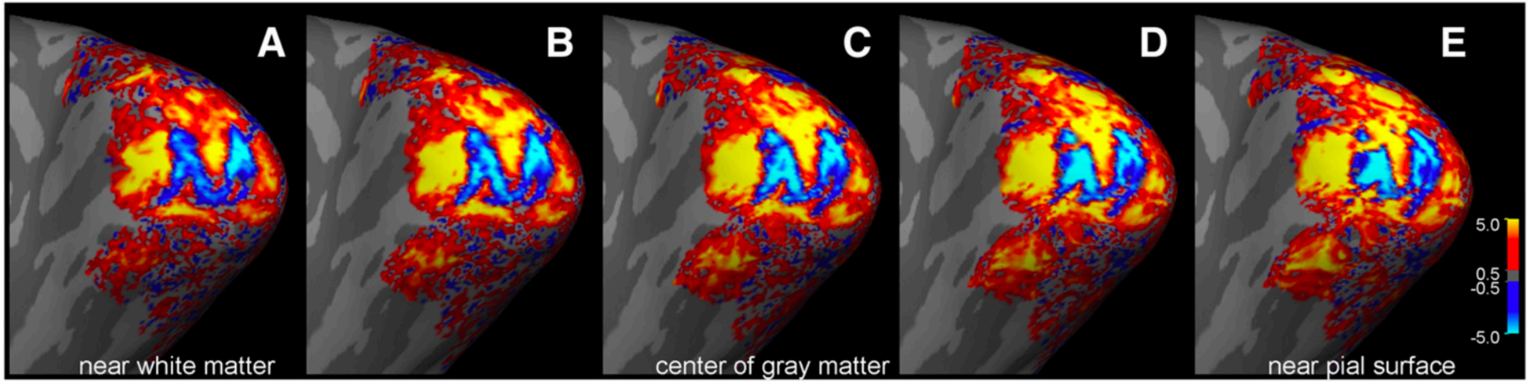
NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

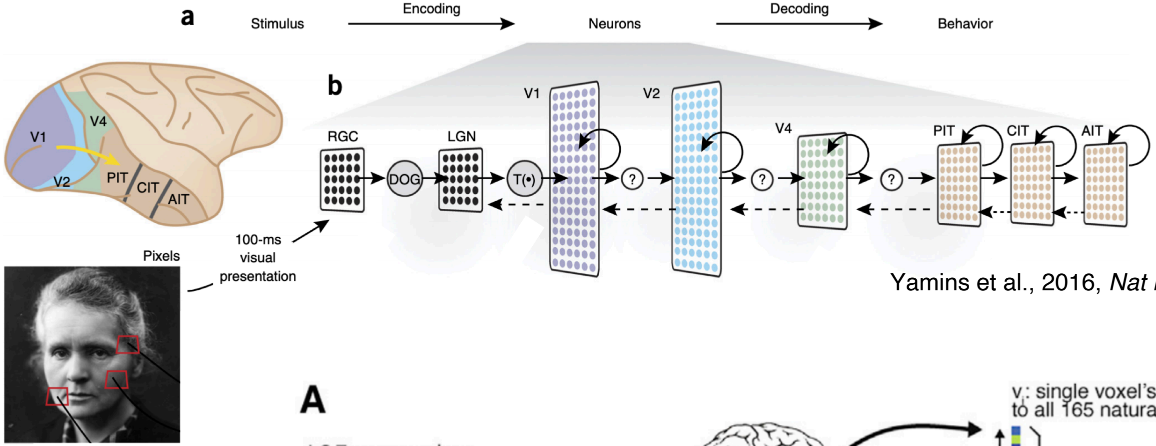
Laminar analysis of 7 T BOLD using an imposed spatial activation pattern in human V1

Jonathan R. Polimeni [a,*], Bruce Fischl [a,b], Douglas N. Greve [a], Lawrence L. Wald [a,c]

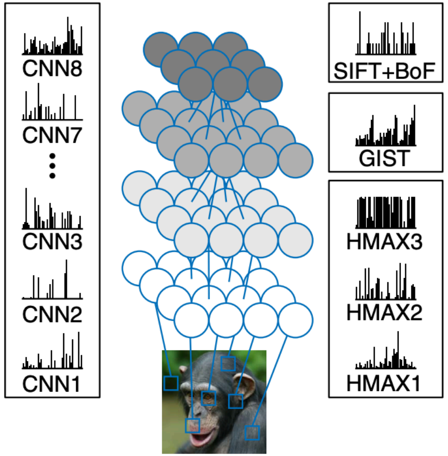
[a] Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Bldg 149 Thirteenth St., Suite 2301, Charlestown, MA 02129, USA
[b] Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, MA, USA
[c] Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA



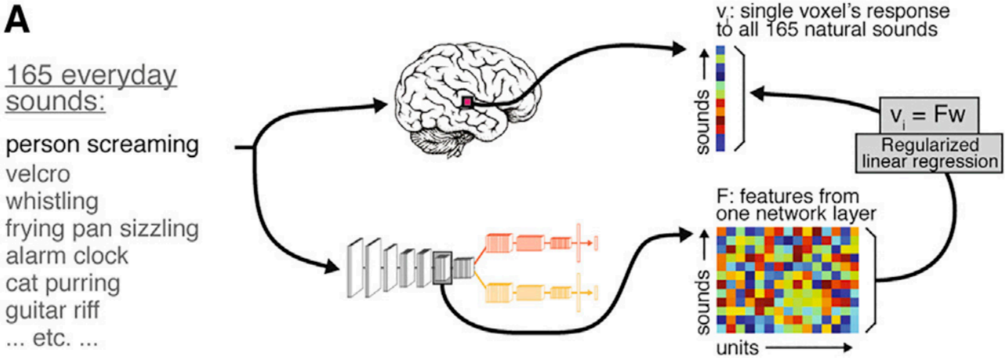
Different systems, different representations



Yamins et al., 2016, *Nat Neurosci*

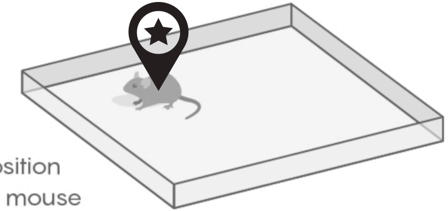


Horikawa et al., 2017, *Nat Comms*



Kell et al., 2018, *Neuron*

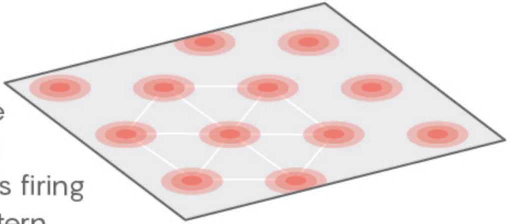
Beyond sensory information processing



Position
of mouse

Mouse entorhinal cortex

One
grid
cell's firing
pattern



A spatial location can be represented
in very different forms across different systems

Maps

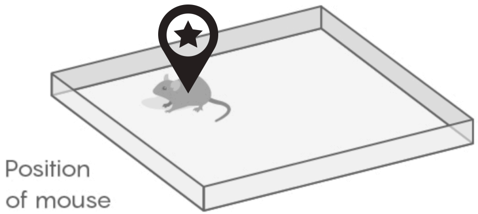
36°41'37.6"N 126°39'21.6"E
(longitude and latitude)

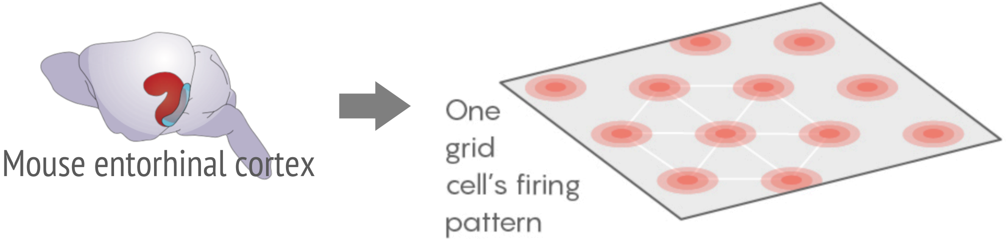
Computer

Binary code

Image sources: https://www.quantamagazine.org/the-brain-maps-out-ideas-and-memories-like-spaces-20190114/
Behrens et al., 2018, *Neuron*

Beyond sensory information processing

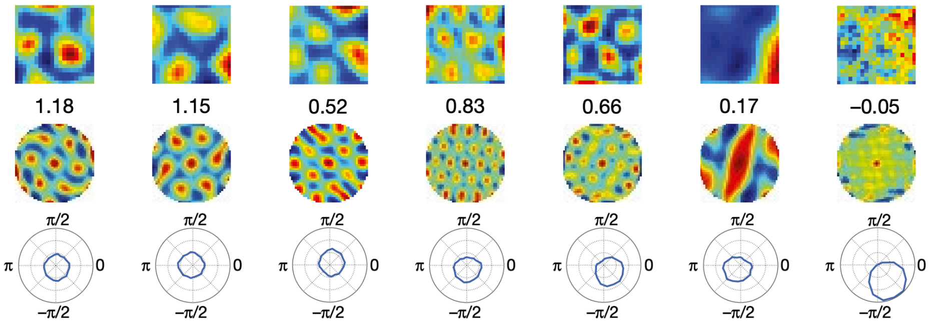




A spatial location can be represented
in very different forms across different systems

Vector-based navigation using grid-like representations in artificial agents

Andrea Banino[1,2,3,5]*, Caswell Barry[2,5]*, Benigno Uria[1], Charles Blundell[1], Timothy Lillicrap[1], Piotr Mirowski[1], Alexander Pritzel[1], Martin J. Chadwick[1], Thomas Degris[1], Joseph Modayil[1], Greg Wayne[1], Hubert Soyer[1], Fabio Viola[1], Brian Zhang[1], Ross Goroshin[1], Neil Rabinowitz[1], Razvan Pascanu[1], Charlie Beattie[1], Stig Petersen[1], Amir Sadik[1], Stephen Gaffney[1], Helen King[1], Koray Kavukcuoglu[1], Demis Hassabis[1,4], Raia Hadsell[1] & Dharshan Kumaran[1,3]*



Banino et al., 2018, *Nature*

Searching for brain representations?

Establishing tighter associations between the mind and the brain:

For example, brain representations of "pain"

a. Pain ⇨ Brain activation (e.g., regions, or patterns)

e.g., P(Brain|Pain), sensitivity, forward inference

b. Large effect size between pain and the brain measure

e.g., Cohen's *d*, explained variance (R^2), etc.

c. Not pain ⇨ No brain activation

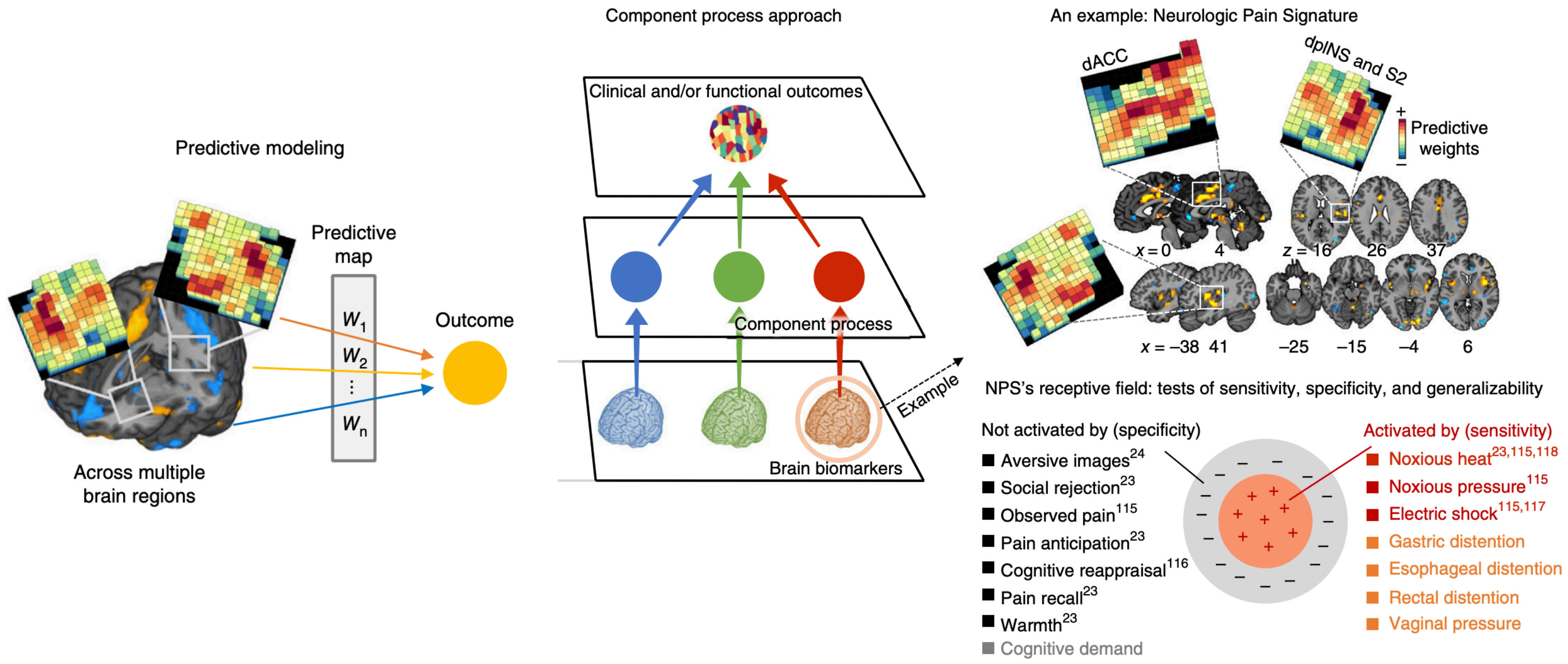
e.g., P(~Brain|~Pain), specificity

d. Brain activation ⇨ Pain

e.g., P(Pain | Brain), positive predictive value, brain decoding, reverse inference

Woo et al., 2017, *Nat Neurosci*; Woo et al., 2015, *Pain*

Identifying good brain representations = Developing brain biomarkers



Woo et al., 2017, *Nat Neurosci*

Some challenges in studying representations

1. Difficult to identify (or develop) good representations
2. Difficult to use (or compare) representations based on different systems (or models)

   e.g., different dimensions, different shapes, etc.

Solution? Use representational distance (or representational similarity) instead

Social network as an example

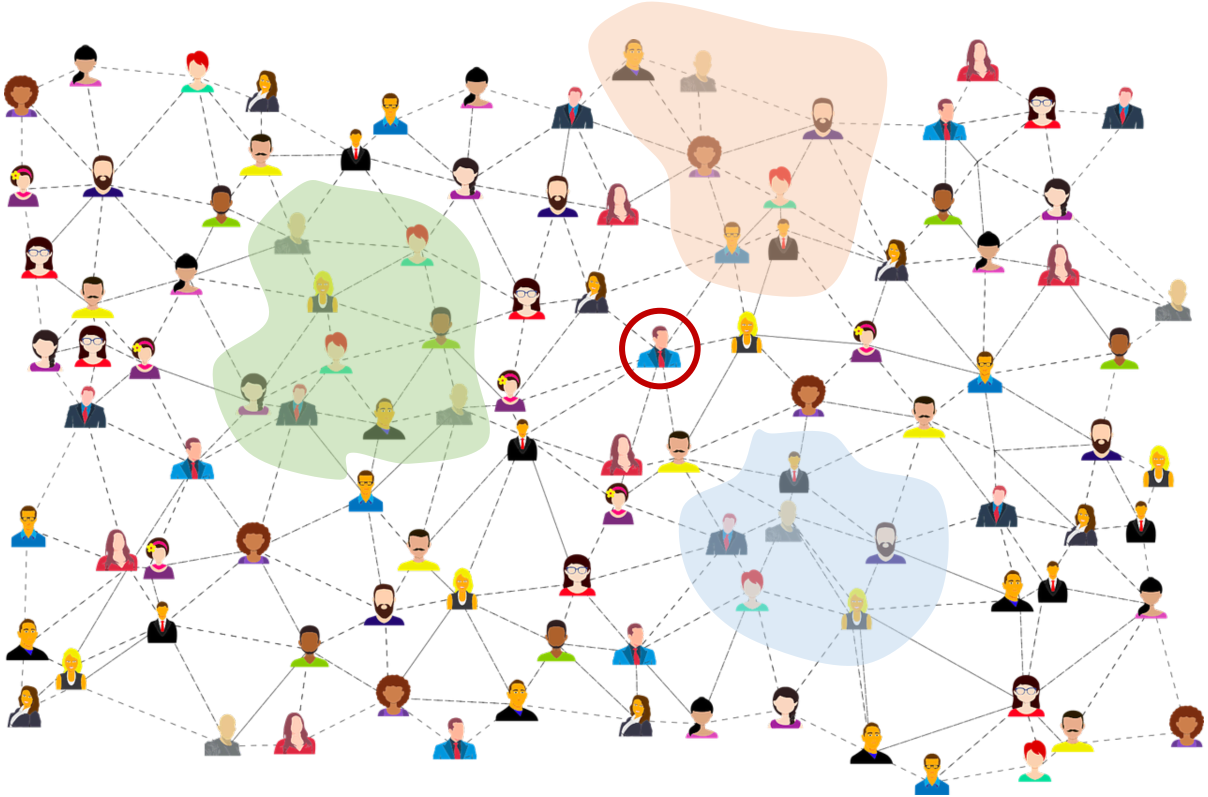
To find potential friends...

We first need to know <u>who I am</u> (let's say its a representation) and find <u>people similar to me</u>

But it is really difficult to know!

Then, we can use the structure of representational similarity!

Social network as an example



For example,
see my friends (i.e., links; e.g., facebook friends)

Compare your links with other people's links

→ I may find my soulmates!

Representational similarity analysis?

Examining the distances among people from <u>my</u> point of view

And comparing it with <u>your</u> distances for people

Limitations of the representational similarity approach

One challenge with the analysis of similarity spaces is that they are fundamentally indeterminate.

There is no single "correct" similarity space within which to compare patterns of activity, just as in general there is no single "correct" decomposition of a dataset into lower dimensionality.

We still need to put our efforts to identify good representations themselves,

while we are using the representational similarity analysis that is useful for answering certain questions

This does not absolve the approach of indeterminacy; it simply pushes that indeterminacy down to a level below the inferences that are being made.

However, neuroscientists are generally comfortable endorsing claims about the similarity of representational spaces, despite the fact that there is no unique underlying space in which they can be defined.

Poldrack, 2020, p. 6

RSA (2): Representational space and Representational similarity analysis (RSA)

Choong-Wan Woo
Director of the Cocoan Lab

Representational space

Poldrack, 2020

“a low-dimensional projection of the responses across stimuli”

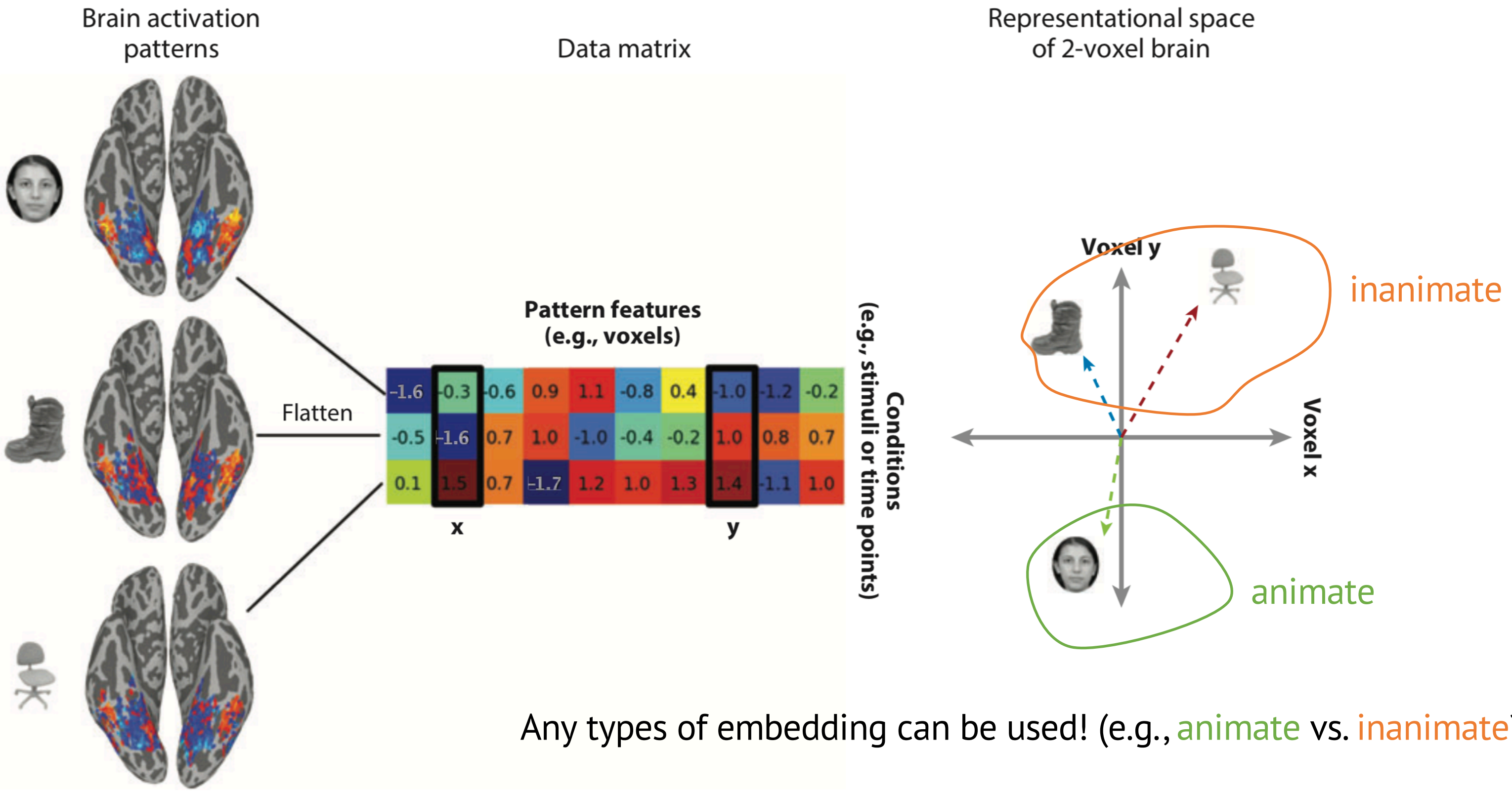
“a low-dimensional embedding”

Haxby et al., 2014, *Annu Rev Neurosci*

REPRESENTATIONAL SPACE

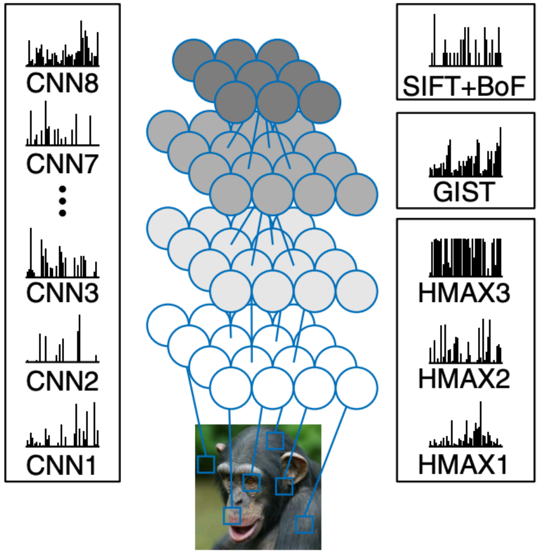
Representational space is a high-dimensional space in which each neural response or stimulus is expressed as a vector with different values for each dimension. In a neural representational space, each pattern feature is a measure of local activity, such as a voxel or a single neuron. In a stimulus representational space, each feature is a stimulus attribute, such as a physical attribute or semantic label.

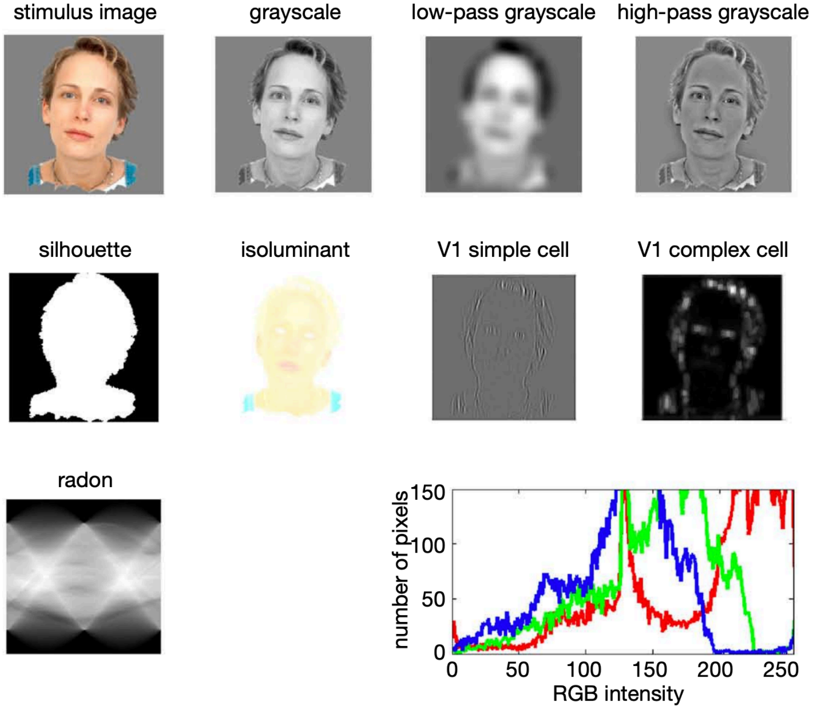
Representational space



Any types of embedding can be used! (e.g., animate vs. inanimate?)

Representational space



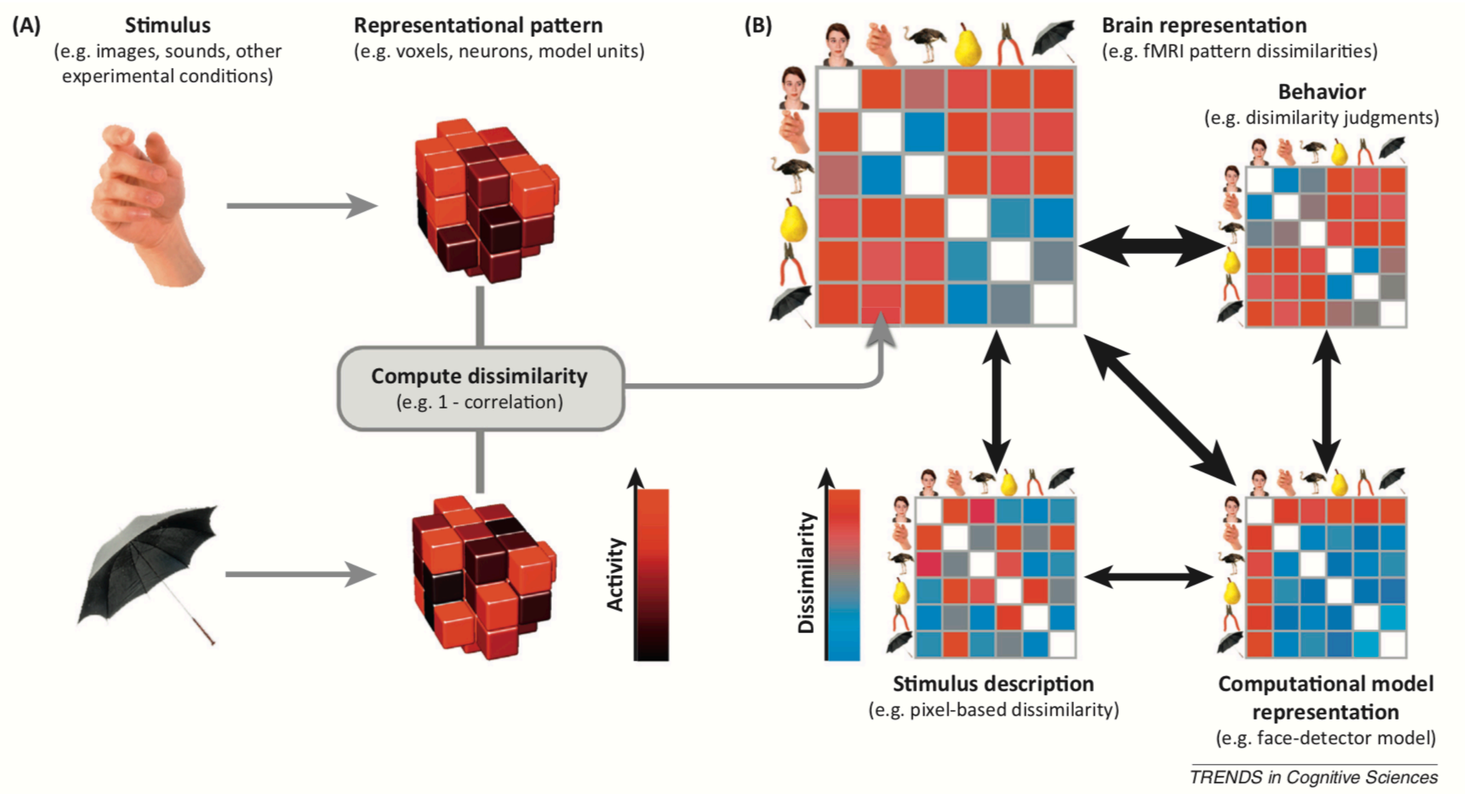


Any embeddings can be used!

Horikawa et al., 2017, *Nat Comms*

Kriegeskorte et al., 2008, *Front Sys Neurosci*

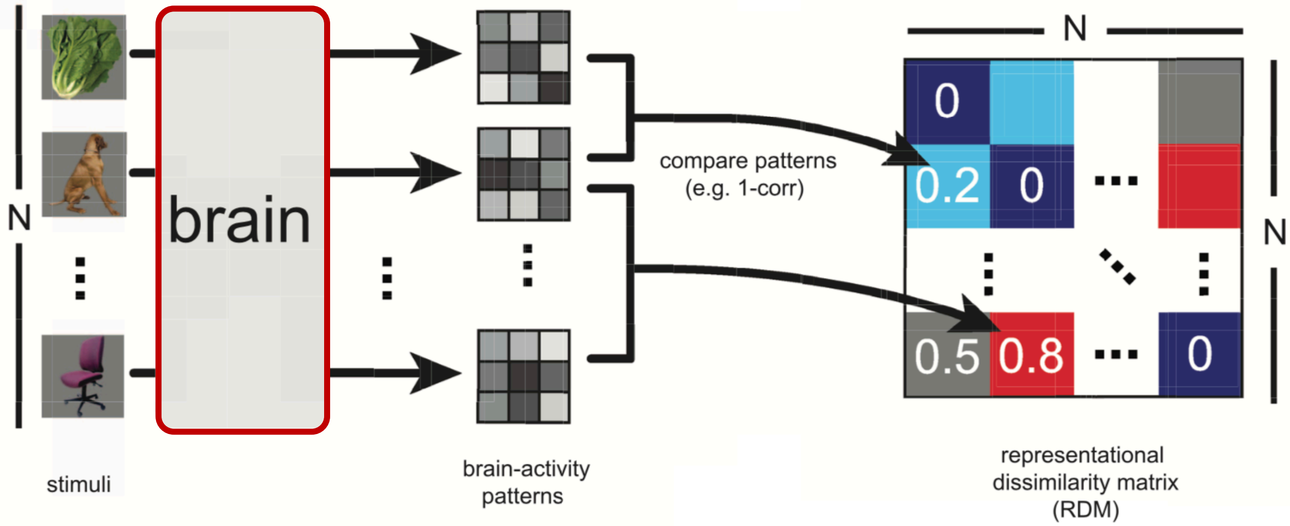
Representational dissimilarity matrix



Kriegeskorte et al., 2013, *TICS*

Representational dissimilarity matrix (in more detail)

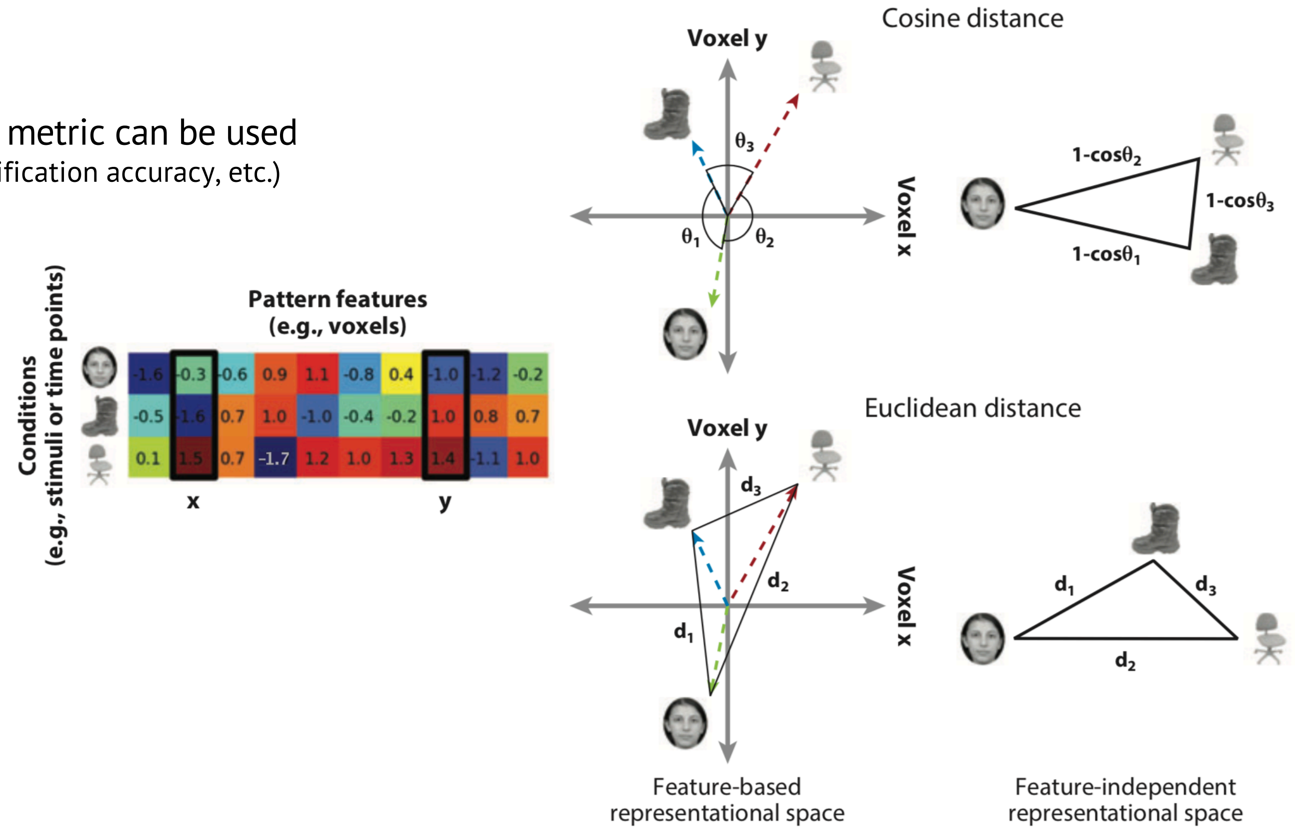
Brain or
Other model representations



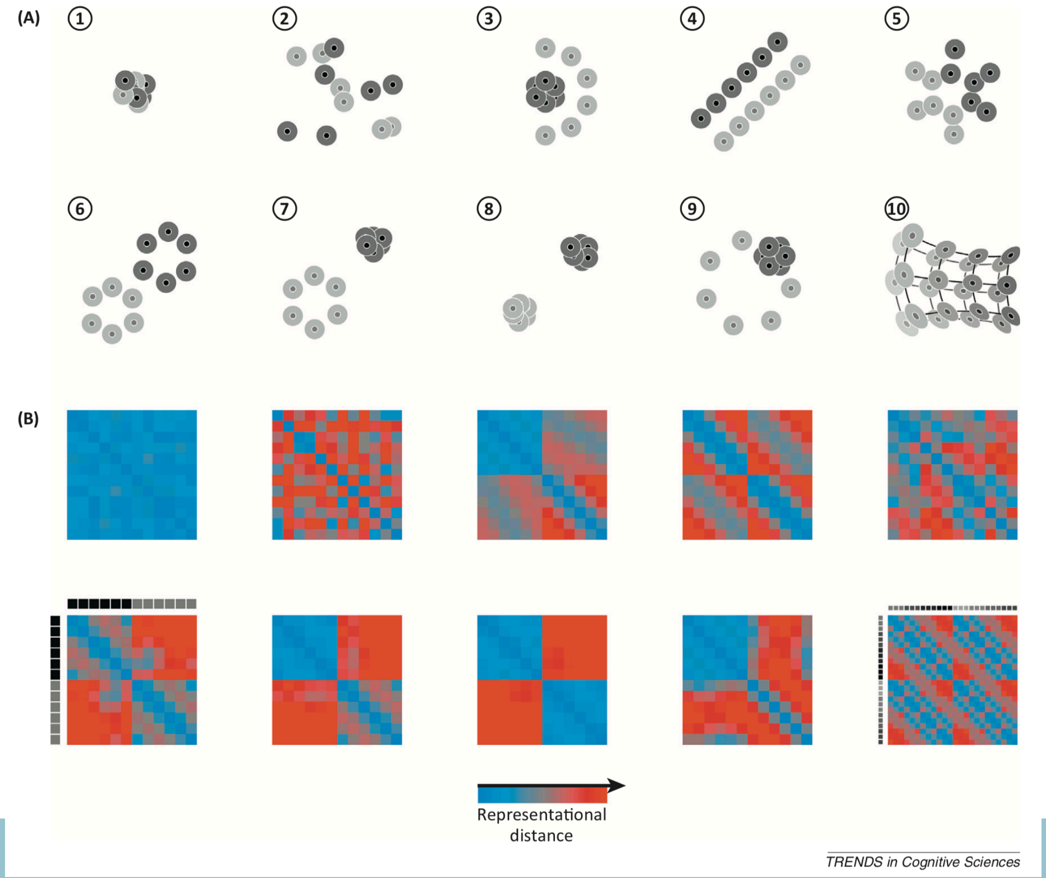
Nili et al., 2014, *PLoS Comp Biol*

Representational dissimilarity matrix (in more detail)

Other types of distance metric can be used
(e.g., cosine, Euclidean, classification accuracy, etc.)



Representational geometry

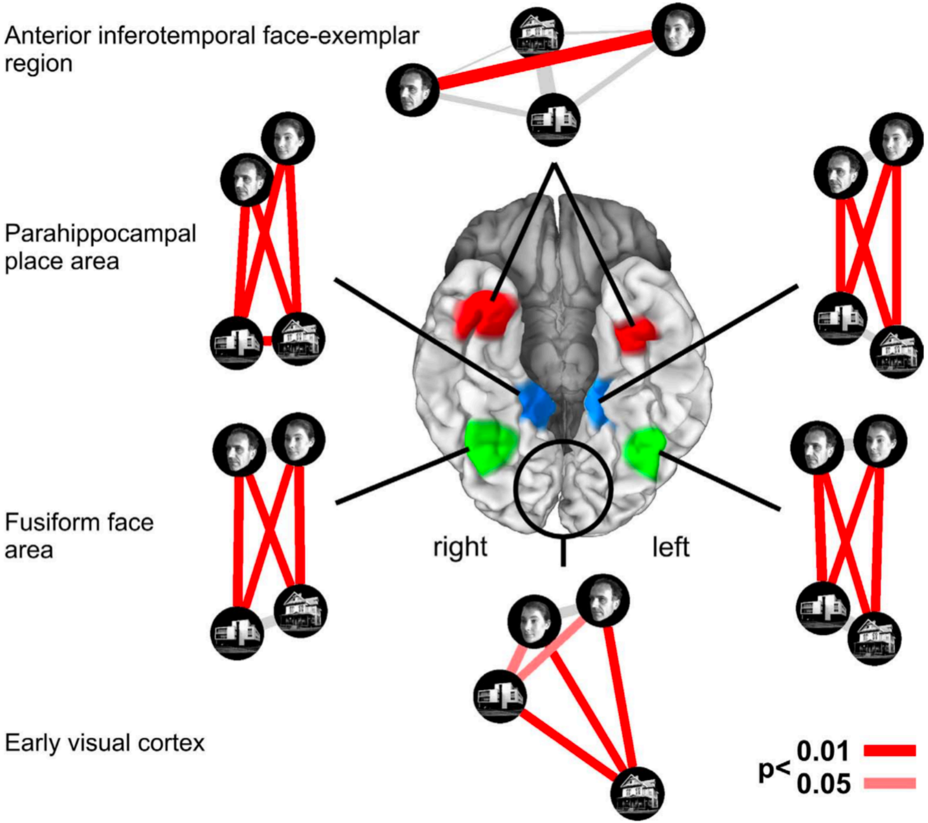


Representational geometry

Representational dissimilarity matrix

Kriegeskorte et al., 2013, *TICS*

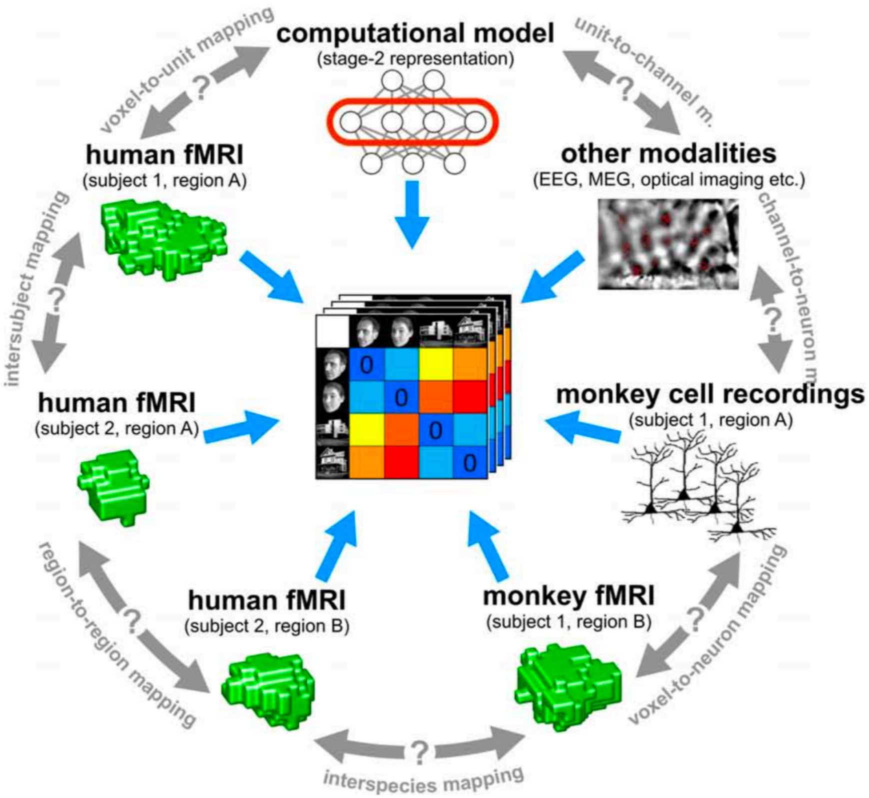
Representational dissimilarity matrix (in more detail)



Different representational geometry from different representational dissimilarity patterns of different brain regions

Kriegeskorte et al., 2008, *Front Sys Neurosci*

Basic idea: You can compare any models/systems using RDM!



Kriegeskorte et al., 2008, *Front Sys Neurosci*

RSA (3): Analysis steps for RSA

Choong-Wan Woo
Director of the Cocoan Lab

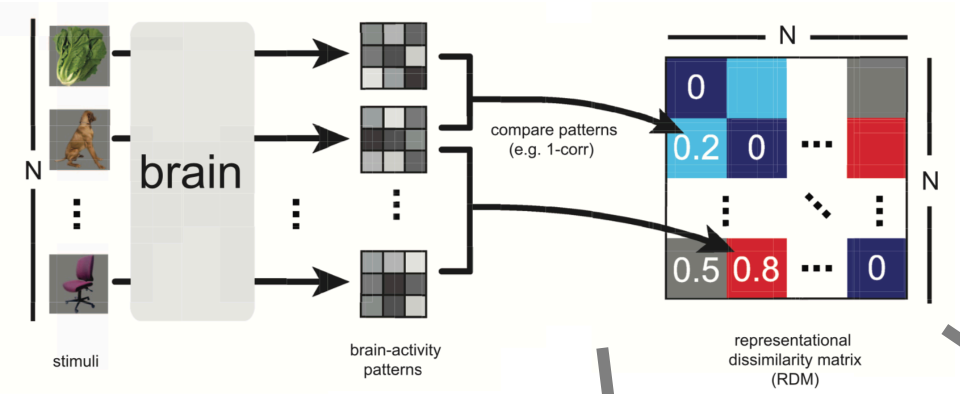
Three steps of representational similarity analysis

Step 1: Computing and visualizing RDMs

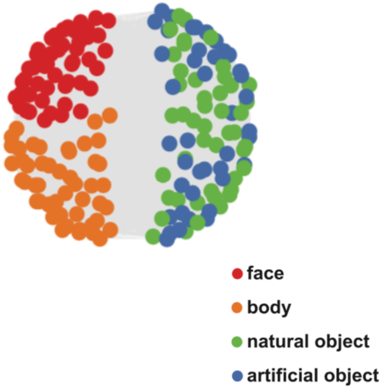
Step 2: Comparing brain and model RDMs

Step 3: Statistical inference

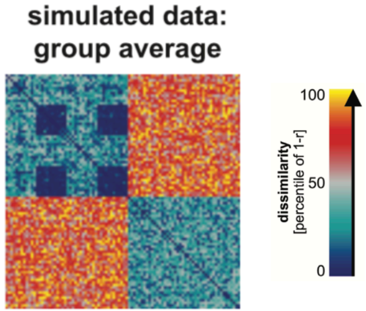
Step 1: Computing and visualizing RDMs



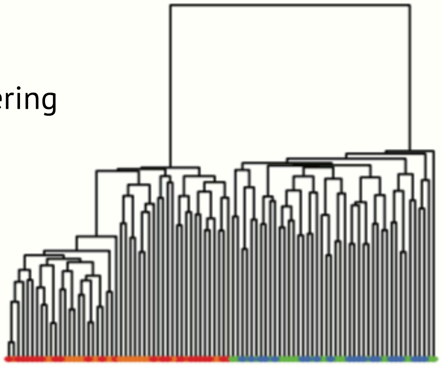
Dimensionality Reduction method for visualization (MDS, t-SNE, PCA, UMAP, etc.)



RDM
(Representational dissimilarity matrix)



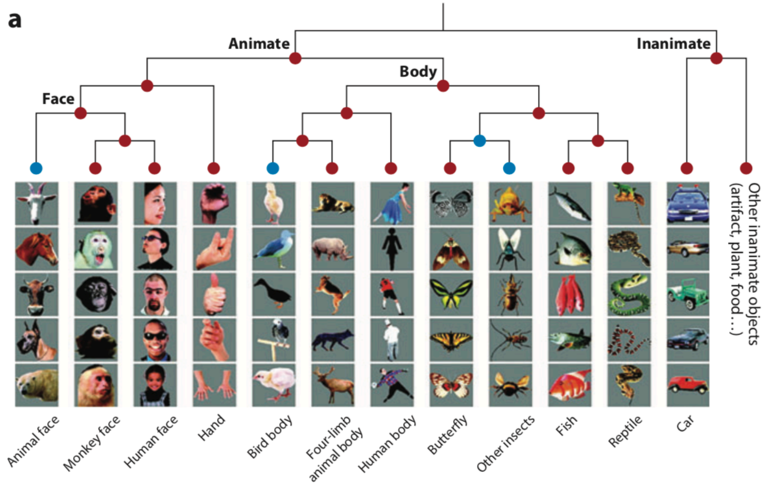
Clustering

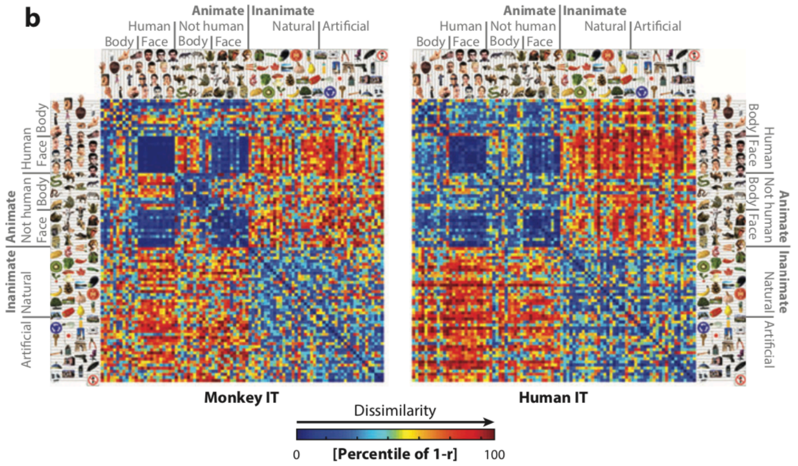


Step 1: Computing and visualizing RDMs

Real study examples:

Dendrogram derived from
multiple single-unit recording
in macaque inferior temporal (IT) cortex
(Kiani et al., 2007, *J. Neurophysiol.*)

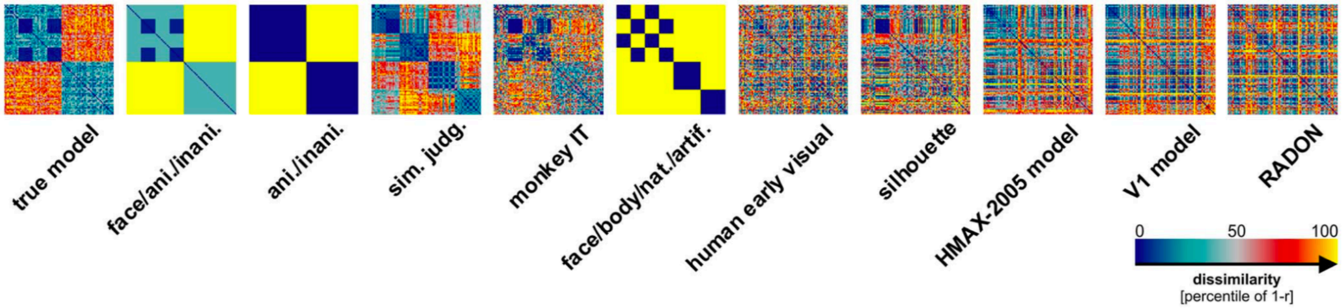




Cross-modal, cross-species comparisons of RDM
for a common set of stimuli
(Kriegeskorte et al., 2008, *Neuron*)

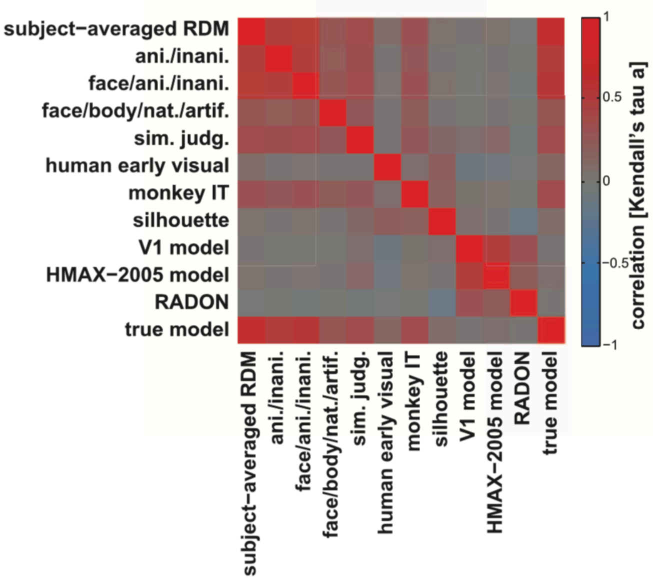
Step 2: Comparing brain and model RDMs

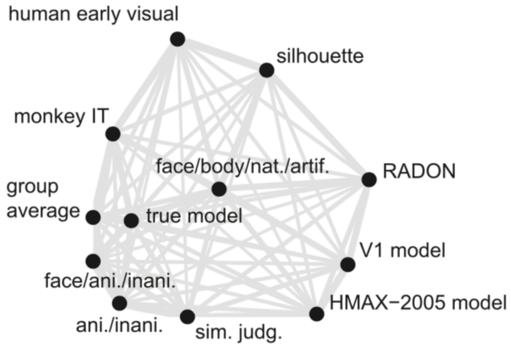
Many model RDMs



Comparing the RDMs: using any metric that quantifies the extent to which two matrices are "in agreement"

e.g., Spearman's rank correlation, Kendall's tau a, etc.

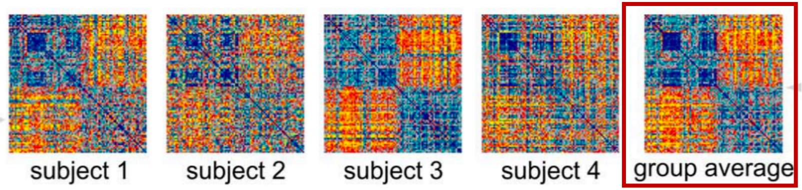




Again, dimensionality reduction for visualization (MDS, PCA, t-SNE, force-directed graph drawing, etc.)

Nili et al., 2014, *PLoS Comp Biol*

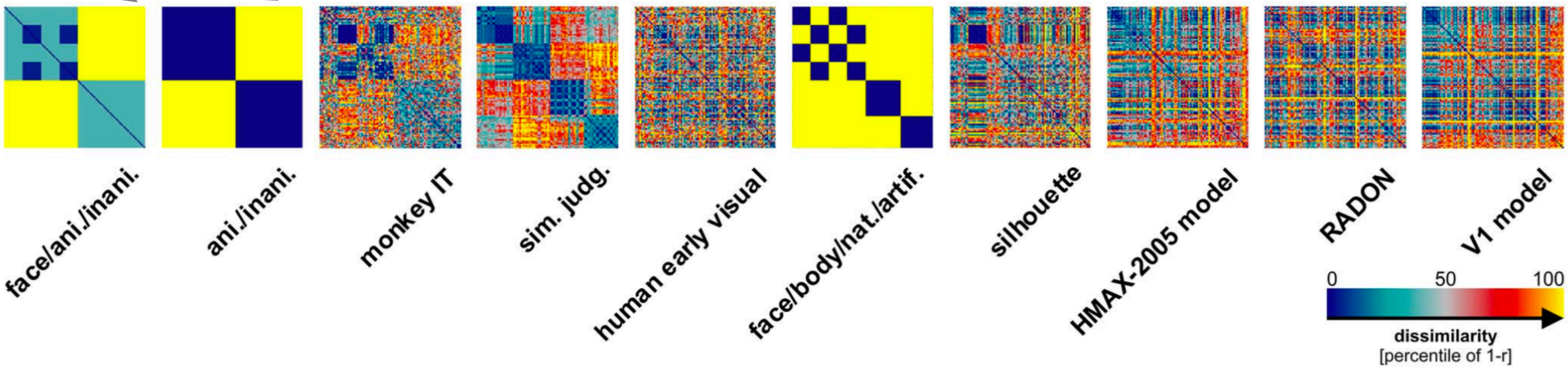
Step 3: Statistical inference



Now you have a reference RDM

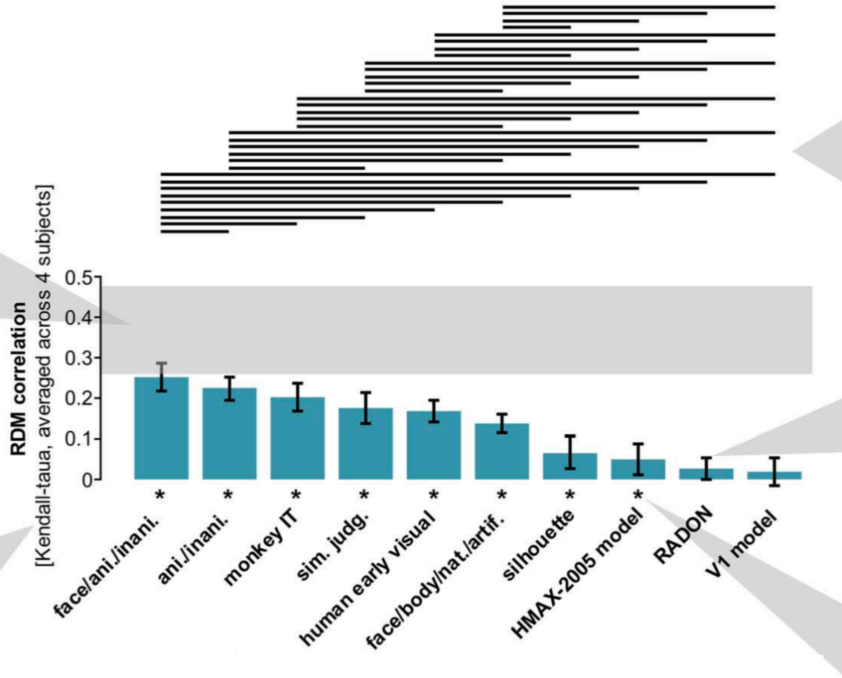
… You can compare them using Kendall's tau a (or Spearman's correlation)

and candidate RDMs



Nili et al., 2014, *PLoS Comp Biol*

Step 3: Statistical inference



Performance is measured, as in Fig. 4, by Kendall's tau a between the reference RDM and the candidate RDMs.

Note: This example uses data from 4 subjects with a large number of stimulus. Thus this uses resampling of the stimulus set. If you have a larger number of subjects, you can also do the resampling of subjects, not stimulus set.

Nili et al., 2014, *PLoS Comp Biol*

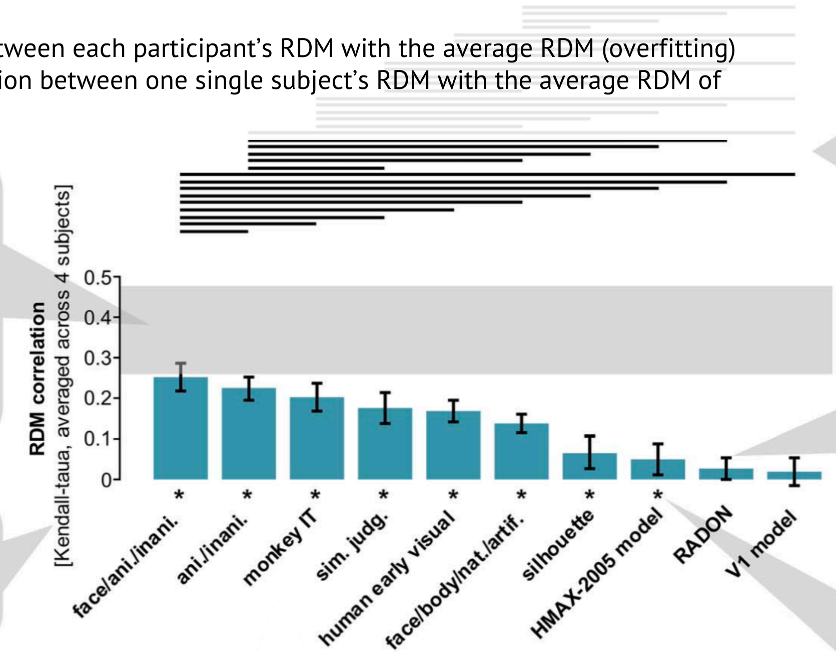
Step 3: Statistical inference

Noise ceiling:

- **Upper bound:** average correlation between each participant's RDM with the average RDM (overfitting)
- **Lower bound:** leave-one-out correlation between one single subject's RDM with the average RDM of others (underfitting)

The **noise ceiling** indicating the expected performance of the true model is much wider here than in the simulated data of Fig. 4. This reflects the fact that only 4 subjects entered this analysis, and the representation of human IT is thus much less precisely defined.

Performance is measured, as in Fig. 4, by Kendall's tau a between the reference RDM and the candidate RDMs.



Pairwise comparisons are based on bootstrap resampling of the stimulus set. This procedure simulates the variability of the estimates across random samples of stimuli (from an imaginary population of stimuli). Multiple testing is accounted for by controlling the expected FDR at 0.05.

Error bars indicate the standard error of the mean based on the bootstrap resampling of the stimulus set.

Each candidate RDM's relatedness to the reference RDM is now tested using a stimulus-label randomization test, which does not require multiple subjects. This time stars instead of p values were chosen to indicate significance. Multiple testing was accounted by controlling the expected FDR at 0.05.

Nili et al., 2014, *PLoS Comp Biol*

RSA (4): Other options

Choong-Wan Woo
Director of the Cocoan Lab

Other RSA options

Other analysis options

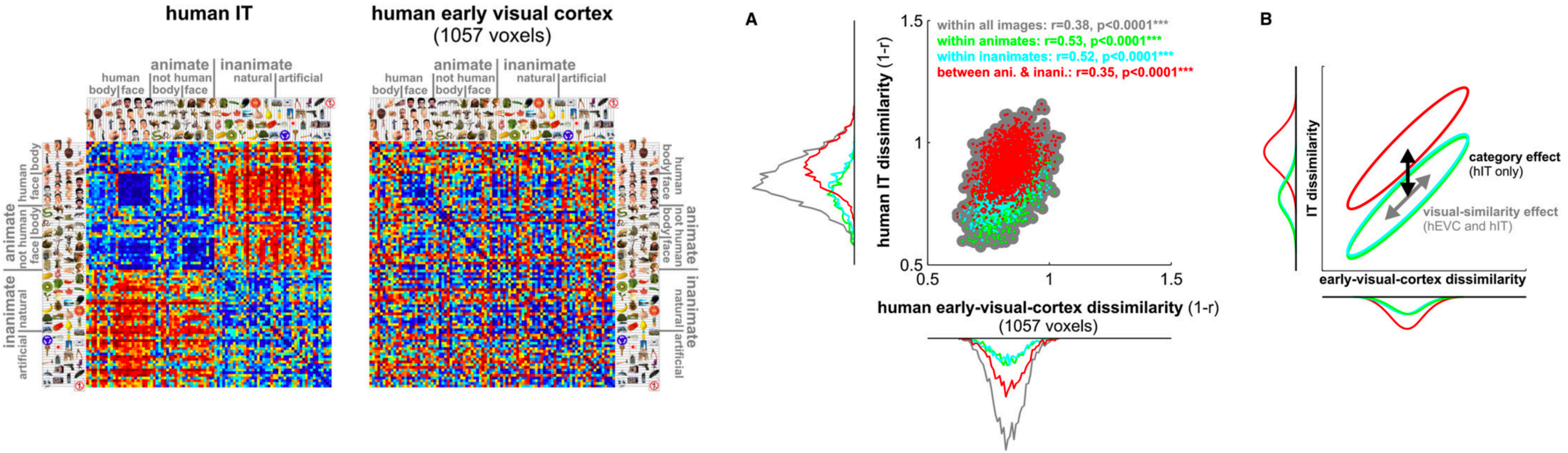
1. Representational connectivity: Comparing brain RDMs among different regions
2. Use classification performance as a distance metric
3. Searchlight RSA: Each region serves as a model
4. Conducting RSA in the GLM context

Other RSA options

Other analysis options

1. Representational connectivity: Comparing brain RDMs among different regions
2. Use classification performance as a distance metric
3. Searchlight RSA: Each region serves as a model
4. Conducting RSA in the GLM context

Example: 1. Representational connectivity



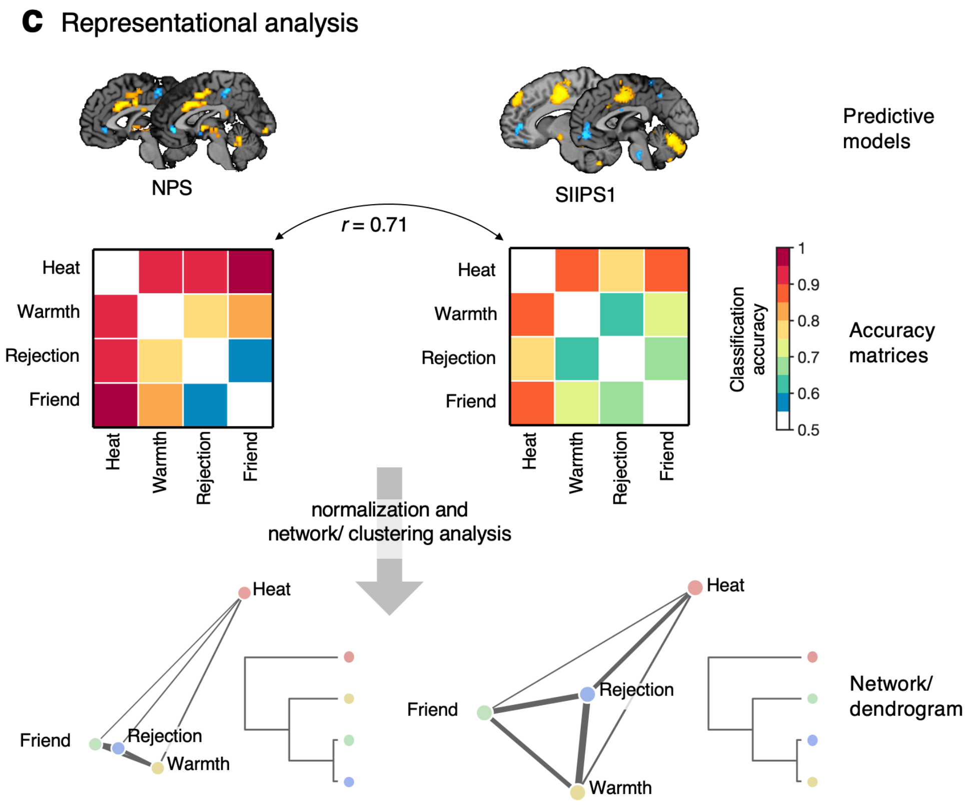
Kriegeskorte et al., 2008, *Neuron*

Other RSA options

Other analysis options

1. Representational connectivity: Comparing brain RDMs among different regions
2. Use classification performance as a distance metric
3. Searchlight RSA: Each region serves as a model
4. Conducting RSA in the GLM context

Example: 2. Accuracy as distance



Kohoutová et al., 2020, *Nat Protocols*

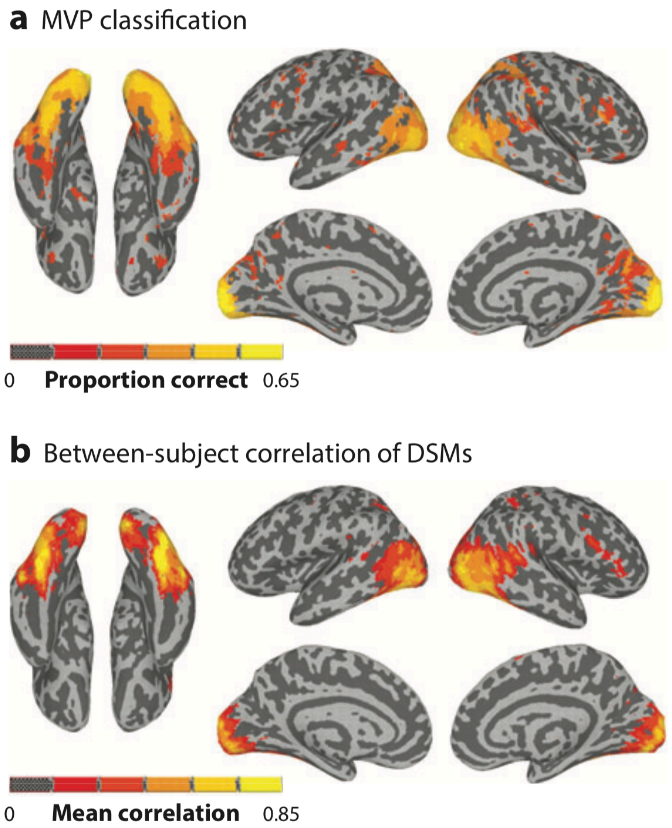
Other RSA options

Other analysis options

1. Representational connectivity: Comparing brain RDMs among different regions
2. Use classification performance as a distance metric
3. Searchlight RSA: Each region serves as a model
4. Conducting RSA in the GLM context

Nili et al., 2014, *PLoS Comp Biol*

Example: 3. Searchlight RSA

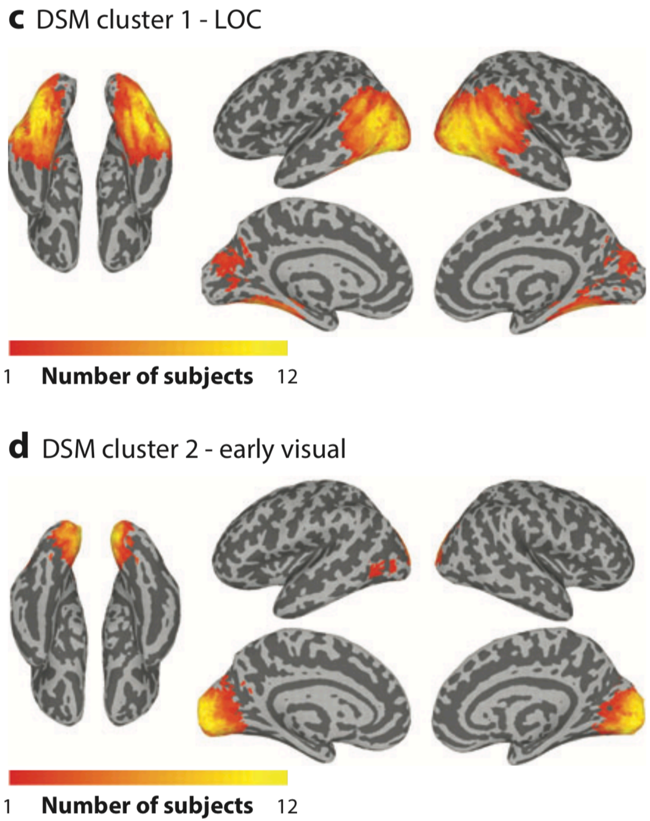


In addition to the MVPA, we can calculate the RDM (in this figure, DSM) for each person, and see between-subject consistency of the RDMs

Haxby et al., 2014, *Annu Rev Neurosci*

Example: 3. Searchlight RSA

We can also cluster the regions based on the RDM patterns (for each individual or for group)



Haxby et al., 2014, *Annu Rev Neurosci*

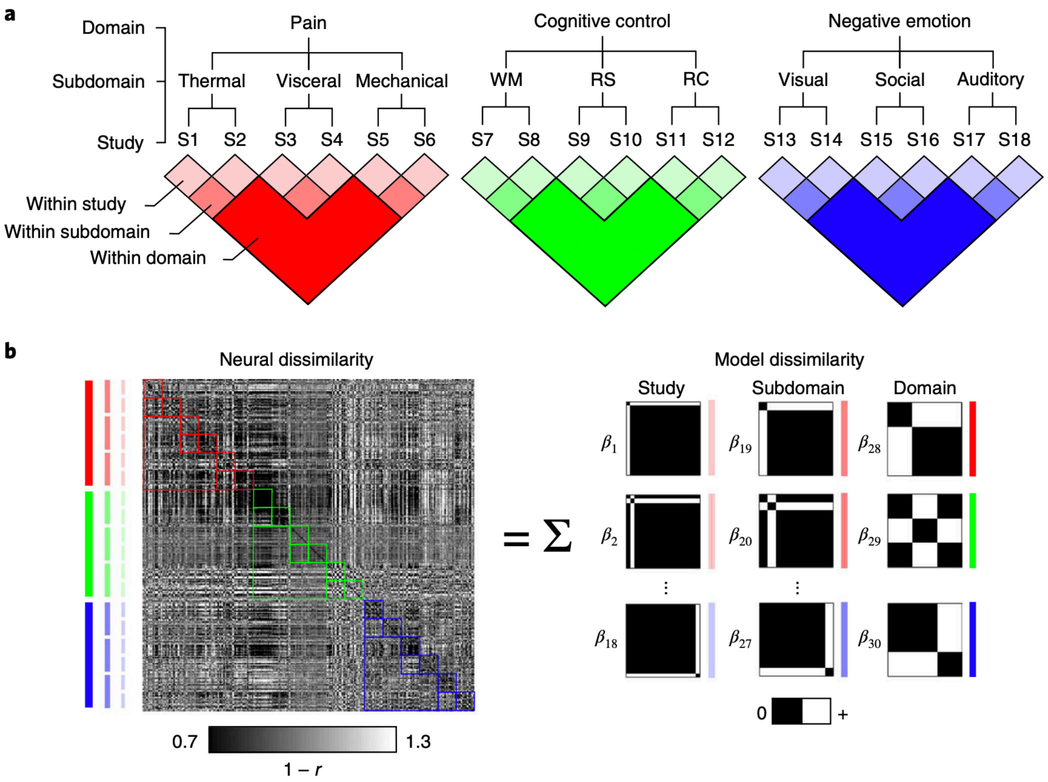
Other RSA options

Other analysis options

1. Representational connectivity: Comparing brain RDMs among different regions
2. Use classification performance as a distance metric
3. Searchlight RSA: Each region serves as a model
4. Conducting RSA in the GLM context

Nili et al., 2014, *PLoS Comp Biol*

Example: 4. RSA in the GLM context

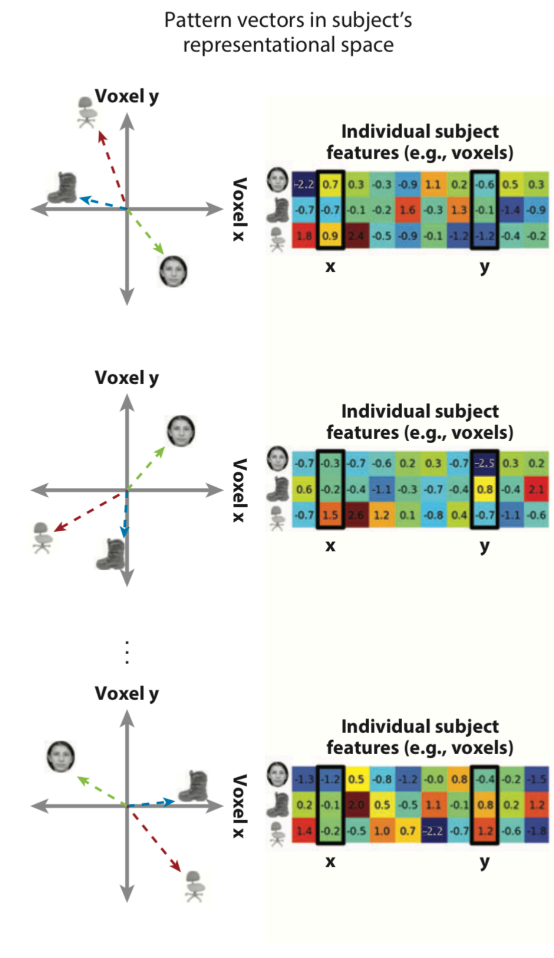


Kragel et al., 2018, *Nat Neurosci*

Additional relevant technique

Hyper-alignment:
aligning individual neural representational spaces into a common model space

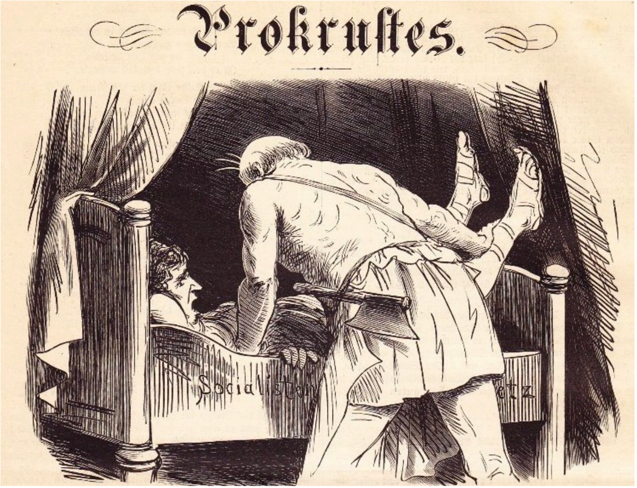
Iterative **procrustean transformation** to the common representational space



Not easy to directly compare the representational spaces across people

Haxby et al., 2014, *Annu Rev Neurosci;* Haxby et al., 2011, *Neuron*

Procrustes



From Greek mythology,

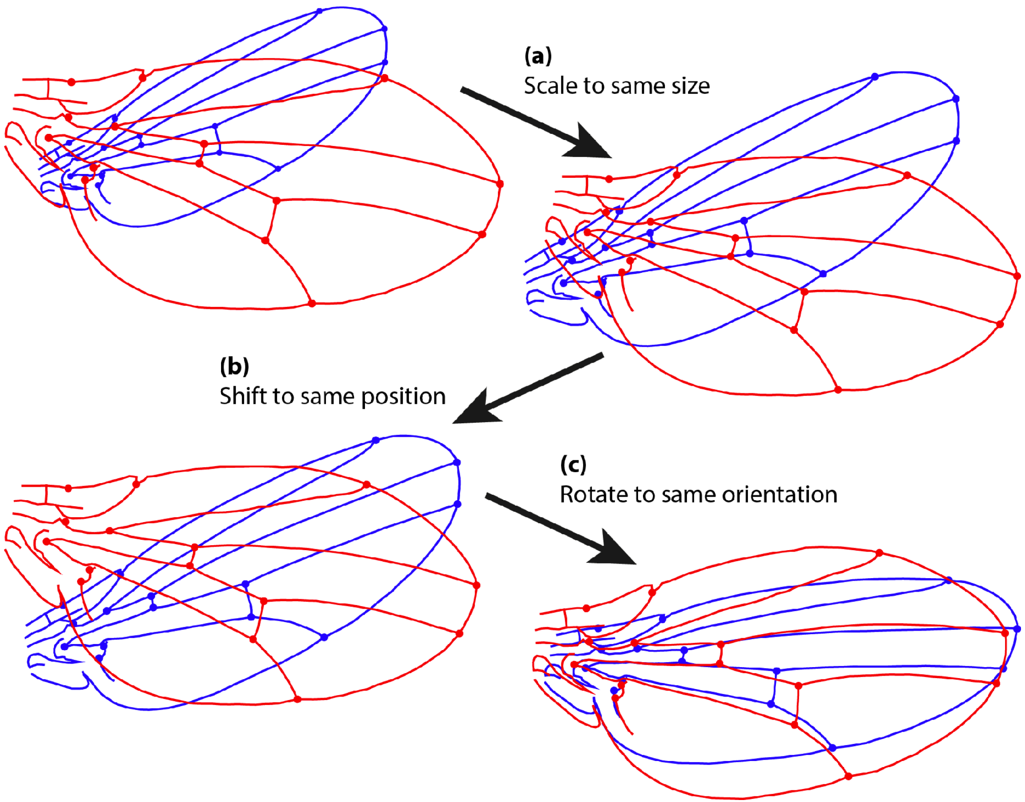
Procrustes (or “the stretcher”) was a bandit from Attica

He attached people by stretching them or cutting off their legs, so as to force them to fit the size of an iron bed

The word "Procrustean" is thus used to describe situations where different lengths or sizes or properties are fitted to an arbitrary standard.

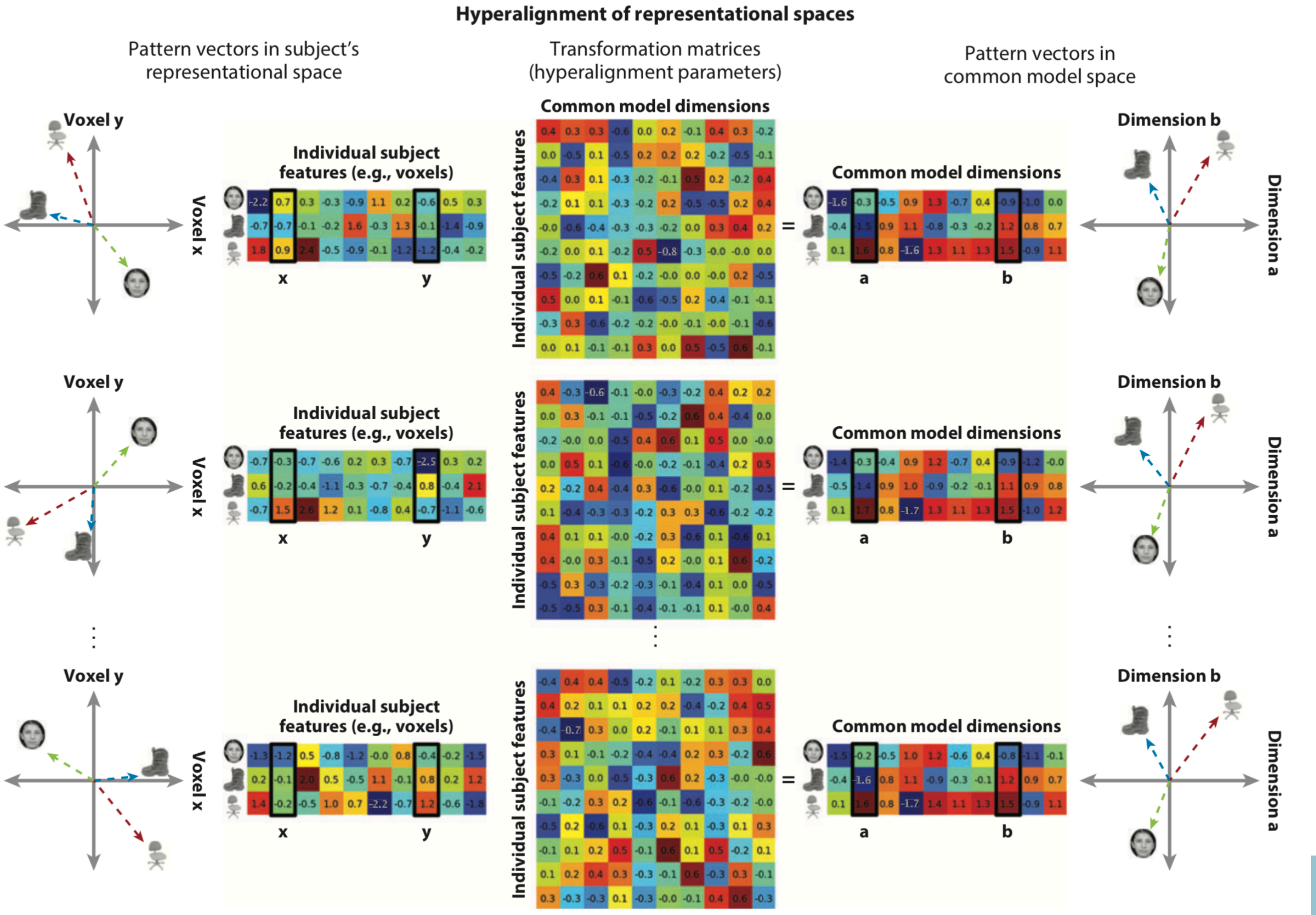
(from Wikipedia)
https://en.wikipedia.org/wiki/Procrustes

Procrustes analysis



Haxby et al., 2014, *Annu Rev Neurosci;* Haxby et al., 2011, *Neuron*

Hyperalign using procrustean transformation



Haxby et al., 2014, *Annu Rev Neurosci*

Thank you for your attention

Hope our representations about “RSA” to be well-aligned

RSA (5): Tutorial

Choong-Wan Woo
Director of the Cocoan Lab

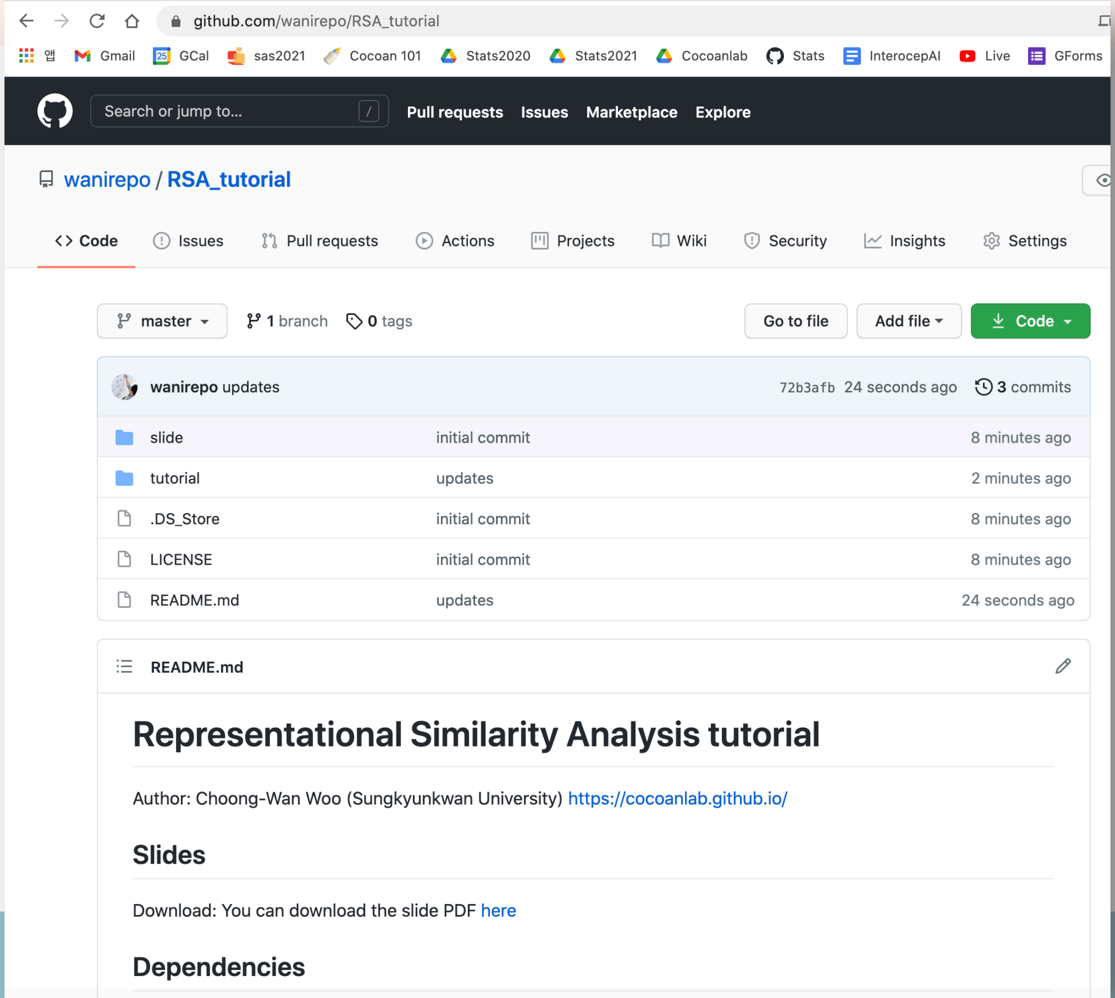
In the tutorial session:

https://github.com/wanirepo/RSA_tutorial

Step 1: Computing and visualizing RDMs

Step 2: Comparing brain and model RDMs

Step 3: Statistical inference



Data and research question:

nature COMMUNICATIONS

ARTICLE

Received 8 May 2014 | Accepted 25 Sep 2014 | Published 17 Nov 2014 DOI: 10.1038/ncomms6380

Separate neural representations for physical pain and social rejection

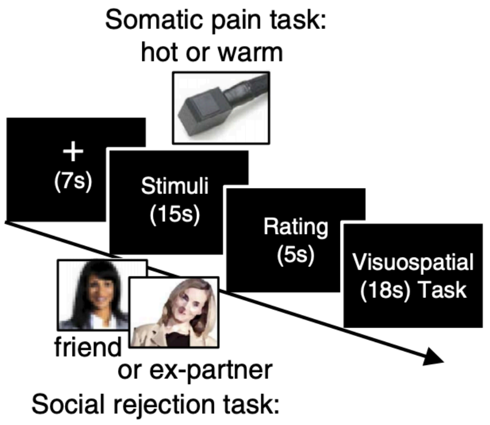
Choong-Wan Woo[1,2], Leonie Koban[1,2], Ethan Kross[3], Martin A. Lindquist[4], Marie T. Banich[1,2], Luka Ruzic[1,2], Jessica R. Andrews-Hanna[2] & Tor D. Wager[1,2]

Current theories suggest that physical pain and social rejection share common neural mechanisms, largely by virtue of overlapping functional magnetic resonance imaging (fMRI) activity. Here we challenge this notion by identifying distinct multivariate fMRI patterns unique to pain and rejection. Sixty participants experience painful heat and warmth and view photos of ex-partners and friends on separate trials. FMRI pattern classifiers discriminate pain and rejection from their respective control conditions in out-of-sample individuals with 92% and 80% accuracy. The rejection classifier performs at chance on pain, and vice versa. Pain- and rejection-related representations are uncorrelated within regions thought to encode pain affect (for example, dorsal anterior cingulate) and show distinct functional connectivity with other regions in a separate resting-state data set ($N = 91$). These findings demonstrate that separate representations underlie pain and rejection despite common fMRI activity at the gross anatomical level. Rather than co-opting pain circuitry, rejection involves distinct affective representations in humans.

Original question: **Can we identify specific patterns of fMRI activity for physical pain and social pain, respectively?**

- We conducted an fMRI experiment (N = 59) using somatic and social pain tasks.
- All 59 individuals (31 females, $M_{age} = 20.8$, $SD_{age} = 3.0$) recently experienced an unwanted break-up with their romantic partners and felt intensely rejected.

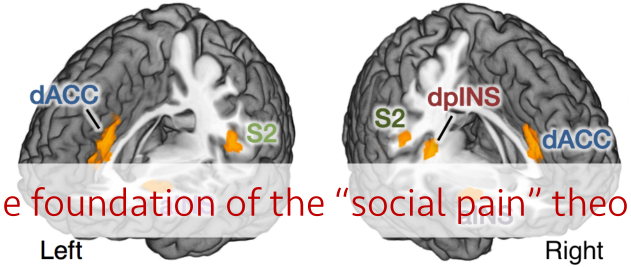
Tasks: Experimental paradigm



Data and research question:

Traditional mapping results

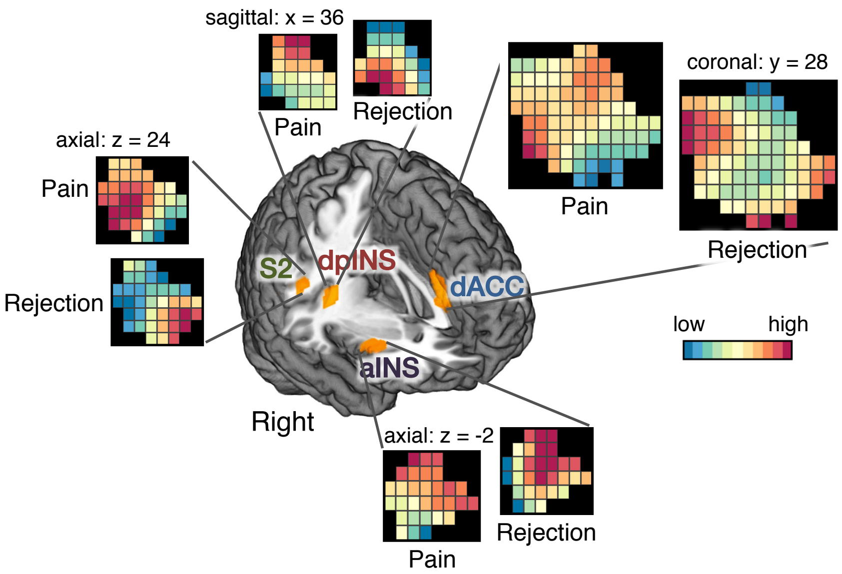
Univariate overlap
between [Heat-pain vs. Warmth] and [Ex-partner vs. Friend]

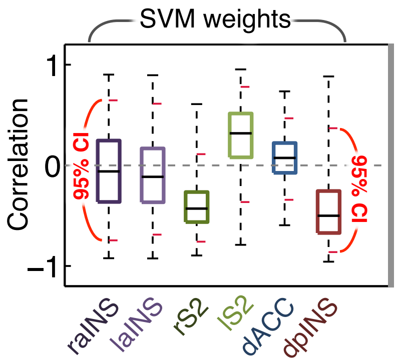


The foundation of the “social pain” theory

Then, we used the predictive mapping approach (linear support vector machines, SVMs) to obtain specific multivariate fMRI pattern for pain and rejection on the same dataset.

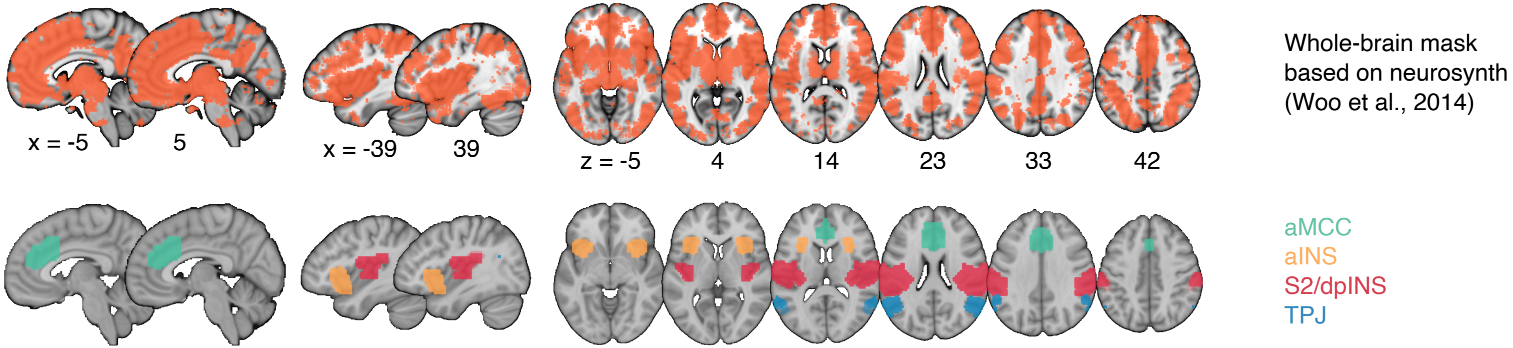
Data and research question:





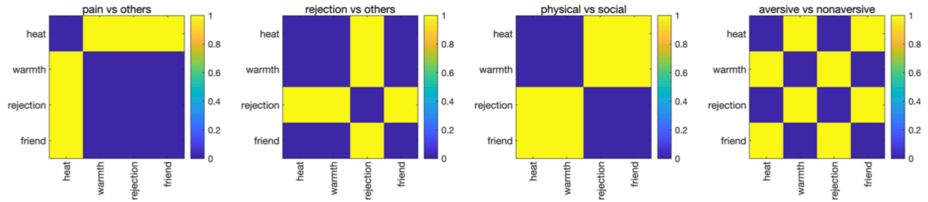
Data and research question:

1. Computing and visualizing RDMs for each participant, for each region



2. Computing four model RDMs with these ROI RDMs

1) Heat vs. others
2) Rejection vs. others
3) Physical vs. social
4) Aversive vs. non-aversive



Question: Which one of these is the best-supported model based on the representational similarity patterns from data?