Multivariate Pattern Analysis & Representational Similarity Analysis

27/08/24

Fernando Ramírez

www.fernando-ramirez.org | *Twitter*: @FRamirez_R2 | fernando.ramirez@nih.gov

National Institute of Mental Health

Goal today is to answer the following questions

➡ 1) What is meant by **MVPA** in *neuroscience*?

1.1) What kind of questions are addressed with MVPA?

1.2) Key limitations of MVPA

2) What is meant by **RSA** in *neuroscience*?

2.1) What kind of questions are addressed with RSA?

2.2) Key limitations of RSA

3) Some ways to mess up your MVPA & RSA analyses

4) Key challenges and future directions

National Institute of Mental Health

Goal today is to answer the following questions

1) What is meant by **MVPA** in *neuroscience*?

   1.1) What kind of questions are addressed with MVPA?

   1.2) Key limitations of MVPA

➡ 2) What is meant by **RSA** in *neuroscience*?

   2.1) What kind of questions are addressed with RSA?

   2.2) Key limitations of RSA

3) Some ways to mess up your MVPA & RSA analyses

4) Key challenges and future directions

Goal today is to answer the following questions

1) What is meant by **MVPA** in *neuroscience*?

    1.1) What kind of questions are addressed with MVPA?

    1.2) Key limitations of MVPA

2) What is meant by **RSA** in *neuroscience*?

    2.1) What kind of questions are addressed with RSA?

    2.2) Key limitations of RSA

➡ 3) Some ways to mess up your MVPA & RSA analyses

4) Key challenges and future directions

National Institute of Mental Health

Goal today is to answer the following questions

1) What is meant by **MVPA** in *neuroscience*?

    1.1) What kind of questions are addressed with MVPA?

    1.2) Key limitations of MVPA

2) What is meant by **RSA** in *neuroscience*?

    2.1) What kind of questions are addressed with RSA?

    2.2) Key limitations of RSA

3) Some ways to mess up your MVPA & RSA analyses

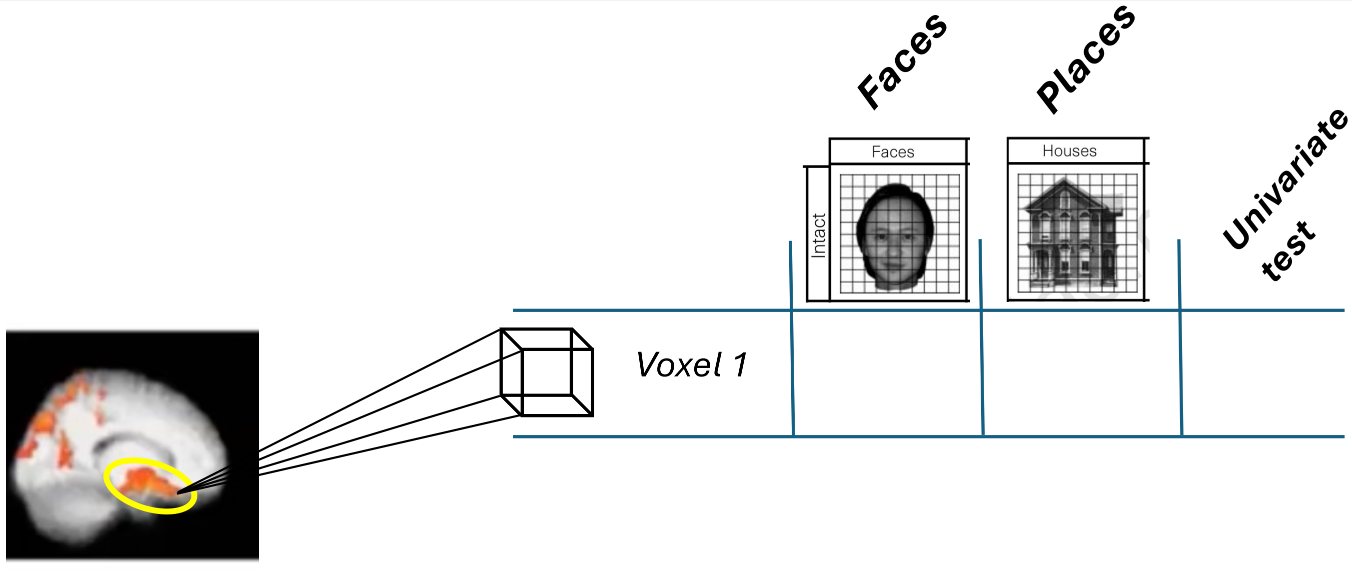
➡ 4) Key challenges and future directions

What is meant by MVPA in *neuroscience*?

First, let's review the concept of a "massively univariate" analysis of fMRI data *(e.g., Friston et al., 1994; 1995)*

Mass univariate analysis of fMRI data

MRI scanner



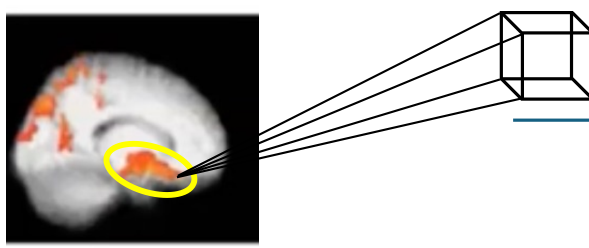
fMRI activation map

GLM: $\boldsymbol{y} = A\beta + \varepsilon$

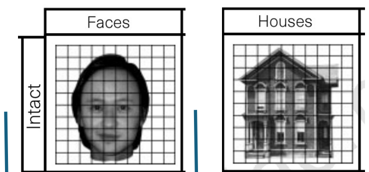
β_1	= *Faces regressor*
β_2	= *Places regressor*
ε	= *Residual error*

Mass univariate analysis of fMRI data

MRI scanner



fMRI activation map



	Faces	Places	Univariate test
Voxel 1	β_1 = 1.2		

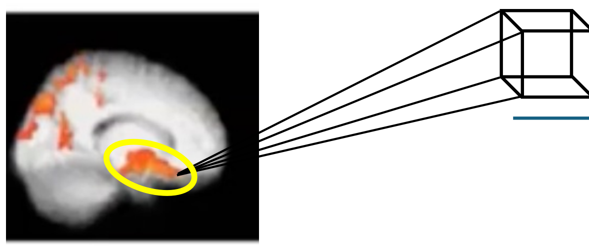
Mean response in voxel 1 for *Faces*

GLM: $y = A\beta + \varepsilon$

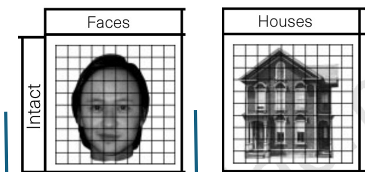
β_1 = *Faces regressor*
β_2 = *Places regressor*
ε = *Residual error*

Mass univariate analysis of fMRI data

MRI scanner



fMRI activation map



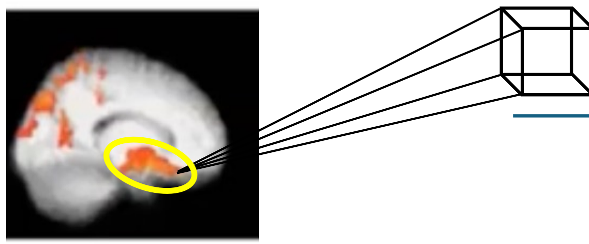
	Faces	Places	Univariate test
Voxel 1	β_1 = 1.2	β_2 = 3.2	
	Mean response in voxel 1 for *Faces*	Mean response in voxel 1 for *Places*	

GLM: $\boldsymbol{y} = A\boldsymbol{\beta} + \varepsilon$

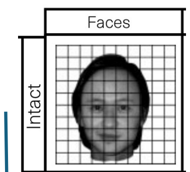
β_1 = *Faces regressor*
β_2 = *Places regressor*
ε = *Residual error*

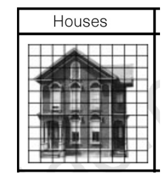
Mass univariate analysis of fMRI data

MRI scanner



fMRI activation map





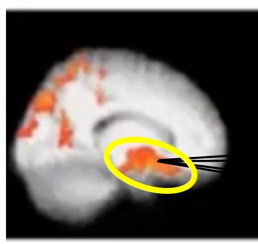
	Faces	Places	Univariate test
Voxel 1	β_1 = 1.2	β_2 = 3.2	p = 0.04
	Mean response in voxel 1 for *Faces*	Mean response in voxel 1 for *Places*	p-value voxel 1

GLM: $\boldsymbol{y} = A\boldsymbol{\beta} + \varepsilon$

β_1 = *Faces regressor*
β_2 = *Places regressor*
ε = *Residual error*

Mass univariate analysis of fMRI data

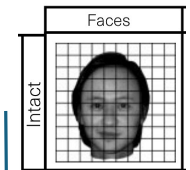
MRI scanner

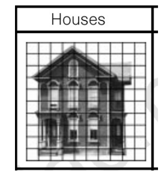


fMRI activation map

GLM: $\boldsymbol{y} = A\boldsymbol{\beta} + \varepsilon$

β_1 = *Faces regressor*
β_2 = *Places regressor*
ε = *Residual error*

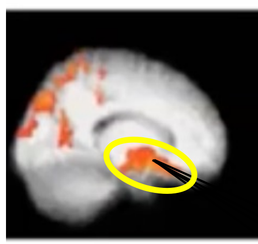




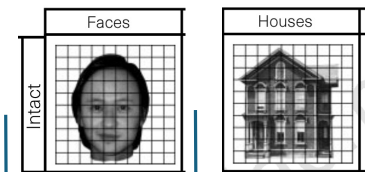
	Faces	**Places**	**t-test (univariate)**
Voxel 1	β_1 = 1.2	β_2 = 3.2	p = 0.05
Voxel 2	0.8	0.62	p = 0.45
	Mean response in voxel 2 for *Faces*	Mean response in voxel 2 for *Places*	p-value voxel 2

Mass univariate analysis of fMRI data

MRI scanner



fMRI activation map



	Faces	Places	t-test (univariate)
Voxel 1	$\beta_1 = 1.2$	$\beta_2 = 3.2$	$p = 0.05$
Voxel 2	0.8	0.62	$p = 0.45$
⋮	⋮	⋮	⋮
Voxel N	2.2	4.5	$p = 0.003$

GLM: $y = A\beta + \varepsilon$

β_1 = *Faces regressor*
β_2 = *Places regressor*
ε = *Residual error*

Although many thousands of voxels are usually analyzed with the univariate approach, ***each voxel*** *is analyzed* ***SEPARATELY***

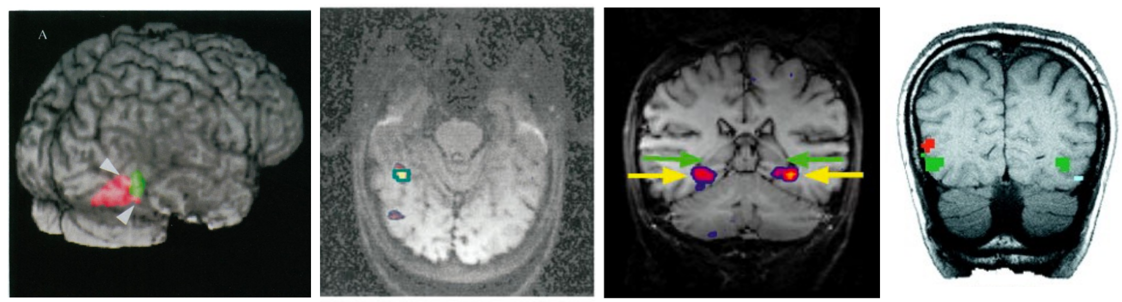
Mass univariate analysis of fMRI data

In the **90's,** this powerful approach to fMRI data analysis revealed multiple **object category selective areas** in the human brain

1995 1997 1998 2001

Mass univariate analysis of fMRI data

In the **90's,** this powerful approach to fMRI data analysis revealed multiple **object category selective areas** in the human brain



1995

LOC:
Lateral Occipital Complex

Malach et al.

1997

FFA:
Fusiform Face Area

Kanwisher et al.
McCarthy et al.

1998

PPA:
Parahippocampal Place Area

Epstein & Kanwisher

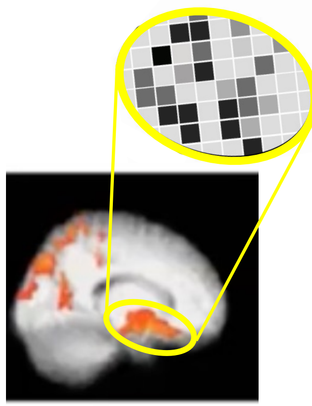
2001

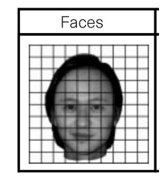
EBA:
Extrastriate Body Area

Downing et al.

What is meant by MVPA in *neuroscience*?

MVPA *stands for* **Multivoxel Pattern Analysis**



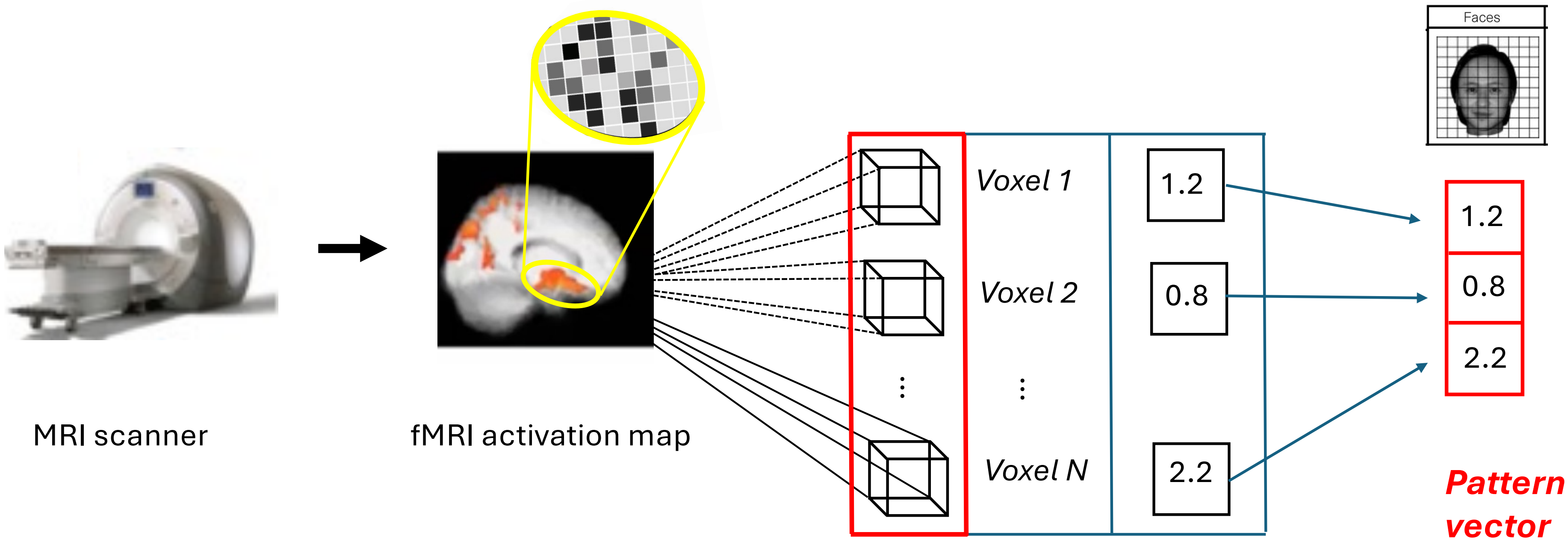


MRI scanner

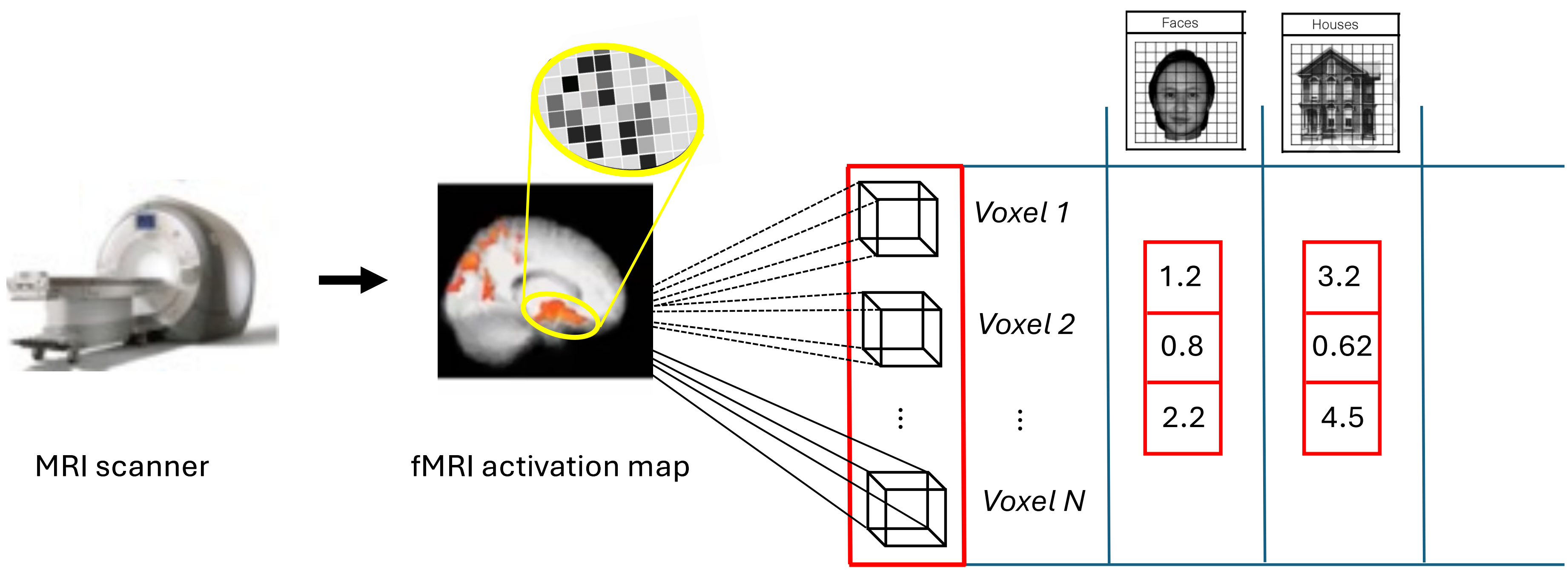
fMRI activation map

What is meant by MVPA in *neuroscience*?

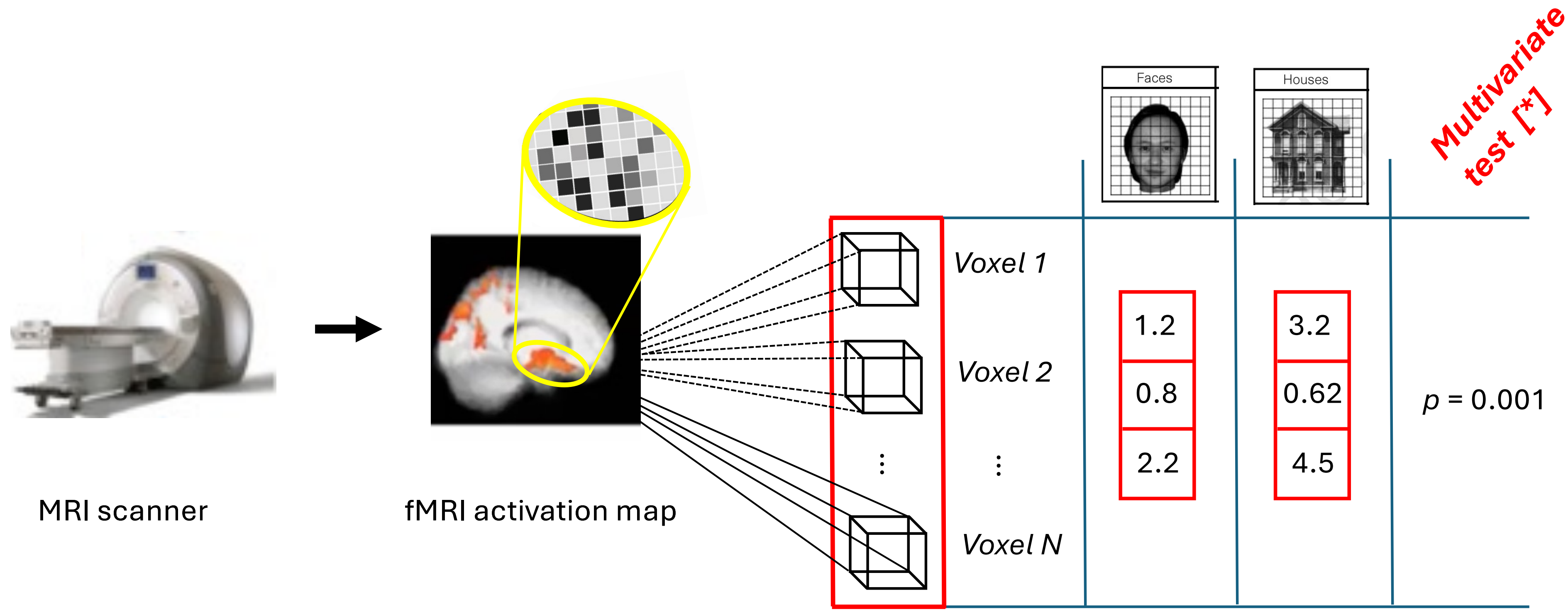
MVPA *stands for* **Multivoxel Pattern Analysis**



What is meant by MVPA in *neuroscience*?

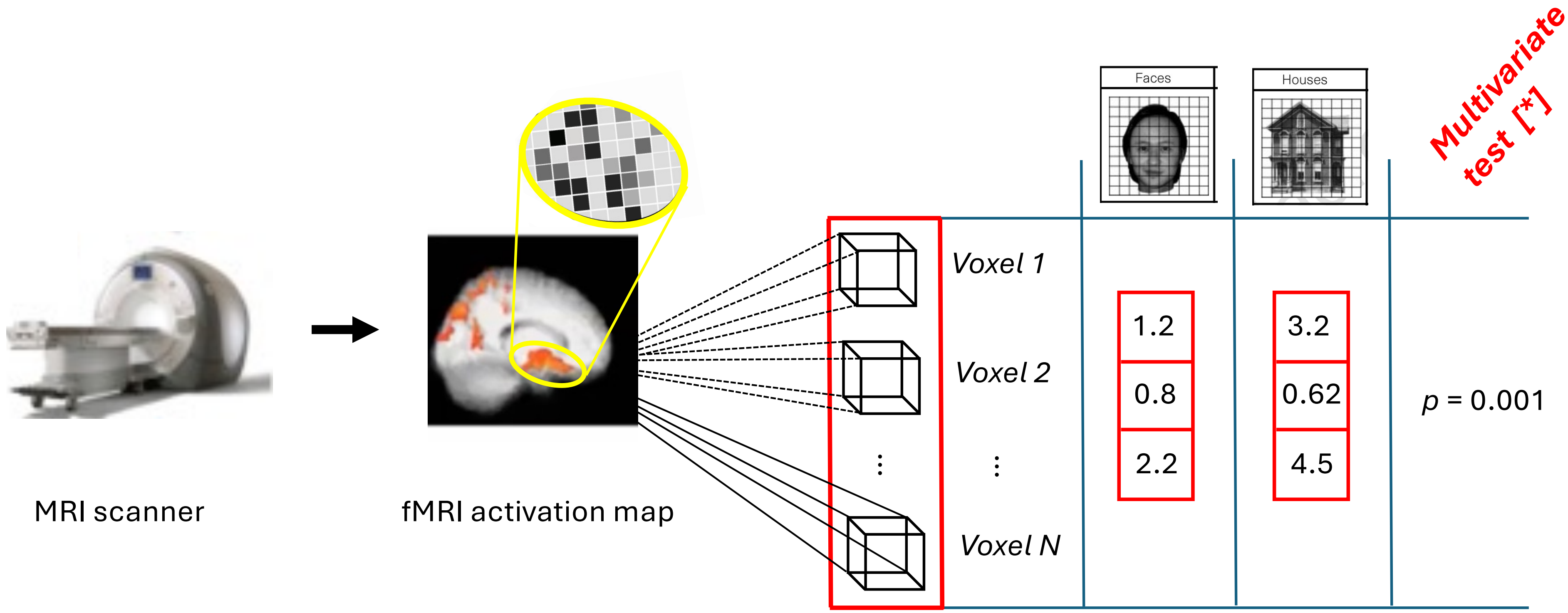


What is meant by MVPA in *neuroscience*?



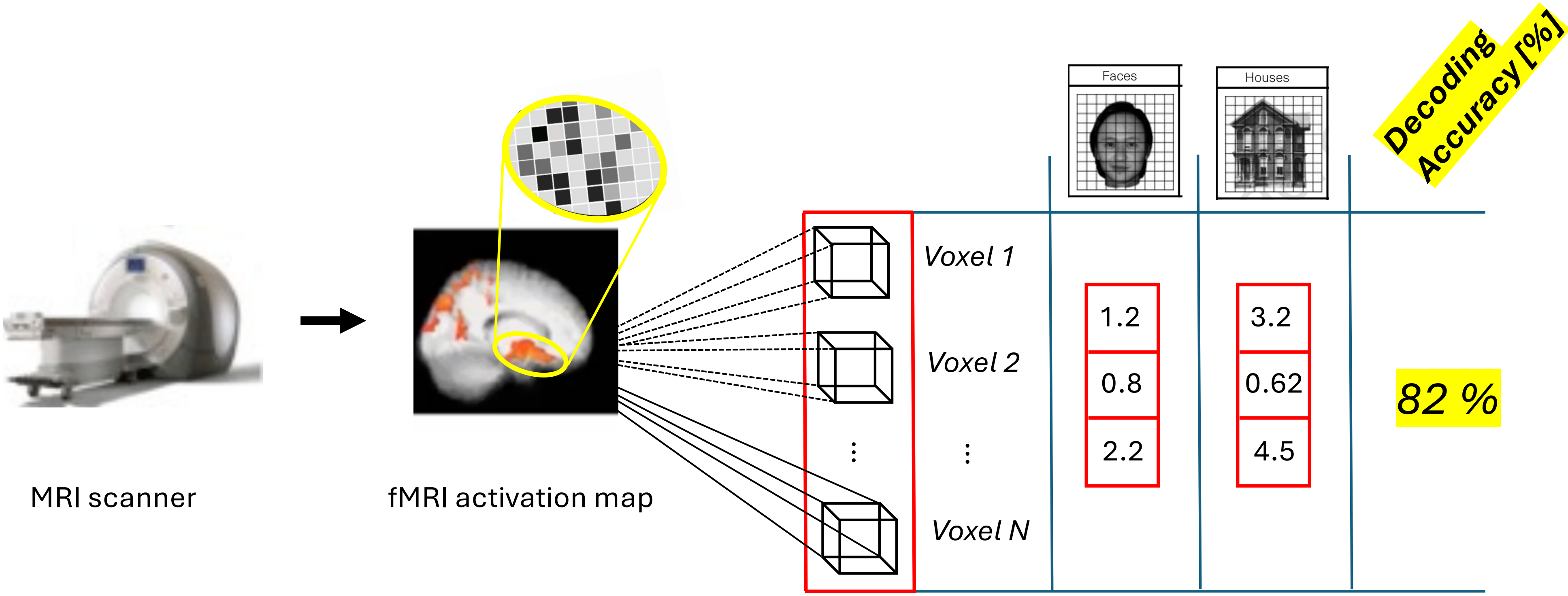
Unlike univariate analyses, in MVPA brain response ***patterns*** *across many voxels are* ***JOINTLY analyzed***

What is meant by MVPA in *neuroscience*?



If patterns differ significantly, we have **decoded** information about the experimental conditions from that set of voxels*

Why are "decoding accuracies" prevalent in MVPA studies?



Haxby et al (2001) *Science*; Spiridon & Kanwisher (2002) *Neuron*; Cox and Savoy (2003) *Neuroimage*
Kamitani and Tong (2005) *Nat Neurosci*, Haynes and Rees (2005) *Nat Neurosci*, O'Toole et al (2005)

Key concept 1: Curse of dimensionality

- The *curse* relates to difficulties that arise when number of datapoints is small relative to the dimensionality of the data
- Volume of a space increases rapidly with dimension, so available data become sparse
- Amount of data to obtain reliable results (e.g., estimate multivariate distributions) grows exponentially with dimension
- Sophisticated ML techniques, like SVMs, exist that perform well under such conditions

But, what is a *"Decoding Accuracy"*?

Decoding accuracies are obtained from a *classification procedure*.

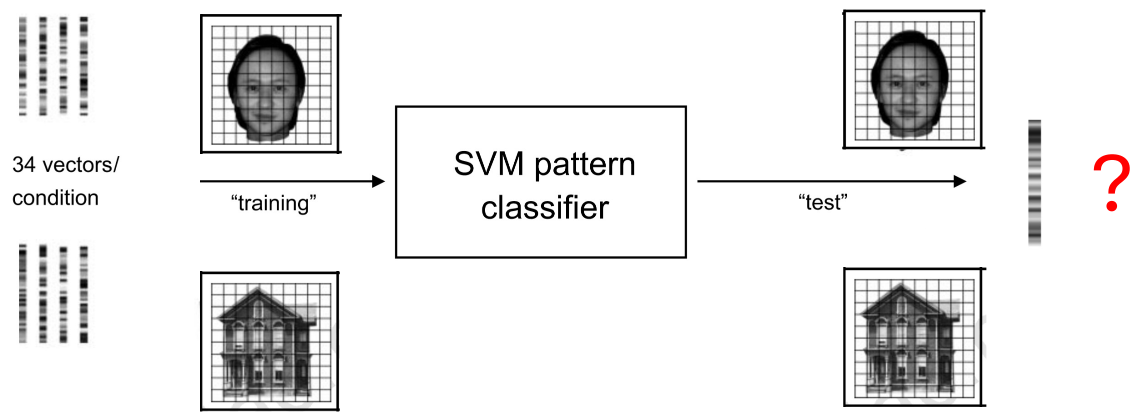
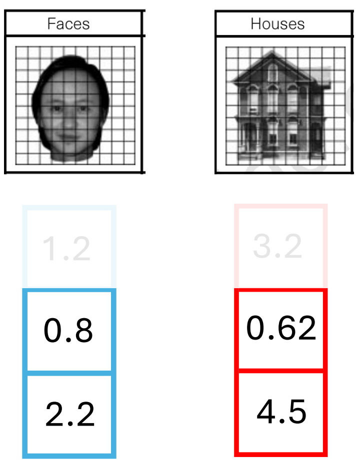


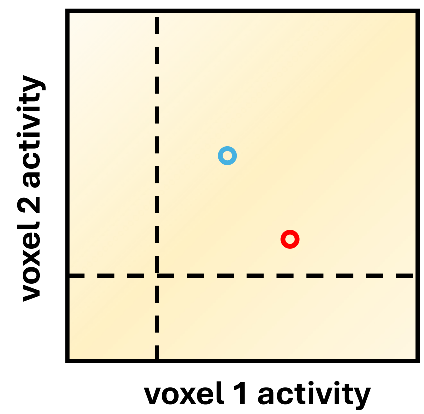
Fig. modified from Eger et al (2008) *J Cog Neurosci*

What is a “*Decoding Accuracy*”?

First, a **pattern classification model** is *trained* using one part of the data

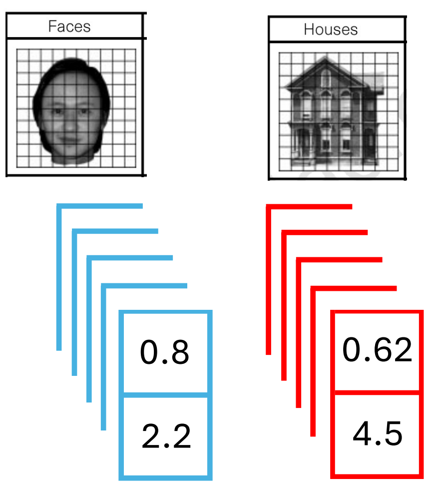


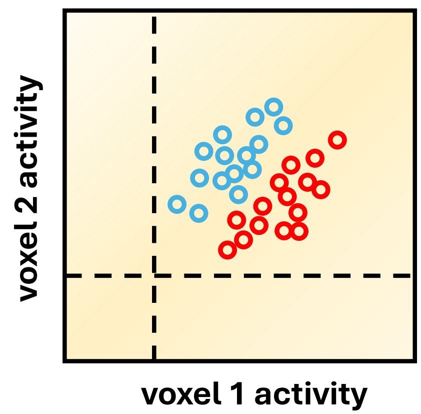
2-D representation of pattern classifier



What is a *“Decoding Accuracy”*?

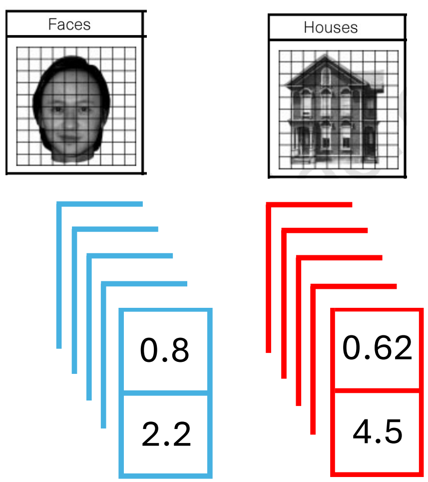
First, a **classification model** is *trained* using one part of the data

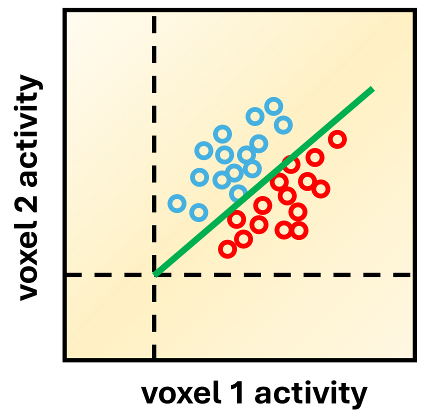




What is a *“Decoding Accuracy”*?

First, a **classification model** is *trained*
using one part of the data





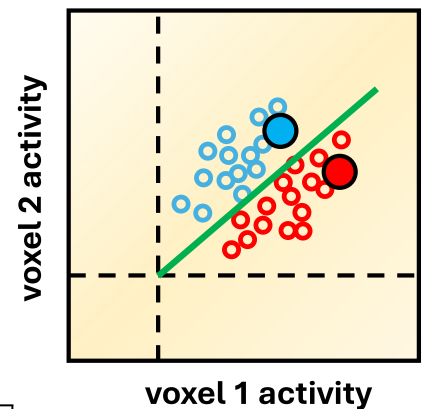
The model:
A hyperplane

What is a *"Decoding Accuracy"*?

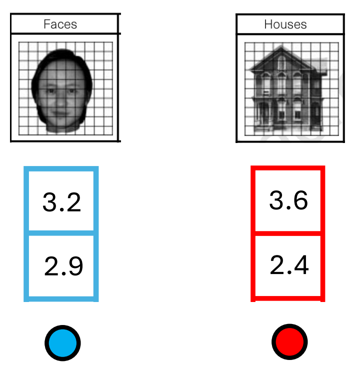
First, a **classification model** is *trained* using one part of the data

Next, the **model** is *tested* on data not used to train the model

Procedure usually iterated with different training and test data partitions
(e.g., "leave-one out" cross-validation)



Yes! *Both patterns correctly classified*



What is a *"Decoding Accuracy"*?

First, a **classification model** is *trained* on one part of the data

Next, the model is *tested* on data not used to train the model

If **model** is any good, *"decoding accuracy"* should be *"above chance"*
(E.g., for a 2-class problem = 50%)

😥

$$\frac{\textit{Correct classifications}}{\textit{All classifications}} = \textit{\textasciitilde 50\% (chance)}$$

What is a *"Decoding Accuracy"*?

First, a **classification model** is *trained* on one part of the data

Next, the model is *tested* on data not used to train the model

If **model** is any good, *"decoding accuracy"* should be *"above chance"* (E.g., for a 2-class problem = 50%)

$$\frac{\textit{Correct classifications}}{\textit{All classifications}} = \textit{100\% (perfect)}$$

What is a *“Decoding Accuracy”*?

First, a **classification model** is *trained* on one part of the data

Next, the model is *tested* on data not used to train the model

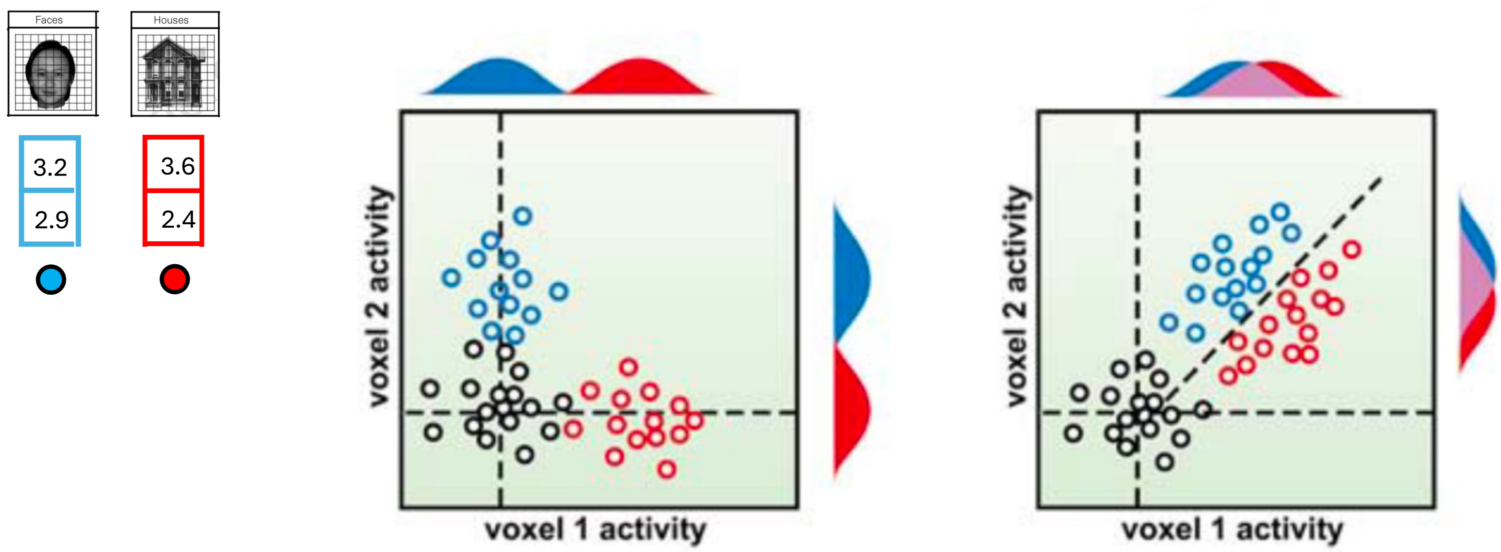
If **model** is any good, *“decoding accuracy”* should be *“above chance”*
(E.g., for a 2-class problem = 50%)

$$\frac{\textit{Correct classifications}}{\textit{All classifications}} = \textit{82.1\% (celebrate)}$$

🕺

Key concept 2: Why is MVPA here to stay?

- <u>Because</u> **multivariate methods** are sensitive to regularities in the data that are *in principle undetectable* (and hence forever hidden) for univariate methods



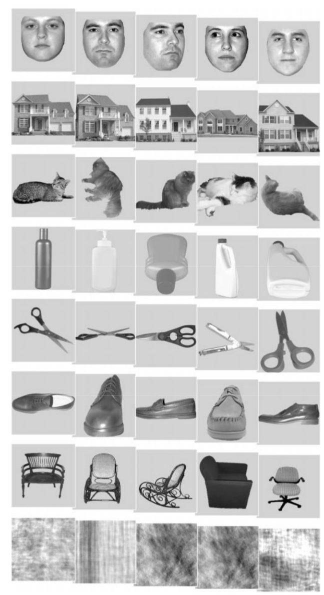
Category-selective *modules* or distributed brain *patterns*?

RESEARCH ARTICLES

Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex

James V. Haxby,[1*] M. Ida Gobbini,[1,2] Maura L. Furey,[1,2] Alumit Ishai,[1] Jennifer L. Schouten,[1] Pietro Pietrini[3]

The functional architecture of the object vision pathway in the human brain was investigated using functional magnetic resonance imaging to measure patterns of response in ventral temporal cortex while subjects viewed faces, cats, five categories of man-made objects, and nonsense pictures. A distinct pattern of response was found for each stimulus category. The distinctiveness of the response to a given category was not due simply to the regions that responded maximally to that category, because the category being viewed also could be identified on the basis of the pattern of response when those regions were excluded from the analysis. Patterns of response that discriminated among all categories were found even within cortical regions that responded maximally to only one category. These results indicate that the representations of faces and objects in ventral temporal cortex are widely distributed and overlapping.



Haxby et al (2001) *Science*

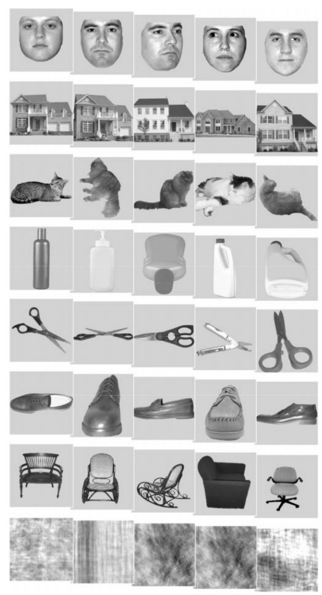
Category-selective *modules* or distributed brain *patterns*?

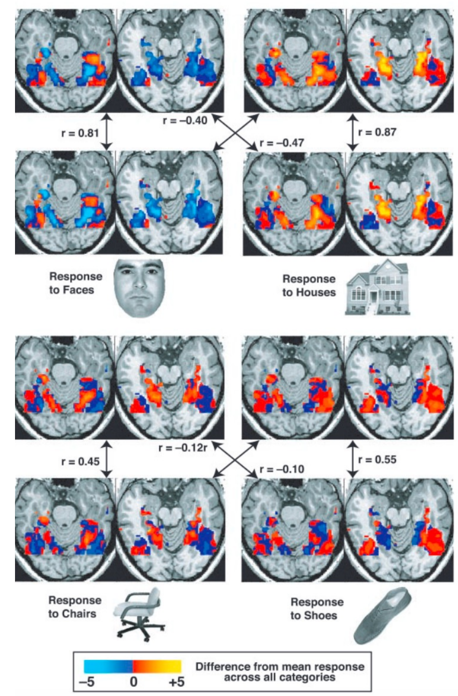
RESEARCH ARTICLES

Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex

James V. Haxby,[1*] M. Ida Gobbini,[1,2] Maura L. Furey,[1,2] Alumit Ishai,[1] Jennifer L. Schouten,[1] Pietro Pietrini[3]

The functional architecture of the object vision pathway in the human brain was investigated using functional magnetic resonance imaging to measure patterns of response in ventral temporal cortex while subjects viewed faces, cats, five categories of man-made objects, and nonsense pictures. A distinct pattern of response was found for each stimulus category. The distinctiveness of the response to a given category was not due simply to the regions that responded maximally to that category, because the category being viewed also could be identified on the basis of the pattern of response when those regions were excluded from the analysis. Patterns of response that discriminated among all categories were found even within cortical regions that responded maximally to only one category. These results indicate that the representations of faces and objects in ventral temporal cortex are widely distributed and overlapping.





Haxby et al (2001) *Science*

Decoding grating orientation from human visual cortex

Predicting the orientation of invisible stimuli from activity in human primary visual cortex

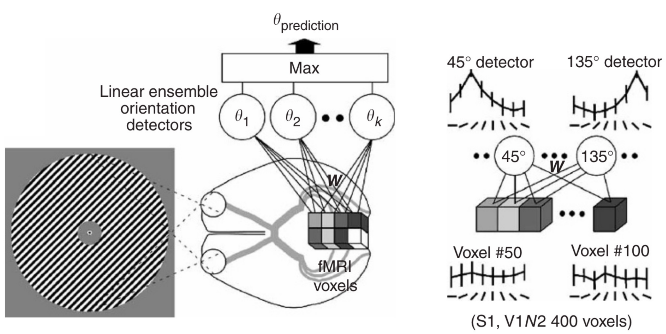
John-Dylan Haynes[1,2] & Geraint Rees[1,2]

Humans can experience aftereffects from oriented stimuli that are not consciously perceived, suggesting that such stimuli receive cortical processing. Determining the physiological substrate of such effects has proven elusive owing to the low spatial resolution of conventional human neuroimaging techniques compared to the size of orientation columns in visual cortex. Here we show that even at conventional resolutions it is possible to use fMRI to obtain a direct measure of orientation-selective processing in V1. We found that many parts of V1 show subtle but reproducible biases to oriented stimuli, and that we could accumulate this information across the whole of V1 using multivariate pattern recognition. Using this information, we could then successfully predict which one of two oriented stimuli a participant was viewing, even when masking rendered that stimulus invisible. Our findings show that conventional fMRI can be used to reveal feature-selective processing in human cortex, even for invisible stimuli.

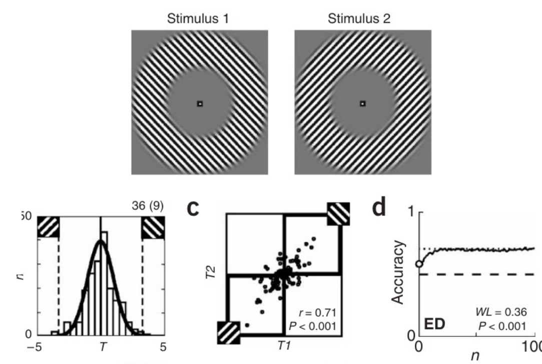
Decoding the visual and subjective contents of the human brain

Yukiyasu Kamitani[1] & Frank Tong[2,3]

The potential for human neuroimaging to read out the detailed contents of a person's mental state has yet to be fully explored. We investigated whether the perception of edge orientation, a fundamental visual feature, can be decoded from human brain activity measured with functional magnetic resonance imaging (fMRI). Using statistical algorithms to classify brain states, we found that ensemble fMRI signals in early visual areas could reliably predict on individual trials which of eight stimulus orientations the subject was seeing. Moreover, when subjects had to attend to one of two overlapping orthogonal gratings, feature-based attention strongly biased ensemble activity toward the attended orientation. These results demonstrate that fMRI activity patterns in early visual areas, including primary visual cortex (V1), contain detailed orientation information that can reliably predict subjective perception. Our approach provides a framework for the readout of fine-tuned representations in the human brain and their subjective contents.



Kamitani and Tong (2005) *Nat Neurosci*



Haynes and Rees (2005) *Nat Neurosci*

Mind reading?

NEUROIMAGING

Decoding mental states from brain activity in humans

John-Dylan Haynes[*‡§] *and Geraint Rees*[‡§]

Abstract | Recent advances in human neuroimaging have shown that it is possible to accurately decode a person's conscious experience based only on non-invasive measurements of their brain activity. Such 'brain reading' has mostly been studied in the domain of visual perception, where it helps reveal the way in which individual experiences are encoded in the human brain. The same approach can also be extended to other types of mental state, such as covert attitudes and lie detection. Such applications raise important ethical issues concerning the privacy of personal thought.

Haynes and Rees (2006) *Nat Rev Neurosci*

Searchlight analysis

Information-based functional brain mapping

Nikolaus Kriegeskorte*†, Rainer Goebel‡, and Peter Bandettini*

*Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, Building 10, Room 1D80B, 10 Center Drive MSC 1148, Bethesda, MD 20892-1148; and ‡Department of Cognitive Neuroscience, Faculty of Psychology, Universiteit Maastricht, Universiteitssingel 40, 6229 ER, Maastricht, The Netherlands

Communicated by Leslie G. Ungerleider, National Institutes of Health, Bethesda, MD, January 10, 2006 (received for review March 19, 2005)

The development of high-resolution neuroimaging and multielectrode electrophysiological recording provides neuroscientists with huge amounts of multivariate data. The complexity of the data creates a need for statistical summary, but the local averaging standardly applied to this end may obscure the effects of greatest neuroscientific interest. In neuroimaging, for example, brain mapping analysis has focused on the discovery of activation, i.e., of extended brain regions whose average activity changes across experimental conditions. Here we propose to ask a more general question of the data: Where in the brain does the activity pattern contain information about the experimental condition? To address this question, we propose scanning the imaged volume with a "searchlight," whose contents are analyzed multivariately at each location in the brain.

neuroimaging | functional magnetic resonance imaging | statistical analysis

coregistered in Talairach space (see ref. 5), where corresponding functional regions can be off by many millimeters between subjects. In a typical group analysis, data are spatially smoothed by convolution with a Gaussian kernel of 8-mm full width at half maximum.

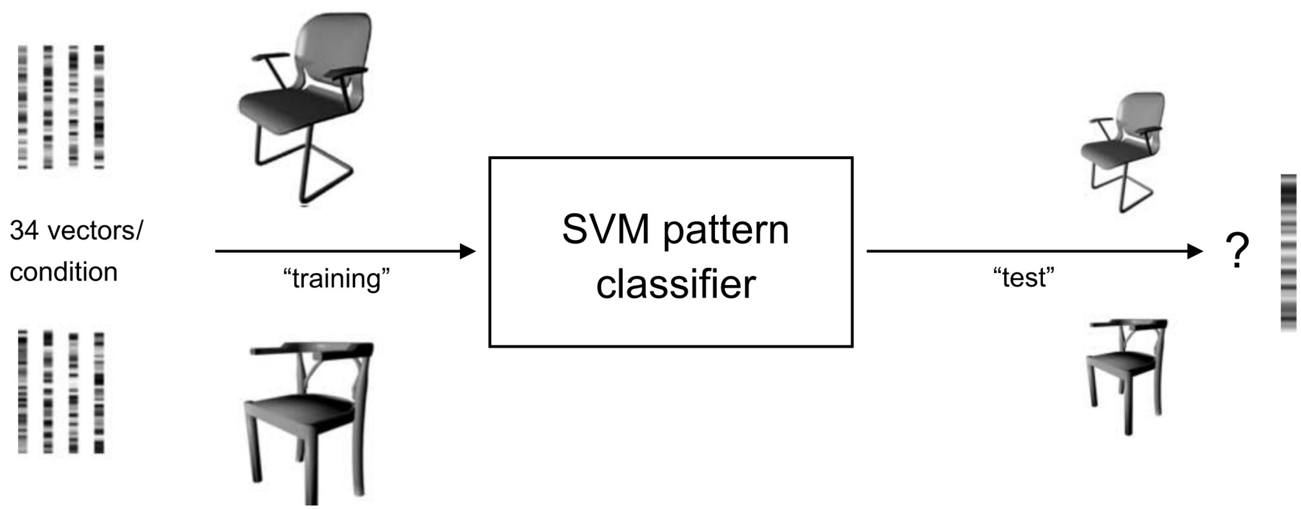
Although smoothing greatly reduces the information content of the data, the local combination of signals it provides is necessary. If smoothing is omitted in a standard voxelwise univariate fMRI analysis, statistical sensitivity suffers, and less of the activated volume is detected (Fig. 2). Upon lowering the threshold, the activation maps show salt-and-pepper patterns, which are hard to distinguish from noise, inconsistent across subjects, and impossible to report verbally. Nevertheless, these fine-scale patterns of weak effects may contain neuroscientifically relevant information.

The amount of information removed by smoothing fMRI data to the scale of functional regions increases with growing spatial

Kriegeskorte, Goebel, and Bandettini (2006) *PNAS*

Cross-decoding

- *Probing for brain representations that generalize across stimulus transformations*



e.g., *Train* BIG, *test* small

Fig. from Eger et al (2008) *J Cog Neurosci*

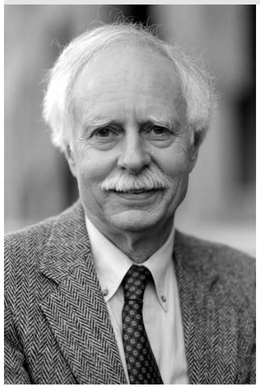
Ramírez et al (2010) *J Vision*; Cichy et al (2010) *J Vision;* Cichy et al (2012) *Neuroimage*; Ramírez et al (2014) *J Neurosci*

MVPA: Interim conclusions:

- MVPA *can* detect information invisible to univariate analysis methods
- MVPA *does not* tell us HOW information encoded in the brain
- MVPA *does* tell that information about our experimental conditions is present in the set of voxels under study

Representational Similarity Analysis

Representational Similarities

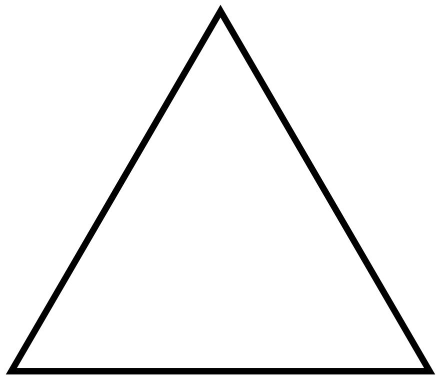


R. Shepard

Psychology & Philosophy

R. Shepard (70s)
- Representational spaces
- 2^{nd} order isomorphism
- Non-metric Multidimensional Scaling (MDS)

P. Churchland vs J. Fodor (80s)
- Connectionist semantics



Neuroscience

S. Ramón y Cajal, 1906
D. Hebb, 1949
Hubel & Wiesel, 1962
C. Gross 1970s

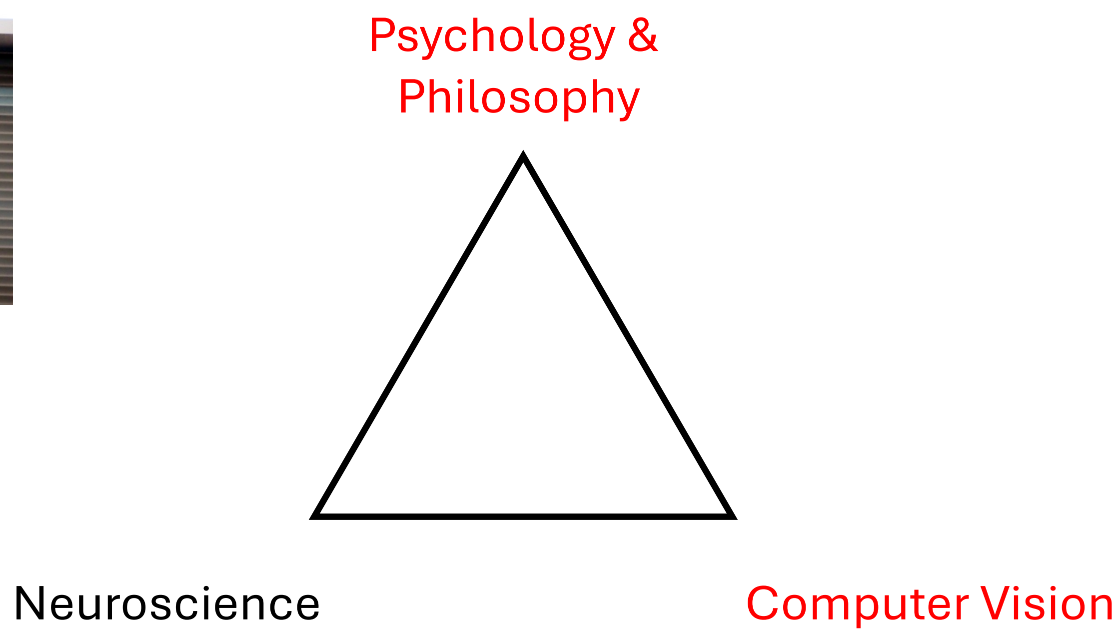
Computer Vision

Artificial Neural Networks (70s)
David Marr (1982) *Vision*
- Object categorization & recognition
- Viewpoint invariance

Multi-layer neural networks (90s)

Representational Similarities

Shimon Edelman



Poggio and Edelman (1990) *Nature*
A network that learns to recognize three-dimensional objects.

Representational Similarities

Shimon Edelman

"Representation is representation of similarities"

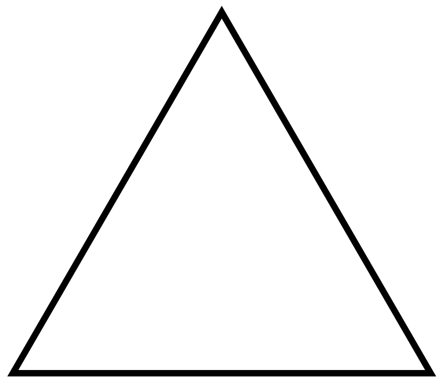
World ⟷ Brain

2nd order isomorphism

Representational Similarities

Shimon Edelman

Psychology & Philosophy



Neuroscience

Computer Vision

Keiji Tanaka 1990s
Shape-tuned cortical columns
in IT cortex (optical imaging)

Psychobiology
1998, 26 (4), 309-321

Toward direct visualization of the internal shape representation space by fMRI

SHIMON EDELMAN
University of Sussex, Falmer, England

KALANIT GRILL-SPECTOR
Weizmann Institute of Science, Rehovot, Israel

TAMMAR KUSHNIR
The Chaim Sheba Medical Center, Tel Hashomer, Israel

and

RAFAEL MALACH
Weizmann Institute of Science, Rehovot, Israel

Reports of columnar organization of the macaque inferotemporal cortex (Tanaka, 1992, 1993a) indicate that ensembles of cells responding to particular objects may be both sufficiently extensive and properly localized to allow their detection and discrimination by means of functional magnetic resonance imaging (fMRI). A recently developed theory of object representation by ensembles of coarsely tuned units (Edelman, 1998; Edelman & Duvdevani-Bar, 1997b) and its implementation as a computer model of recognition and categorization (Cutzu & Edelman, 1998; Edelman & Duvdevani-Bar, 1997a) provide a computational framework in which such findings can be interpreted in a straightforward fashion. Taken together, these developments in the study of object representation and recognition suggest that direct visualization of the internal representations may be easier than was previously thought. In this paper, we show how fMRI techniques can be used to investigate the internal representation of objects in the human visual cortex. Our initial results reveal that the activation of most voxels in object-related areas remains unaffected by a coarse scrambling of the natural images used as stimuli and that a map of the representation space of object categories in individual subjects can be derived from the distributed pattern of voxel activation in those areas.

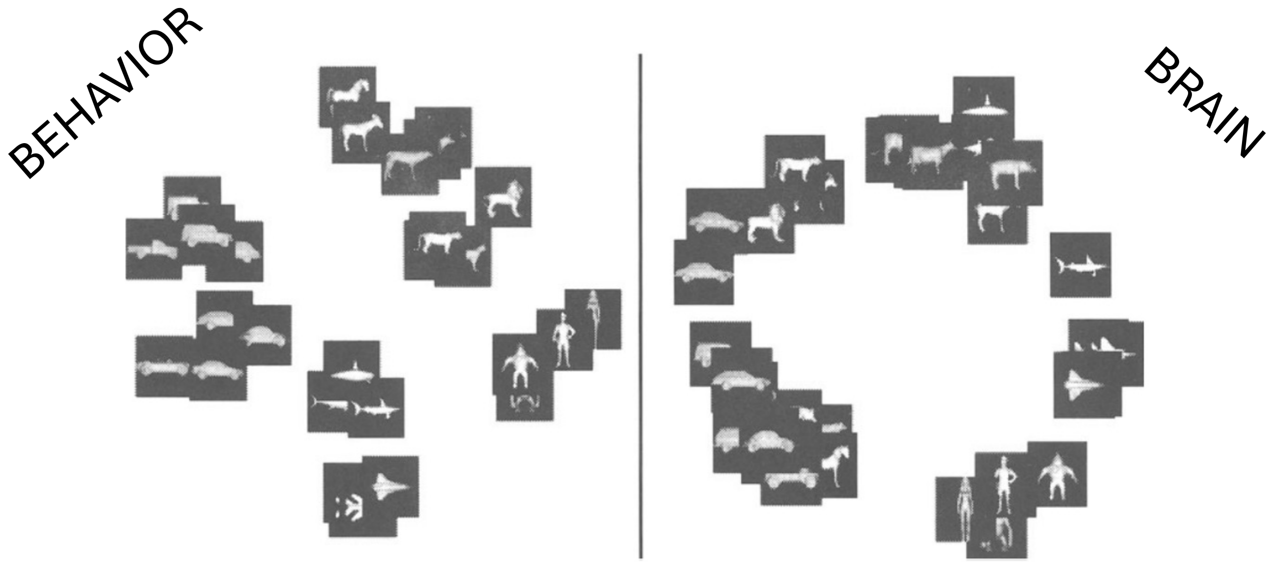
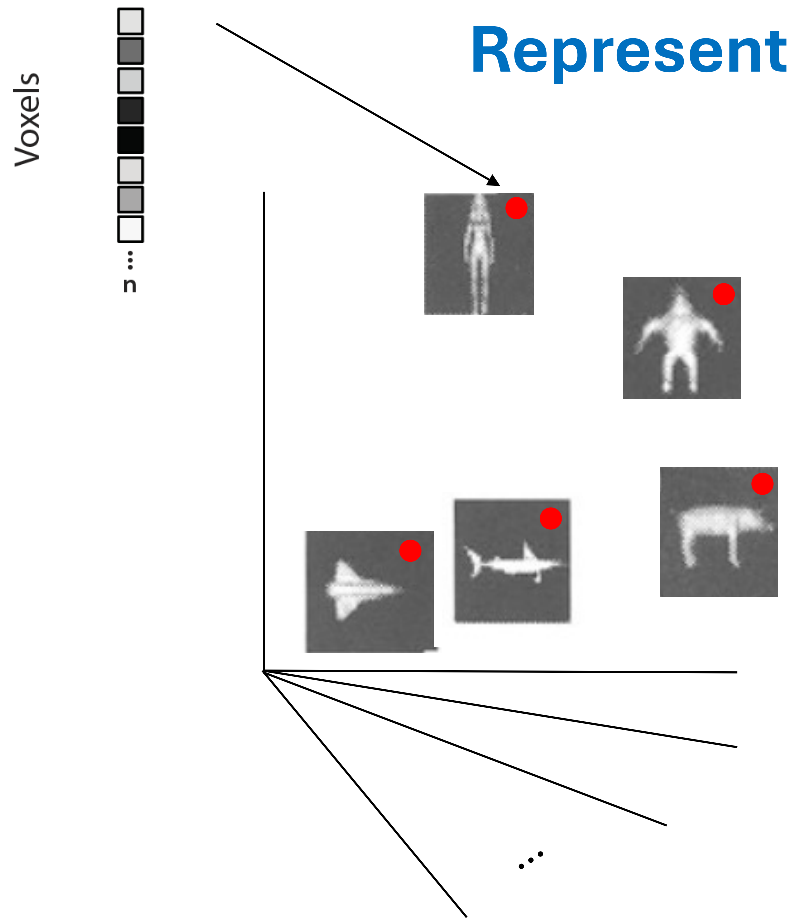
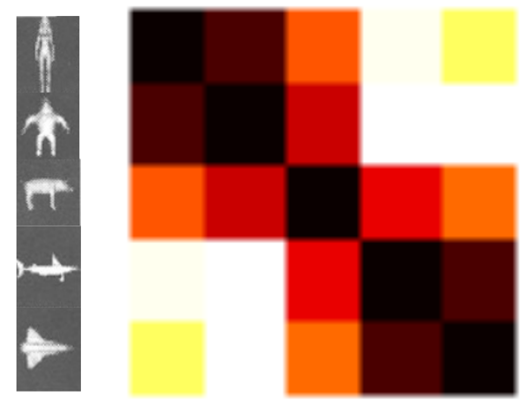


Figure 6. A comparison between the layout of the object representation space derived by multidimensional scaling (MDS; see Shepard, 1980) from perceptual judgment of similarities among objects (left) and that from the similarities among activation patterns measured by fMRI (right). Left: A two-dimensional MDS configuration of the 32 objects, recovered from the *psychophysically determined* dissimilarity matrix combined from all subjects. Right: A two-dimensional MDS configuration of the 32 objects, recovered from the combined *voxel-space representation* derived from the most significant object-related voxels in all 7 subjects. Although the MDS configuration derived from the voxel-space representation is noisier than the configuration retrieved from the psychophysical test, there is some clustering according to object categories. Note that the MDS configuration space does not necessarily correspond in a simple fashion to the physical (anatomical) space in the cortex.

Representational Similarities

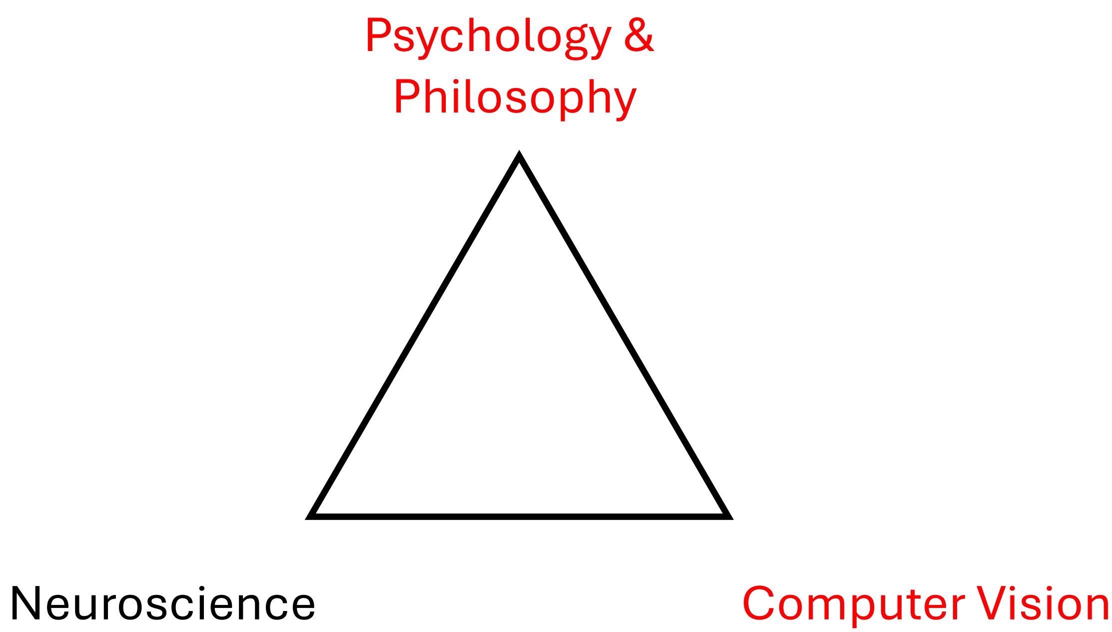




MDS receives as input either the “patterns” or, directly, a DISTANCE MATRIX

Representational Similarities

Garrison Cottrell



PHILOSOPHICAL PSYCHOLOGY, VOL. 13, NO. 1, 2000

Content and cluster analysis: assessing representational similarity in neural systems

AARRE LAAKSO & GARRISON COTTRELL

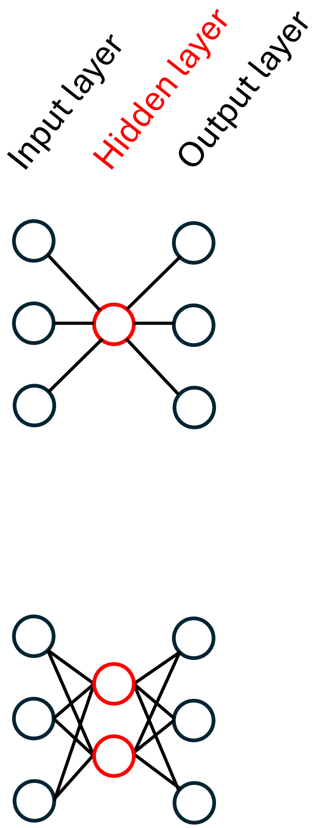
ABSTRACT *If connectionism is to be an adequate theory of mind, we must have a theory of representation for neural networks that allows for individual differences in weighting and architecture while preserving sameness, or at least similarity, of content. In this paper we propose a procedure for measuring sameness of content of neural representations. We argue that the correct way to compare neural representations is through analysis of the distances between neural activations, and we present a method for doing so. We then use the technique to demonstrate empirically that different artificial neural networks trained by backpropagation on the same categorization task, even with different representational encodings of the input patterns and different numbers of hidden units, reach states in which representations at the hidden units are similar. We discuss how this work provides a rebuttal to Fodor and Lepore's critique of Paul Churchland's state space semantics.*

Vector coding of Representational Similarities

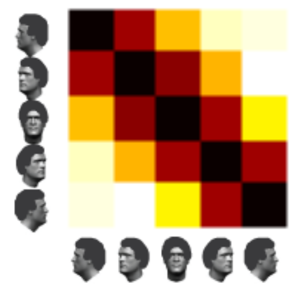
A modest proposal

We have argued that having a method for comparing the relative positions of concepts in one state space to the relative positions of concepts in another state space is critical for state space semantics. The method we propose here works well for neural networks, and may be generalizable to animals and robots. The basic idea is to collect the activation patterns evoked by inputs and compute all possible distances between these representations. The distances between representations capture the structure of representational space. We then compute the correlation between the distances between representations in one state space and the distances between representations in the other state space. This procedure can be used to measure the similarity between any two neural representations (be they from natural or artificial networks, from input, output, or hidden unit representations, from the same or different networks, with the same or different numbers of units).

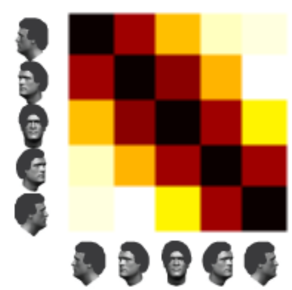
Relating Representations using Distance Matrices



1-D

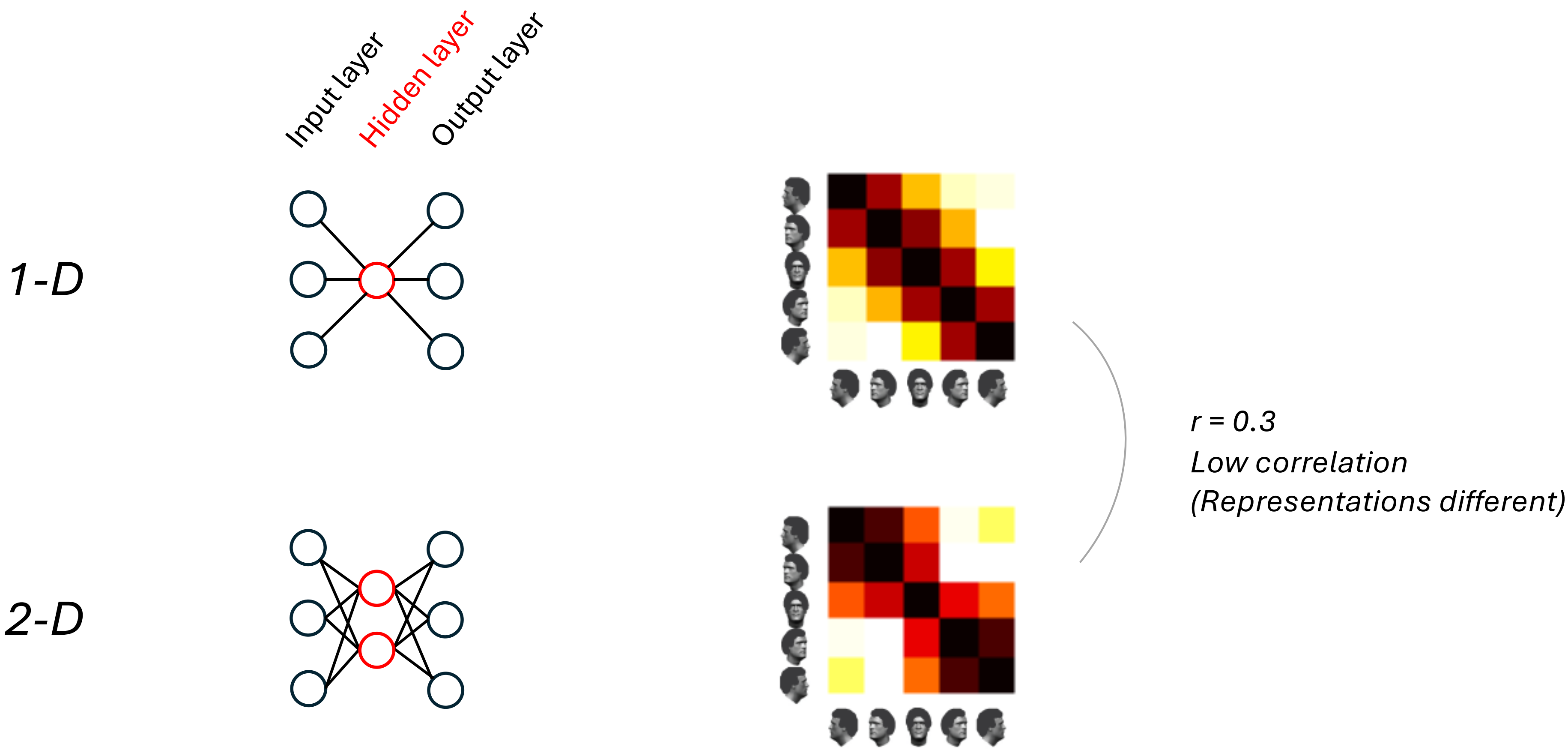


2-D



r = 1
Perfect correlation!
(Representations are the same)

Relating Representations using Distance Matrices



Representational Similarity Analysis

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

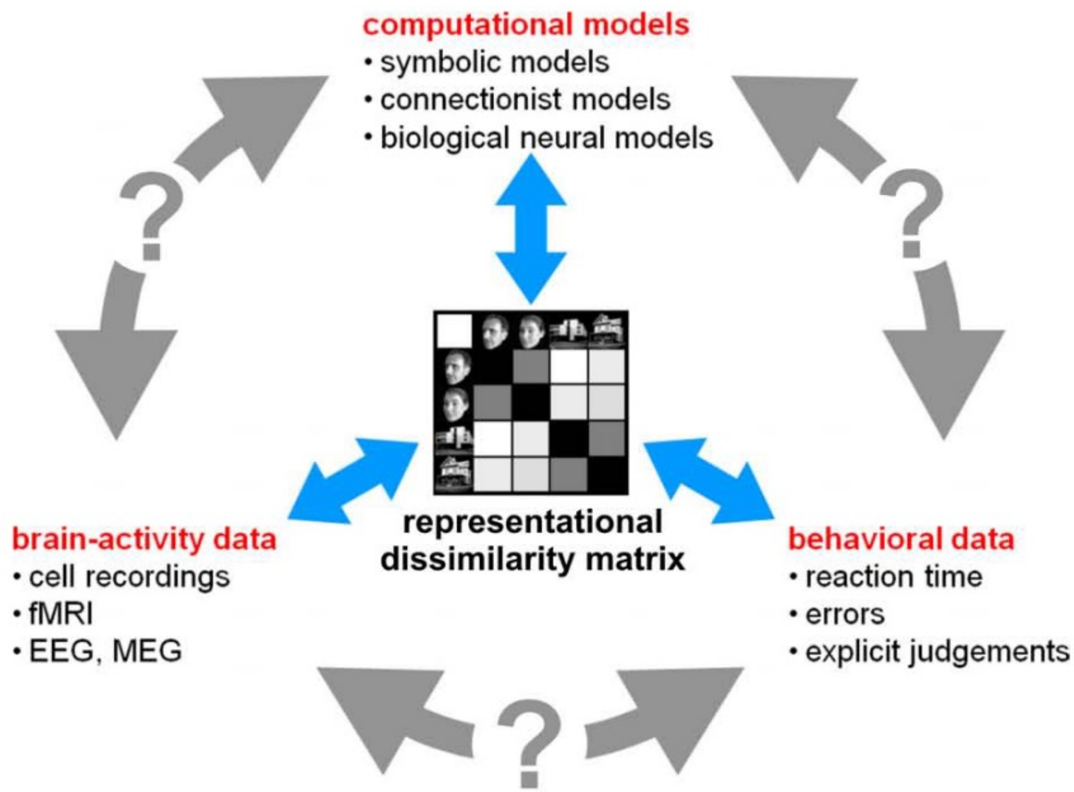
Edited by:
Mriganka Sur, Massachusetts Institute of Technology, USA

Reviewed by:
Michael A. Silver, University of California, USA
Doris Y. Tsao, University of Bremen, Germany

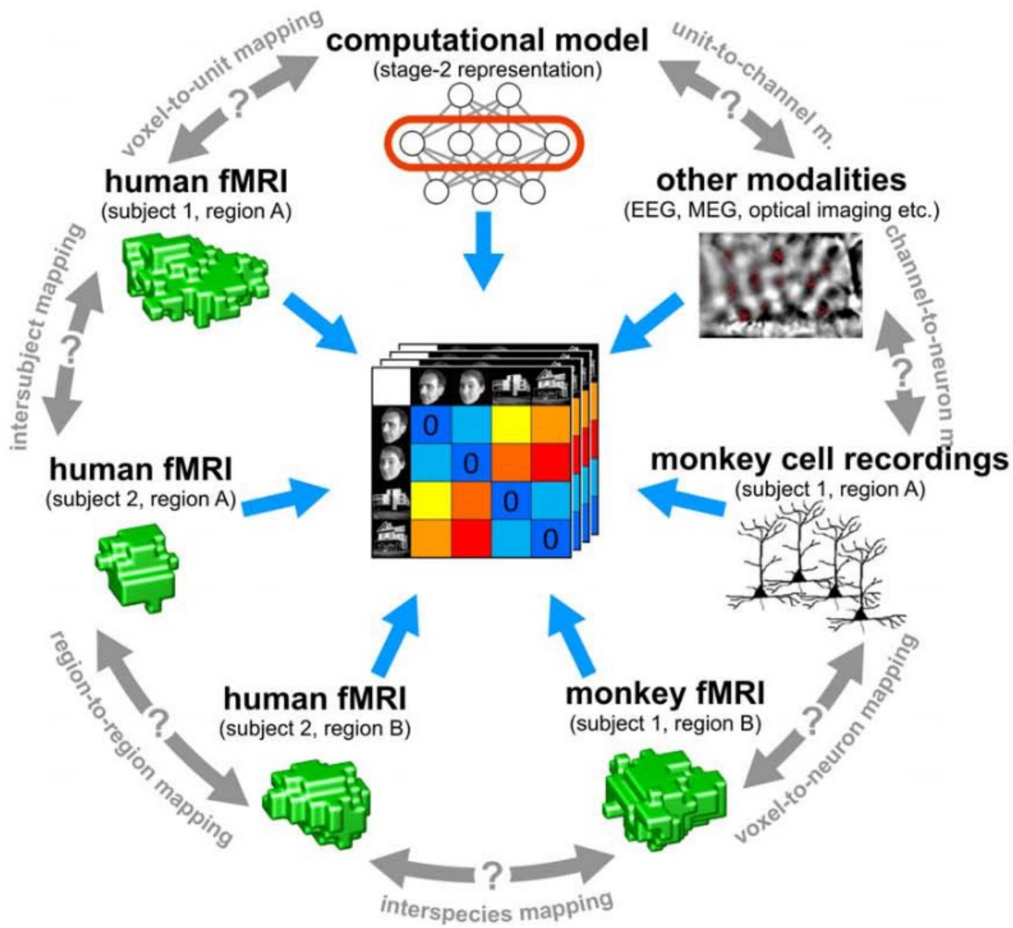
****Correspondence:***
Nikolaus Kriegeskorte, Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Building 10, Room 1D80B, 10 Center Dr. MSC 1148, Bethesda, MD 20892-1148, USA.
e-mail: kriegeskorten@mail.nih.gov

A fundamental challenge for systems neuroscience is to quantitatively relate its three major branches of research: brain-activity measurement, behavioral measurement, and computational modeling. Using measured brain-activity patterns to evaluate computational network models is complicated by the need to define the correspondency between the units of the model and the channels of the brain-activity data, e.g., single-cell recordings or voxels from functional magnetic resonance imaging (fMRI). Similar correspondency problems complicate relating activity patterns between different modalities of brain-activity measurement (e.g., fMRI and invasive or scalp electrophysiology), and between subjects and species. In order to bridge these divides, we suggest abstracting from the activity patterns themselves and computing representational dissimilarity matrices (RDMs), which characterize the information carried by a given representation in a brain or model. Building on a rich psychological and mathematical literature on similarity analysis, we propose a new experimental and data-analytical framework called representational similarity analysis (RSA), in which multi-channel measures of neural activity are quantitatively related to each other and to computational theory and behavior by comparing RDMs. We demonstrate RSA by relating representations of visual objects as measured with fMRI in early visual cortex and the fusiform face area to computational models spanning a wide range of complexities. The RDMs are simultaneously related via second-level application of multidimensional scaling and tested using randomization and bootstrap techniques. We discuss the broad potential of RSA, including novel approaches to experimental design, and argue that these ideas, which have deep roots in psychology and neuroscience, will allow the integrated quantitative analysis of data from all three branches, thus contributing to a more unified systems neuroscience.

Representational Similarity Analysis

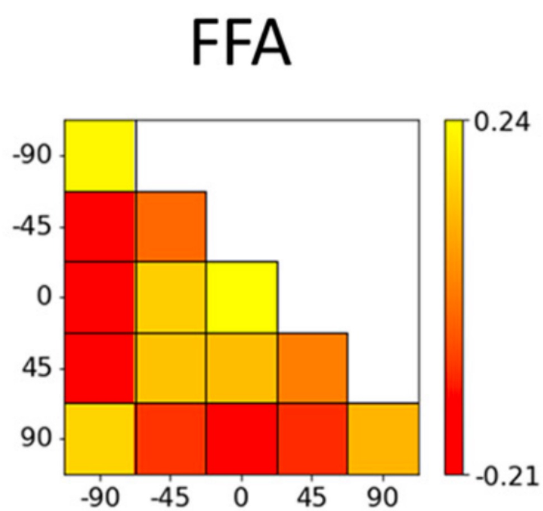


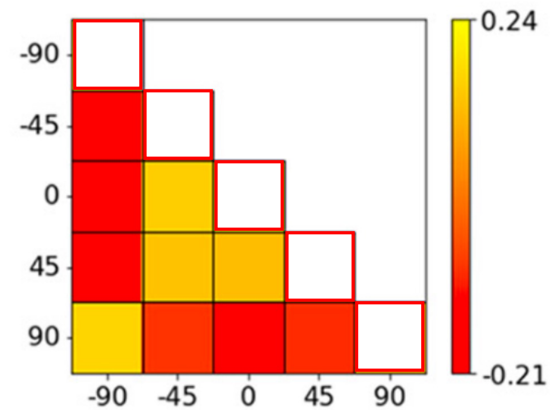
Representational Similarity Analysis



Representational Similarity Analysis

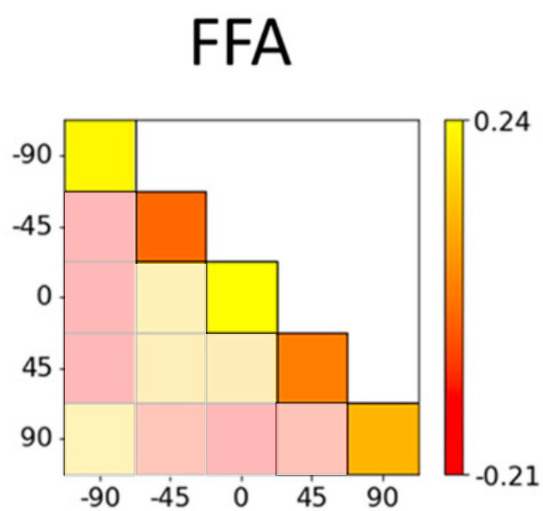
Not so fast...





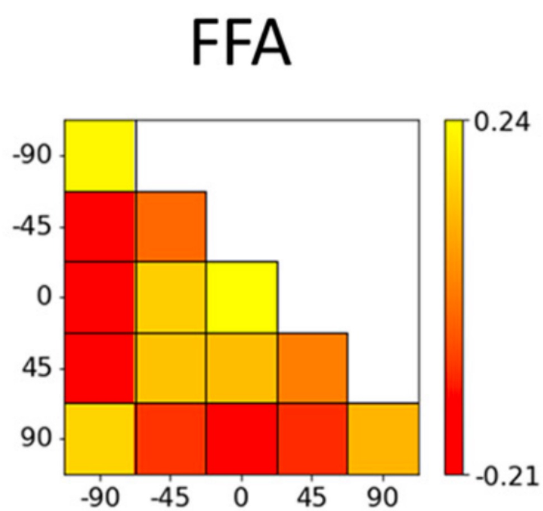
Matrix from Flack et al. (2019) *J Neurosci* (Figure 2)

Signal, noise, and the diagonal



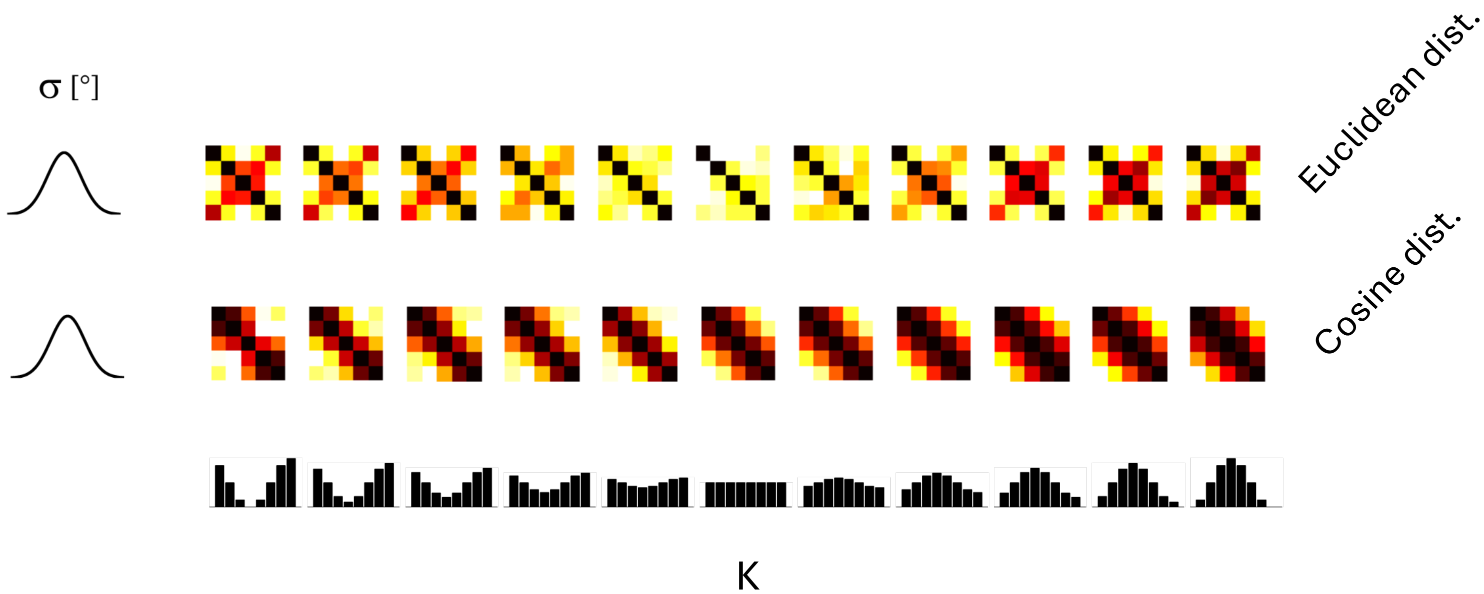
Matrix from Flack et al. (2019) *J Neurosci* (Figure 2)

Lower triangular entries depend on the main diagonal



Matrix from Flack et al. (2019) *J Neurosci* (Figure 2)

Expected pattern similarities



Ramírez et al. (2014) *J Neurosci*
Ramírez (2018) *Neuroscientist*

Ramírez et al., (2020) *Neuropsychologia*
Ramírez and Merriam (2020) *Nature Comms*

Revsine et al., (2024) *J Neurosci*

Influence of mean centering on RSA conclusions

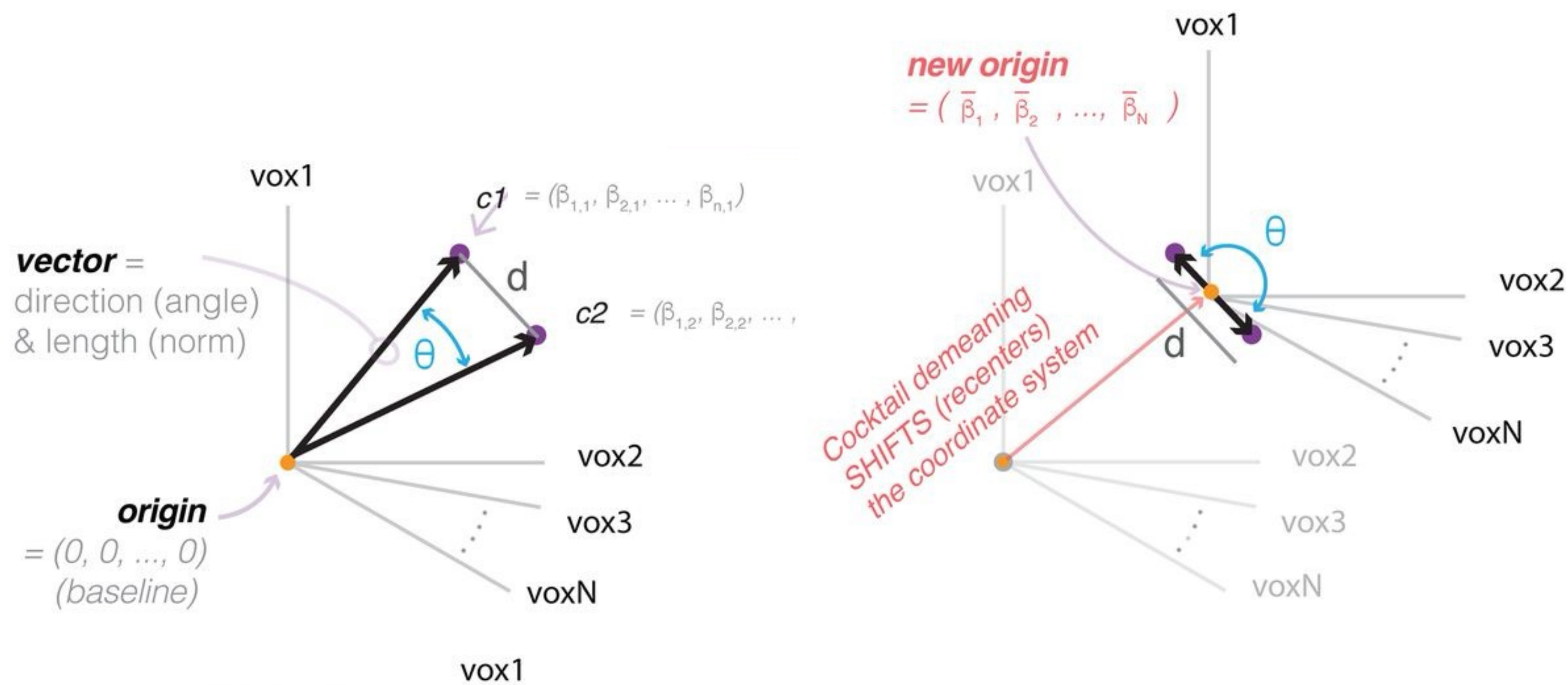
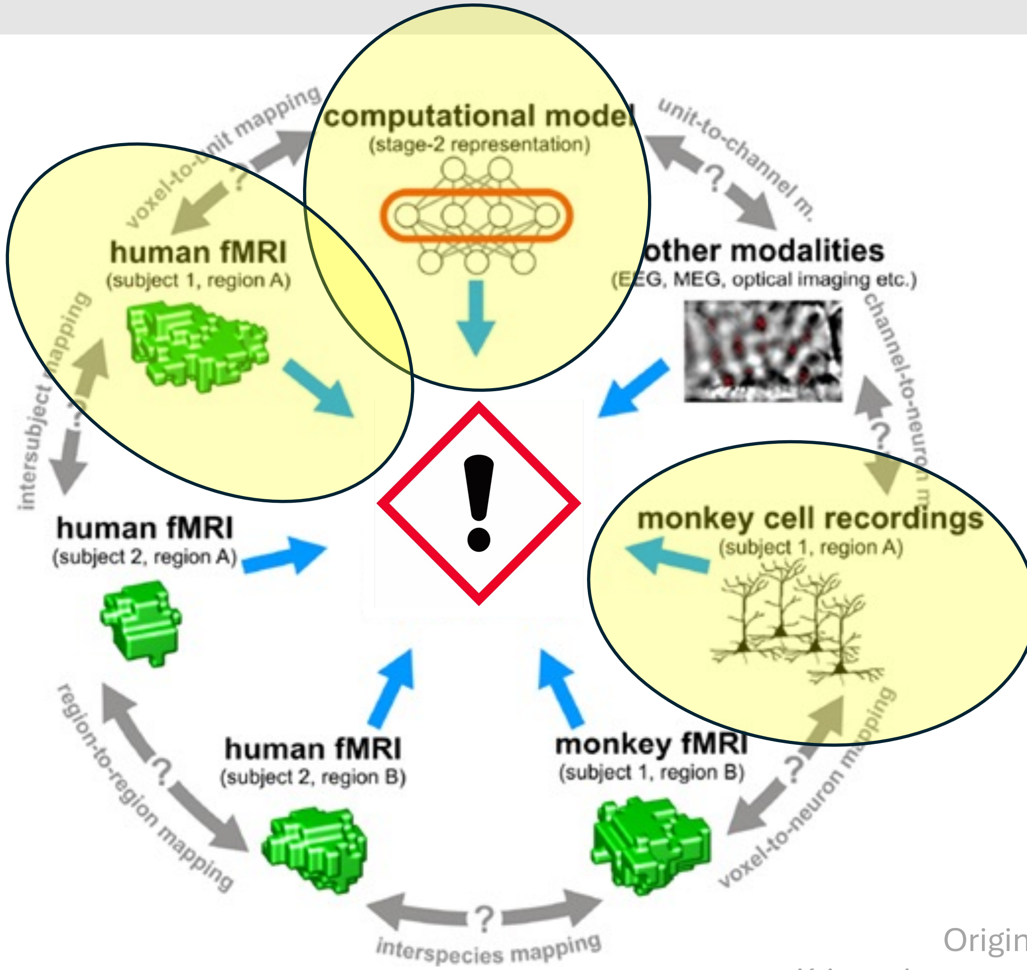


Fig. from Revsine et al., (2024) *J Neurosci*

Garrido et al (2013) *Front in Neurosci*
Ramírez (2017) *BioRxiv*
Ramírez et al (2020) *Neuropsychologia*

Representational Similarity Analysis



Scannell and Young (1999)
Diedrichsen et al (2011) Neuroimage
Ramírez et al (2014) J Neurosci
Ramírez and Merriam (2020) Nat Comms

Original Fig. (intervened here) from
Kriegeskorte et al. (2008) Front Syst Neurosci

"*The most general approach to understanding the challenges facing perceptual systems is to cast them as problems in data analysis. For theorists, this perspective immediately raises the question of the dimensionality of the data, which strongly influences the appropriate mathematical approaches to making sense of it.*"

Shimon Edelman (2002) *Nat Neurosci*

Conclusions

The School of Athens, Raphael

- The study of representational similarities has a rich history in cognitive sciences
- Representational Similarity Analysis addresses the key challenge of putting data in a format amenable to comparisons between species, data modalities, and computational models
- The recommendation to abstract from the measurement process, however, is impractical.
- Incorporating information about signal to noise ratios and the measurement process can help mitigate biases
- Data demeaning across conditions not recommended and previously shown to lead to erroneous conclusions

Conclusions

- **Multivariate Pattern Analysis (MVPA)**
  - Can detect information invisible to mass univariate methods
  - *Do not* tell us HOW information is encoded
  - Can tell us that information about our conditions is present in a set of voxels
- **Representational Similarity Analysis (RSA)**
  - Can help understand HOW information is encoded in the brain
  - May be a way forward when the "correspondence" between data channels and model units is unknown
  - However, the method's flexibility comes at a cost
  - RSA conclusions depend on analysis choices, such as pattern dissimilarity measure and data normalization strategies
  - It is not generally advisable to "abstract from the measurement process"
- **Model guided approaches to RSA can help mitigate some of the method's limitations, as well as improve the interpretability of empirical findings**

The School of Athens, *Raphael*

Thank you for your attention!

www.fernando-ramirez.org | *Twitter*: @FRamirez_R2 | fernando.ramirez@nih.gov