Encoding and decoding with voxel-wise models

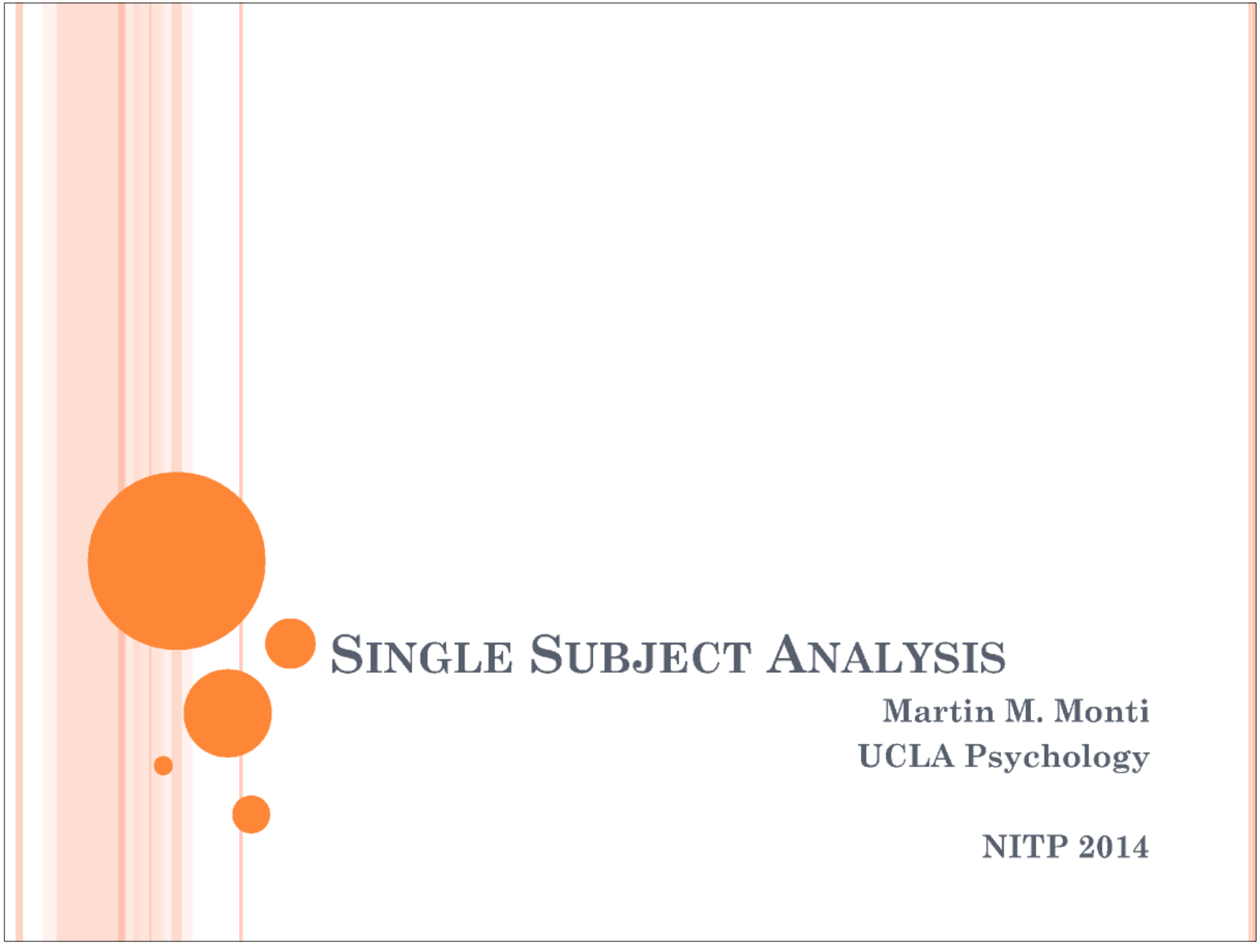
Jack Gallant
University of California at Berkeley

Main messages for today

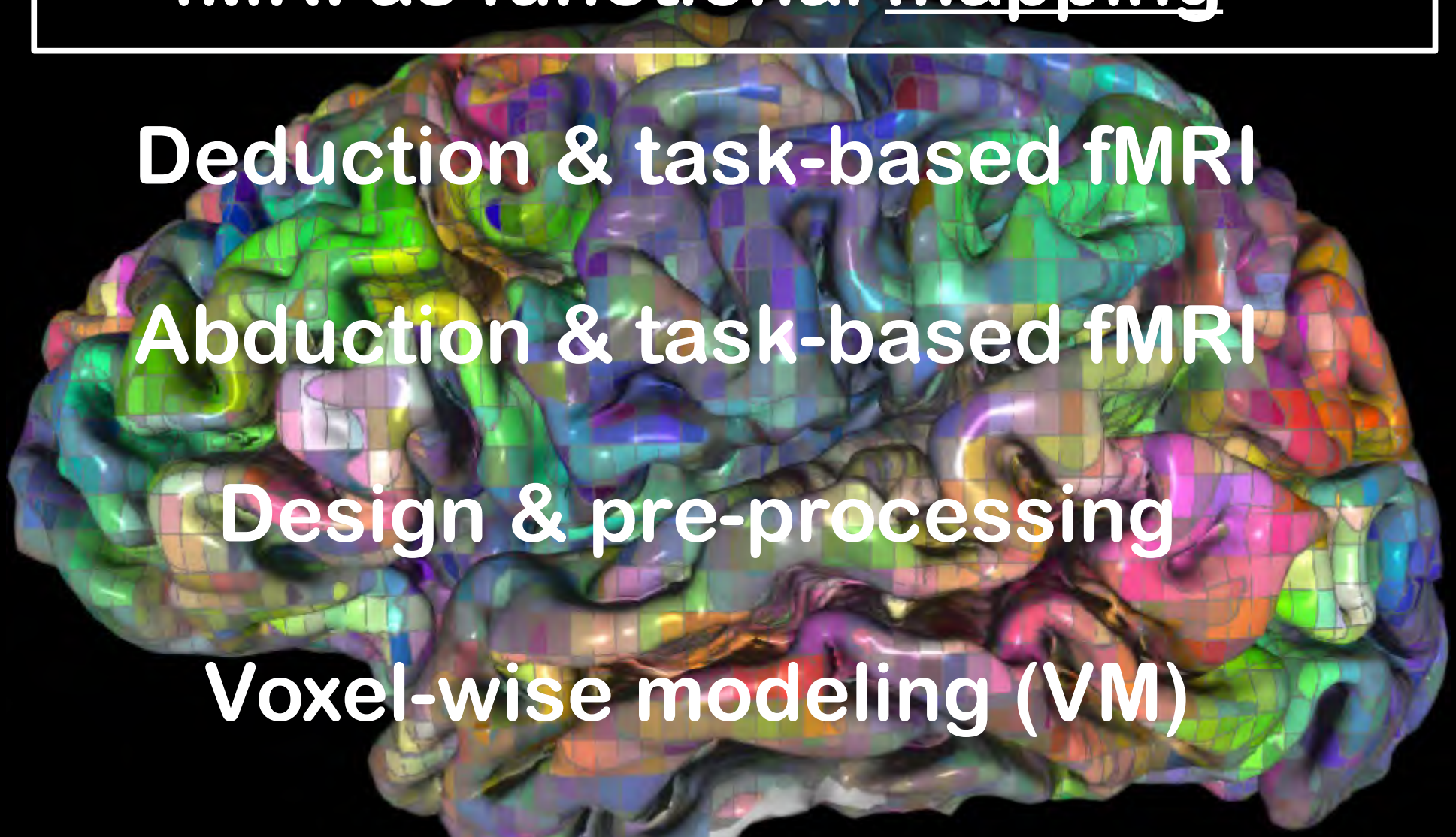
- The classical deductive, null-hypothesis testing approach used in cognitive neuroscience is weak, inefficient and often produces misleading results.
- A more open-ended, abductive approach can be much more efficient than the classical approach, and it can quickly produce powerful predictive models.

- The standard SPM pipeline discards much useful information because it focuses almost solely on Type I error and ignores Type II error.
- A voxel-wise modeling approach preserves much more of the useful information in the data and so minimizes both Type I and Type II error.

For further background check Martin's lectures



fMRI as functional mapping



Deduction & task-based fMRI

Abduction & task-based fMRI

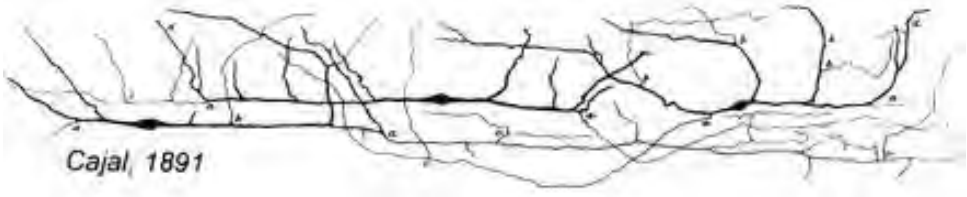
Design & pre-processing

Voxel-wise modeling (VM)

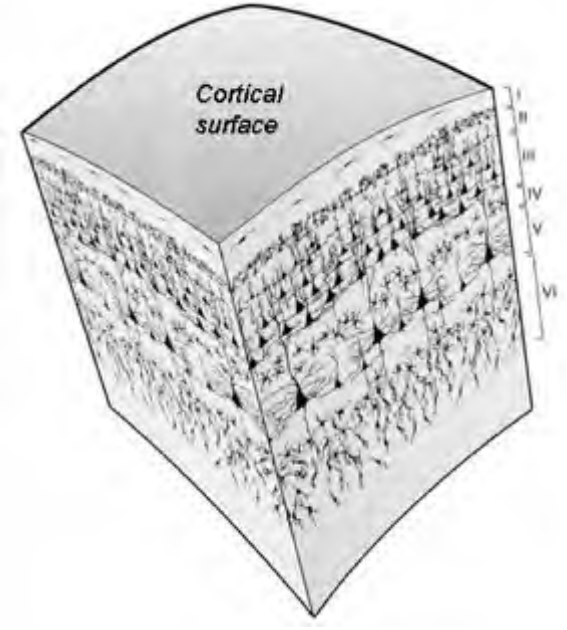
Using VM to Decode

The brain is organized at multiple scales

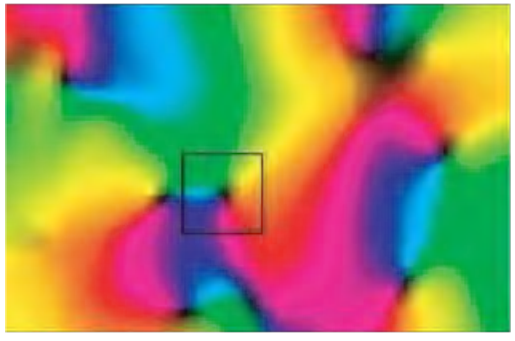
Neurons



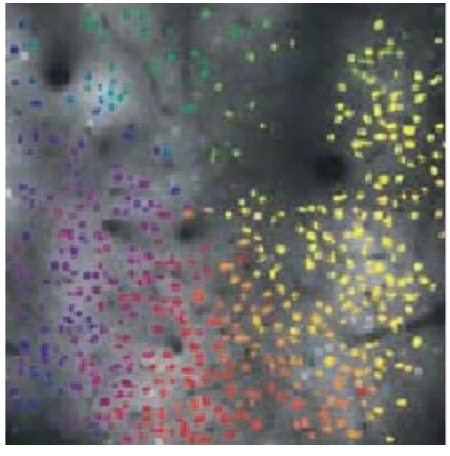
Layers and columns



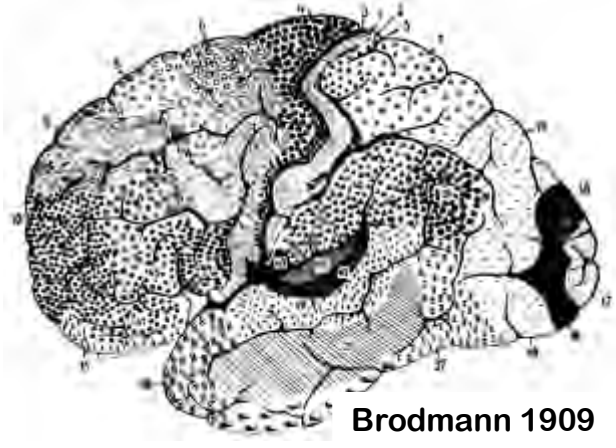
Maps



Ohki et al., 2006



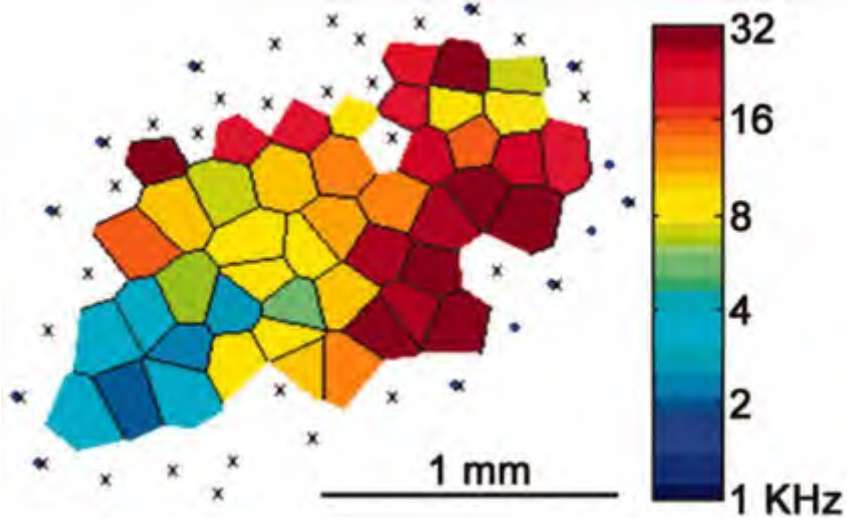
Areas



Brodmann 1909

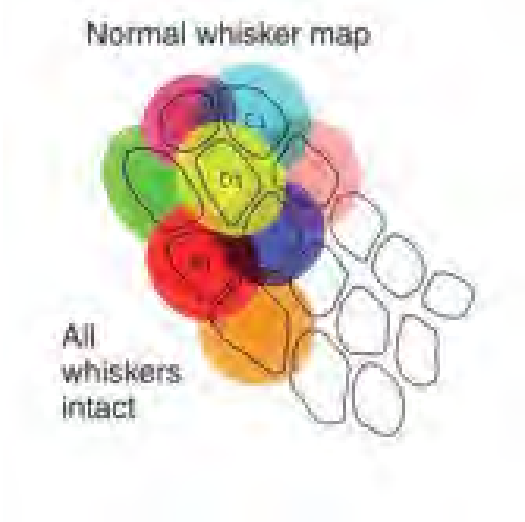
Information is represented in functional maps

Tonotopic map in A1



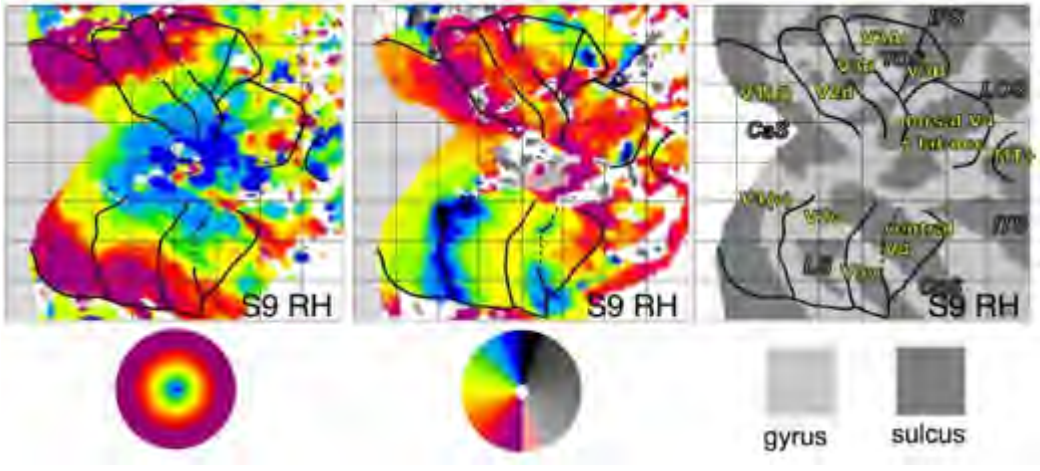
Kenet et al. (2007)

Whisker map in S1 (rodent)



Feldman & Brecht, 2005

Retinotopic maps

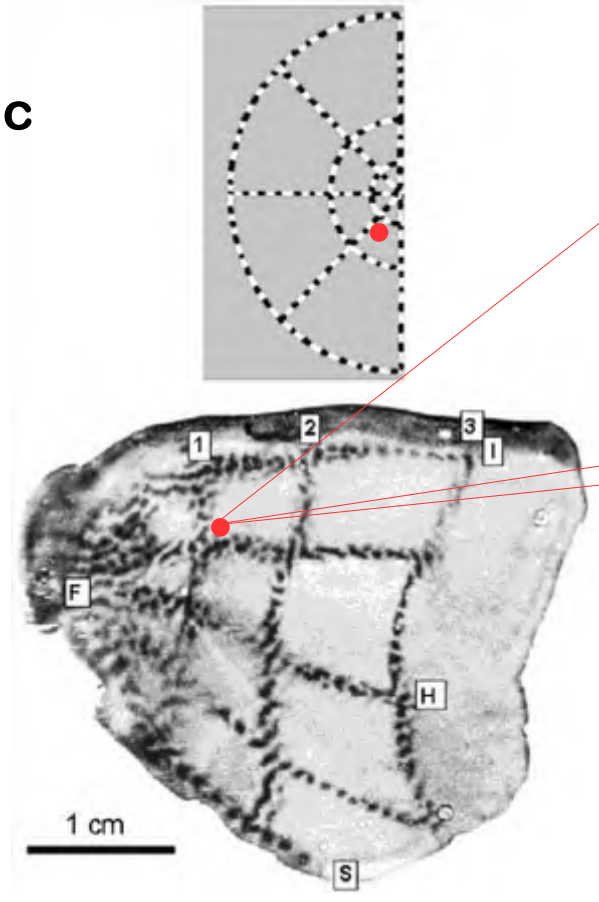


Hansen, Kay & Gallant, *J. Neurosci* (2007)

Maps are organized both globally and locally

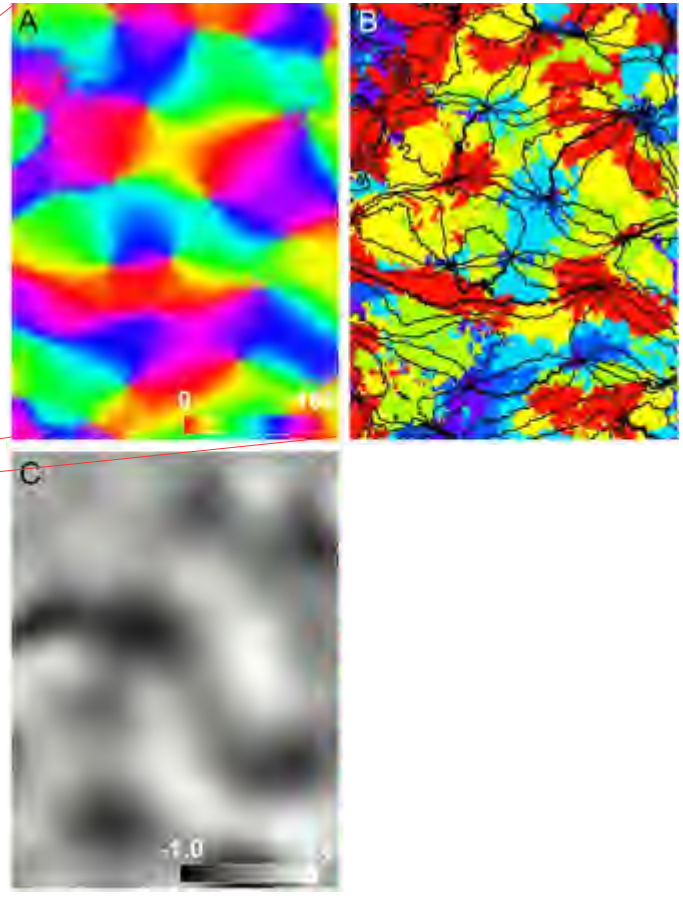
Retinotopic Mapping Stimulus

2DG map on flattened Macaque cortex



Orientation

Spatial frequency

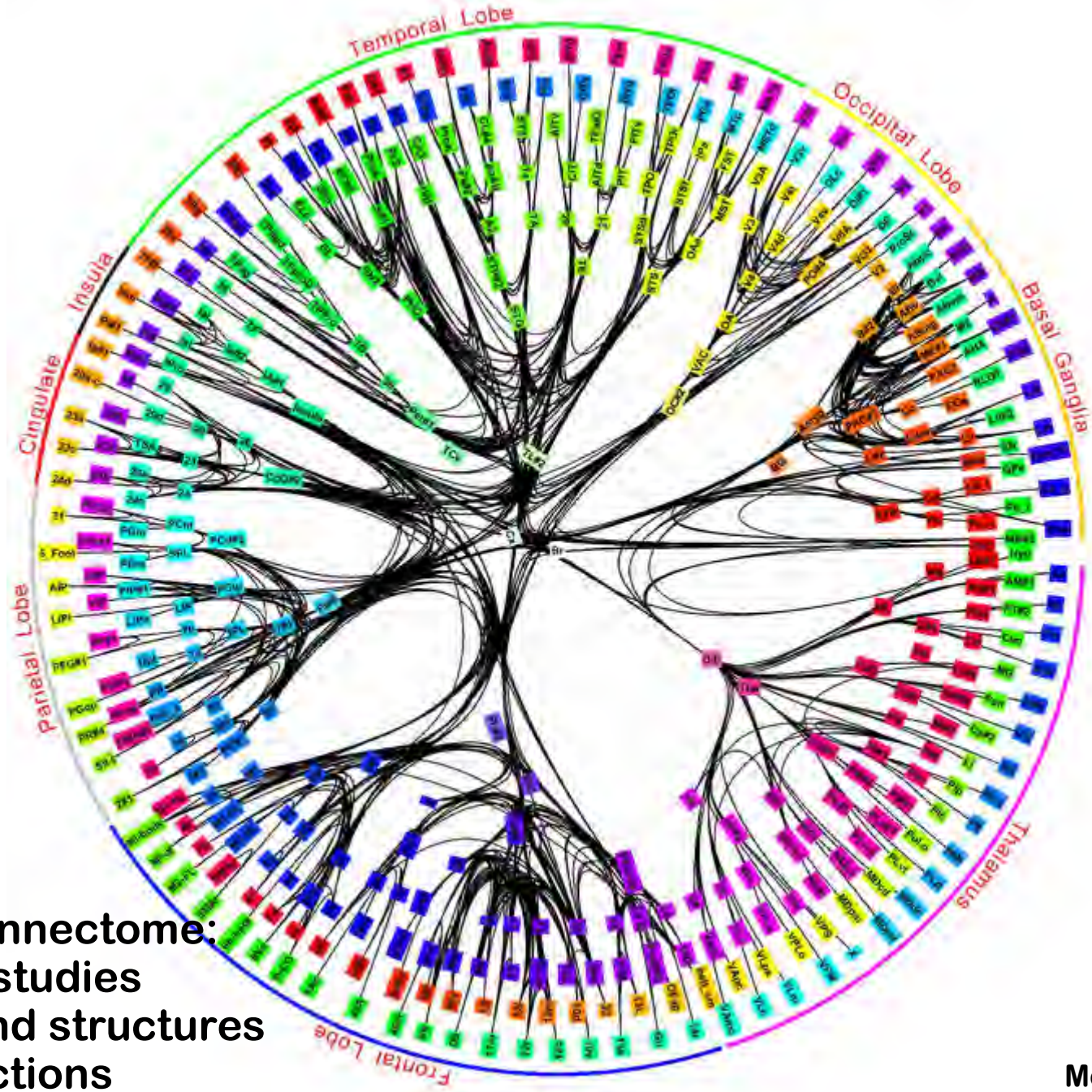


Ocular dom.

Tootel et al., *J. Neurosci* (1988)

Issa, Trepel & Stryker, *J. Neurosci* (2000)

Human cortex could contain hundreds of areas

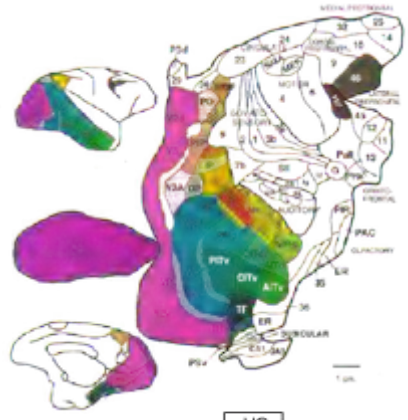


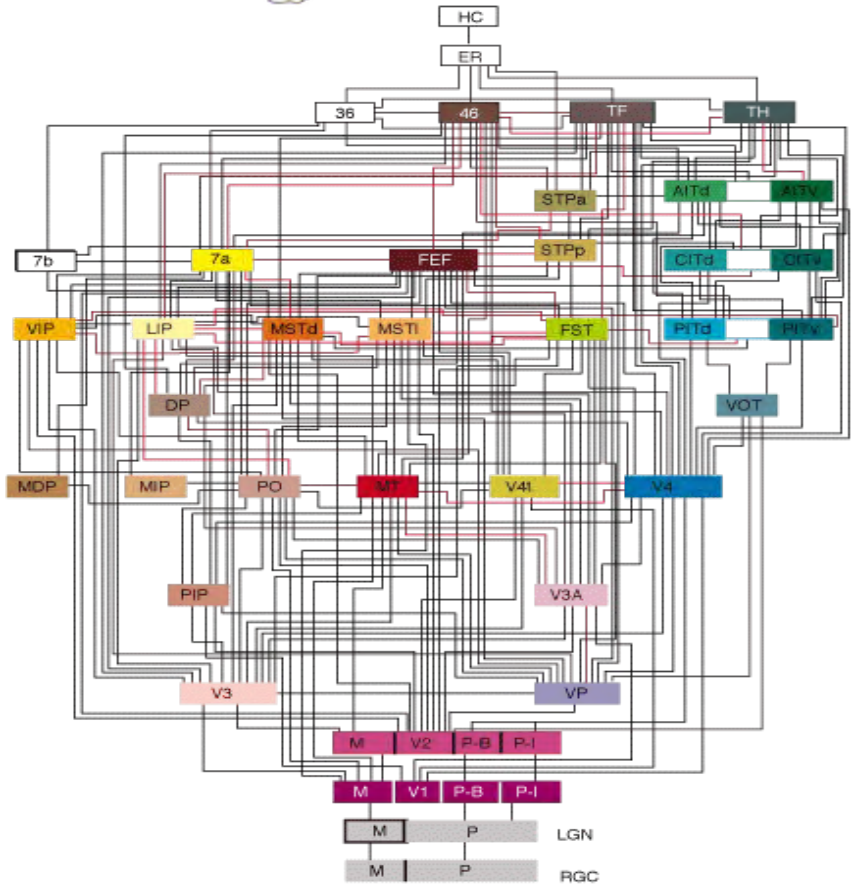
Macaque connectome:
410 tracing studies
383 areas and structures
6602 connections

Modha & Singh 2010

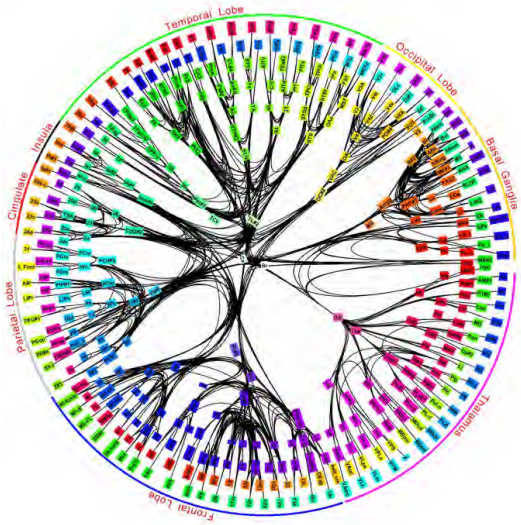
Mammalian vision as a model system

- Dozens of distinct areas.
- Areas arranged in a hierarchical, parallel network.
- Transformations between areas are nonlinear.
- Areas contain systematic, high-dimensional maps.
- Each area represents different visual information.



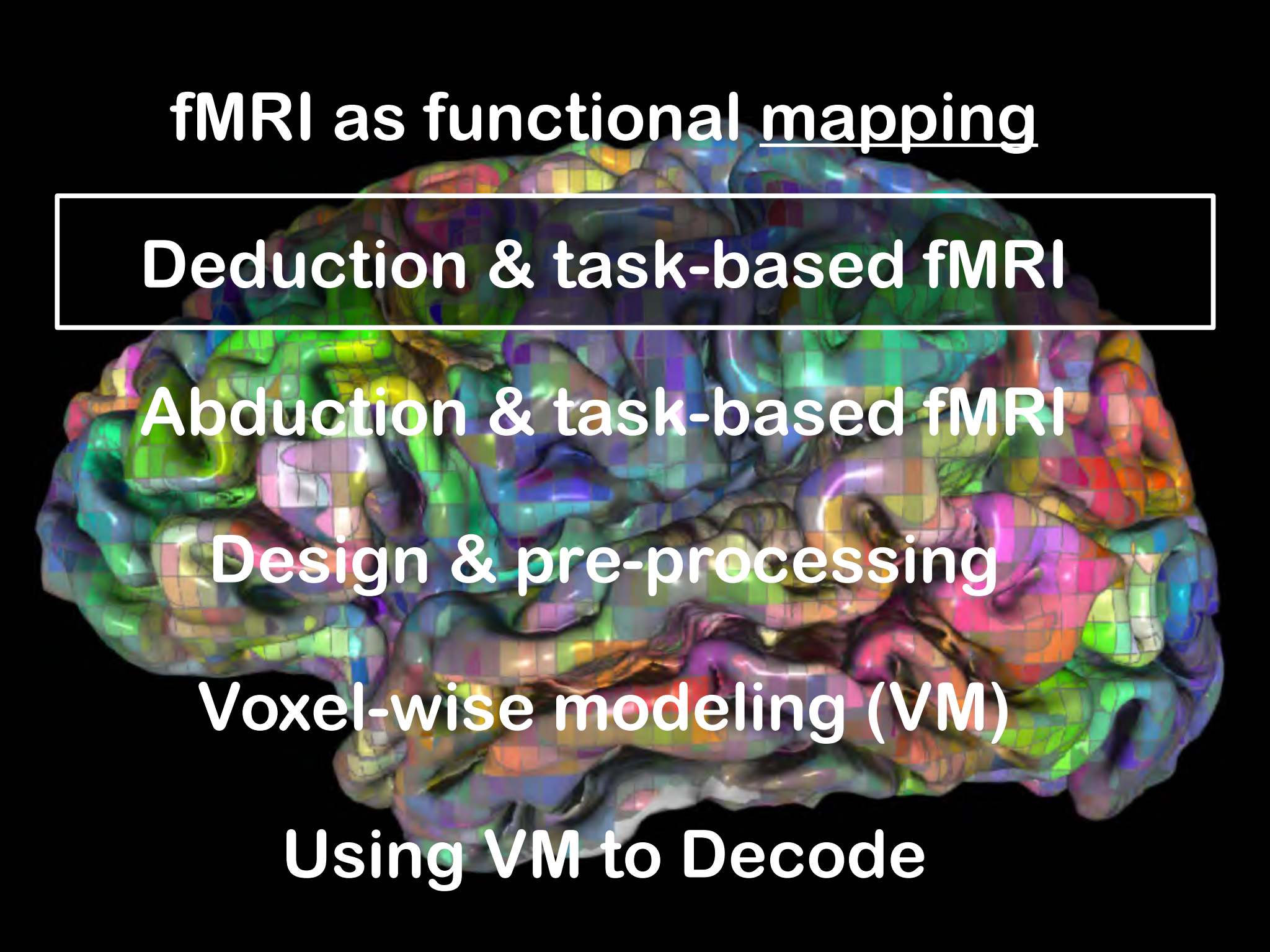


383 areas



Modha & Singh, *PNAS*, 2010

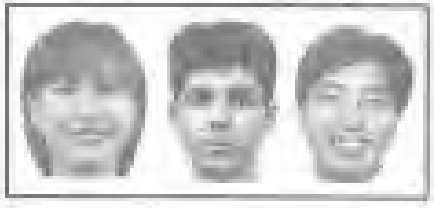
Felleman and Van Essen, *Cerebral Cortex*, 1992

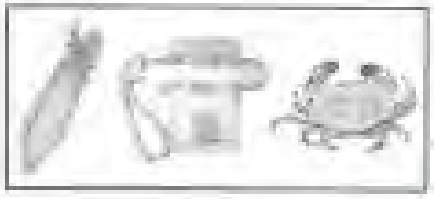

fMRI as functional mapping
Deduction & task-based fMRI
Abduction & task-based fMRI
Design & pre-processing
Voxel-wise modeling (VM)
Using VM to Decode

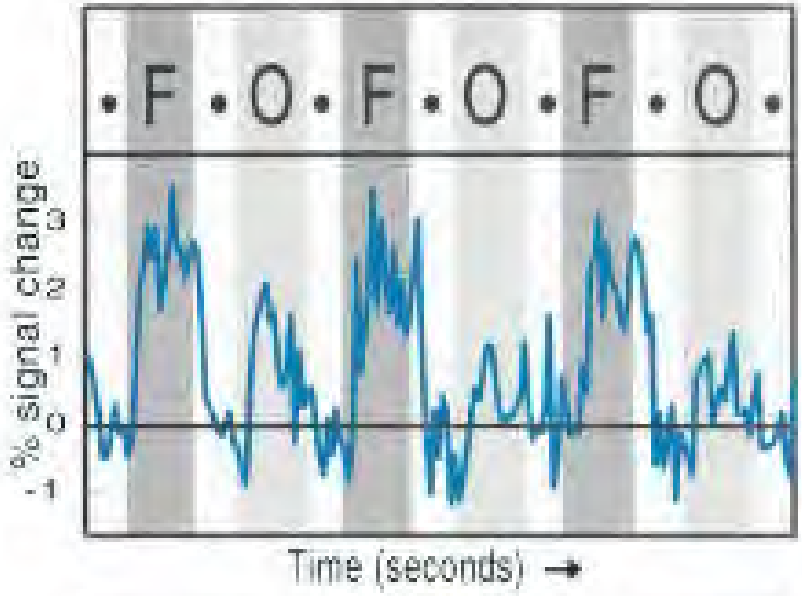
The deductive approach to task-based fMRI

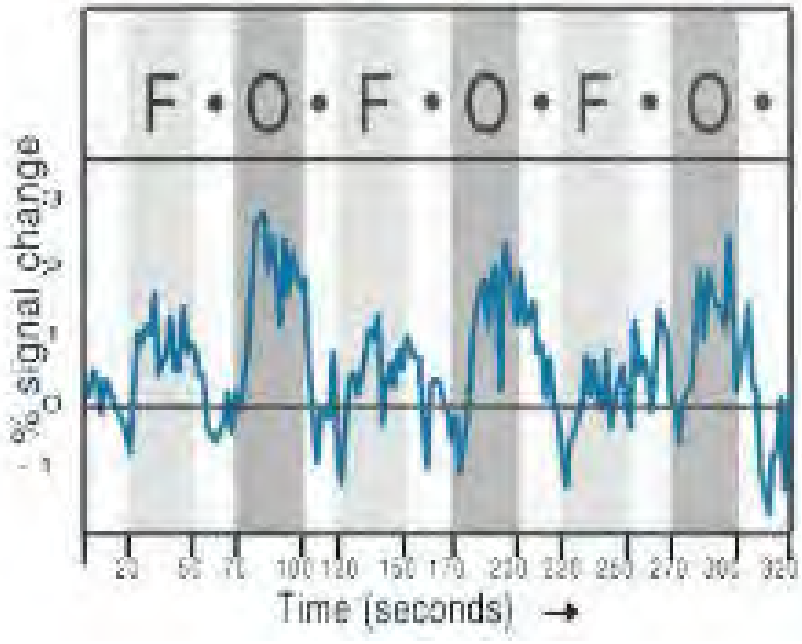
- Find out how the brain mediates behavior.
  - Use simple stimulus or task with few conditions.
- Find out how the brain is organized into areas.
  - Use anatomy, localizers or a searchlight to discover ROIs.
- Find out what information is mapped within each area.
  - Test for statistically significant differences in responses across conditions, or use a classifier.
- Find out how these maps vary across individuals.
  - Map individual brains into standardized anatomical coordinates and do analysis at group level.

The deductive approach to task-based fMRI

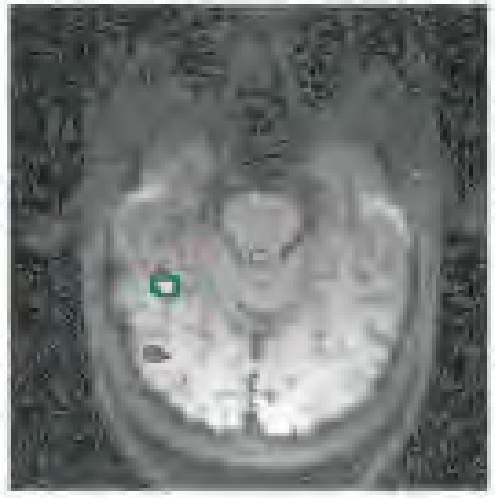




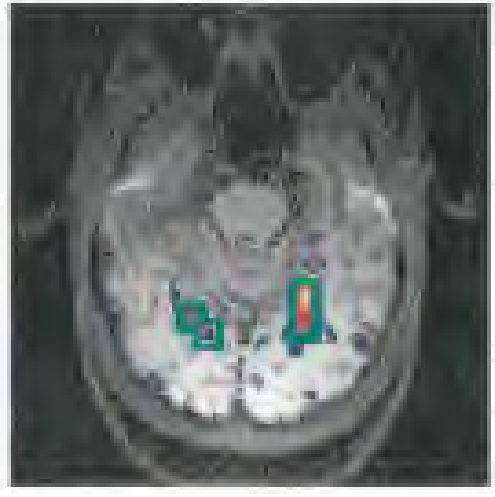




Faces > Objects



Objects > Faces

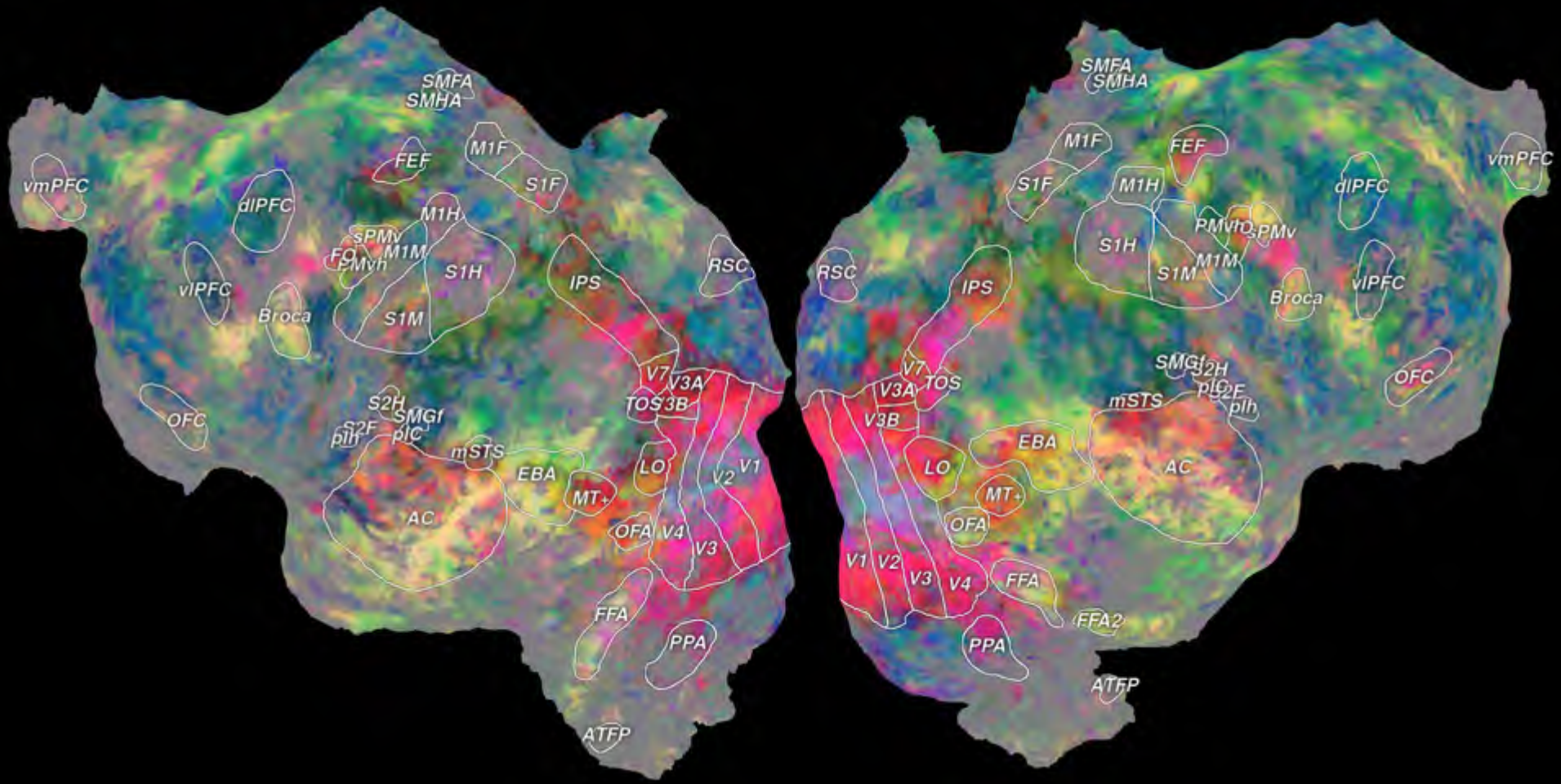


Kanwisher, McDermott & Chun, *Journal of Neuroscience*, 1997

Some visual ROIs revealed by this approach

Name	Location	Contrast	References
FFA (fusiform face area)	Posterior fusiform gyrus	Faces – Objects	Kanwisher et al, 1997 McCarthy et al, 1997
OFA (occipital face area)	Just anterior to V4v/VO	Faces – Objects	Kanwisher et al, 1997 Halgren et al, 1999
IFSFP (inferior frontal sulcus face patch)	IFS anterior to precentral sulcus	Faces – Objects	Avidan et al, 2005 Tsao et al, 2008
ATFP (anterior temporal face patch)	Temporal pole	Faces – Objects	Rajimehr et al, 2009
STSFP (superior temporal sulcus face patch)	Posterior superior temporal sulcus	Faces – Objects	Clark et al, 1996 Kanwisher et al, 1997
EBA (extrastriate body area)	Anterior to MT+ on the medial temporal gyrus	Bodies – Objects	Downing et al, 2001
FBA (fusiform body area)	Fusiform sulcus/gyrus anterior to FFA	Bodies - Objects	Peelen & Downing, 2005 Schwarzlose et al, 2005
PPA (parahippocampal place area)	Collateral fissure	Scenes – Objects	Epstein & Kanwisher, 1998
TOS (transverse occipital sulcus)	Just inferior to V7	Scenes – Objects	Nakamura et al, 2000 Hasson et al, 2003
RSC (retrosplenial cortex)	Medial wall just superior to PPA	Scenes – Objects	Aguirre et al, 1996
FEF (frontal eye field)	Precentral sulcus adjoining superior frontal sulcus	Saccades – Fixation	Luna et al, 1998
iFEF/FO (inferior frontal eye field)	Inferior portion of precentral sulcus	Saccades – Fixation	Berman et al, 1999 Corbetta et al, 1998

Localizers produce misleading results



The localizer/MVPA approach produce misleading estimates of both functional and spatial specificity. This is caused by the experimental designs and the analysis pipeline.

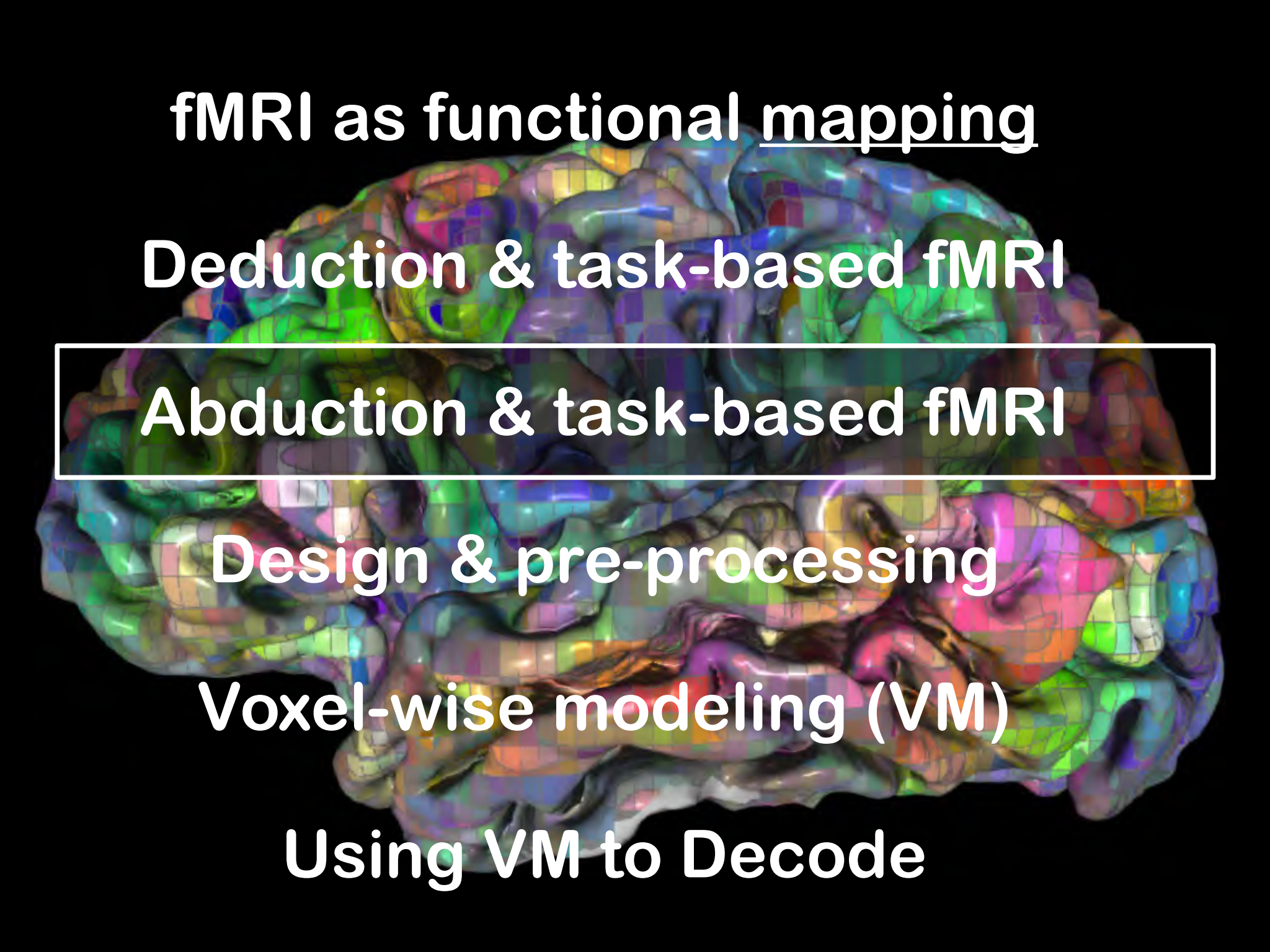
Nunez-Elizalde, Huth & Gallant, *in preparation*

Problems with the classical approach

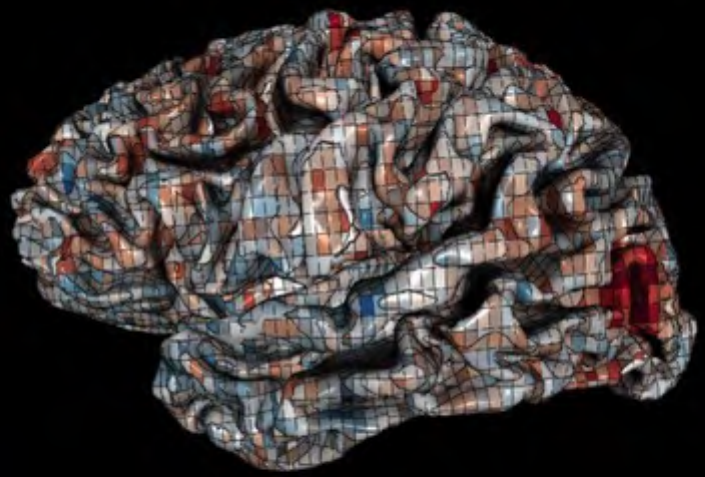
- Many assumptions are implicit in the operational definitions and selection of task conditions.
- It cannot recover detailed information about representations within areas.
- It is both too conservative (Type 1 error control too strict) and insufficiently sensitive (Type 2 error control insufficient).
- Results often generalize poorly beyond the tested subspace.
- It doesn't offer any method for determining when you should be satisfied with a model.

What we would like from an approach

- Provides a method to test alternative hypotheses quickly.
- Provides rich behavioral & brain data for low cost.
- Makes all assumptions and operational definitions quantitatively explicit.
- Easily bridges between psychological concepts and brain measurements.
- Recovers fine detail in cortical maps.
- Minimizes Type I error.
- Minimizes Type II error.
- Provides objective measures of significance and effect size (i.e., importance).


fMRI as functional mapping
Deduction & task-based fMRI
Abduction & task-based fMRI
Design & pre-processing
Voxel-wise modeling (VM)
Using VM to Decode

What analysis would be optimal for these data?

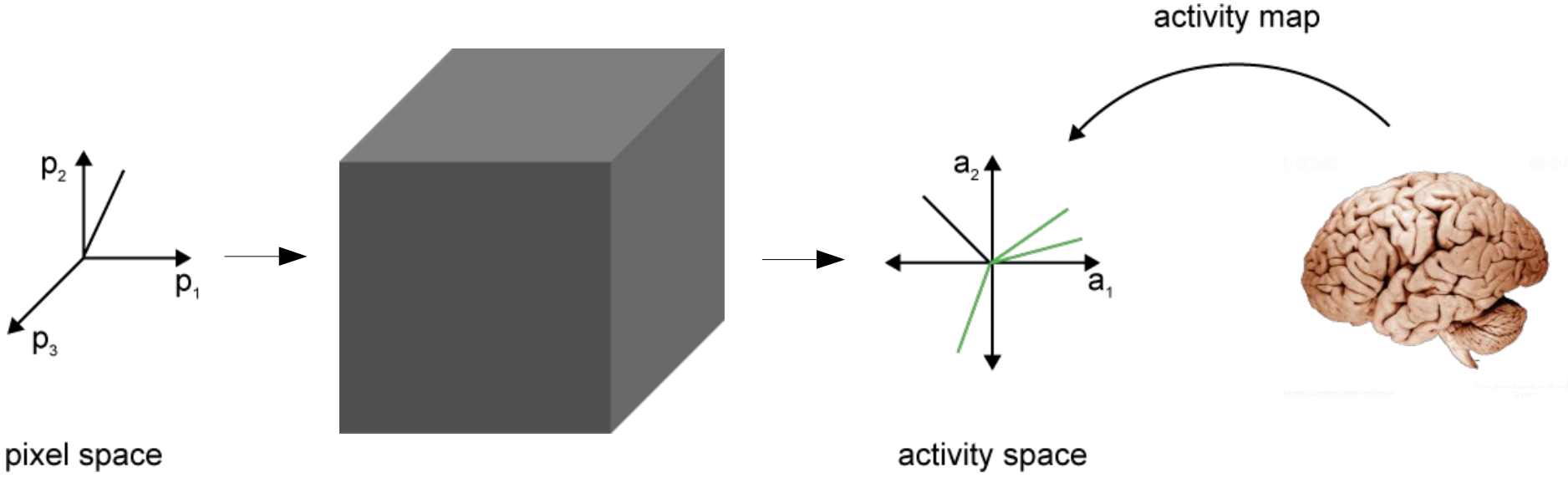


Less activity

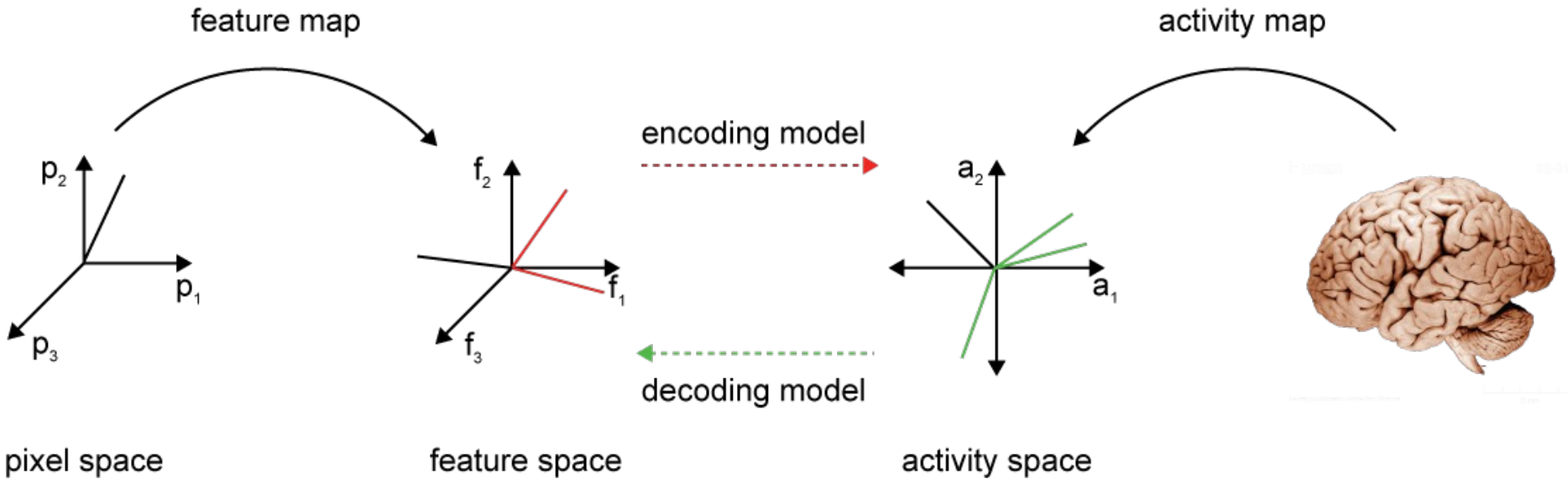
Average activity

More activity

The system identification view of neurosicence



The system as a nonlinear feature space



Nonlinear Transform

(Hypothesis)

Linear Transform

(Regression)

Linear or Nonlinear Transform

(Measurement)

Voxel-wise modeling (VM)

- Find out how the brain mediates behavior.
  - Use broad range of stimulus/task conditions.
- Find out how the brain is organized into areas.
  - For each voxel in each subject, fit competing linearized models that embody different feature spaces and compare model predictions.
- Find out what information is mapped within each area.
  - Visualize voxel tuning and find feature subspace that best describes tuning of population.
  - Find out how these maps vary across individuals.
  - Define maps and areas in individual subjects and aggregate across subjects.

VM differs from the classical approach in that...

- The stimuli & tasks can be complicated/high-dimensional.
- Two separate, interleaved data sets are acquired: one for fitting, one for testing.
- Regression occurs within a high- dimensional feature space that mediates between stimulus/task variables and BOLD responses.
- A separate spatio-temporal HRF is estimated for each voxel each feature and each delay.
- No spatial smoothing is performed.
- No cross-subject averaging is performed.
- Predictions are used to evaluate and compare models.
- Interpretation involves visualizing voxels and maps.

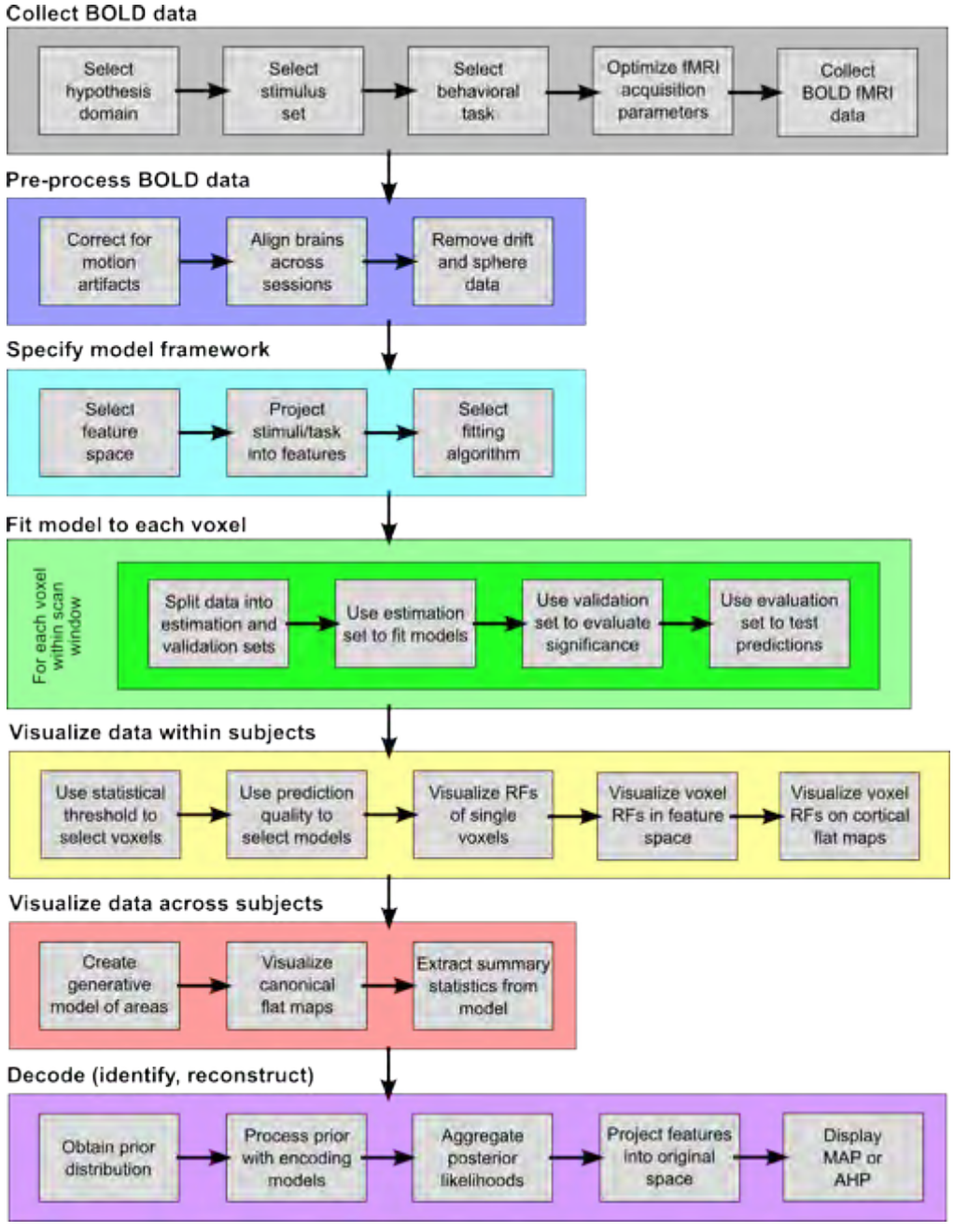
Voxel-wise modeling (VM)

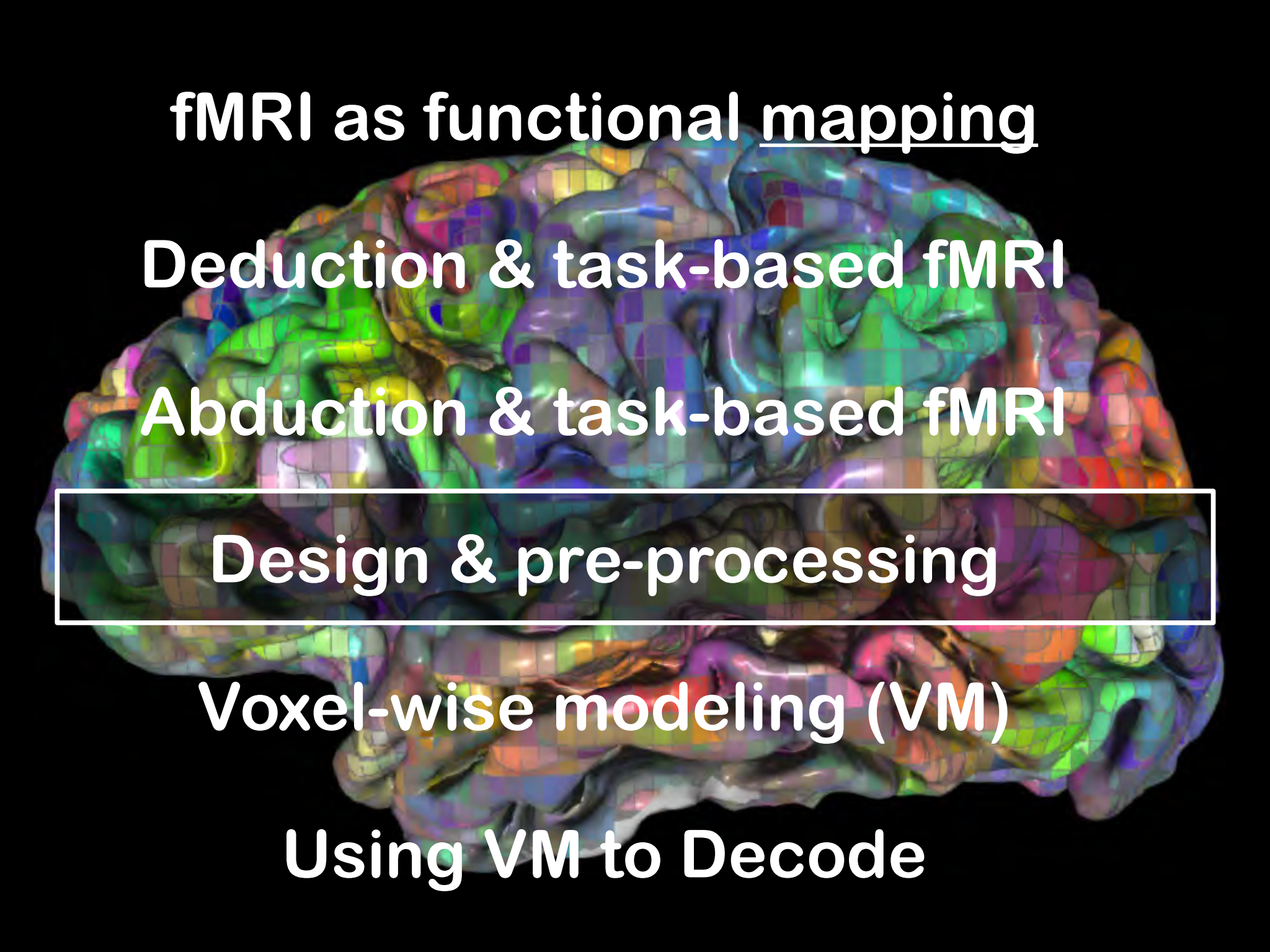
Collect functional data

Estimate voxel-wise models

Visualize & interpret results

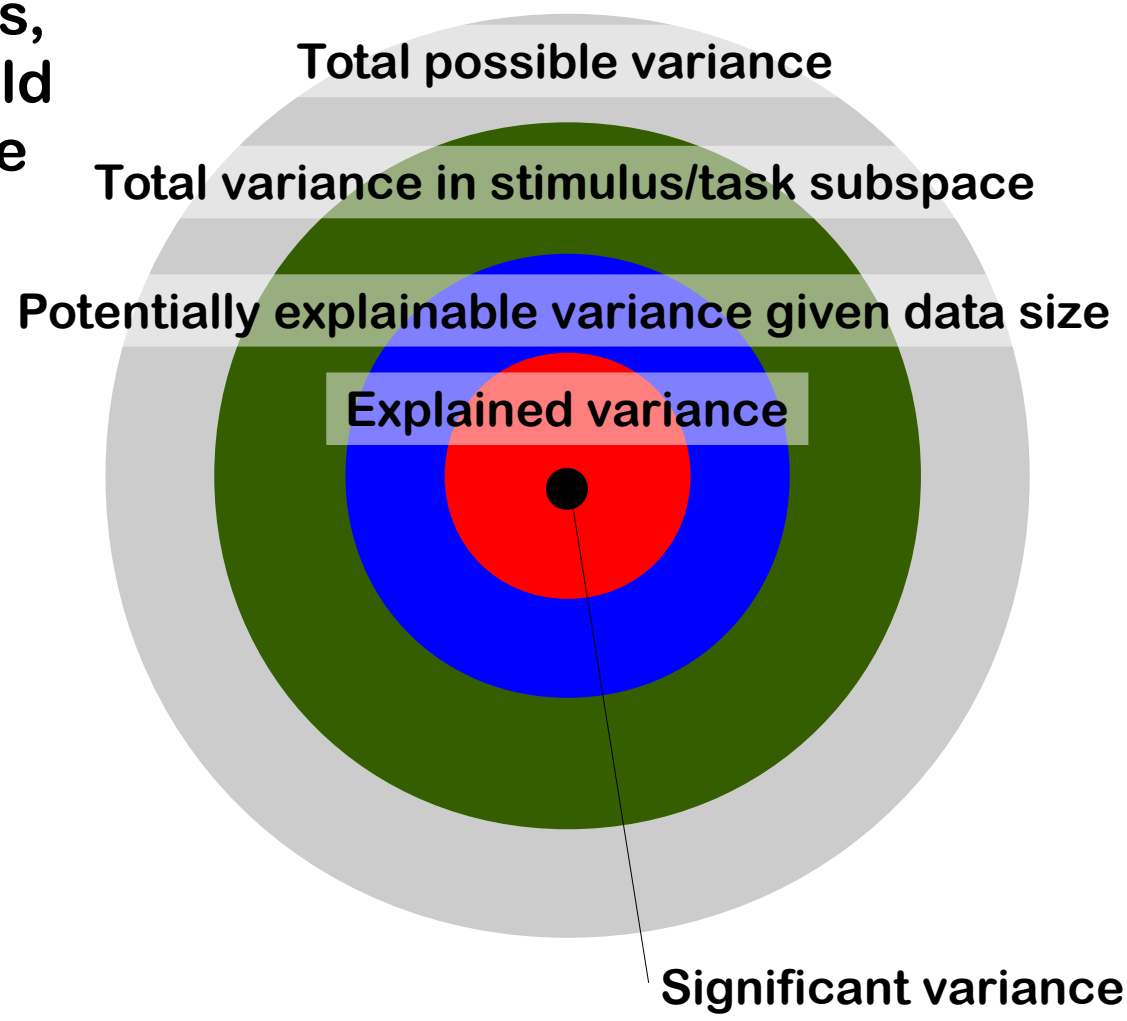
Decode information




fMRI as functional mapping
Deduction & task-based fMRI
Abduction & task-based fMRI
Design & pre-processing
Voxel-wise modeling (VM)
Using VM to Decode

Sampling the stimulus and task space

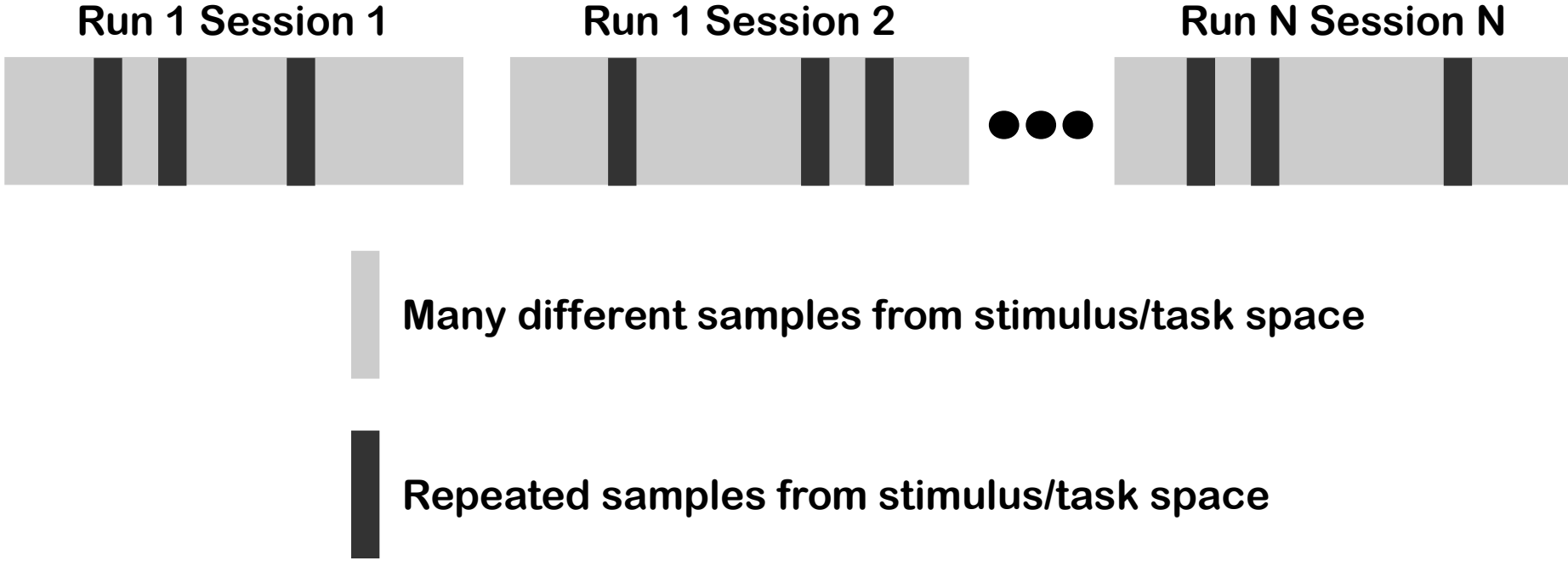
- The fit set is used to estimate voxel-wise models, so the stimulus/tasks should sample the relevant feature spaces as completely as possible. Optimize by collecting few trials from many different states.
- The test set is used to validate fit models, so the signals must be measured accurately. Optimize by collecting many repeated trials from a few states.
- Make sure the stimuli/task spaces for fit and test sets overlap!



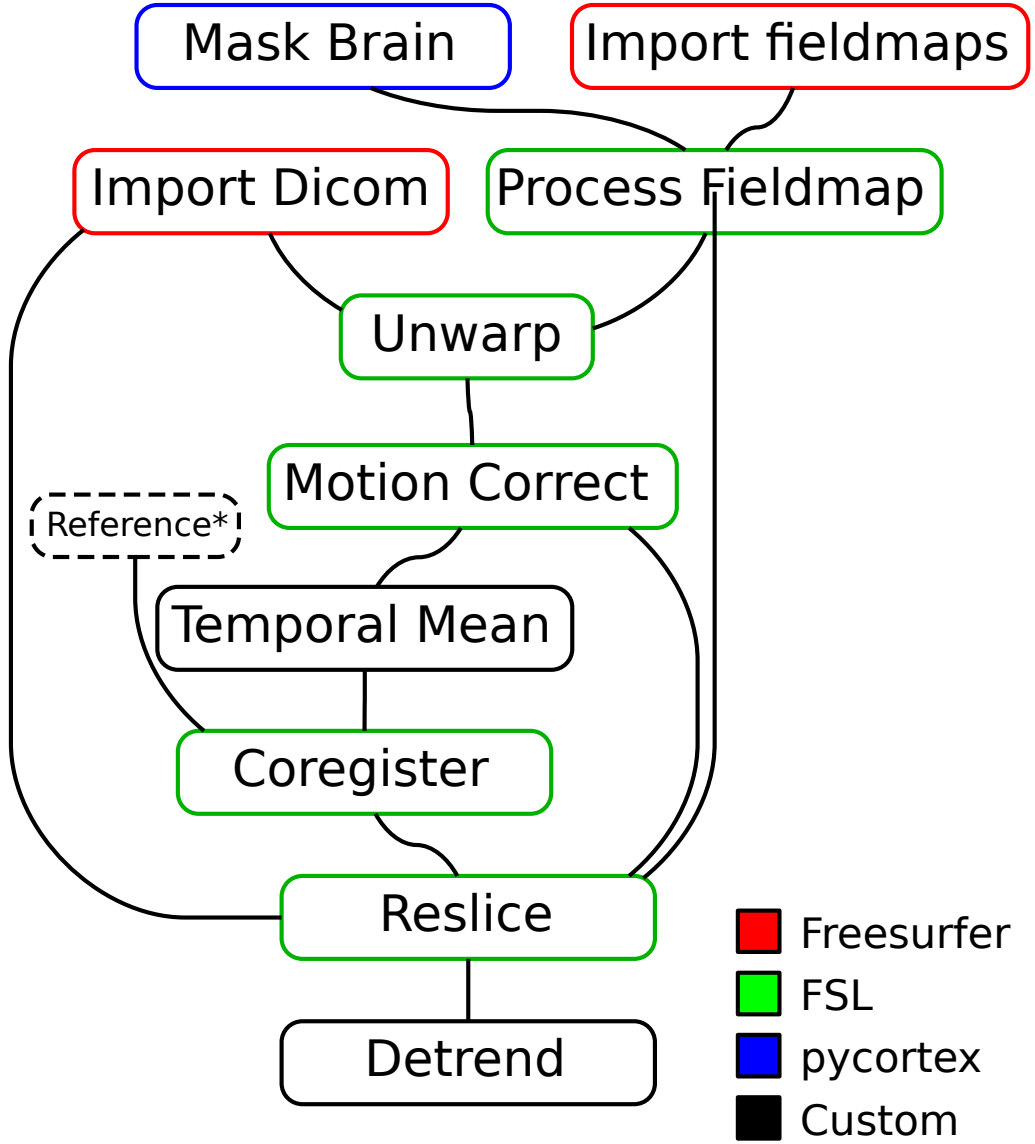
Wu, David and Gallant, *Annual Reviews of Neuroscience,* 2006

Sampling the stimulus and task space

- Sample as much of the stimulus/task space as possible.
- Obtain a good estimate of responses to the validation data.
- Make sure non-stationary responses are evenly distributed across both the estimation and validation data.

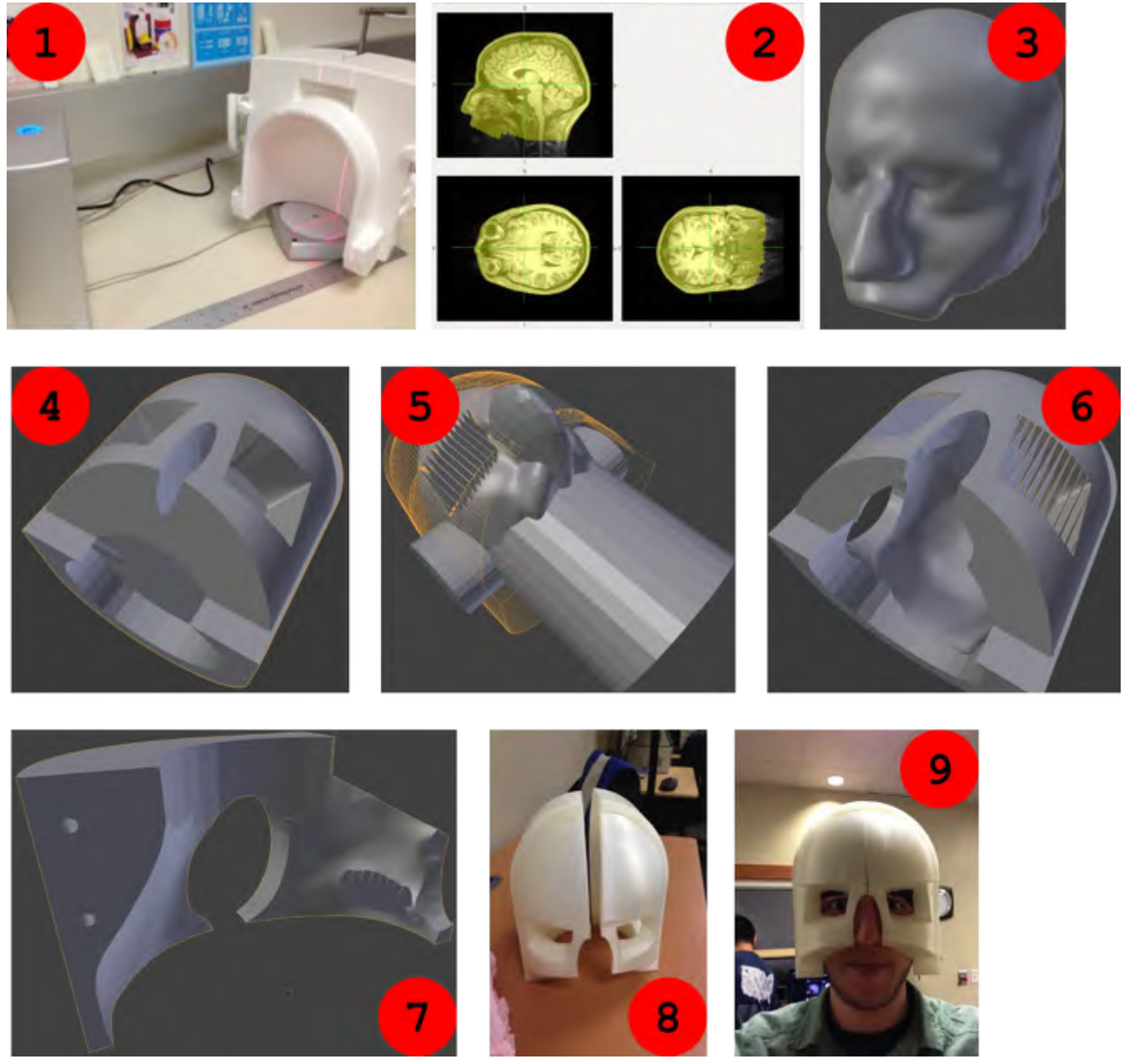


Our pre-processing pipeline



* Reference volume from first scan
or alternate session

Minimizing head motion



Gao, Huth & Gallant, *in preparation*

Maximizing signal while detrending

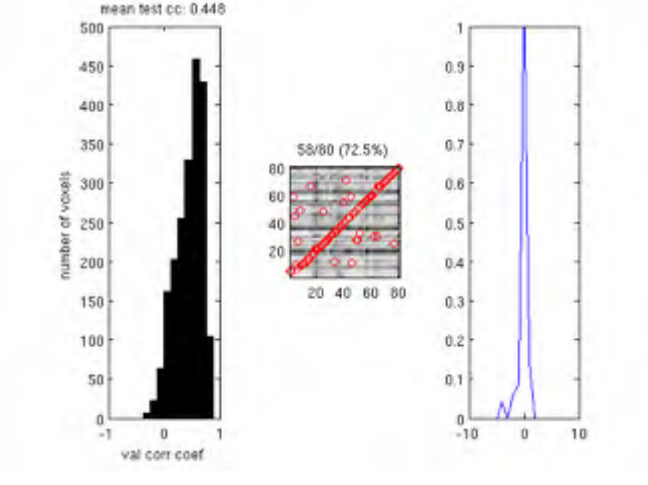
Siemens Tim Trio DACs: 20 bits

DICOM files: 16 bits

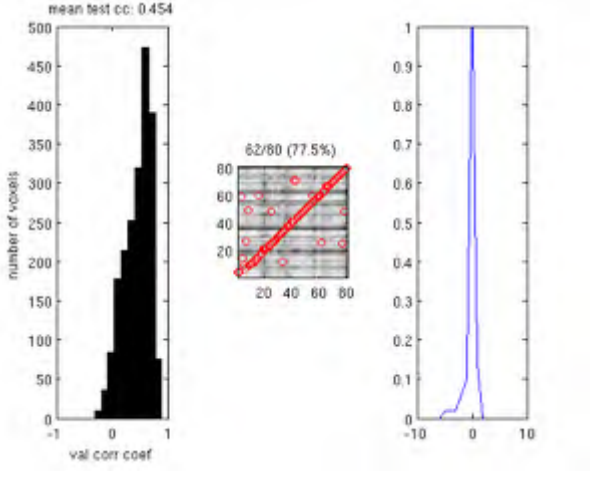
DICOM specification: 12 bits

Siemens DICOMs: 8 bits

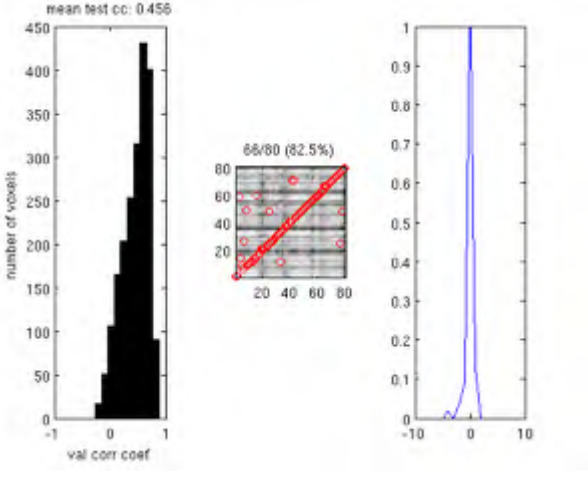
Polynomial



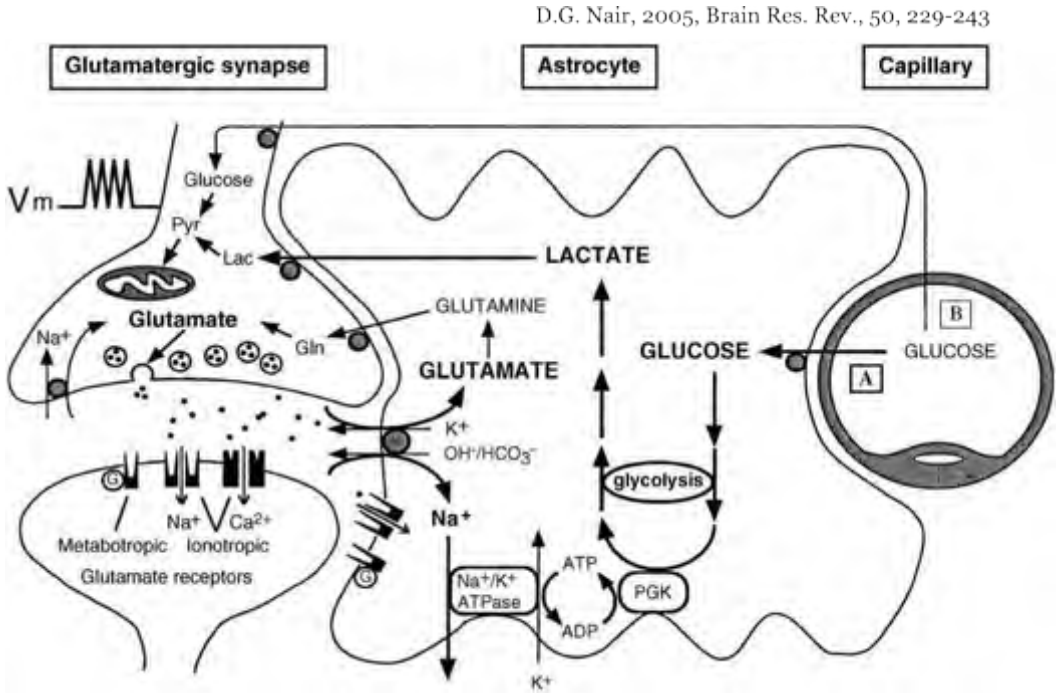
Median

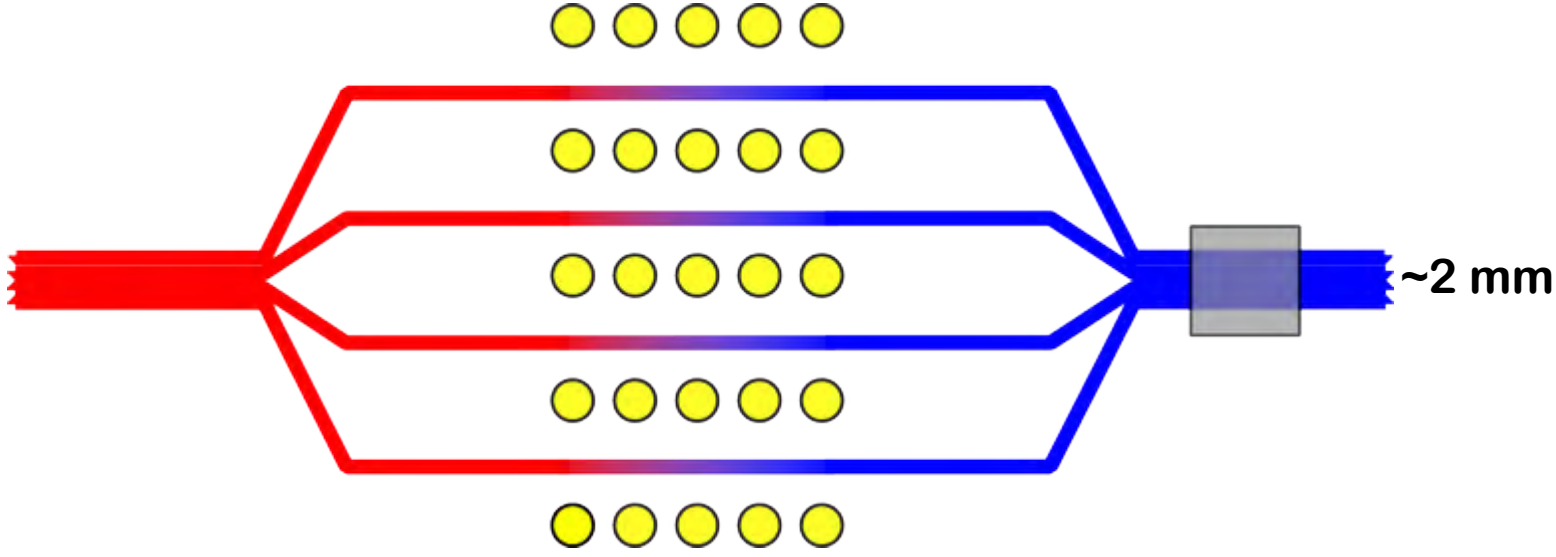


Savitsky-Golay

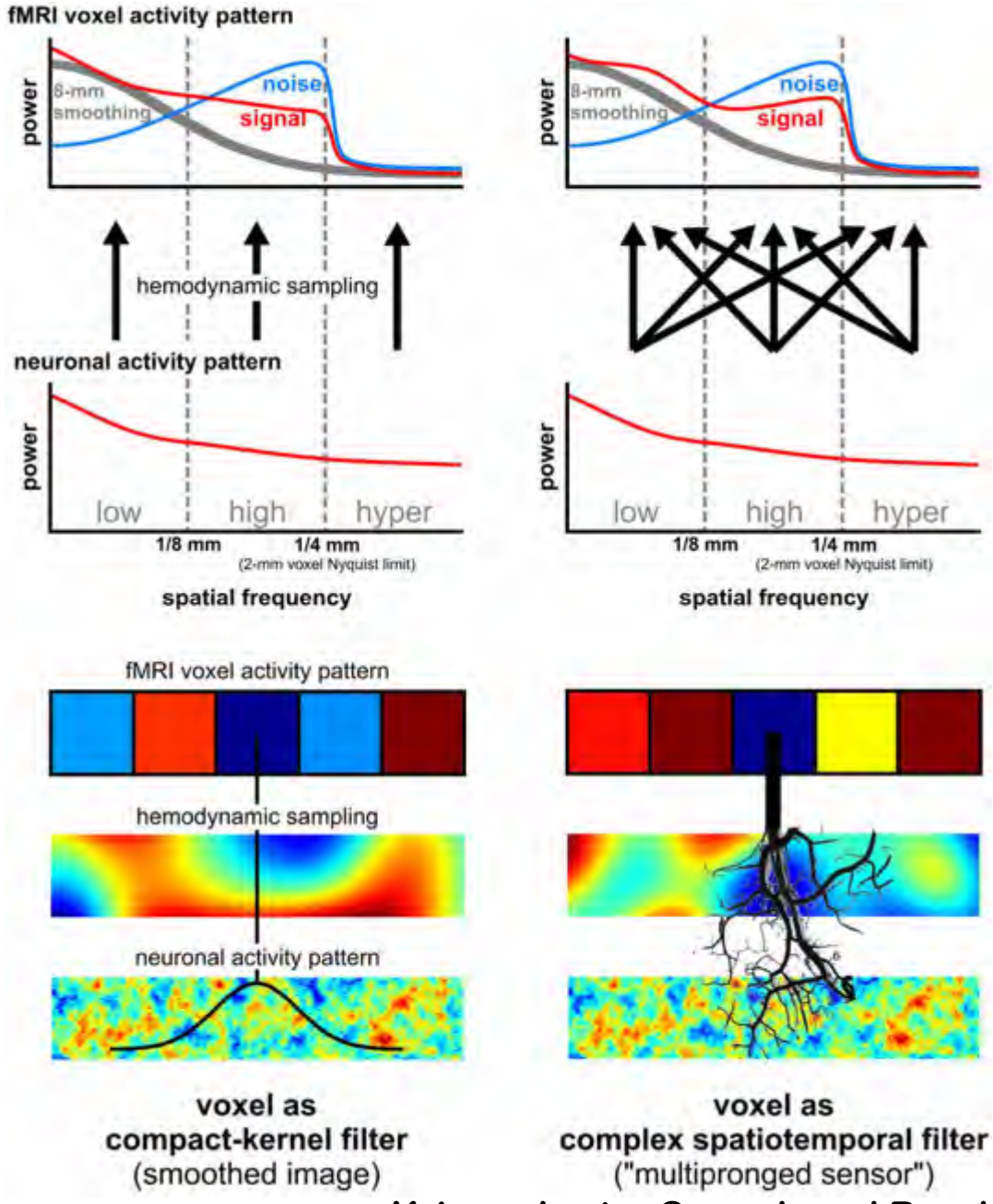


FMRI does NOT measure neural activity!





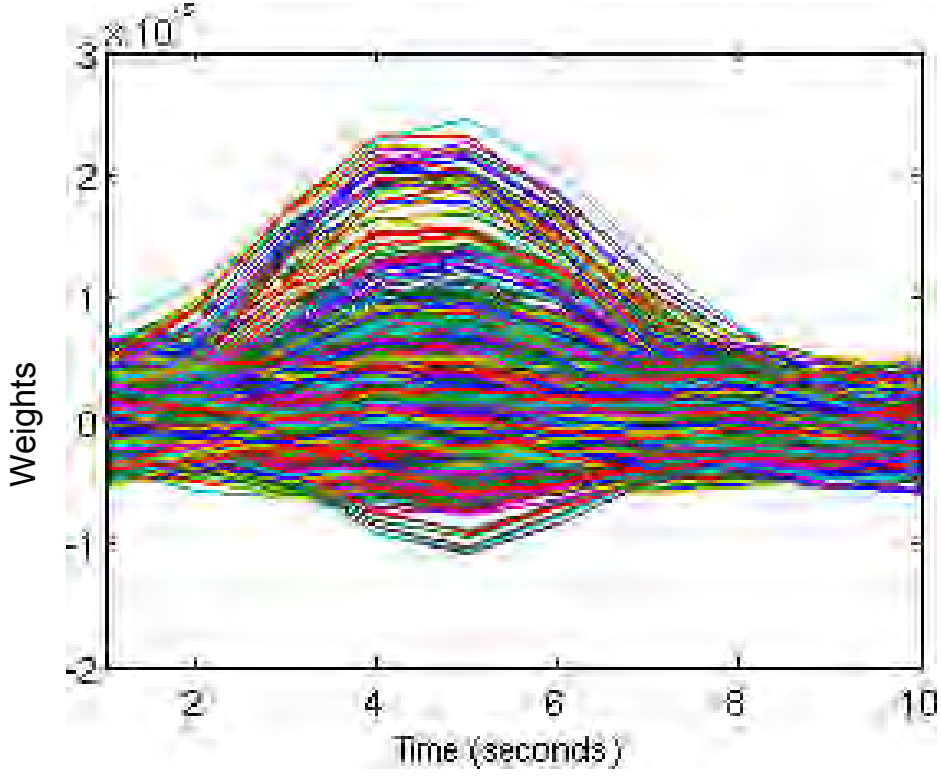
Complexities of hemodynamic coupling



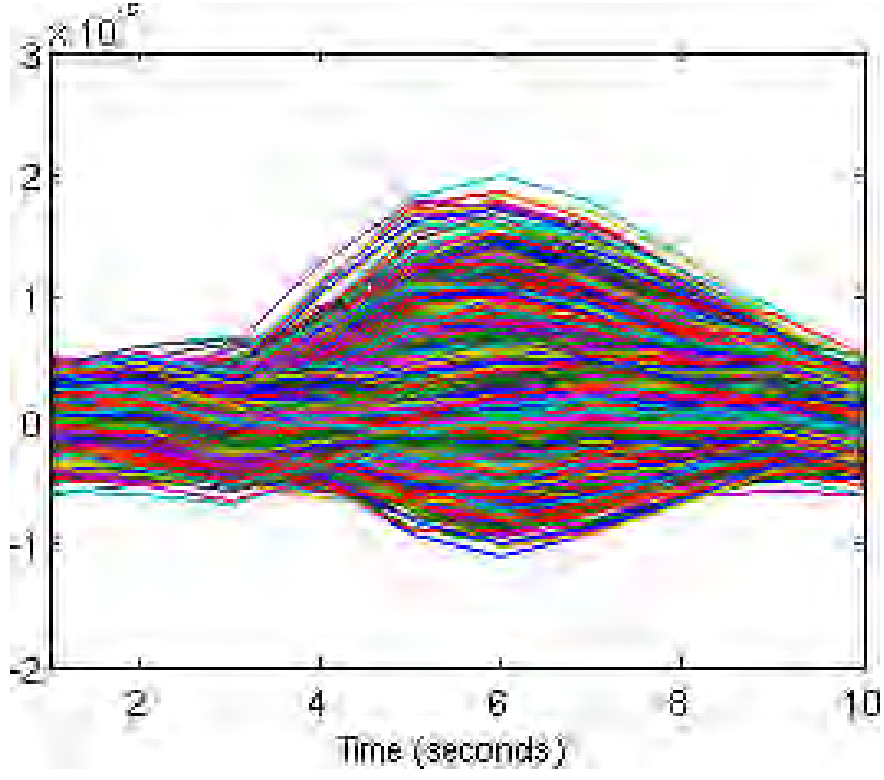
Kriegeskorte, Cusack and Bandettinni, *NeuroImage*, 2010

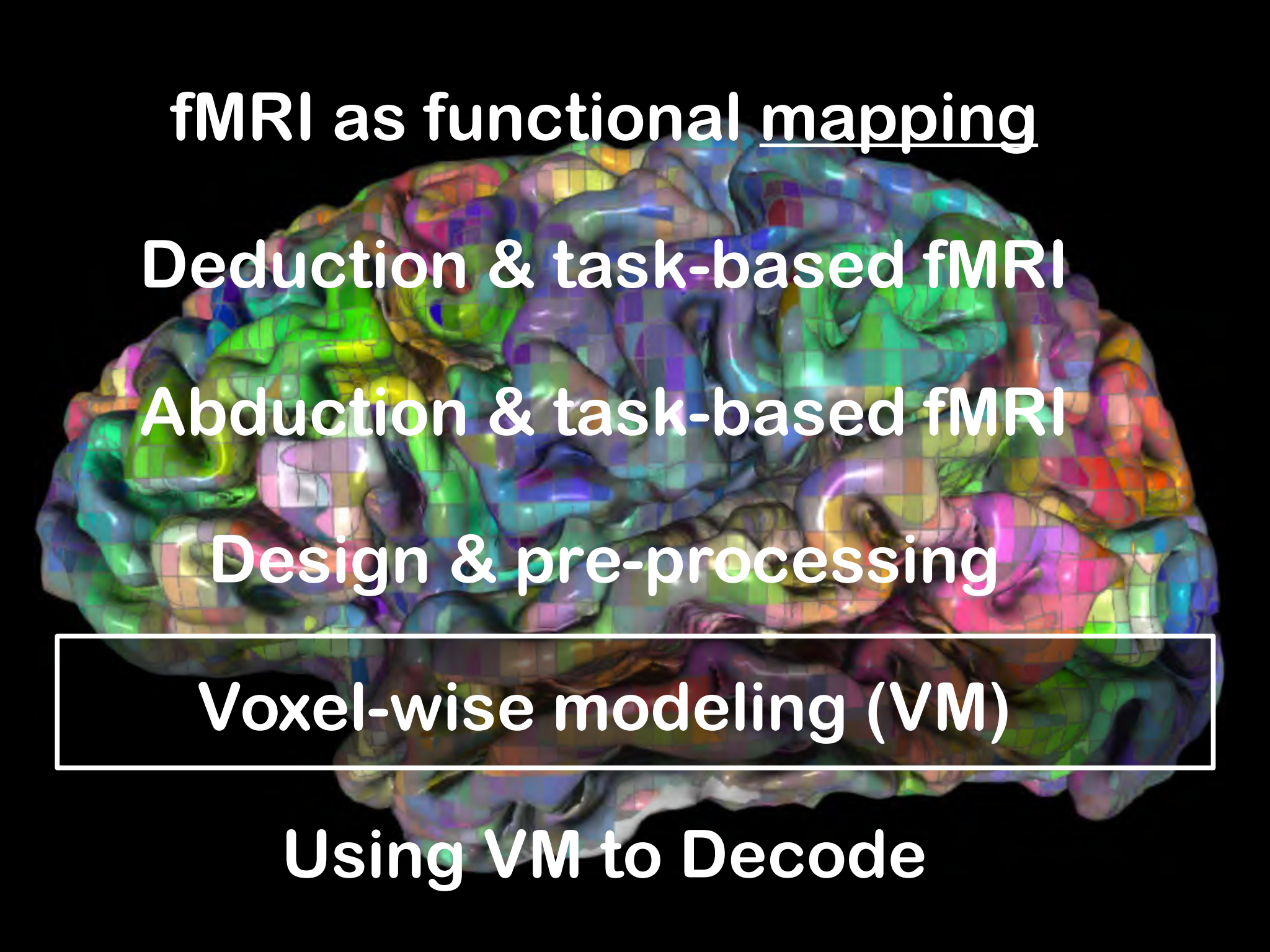
The HRF varies across voxels & features

Feature-dependent HRFs

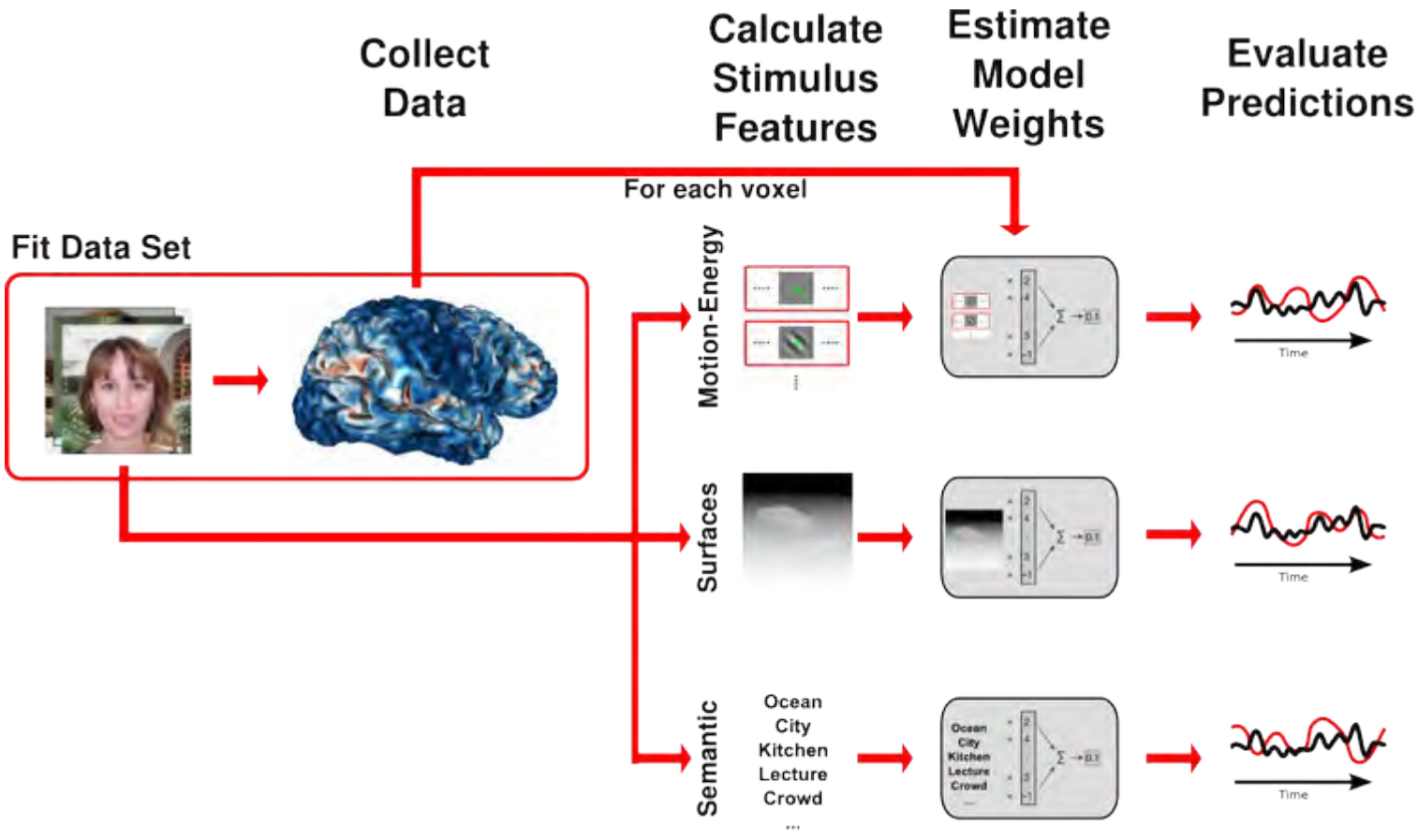


Cannonical HRFs

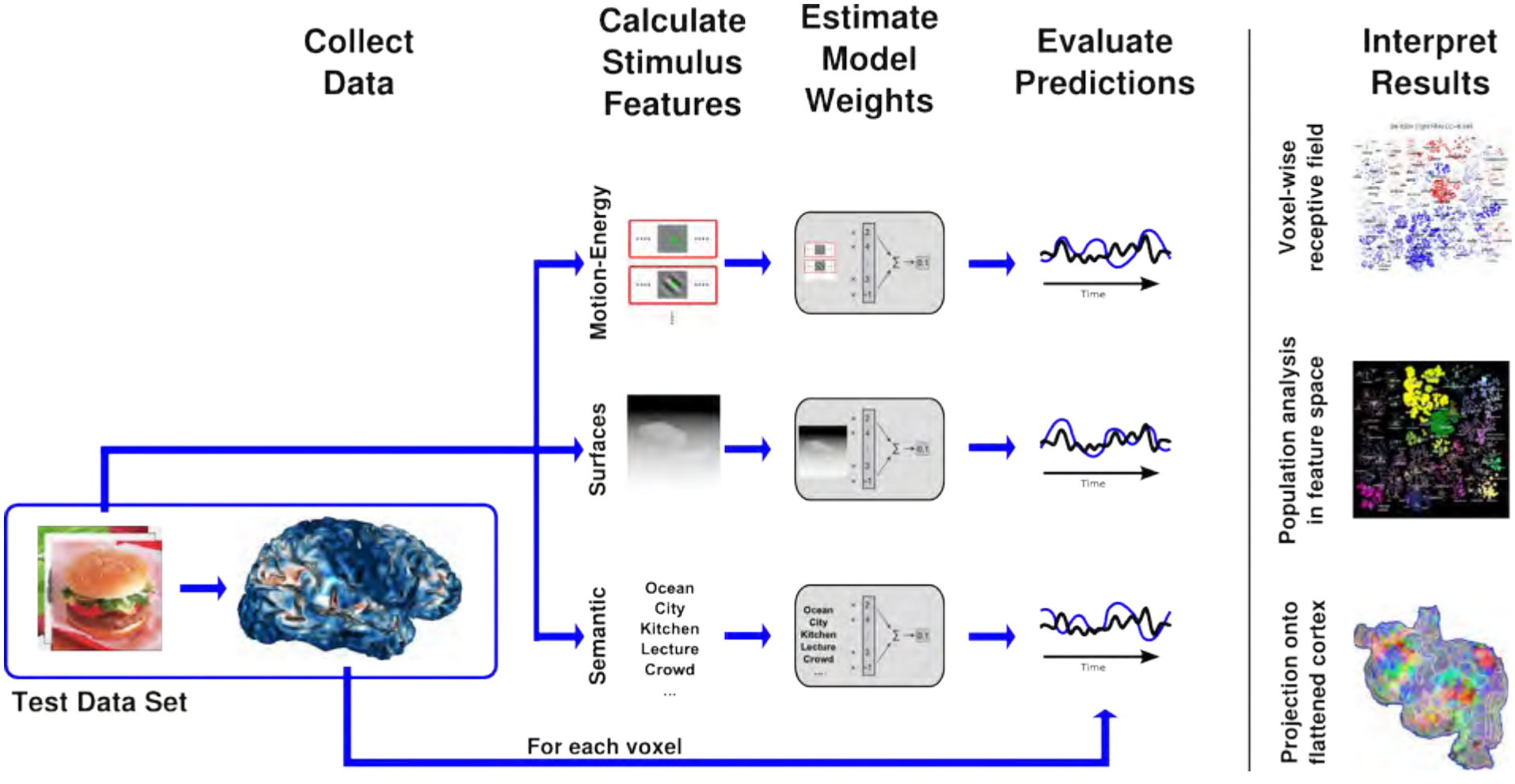



fMRI as functional mapping
Deduction & task-based fMRI
Abduction & task-based fMRI
Design & pre-processing
Voxel-wise modeling (VM)
Using VM to Decode

Typical example: VM for silent movie data

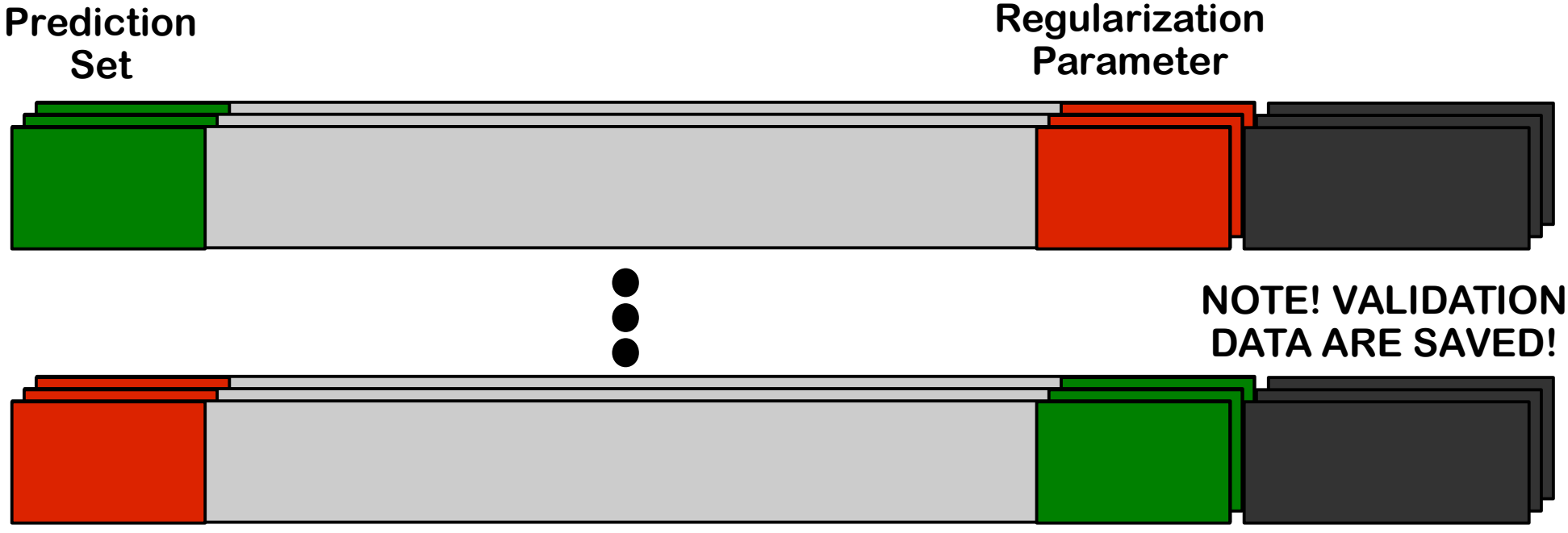


Typical example: VM of silent movie data



Fitting the models

- Replicate the data at time lags covering the HRF.
- Separate the estimation set into 3 subsets: 80% to fit the weights, 10% to fit the regularization parameter and 10% to evaluate predictions.
- Bootstrap the regularization and prediction sets.
- Average model weights across bootstrap samples.



A category model for high-level vision

Alex Huth

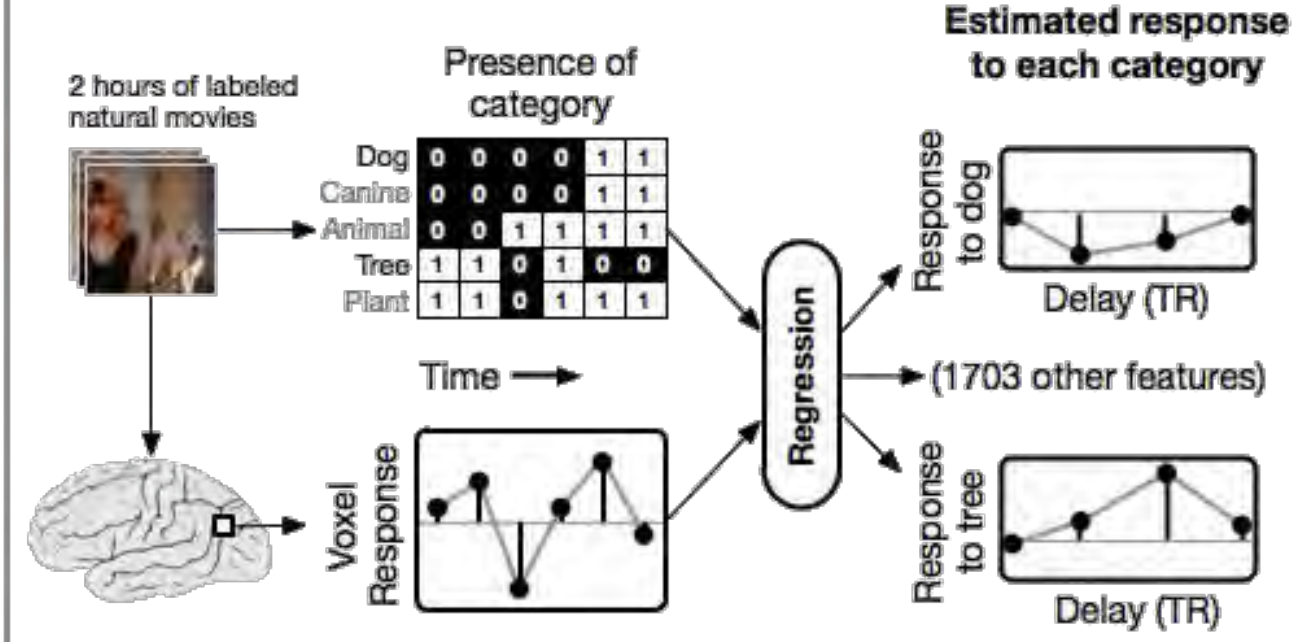
Labels:

manually assigned label		inferred labels
dog		woman
canine		adult
animal		person
⋮		⋮

Video clips labeled every 1s

Labels:

jaguar	tree
feline	vasc. pl.
animal	plant
organism	organism
object	object
⋮	⋮

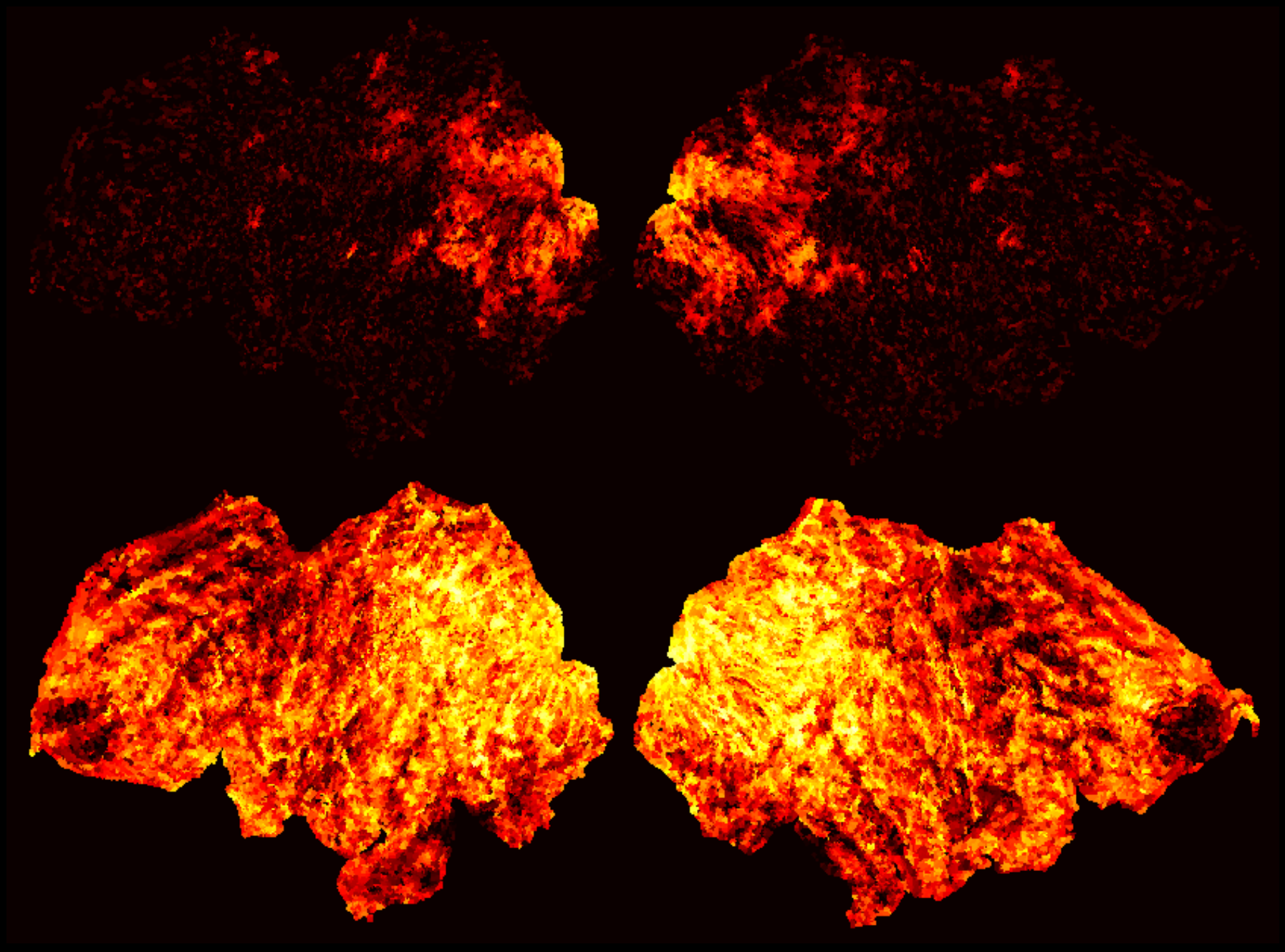


Huth, Nishimoto, Vu & Gallant, *Neuron*, 2012

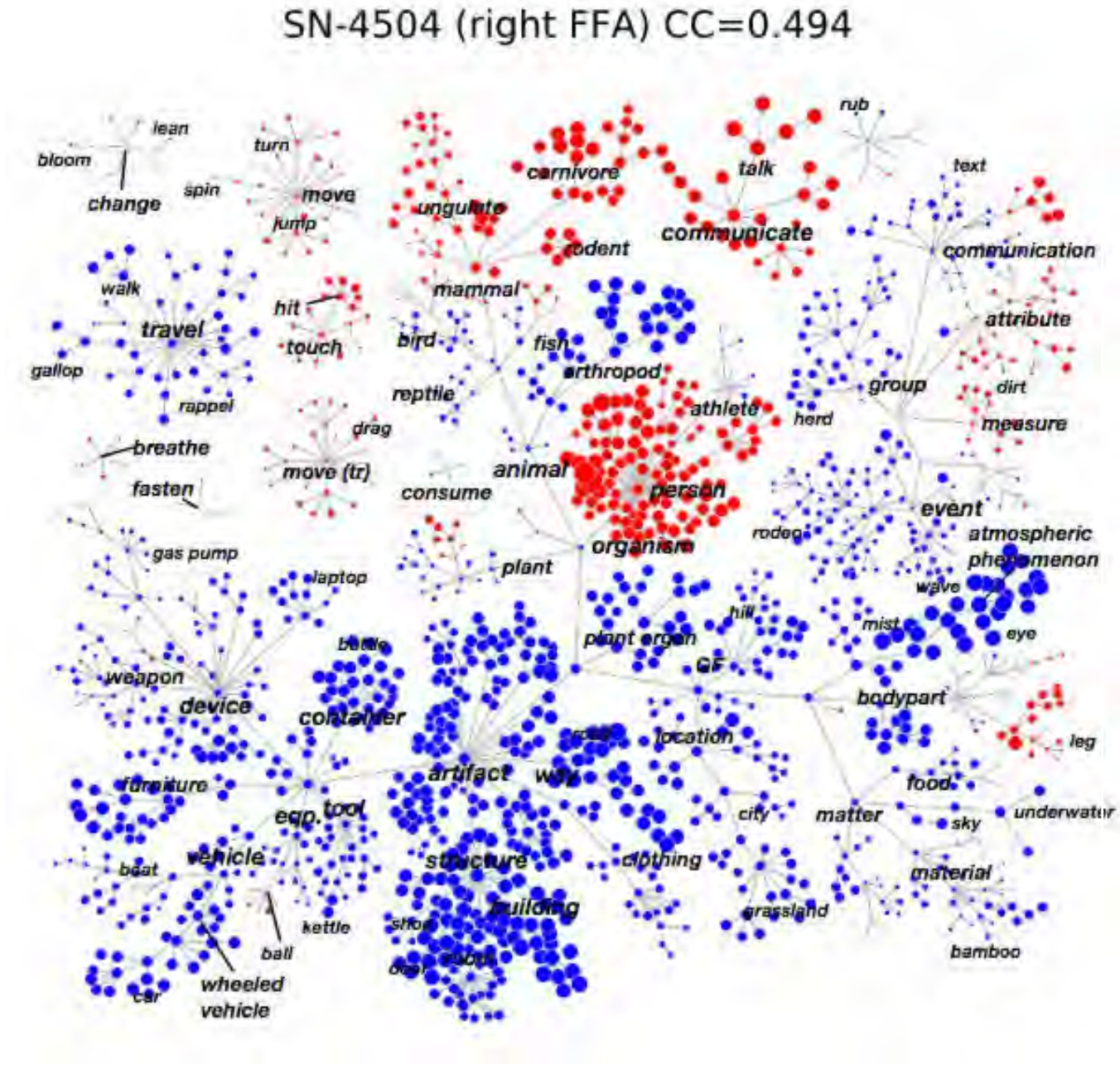
Predictions of the semantic category model

Huth, Nishimoto, Vu & Gallant, Neuron, 2012

Task-based signals are a fraction of all signals

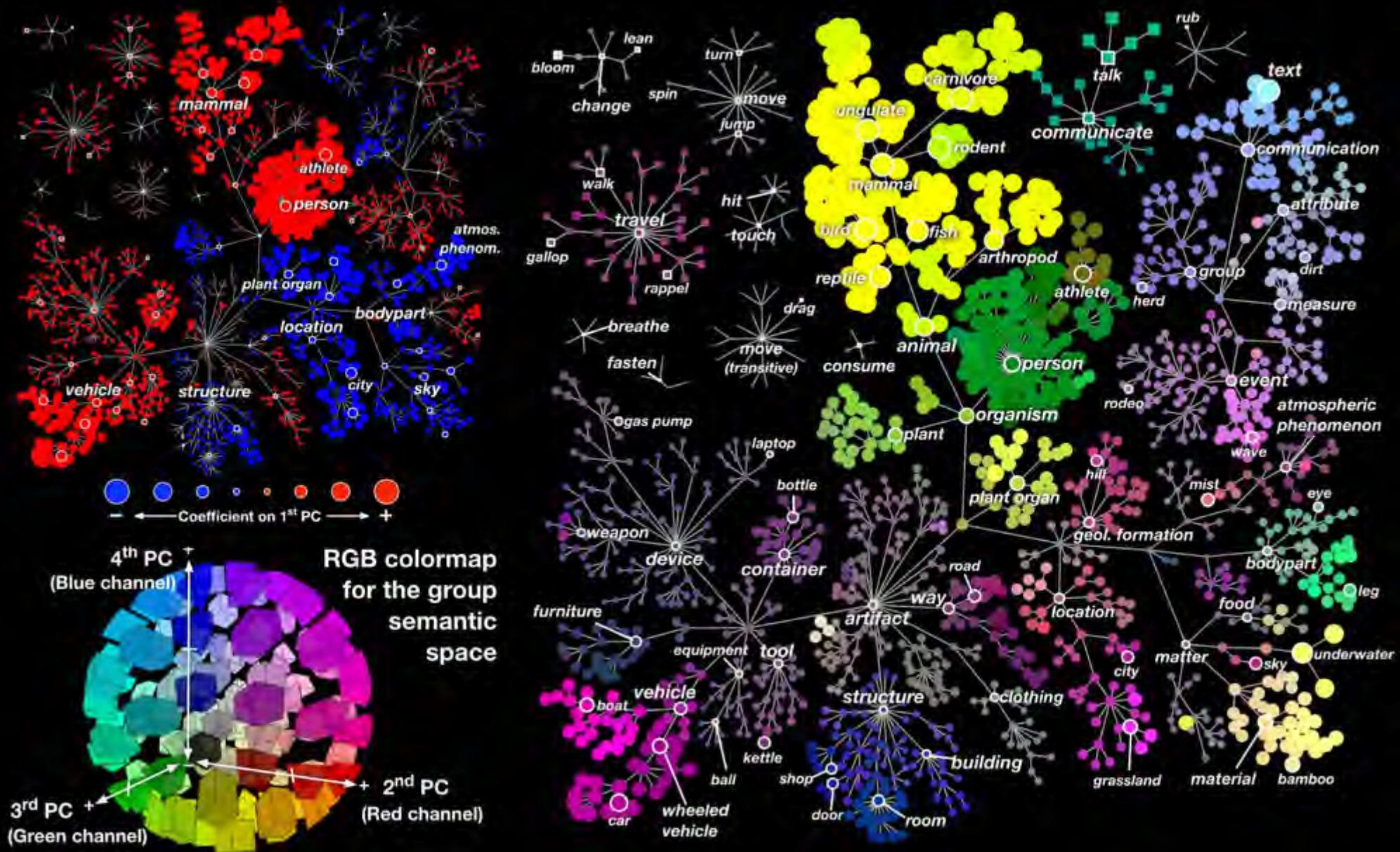


The category model for one FFA voxel



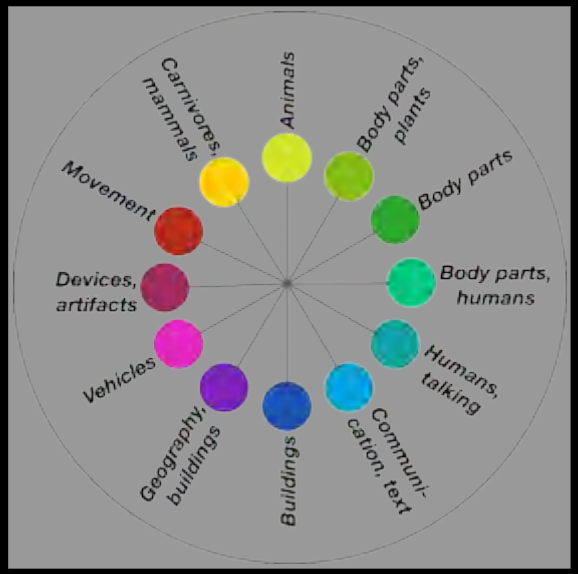
Huth, Nishimoto, Vu & Gallant, *Neuron*, 2012

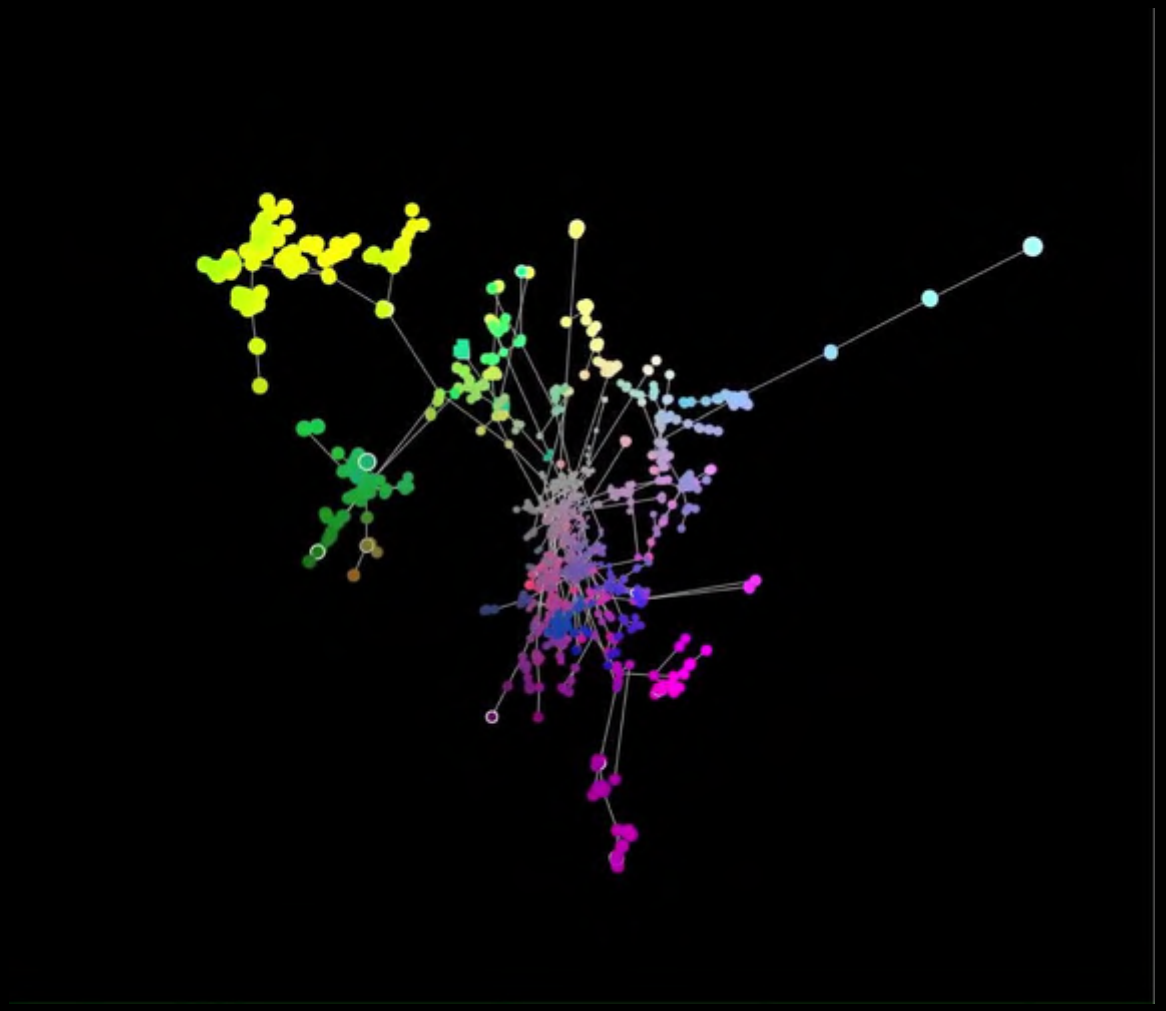
Representation of object and action categories



Huth, Nishimoto, Vu & Gallant, *Neuron*, 2012

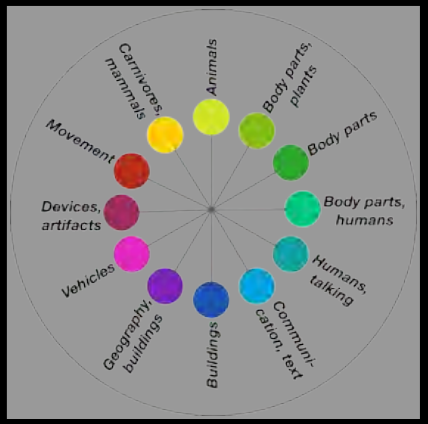
Representation of object and action categories





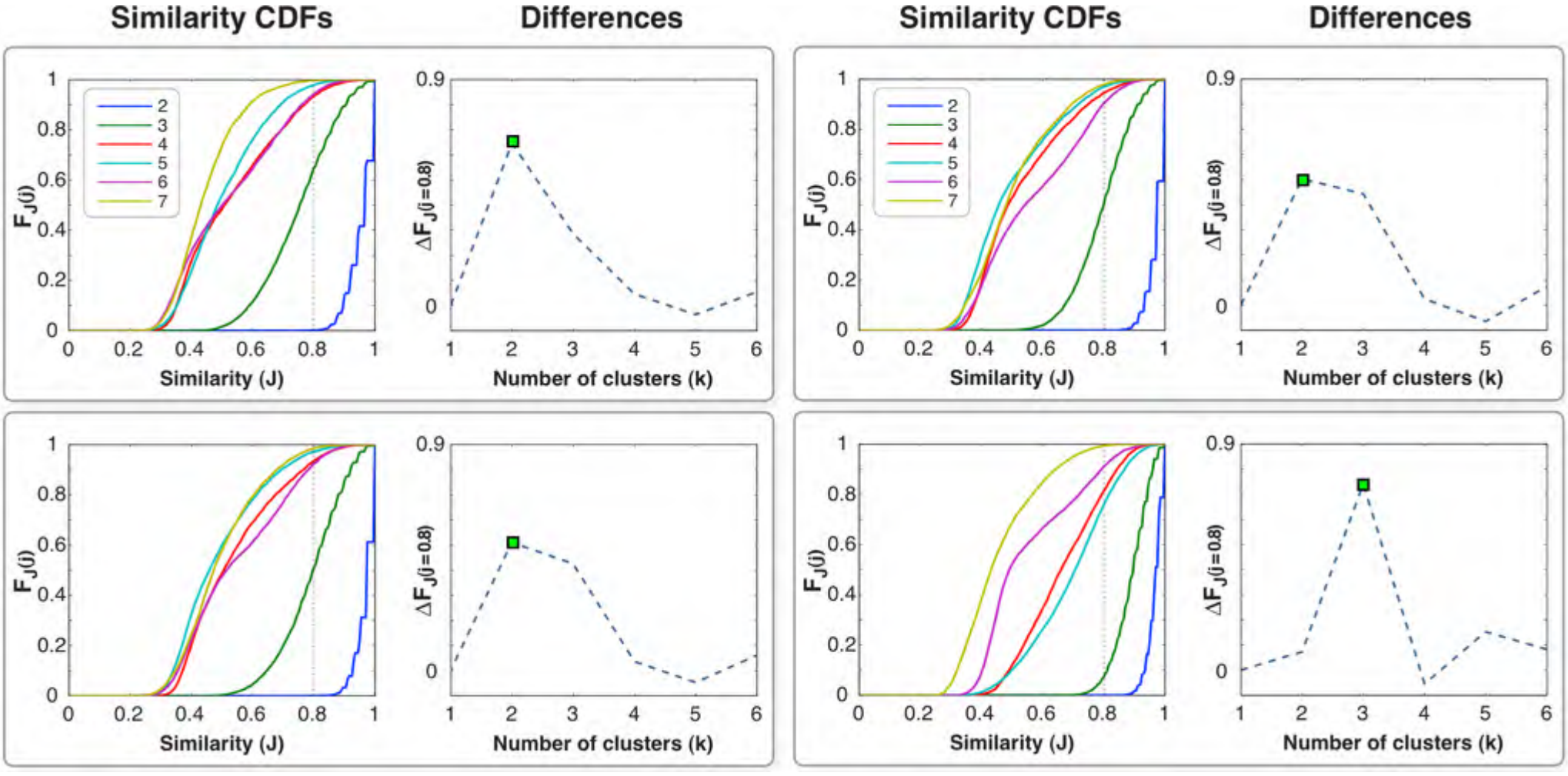
Huth, Nishimoto, Vu & Gallant, *Neuron*, 2012

Object and action category maps



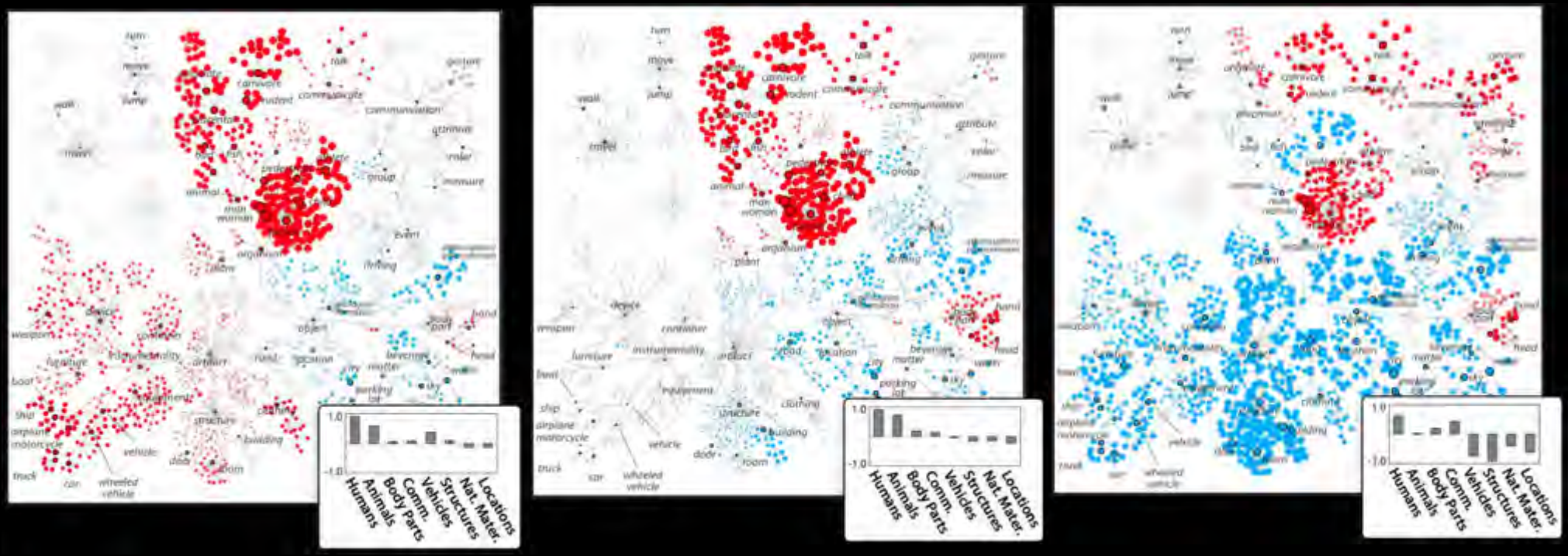
Huth, Nishimoto, Vu & Gallant, *Neuron*, 2012

Functional clusters within FFA?



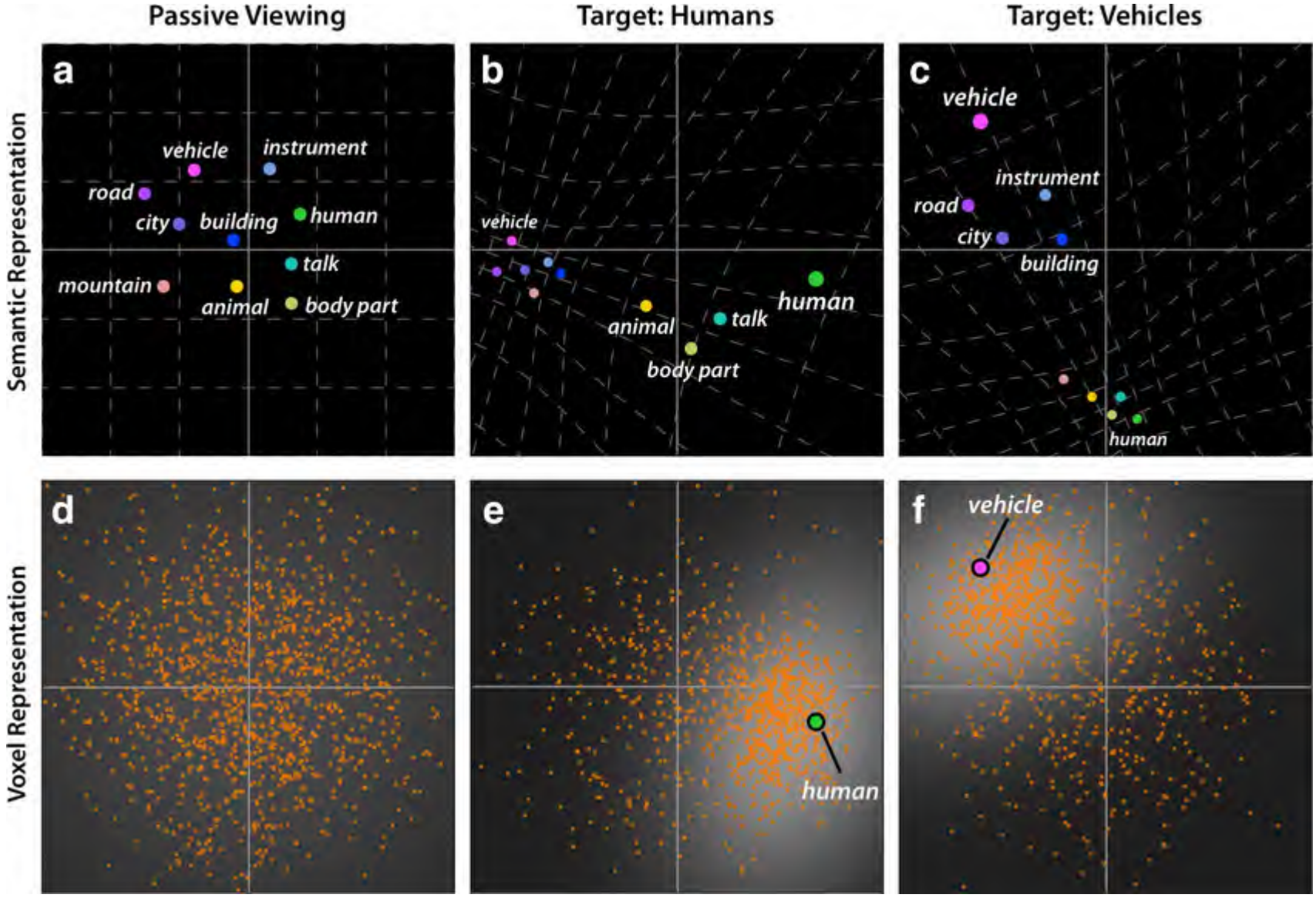
Cukur, Huth, Nishimoto & Gallant, *Journal of Neuroscience*, 2013

FFA consists of 3 separate functional clusters



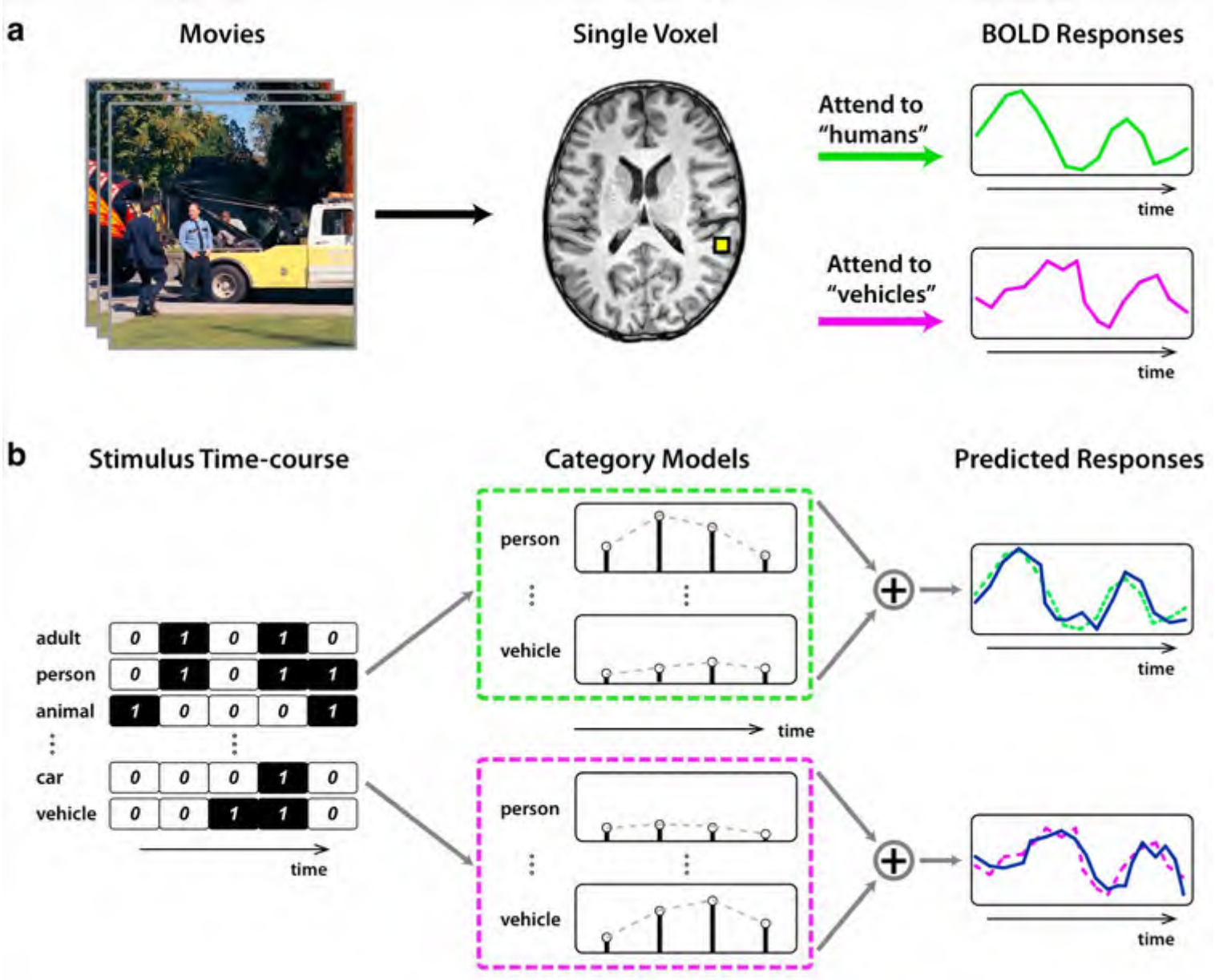
Cukur, Huth, Nishimoto & Gallant, *Journal of Neuroscience.*, 2013

Attention alters the semantic space



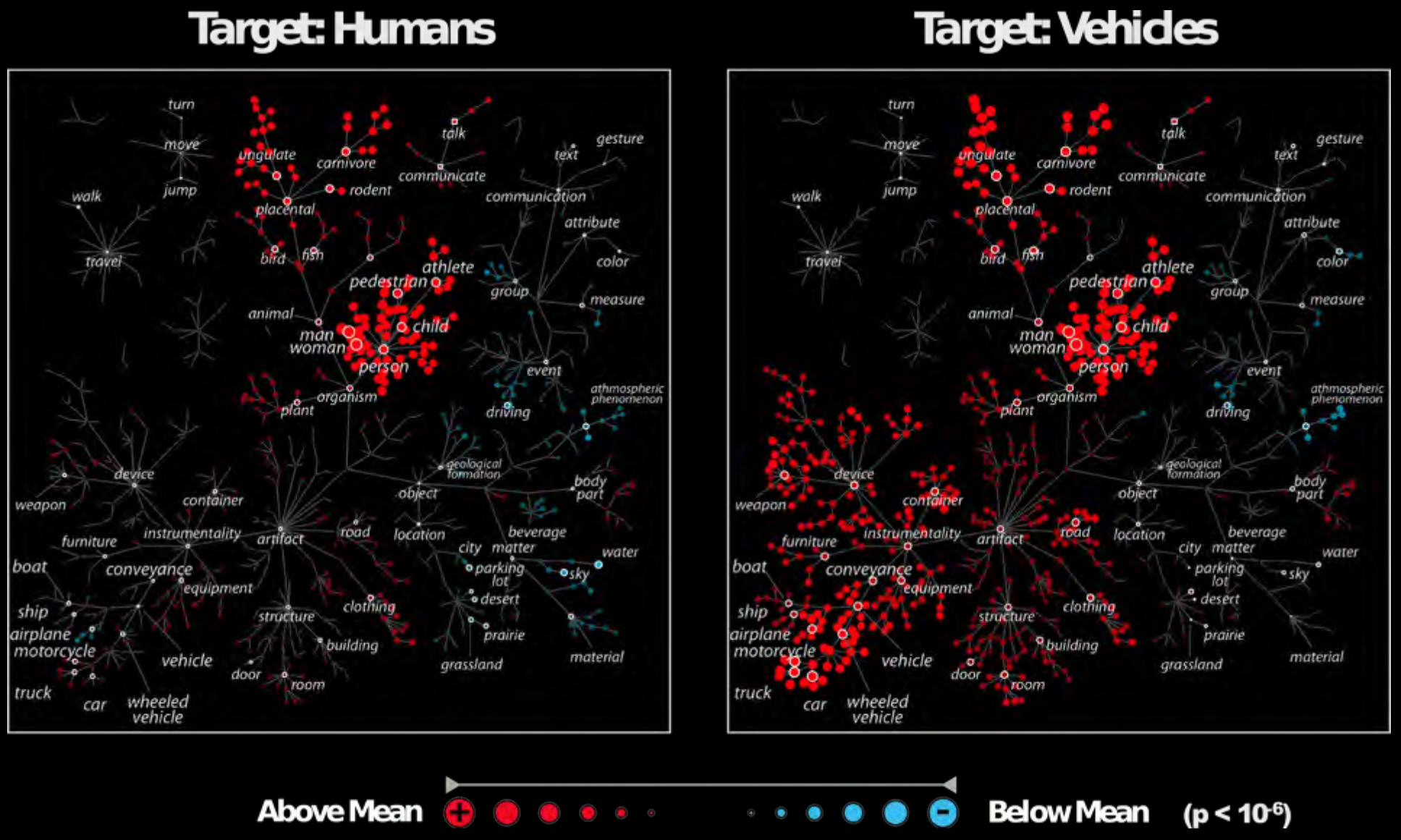
Cukur, Nishimoto, Huth & Gallant, *Nature Neuroscience*, 2013

Fitting the category attention model



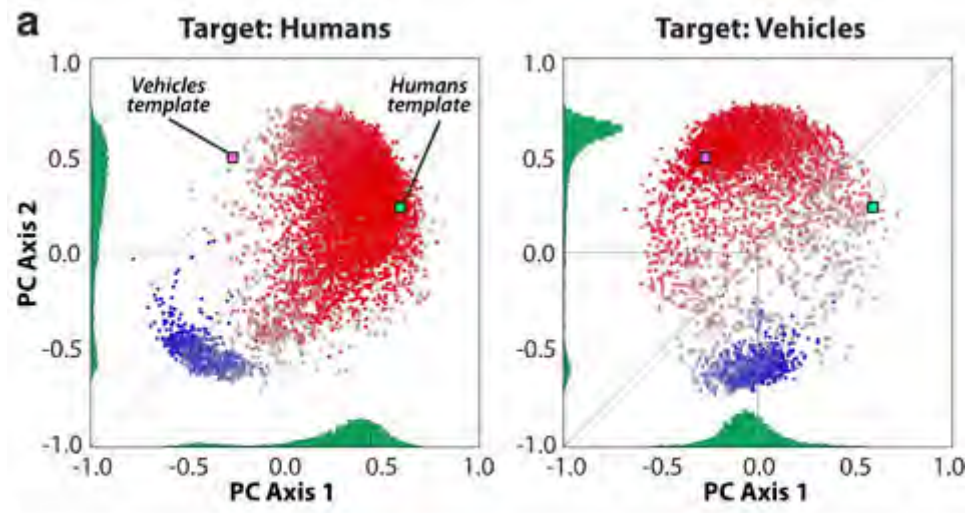
Cukur, Nishimoto, Huth & Gallant, *Nature Neuroscience*, 2013

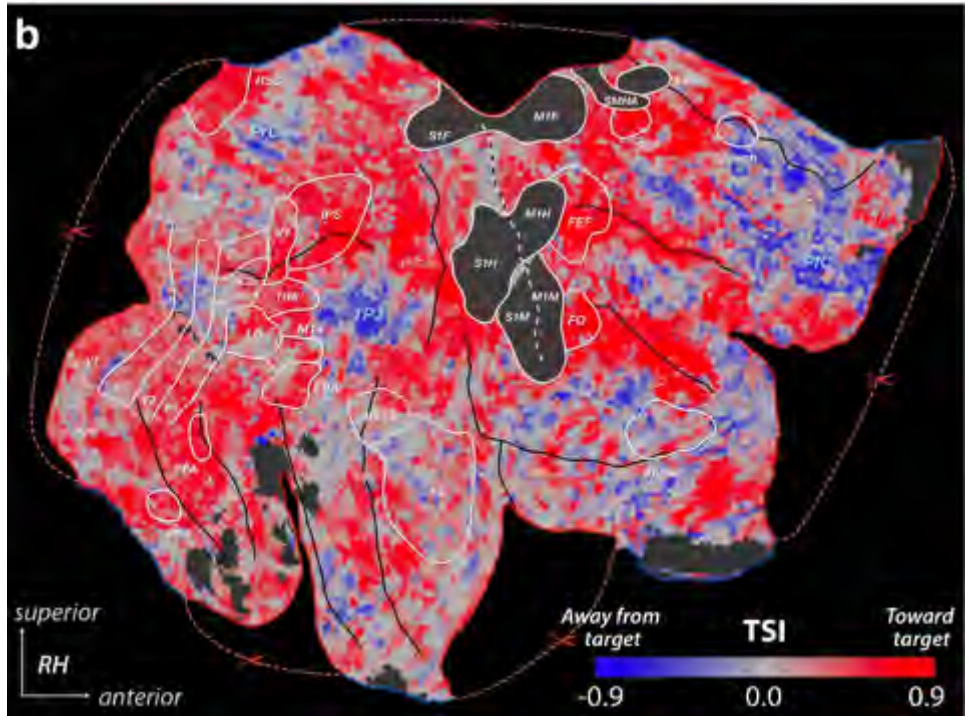
Attention shifts voxel tuning



Cukur, Nishimoto, Huth & Gallant, *Nature Neuroscience*, 2013

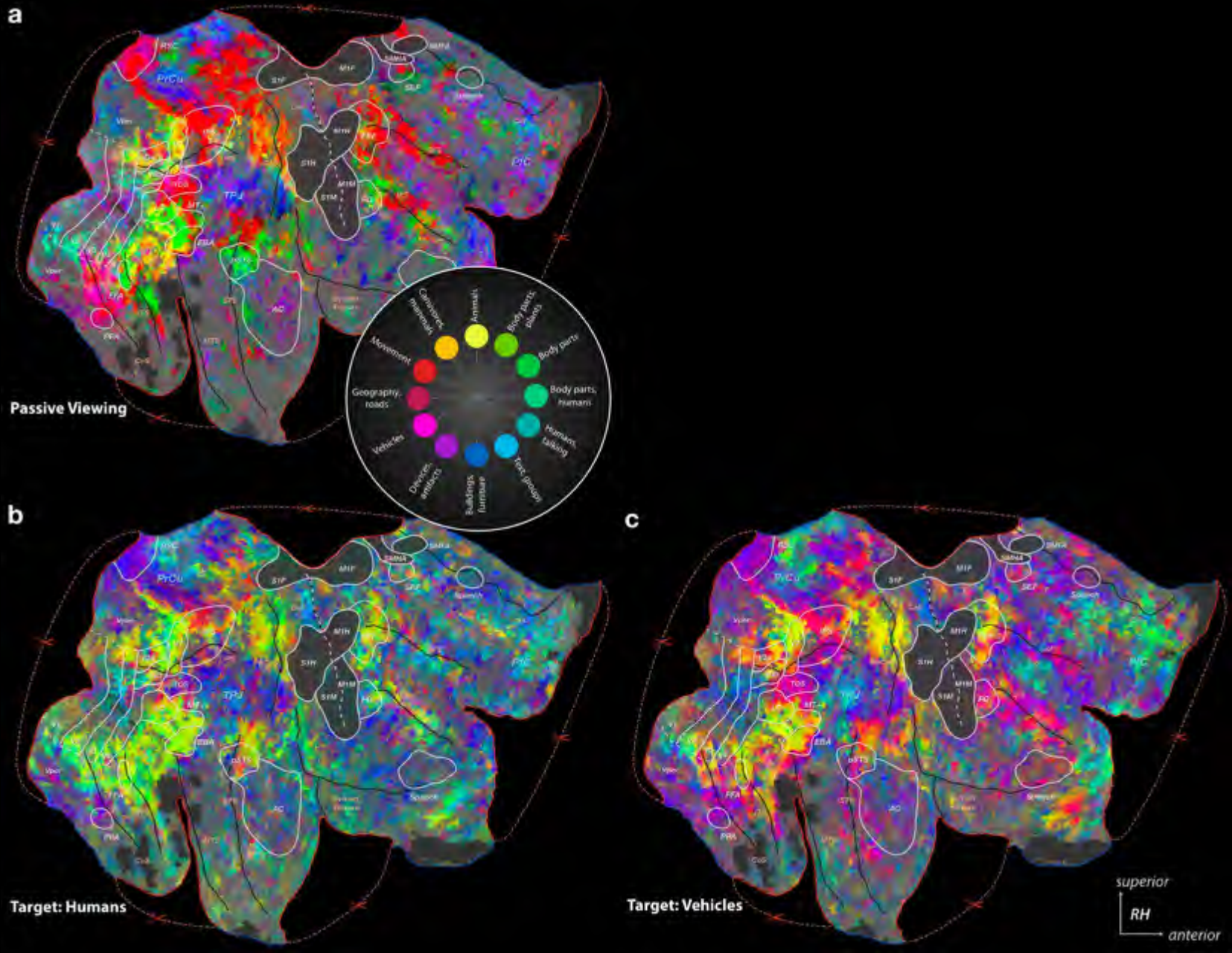
Attention shifts category tuning in single voxels





Cukur, Nishimoto, Huth & Gallant, *Nature Neuroscience*, 2013

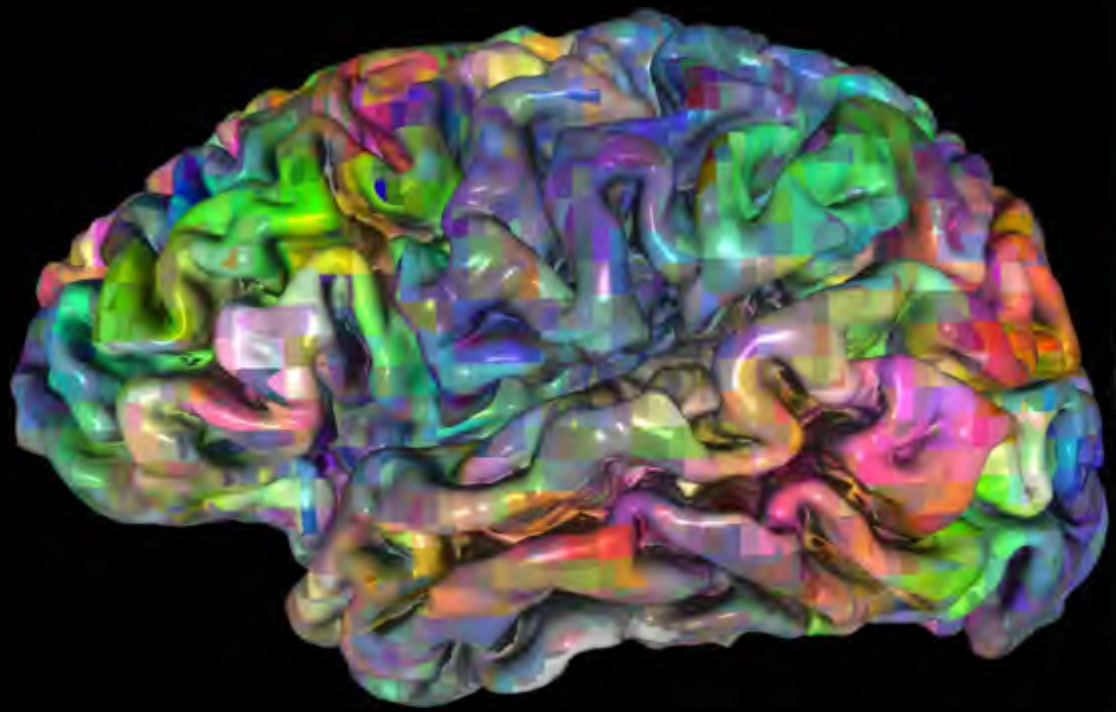
Attention changes cortical maps



Cukur, Nishimoto, Huth & Gallant, *Nature Neuroscience*, 2013

VM is more efficient than localizers/MVPA

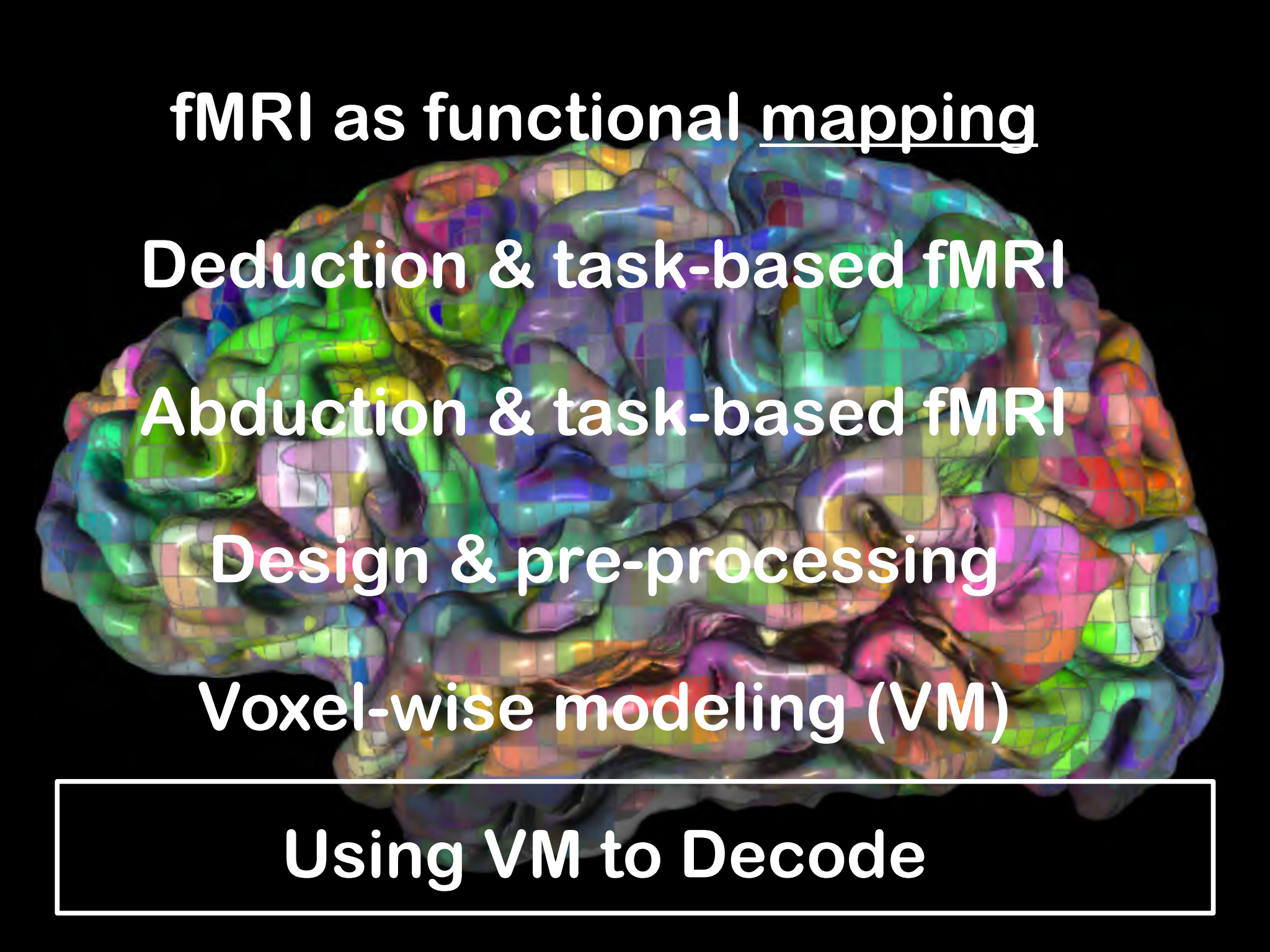
Voxel-wise modeling
(> 6-10 bits / voxel)



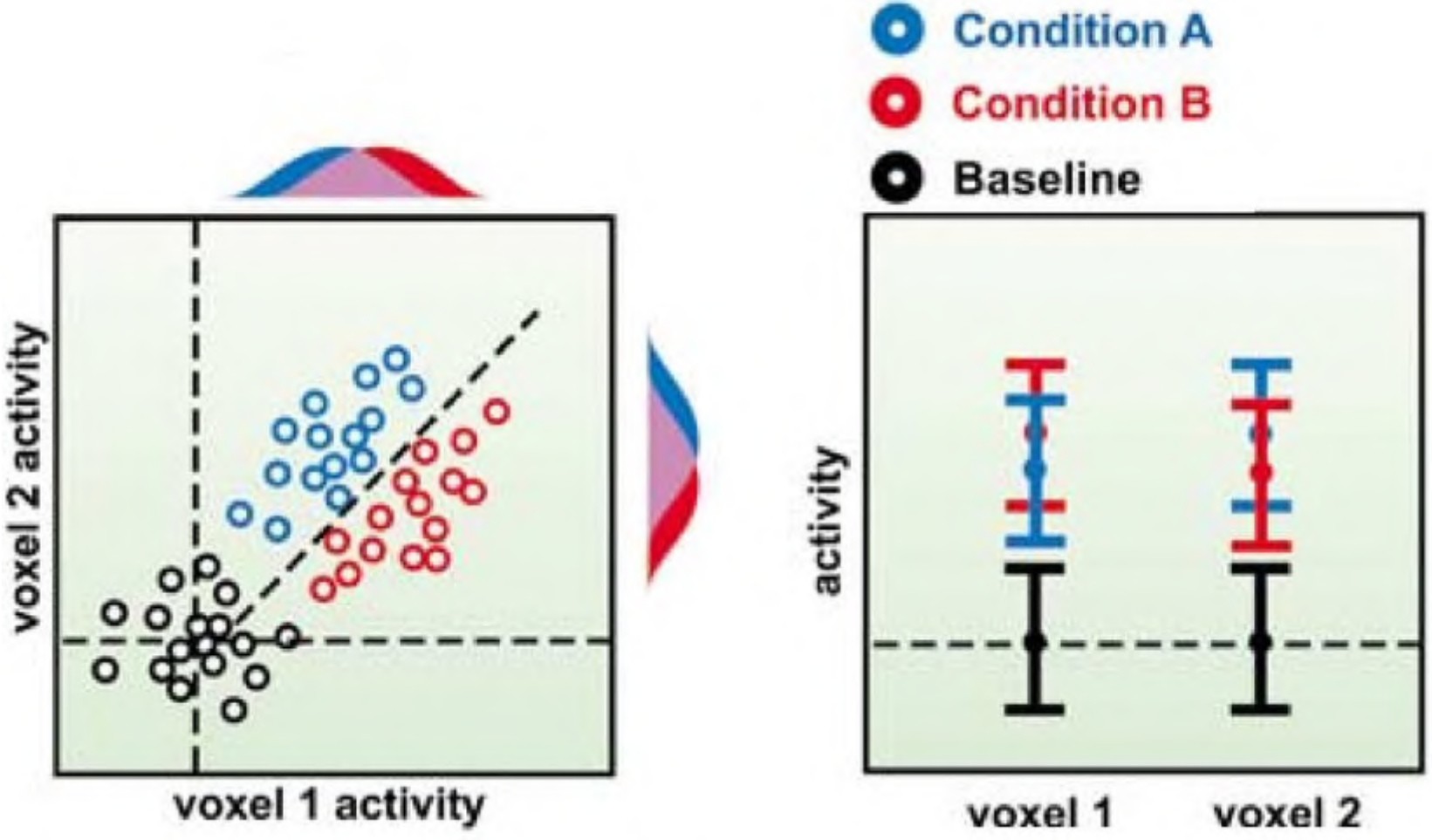
Huth et al, *in review*

Semantic localizer
(< 1 bit / voxel)

Fedorenko & Kanwisher,
J. Neurophys., 2010

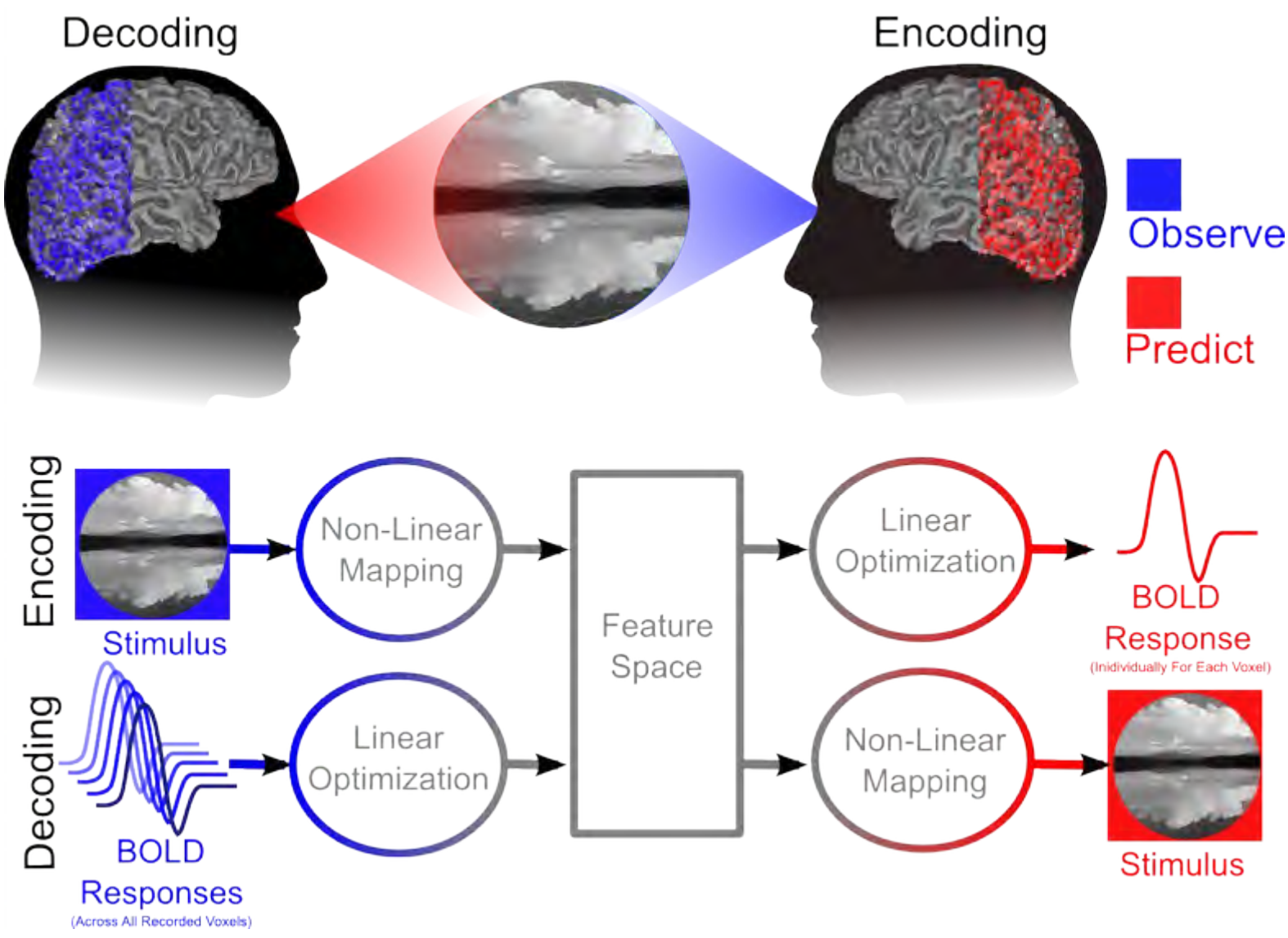

fMRI as functional mapping
Deduction & task-based fMRI
Abduction & task-based fMRI
Design & pre-processing
Voxel-wise modeling (VM)
Using VM to Decode

The most common decoding method is MVPA



Cox & Savoy, 2003

Encoding versus decoding

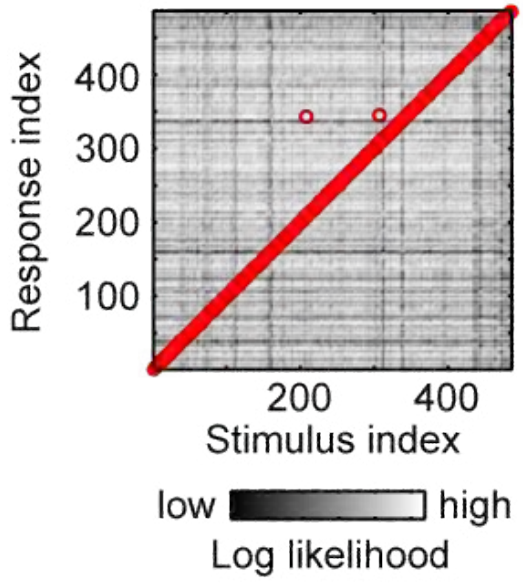


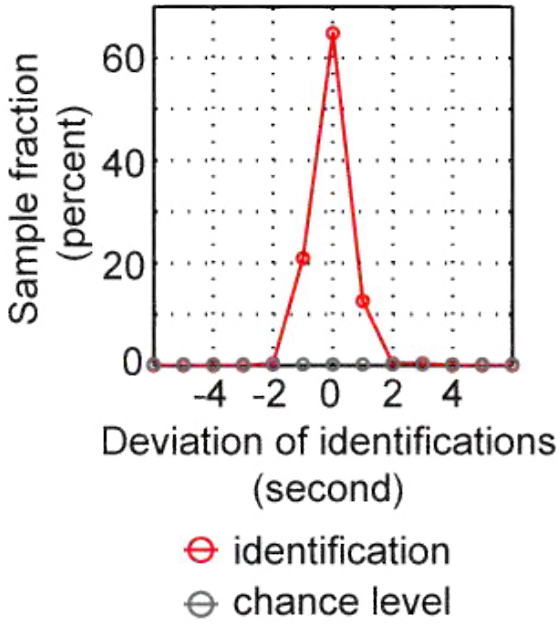
$$P(f(S)|R) \propto P(R|f(S))\,P(f(S))$$

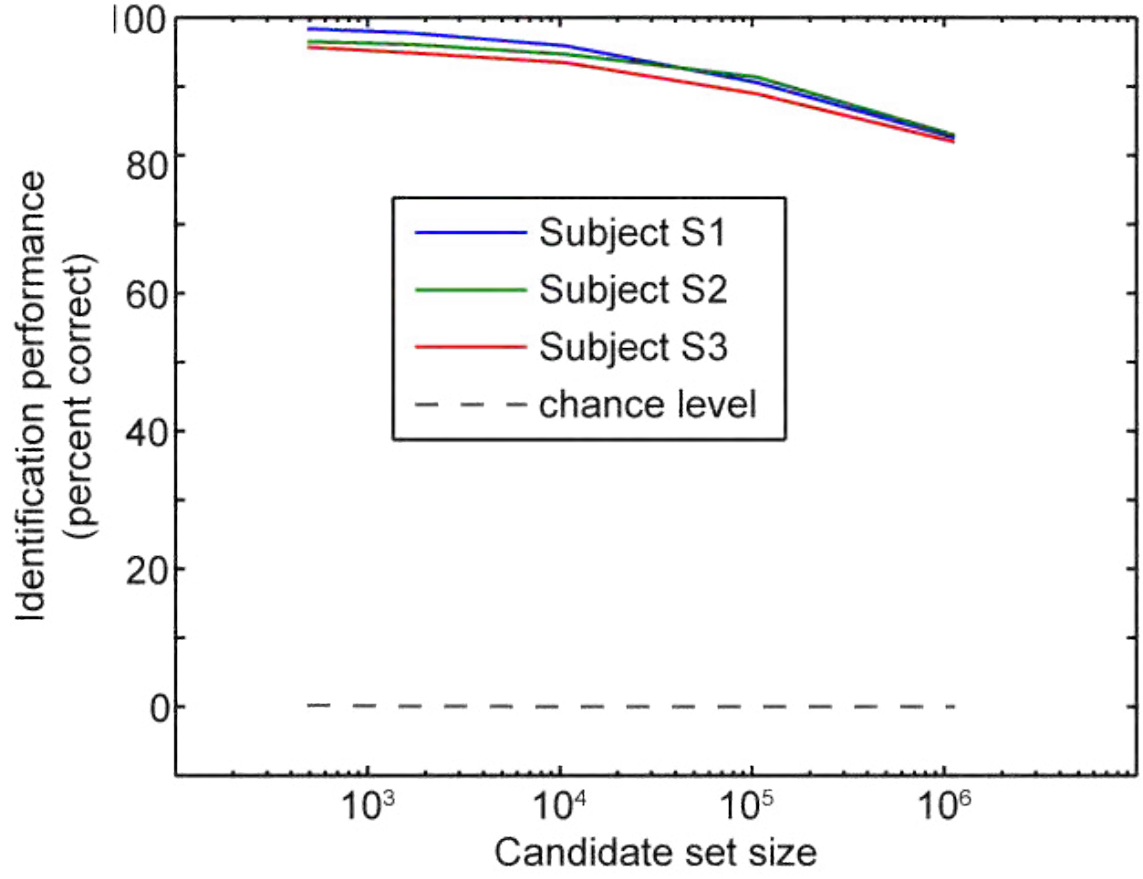
Encoding versus decoding

Encoding and decoding are scientifically equivalent EXCEPT that you cannot estimate the noise ceiling for a decoder.

Movie identification by the motion energy model







Nishimoto, Vu, Naselaris, Benjamini, Yu and Gallant, *Current Biology*, 2011

The motion-energy model decodes movies

Nishimoto, Vu, Naselaris, Benjamini, Yu and Gallant, *Current Biology*, 2011

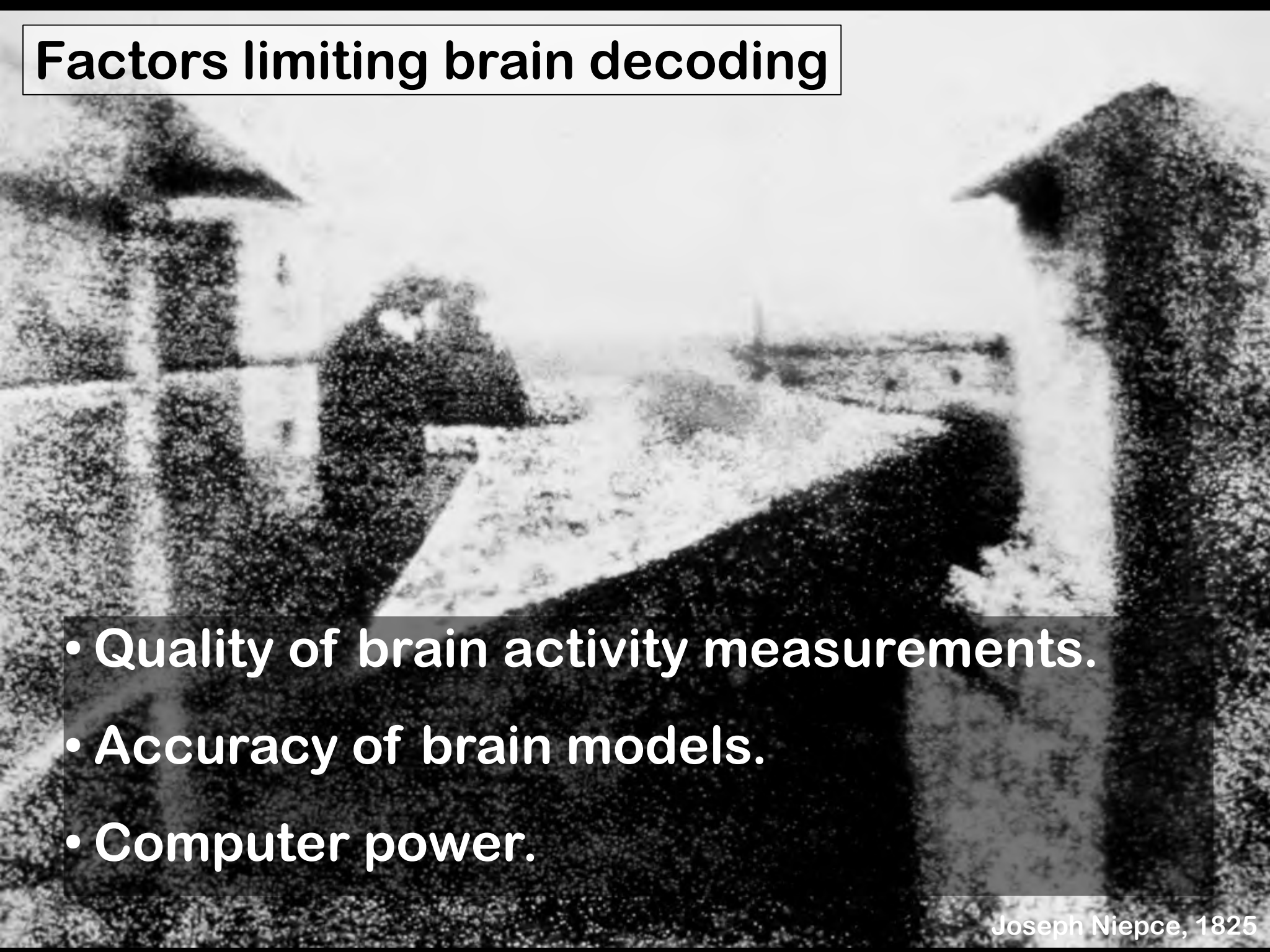
The category model decodes objects & actions

Movie

Likely Objects and Actions

Huth, Lee, Nishimoto & Gallant, *in preparation*

Factors limiting brain decoding



- Quality of brain activity measurements.
- Accuracy of brain models.
- Computer power.

Joseph Niepce, 1825

Advantages of voxel-wise modeling

- More sensitive and specific than any other method.
- Produces useful results in single subjects.
- Produces maps at the finest scale of detail available.
- Does not require defining ROIs, but can be used to discover ROIs and gradients.
- Reveals substructure and detailed tuning within ROIs.
- Produces estimates of both significance AND effect size.
- Makes visualization and interpretation simple.
- Allows predictions out of the fit set, and provides a principled platform for decoding.
- Can be generalized to include voxel cross-correlations or group-level analysis.
- Can be used to decode brain activity with the highest accuracy currently attainable.

Current Lab Members

Natalia Bilenko
James Gao
Alex Huth
Fatma Imamoglu
Mark Lescroart
Lydia Majur
Anwar Nunez
Michael Oliver
Dustin Stansbury

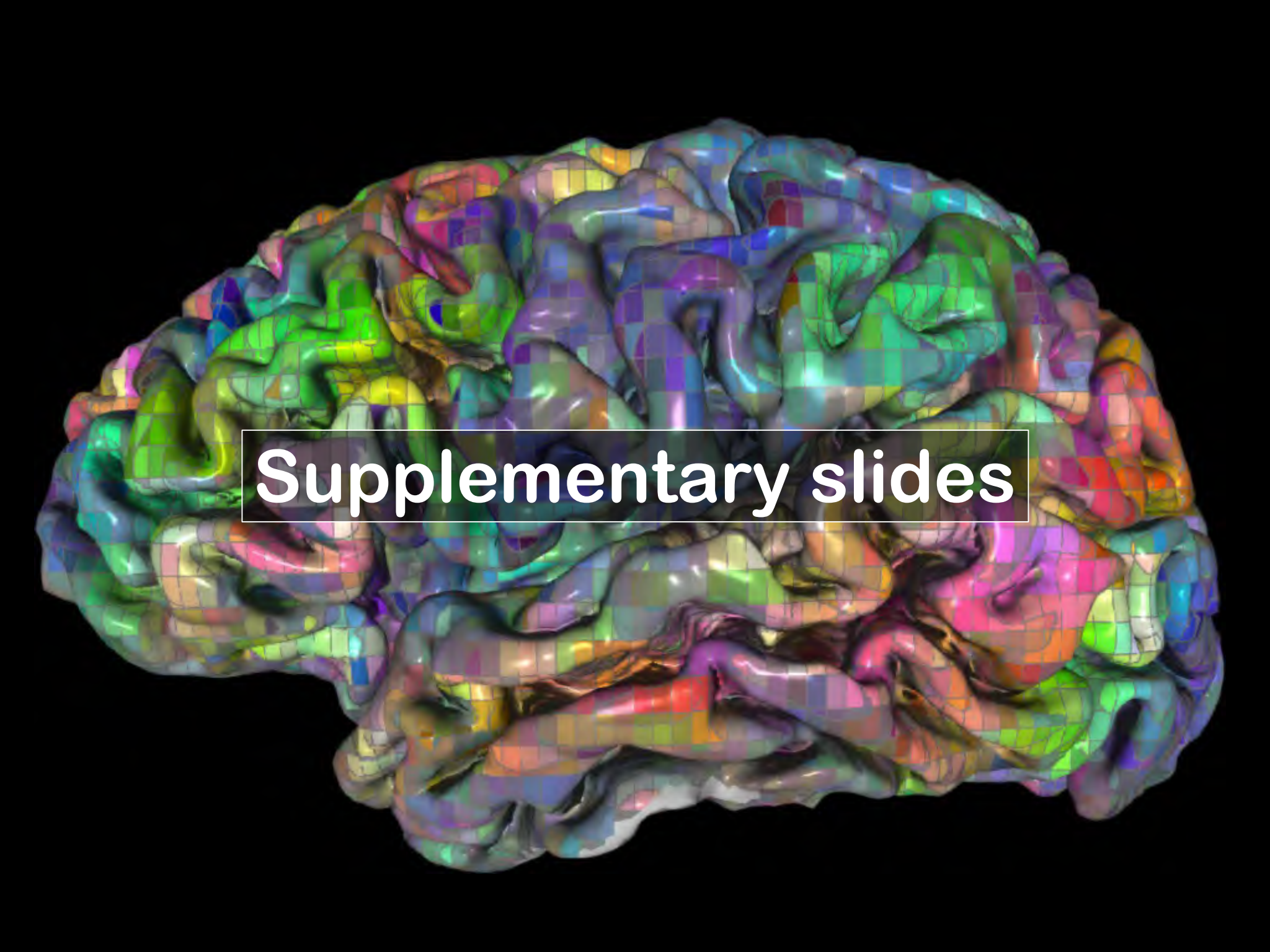
Collaborating labs

Frederic Theunissen, Bin Yu,
Tom Griffiths, Cheryl Olman
Bertrand Thirion,
Essa Yacoub, Kamil Ugurbil

Past Lab Members

Tolga Cukur (Bilkent U)
Stephen David (OHSU))
Kate Gustavsen (UTVM)
Kathleen Hansen (NIH)
Ben Hayden (URochester)
Kendrick Kay (WashU)
James Mazer (Yale)
Thomas Naselaris (MUSC)
Shinji Nishimoto (NIST Japan)
Ryan Prenger (LANL))
Ben Willmore (Cambridge)
Bill Vinje (Salk)
An Vu (UCB)
Michael Wu (Lithium Inc)

This work was supported by NEI, NIMH and NSF


Supplementary slides

Opinions: Science & cognitive neuroscience

- The 1st law of science: There is no free lunch.
- The 2nd law of science: One person's signal is another person's noise.
- The 1st law of neuroscience: No matter what your theory is, it is insufficient to explain the brain.
- The 2nd law of neuroscience: The brain doesn't care what you think about the brain.
- Statistical significance is necessary but not sufficient for doing science.
- The goal of science is to formulate an intelligible explanation the system that predicts accurately.
- You can learn a lot about a little, or a little about a lot, but the amount learned is determined by the size of the data.

Opinions: fMRI

- The biggest problem with fMRI data isn't Type I error, its Type II error.
- If you don't have a cortical mapping question, you shouldn't be using fMRI.
- All fMRI studies measure an entangled combination of representation and intention information.
- The biggest factors determining individual variability in BOLD signal quality are (1) the size of the brain relative to the receive coil, (2) head and body motion, (3) attention.
- Remember that the people who built your magnet were trying to make clinical radiologists happy.

Opinions: fMRI

- Many fMRI studies make implicit assumptions of linearity. (e.g., hemodynamic coupling or cognitive superposition). These are almost always wrong.
- Virtually every fMRI study spheres the data to remove non-stationary components. This is the wrong thing to do, but no one knows what the right thing is.
- Flowchart models developed in cognitive psychology often have little to do with cortical organization
- Functional connectivity has nothing to do with connectivity and little to do with function.
- MVPA decoding has nothing whatsoever to do with decoding.
- Granger causality has nothing to do with causality.
- It takes more data to accurately estimate functional connectivity than to estimate task-related effects.

Opinions: Design and data collection

- It is usually better to collect more data from fewer subjects than to collect fewer data from more subjects.
- Optimize fMRI data acquisition for every experiment.
- Always collect separate, interleaved data sets for estimation (fit) and validation (test).
- Use well trained subjects who attend and who do not move.
- Place subjects consistently in the magnet and collect field maps.
- Measure field distortion caused by your peripherals and place them consistently.
- Collect physiological data and all other telemetry possible.
- Collect field maps.

Opinions: Pre-processing

- Check for artifacts and alignment BY HAND in every single run.
- Detrend with a Savitsky-Golay filter (or at least a median filter).
- Z-score data within voxels and within runs.
- Estimate a separate HRF for every feature, every voxel and every subject.
- Never smooth the data blindly. Avoid smoothing at all if possible.
- Be very careful when aggregating data across runs or sessions.
- Whatever automated pipeline you are using, it doesn't work well enough.

Opinions: Data analysis and modeling

- Smoothing is usually bad, blind smoothing is always bad.
- If you are discarding data to make your statistics work, you are doing the wrong statistics.
- Focus on single subjects first. Only proceed to group-level analysis after you thoroughly understand the single subjects.
- Focus on prediction and effect size, not significance.

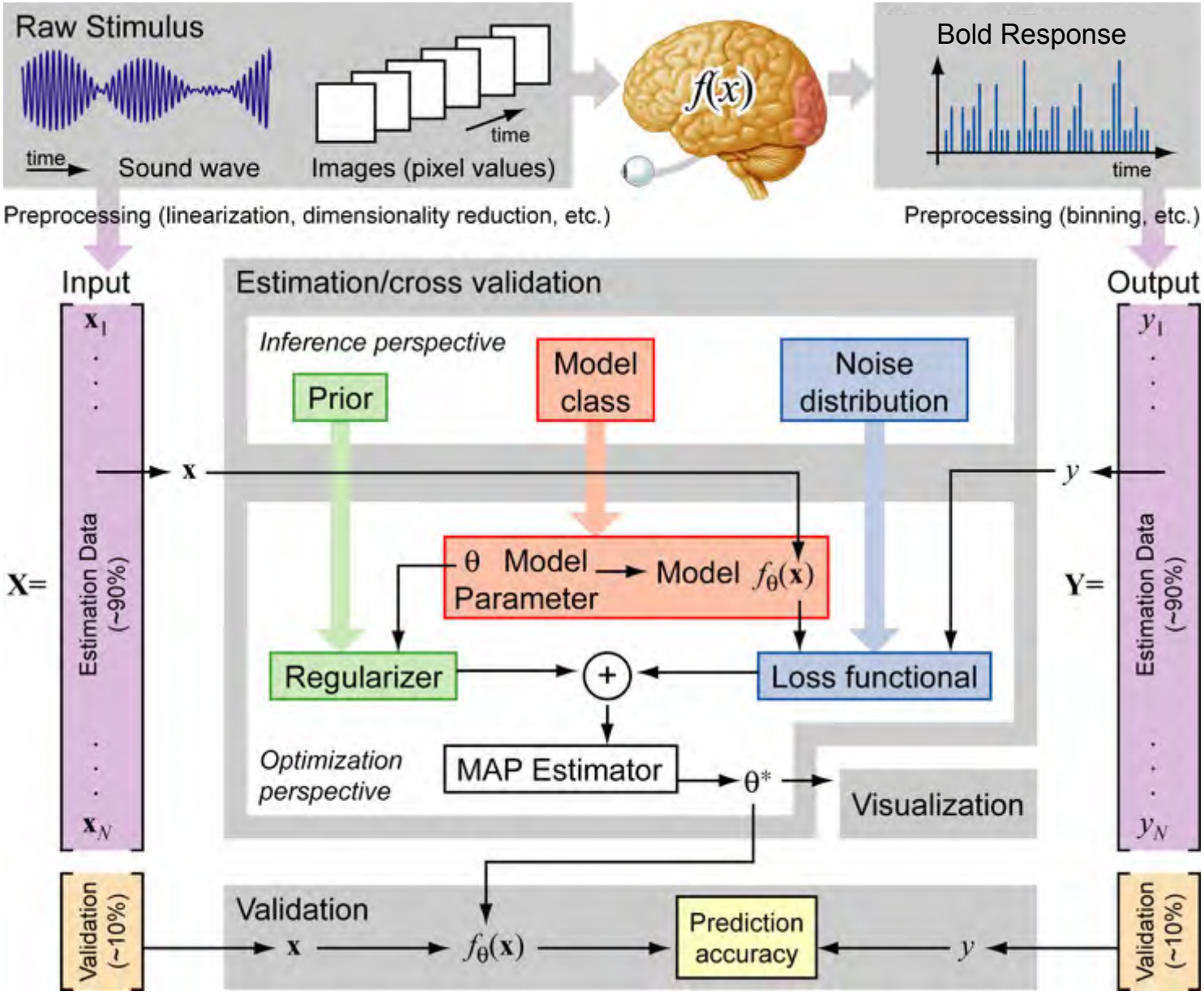
Opinions: Interpretation and visualization

- Comparisons of activity/correlations between conditions/areas are not valid unless the SNR across the conditions/areas is equal.
- If you are running a study on cortical activation, show your data on both inflated hemispheres and flat maps!
- If you show thresholded data, show the un-thresholded data as well.
- If you are showing group-level results, show the individual results as well (and report how often the phenomenon was seen in individual subjects).
- Always report variance explained as a portion of the potentially explainable variance.

Opinions: Decoding

- Decoding is a good way to do engineering, but it is generally a bad way to do science.
- The MVPA classifier approach is not really decoding.
- The best encoding model will create the best decoding model.
- There are a few special cases where scientific issues can be addressed with decoding.

Voxel-wise modeling as statistical inference



Wu, David & Gallant, *Ann. Rev. Neuro.* (2006)

Parametric models

System Identification Algorithm	Model class	Loss (Noise model)	Regularizer	Assumptions & Priors on param.	Objective function	# of model param.	# of free param.
Spike-triggered average	Linear	L_2 loss (Gaussian)	constant	All values are equally likely: Flat prior $\mathbf{A}^{-1}=\mathbf{0}$	No local minima & smooth	d+1	0
Normalized reverse correlation	Linear	L_2 loss (Gaussian)	constant	All values are equally likely: Flat prior $\mathbf{A}^{-1}=\mathbf{0}$	No local minima & smooth	d+1	0
Ridge Regression	Linear	L_2 loss (Gaussian)	$\sigma_\beta^{-2}\|\beta\|_2^2$	Independent w/ equal variance: Gaussian $\mathbf{A}=\sigma_\beta^2\mathbf{I}$	No local minima & smooth	d+1	1
Linear model + auto. relevancy determination	Linear	L_2 loss (Gaussian)	$\sum_{k=1}^{d}\sigma_{\beta_k}^{-2}\beta_k^2$	Independent: Gaussian $\mathbf{A}=\mathrm{diag}(\sigma_{\beta_k}^2)$	No local minima & smooth	d+1	d
Linear model w/ smooth prior	Linear	L_2 loss (Gaussian)	$\delta^{-1}\|\mathbb{D}\beta\|_2^2$	Smooth over input space: Gaussian $\mathbf{A}=\mathbb{D}^2$	No local minima & smooth	d+1	1

Bold Response

Wu, David & Gallant, *Ann. Rev. Neuro.* (2006)

Non-parametric models

System Identification Algorithm	Model class	Loss (Noise model)	Regularizer	Assumptions & Priors on param.	Objective function	# of model param.	# of free param.
Spike-triggered covariance	2nd Order	L_2 loss (Gaussian)	constant	All values are equally likely: Flat prior $\mathbf{A}^{-1}=\mathbf{0}$	No local minima & smooth	d^2+d+1	0
Linearized reverse correlation	Linearized reverse correlation	L_2 loss (Gaussian)	$\sum_{k,D_k>\lambda} D_k\left(\mathbf{Q}_k^T\beta\right)^2$	In a stimulus PC subspace $\mathbf{A}^{-1}=\mathbf{QD}(D_k>\lambda)\mathbf{Q}^T$	No local minima & smooth	kd+1	2
Neural network (+auto. relevancy determination)	Neural network	L_2 loss (Gaussian)	$\sum_{k=1}^{(d+2)h+1} \sigma_{\theta_k}^{-2}\theta_k^2$	Independent: Gaussian $\mathbf{A}=\mathrm{diag}(\sigma_{\theta_k}^2)$	Has local minima, but smooth	$(d+2)h+1$	$d+3$ & h
Kernel ridge regression	Kernel regression	L_2 loss (Gaussian)	$\sigma_\beta^{-2}\|\beta\|_2^2$	Independent w/ equal variance: Gaussian $\mathbf{A}=\sigma_\beta^2\mathbf{I}$	No local minima & smooth	$\leq N$	1 & Φ
Support vector regression	Kernel regression	ε-insensitive loss	$\sigma_\beta^{-2}\|\beta\|_2^2$	Independent w/ equal variance: Gaussian $\mathbf{A}=\sigma_\beta^2\mathbf{I}$	No local minima, but not smooth	$<N$	2 & Φ

Bold Response

Wu, David & Gallant, *Ann. Rev. Neuro.* (2006)

Voxel-wise encoding model

A voxel encoding model for a population of voxels R is given by the following distribution:

$$p(R|S) \sim MVN(\mu(S), \Gamma) \propto e^{(-(R-\mu(S))\Gamma^{-1}(R-\mu(S))^T)}$$

$MVN(\mu(S), \Gamma)$ is a multivariate normal distribution with mean response vector $\mu(S)$ and covariance Γ.
$\mu(S) = (\mu_1(S), \ldots, \mu_N(s))$ gives the mean responses of each voxel to stimulus features S.
$R = (r_{1,} \ldots, r_N)$ is a sample of the population response vector.
S is a vector of stimulus features.

The linearized encoding model

The encoding model seeks to predict the stimulus-dependent mean response, $\hat{r}_i = \mu_i(S)$.
$\hat{r}_i$ is the predicted scalar response of the *i*-th voxel.

In a linearized encoding model the stimulus vector *S* is a nonlinear function g of the underlying stimulus pixels *P:*

$$S = g(P)$$

Then the predicted response for each voxel, $\hat{r}_i$, is a linear combination of the features, $\hat{r}_i = S w_i$.
w_i is the feature weight vector for the i-th voxel.

The linearized encoding model with delays

For continuous stimuli we include feature weights at different delays. The voxel-wise encoding model is then:

$$\hat{r}_i = \sum_{t \subset T} S^t w_{i,t}$$

T gives causal delays (usually 0...20 seconds)
S^t corresponds to S but with temporal shifts of size t .

Here S' is a horizontally-concatenated matrix of S with temporal shifts (e.g., $S' = [\, S^0, S^1, S^2 \,]$), so the model is:

$$\hat{r}_i = S' w_i'$$

where the length of w_i' is [# of delays] times longer than the original w_i .

Fitting the encoding model

w_i for each voxel are obtained using either:

L2-penalized least-squares (i.e. ridge) regression:

$$\min_{w_i} \sum (Sw_i - R_i)^2 + \lambda \sum w_i^2$$

L1-penalized least-squares (i.e. Lasso) regression:

$$\min_{w_i} \sum (Sw_i - R_i)^2 + \lambda \sum \|w_i\|$$

R_i is the response time course of the i-th voxel.
λ is a scalar hyperparameter.

Using the multi-voxel likelihood to decode

To decode we find the S that maximizes the multi-voxel likelihood. Assuming additive Gaussian noise and a stick prior then this can be expressed as a MVN distribution:

$$p(S|R) \propto p(R|S)\, p(S) \propto e^{[(R-SW)\Gamma^{-1}(R-SW)^T]}$$

W is a matrix whose rows are encoding model weights.
Γ is the estimated noise covariance matrix:

$$\Gamma_{ij} = (R_i - SW_i)^T (R_j - SW_j)$$

Because the inverse of Γ_{ij} is typically unstable regularization must be used to obtain an estimate.

Estimating the decoding model directly

Alternatively, the posterior distribution can be calculated directly by exchanging R and S in the equations above:

$$\hat{s}_i = R\, w_i^d$$

and

$$p\left(S|R\right) \sim MVN\left(RW^d, \Gamma^d\right)$$

where w_i^d, W^d and Γ^d are the equivalent variables for the direct model.
However, this is not advisable for several reasons.

Where to learn more...

ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 33 (2006) 1104–1116

Inverse retinotopy: Inferring the visual content of images from brain activation patterns

Bertrand Thirion,[a,*] Edouard Duchesnay,[b] Edward Hubbard,[d] Jessica Dubois,[c] Jean-Baptiste Poline,[c] Denis Lebihan,[c] and Stanislas Dehaene[d]

[a]INRIA Futurs, Service Hospitalier Frédéric Joliot, 4, Place du Général Leclerc 91401 Orsay Cedex, France
[b]Unité INSERM ERM 0205, Service Hospitalier Frédéric Joliot, 4, Place du Général Leclerc, 91401 Orsay Cedex, France
[c]CEA, DSV, DRM, SHFJ, 4, Place du Général Leclerc, 91401 Orsay Cedex, France
[d]Unité INSERM 562 "Neuroimagerie Cognitive", Service Hospitalier Frédéric Joliot, 4 Place du Général Leclerc, 91401 Orsay Cedex, France

Received 12 January 2006; revised 26 June 2006; accepted 28 June 2006
Available online 9 October 2006

Traditional inference in neuroimaging consists in describing brain activations elicited and modulated by different kinds of stimuli. Recently, however, paradigms have been studied in which the converse operation is performed, thus inferring behavioral or mental states associated with activation images. Here, we use the well-known retinotopy of the visual cortex to infer the visual content of real or imaginary scenes from the brain activation patterns that they elicit. We present two decoding algorithms: an explicit technique, based on the current knowledge of the retinotopic structure of the visual areas, and an implicit technique, based on supervised classifiers. Both algorithms predicted the stimulus identity with significant accuracy. Furthermore, we extend this principle to mental imagery data: in five data sets, our algorithms could reconstruct and predict with significant accuracy a pattern imagined by the subjects.

Introduction

The neuroimaging inverse problem

Validation of anatomo-functional knowledge produced from neuroimaging data is a difficult task. While statistical significance, reproducibility and multi-modal coherence are well-accepted proofs of consistency, neuroscientists lack a gold standard to assess the significance of their findings. A possible way to solve this issue is to reason as follows: understanding a cognitive subsystem of the brain means that the stimulus-to-activation chain has been identified. More precisely, although the detailed mechanisms of neural and hemodynamic activation are not fully understood, we can expect that a controlled stimulus (e.g. a flashing checkerboard) will produce a known pattern of activation. When this holds, the processing chain can be inverted, leading to activation-to-stimulus inference. When possible, this inverse inference allows good performance characterization, since the results are expressed in terms of predicted versus true stimulus, in the well-known (and controlled) stimulus space.

This point of view has already been investigated in the case of motor experiments (Dehaene et al., 1998), mental imagery (O'Craven and Kanwisher, 2000), counting/subitizing (Piazza et al., 2003), the notion of object categories (Haxby et al., 2001; Carlson et al., 2003; Cox and Savoy, 2003), the orientation of visual stimuli (Haynes and Rees, 2005; Kamitani and Tong, 2005) and lie detection (Davatzikos et al., 2005). It has been popularized under the concept of brain reading. This novel approach in neuroimaging has been facilitated by the use of data classification techniques such as Linear Discriminant Analysis (LDA) (Carlson et al., 2003) and more recently, Support Vector Machines (SVM) (Cox and Savoy, 2003; LaConte et al., 2005) that can take functional images as input and classify them into categories (supervised classification). But in that case the activation-to-stimulus function remains implicit, i.e. it is embedded in a set of learning samples, each one being associated with a known stimulus. An important question is whether this binding may be made explicit.

There is at least one system in which the stimulus-to-activation coding is known explicitly: this is the case of retinotopy, where the spatial layout of an image is in the visual field also spatially encoded in the primary visual cortex (Sereno et al., 1995). The inverse problem consists in predicting the spatial layout of an activation pattern (stimulus) given a functional activation image. We address this problem with two kinds of analysis tools: supervised classification (based on SVMs) and an explicit inversion of the stimulus-to-activation function (inverse retinotopy). In this paper, we study two different situations: a visual stimulation experiment, in which the subject passively views a sequence of stimuli chosen among a discrete set and a mental imagery experiment in which the subject is asked to imagine a self-selected pattern chosen among the presented stimuli.

* Corresponding author.
E-mail address: bertrand.thirion@inria.fr (B. Thirion).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter
doi:10.1016/j.neuroimage.2006.06.062

ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 39 (2008) 181–205

Bayesian decoding of brain images

Karl Friston,[a,*] Carlton Chu,[a] Janaina Mourão-Miranda,[b] Oliver Hulme,[a] Geraint Rees,[a] Will Penny,[a] and John Ashburner[a]

[a]Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, 12 Queen Square, London WC1N 3BG, UK
[b]Biostatics Department, Centre for Neuroimaging Sciences, Institute of Psychiatry, King's College London, UK

Received 9 July 2007; revised 7 August 2007; accepted 12 August 2007
Available online 24 August 2007

This paper introduces a multivariate Bayesian (MVB) scheme to decode or recognise brain states from neuroimages. It resolves the ill-posed many-to-one mapping, from voxel values or data features to a target variable, using a parametric empirical or hierarchical Bayesian model. This model is inverted using standard variational techniques, in this case expectation maximisation, to furnish the model evidence and the conditional density of the model's parameters. This allows one to compare different models or hypotheses about the mapping from functional or structural anatomy to perceptual and behavioural consequences (or their deficits). We frame this approach in terms of decoding measured brain states to predict or classify outcomes using the rhetoric established in pattern classification of neuroimaging data. However, the aim of MVB is not to predict (because the outcomes are known) but to enable inference on different models of structure-function mappings, such as distributed and sparse representations. This allows one to optimise the model itself and produce predictions that outperform standard pattern classification approaches, like support vector machines.

Technically, the model inversion and inference uses the same empirical Bayesian procedures developed for ill-posed inverse problems (e.g., source reconstruction in EEG). However, the MVB scheme used here extends this approach to include a greedy search for sparse solutions. It reduces the problem to the same form used in Gaussian process modelling, which affords a generic and efficient scheme for model optimisation and evaluating model evidence. We illustrate MVB using simulated and real data, with a special focus on model comparison; where models can differ in the form of the mapping (i.e., neuronal representation) within one region, or in the (combination of) regions per se.

Keywords: Parametric empirical Bayes; Expectation maximisation; Gaussian process; Automatic relevance determination; Relevance vector machines; Classification; Multivariate; Support vector machines; Classification

Introduction

The purpose of this paper is to describe an empirical Bayesian approach to the multivariate analysis of imaging data that brings pattern classification and prediction approaches into the conventional inference framework of hierarchical models and their inversion. The past years have seen a resurgence of interest in the multivariate analysis of functional and structural brain images. These approaches have been used to infer the deployment of distributed representations and their perceptual or behavioural correlates. In this paper, we try to identify the key issues entailed by these approaches and use these issues to motivate a better approach to estimating and making inferences about distributed neuronal representations.

This paper comprises three sections. In the first, we review the development of multivariate analyses with a special focus on three important distinctions; the difference between mass-univariate and multivariate models, the difference between generative and recognition models and the distinction between inference and prediction. The second section uses the conclusions of the first section to motivate a simple hierarchical model of the mapping from observed brain responses to a measure of what those responses encode. This model allows one to compare different forms of encoding, using conventional model comparison. In the final section, we apply the multivariate Bayesian model of the second section to real fMRI data and ask where and how visual motion is encoded. We also show that the ensuing model outperforms simple classification devices like linear discrimination and support vector machines. We conclude with a discussion of generalisations; for example, nonlinear models and the comparison of multiple conditions to disambiguate between functional selectivity and segregation in the cortex.

Multivariate models and classification

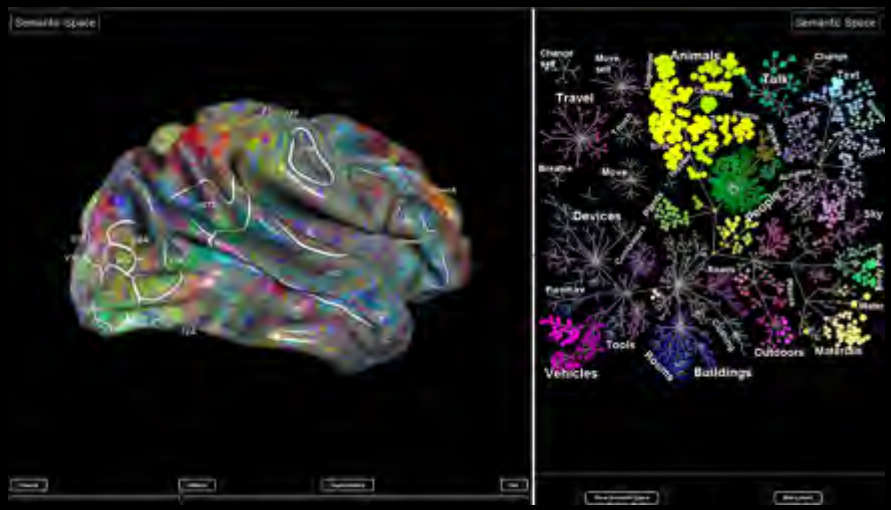
Mappings and models

In this section, we review multivariate approaches and look at the distinction between inference and prediction. This section is written in a tutorial style in an attempt to highlight some of the

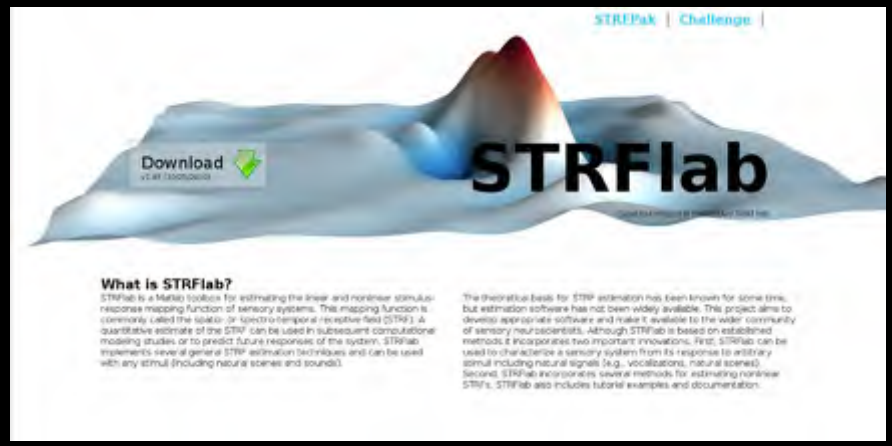
* Corresponding author. Fax: +44 207 813 1445.
E-mail address: k.friston@fil.ion.ucl.ac.uk (K. Friston).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter
doi:10.1016/j.neuroimage.2007.08.013

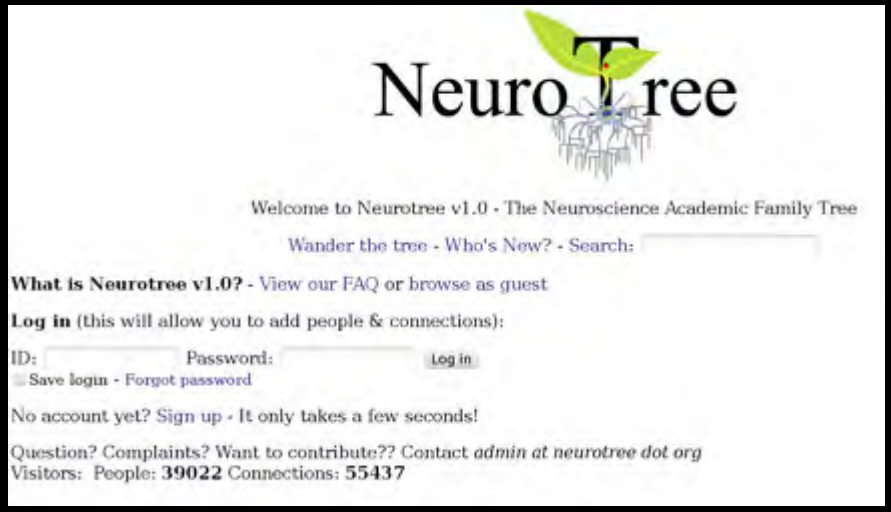
Gallant lab open source and web initiatives



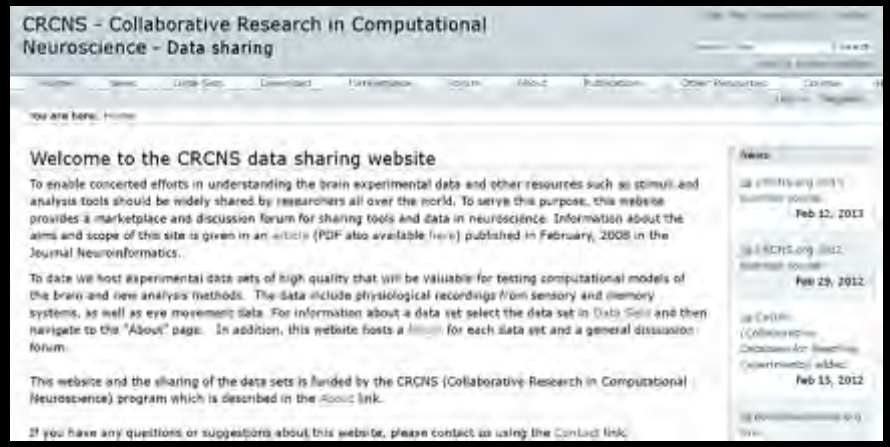
gallantlab.org



strflab.berkeley.edu



neurotree.org



crcns.org