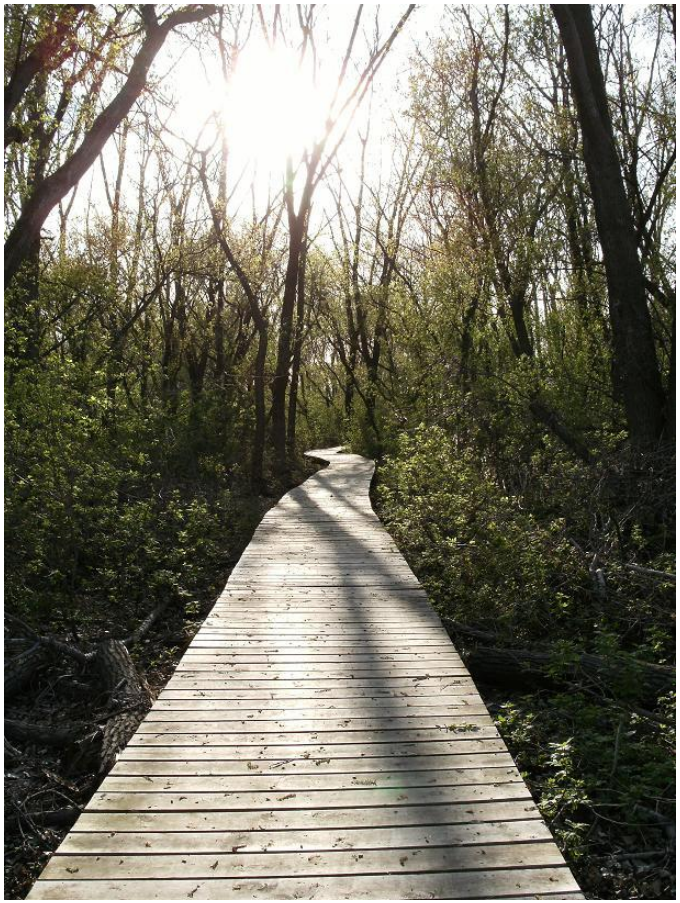
More machine learning for fMRI

Francisco Pereira

[many slides adapted from slides by Ken Norman]

Botvinick Lab
Princeton Neuroscience Institute
Princeton University



More machine learning for fMRI

"Wanderer, your footsteps are the road, and nothing more;
wanderer, there is no road, the road is made by walking.
By walking one makes the road, and upon glancing behind
one sees the path that never will be trod again.
Wanderer, there is no road-- only wakes upon the sea."

Antonio Machado

what questions can be tackled?

- is there information?
  (pattern discrimination)
- where/when is information present?
  (pattern localization)
- how is information encoded?
  (pattern characterization)

where are we?

stimulus → mind → brain → fMRI data

(task)

where are we?

stimulus → mind → brain → fMRI data

(task)

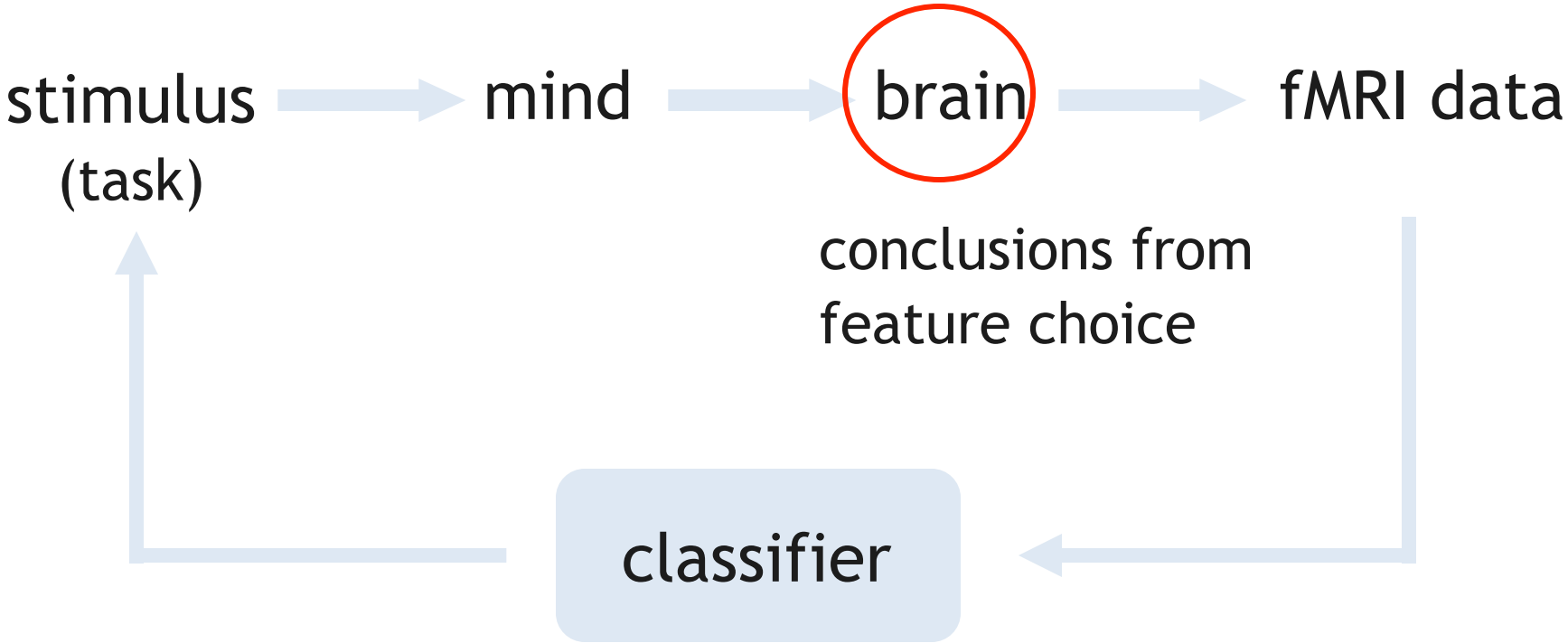
no dualism implied!

where are we?

stimulus → mind → brain → fMRI data

(task)

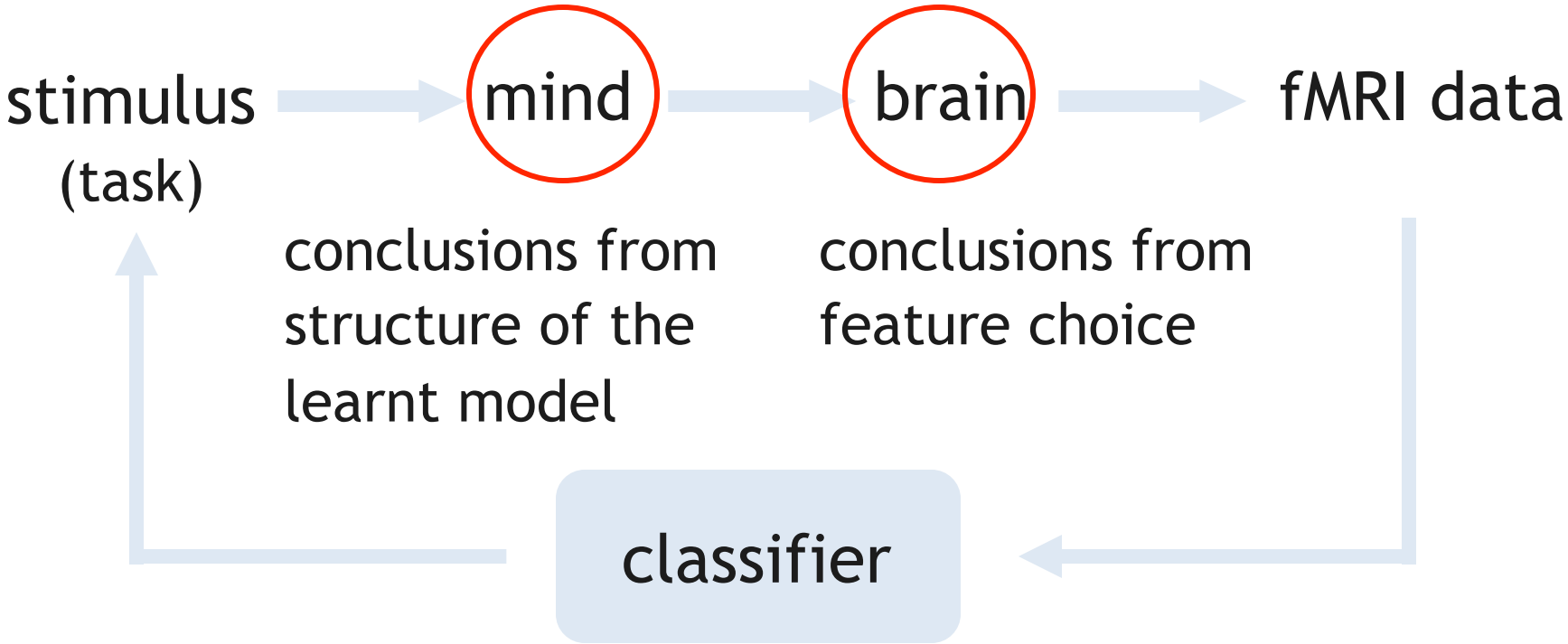
where are we?



- voxel location
- voxel behaviour
- time within trial

dependent on experiment

where are we?



- weights on features
- hidden layer activations

dependent on prediction model

- voxel location
- voxel behaviour
- time within trial

dependent on experiment

what do we want?

stimulus (task) → mind → brain → fMRI data

mind ↓ what is present in the mind as the task is performed?

where can it come from?

stimulus (task) → mind → brain → fMRI data

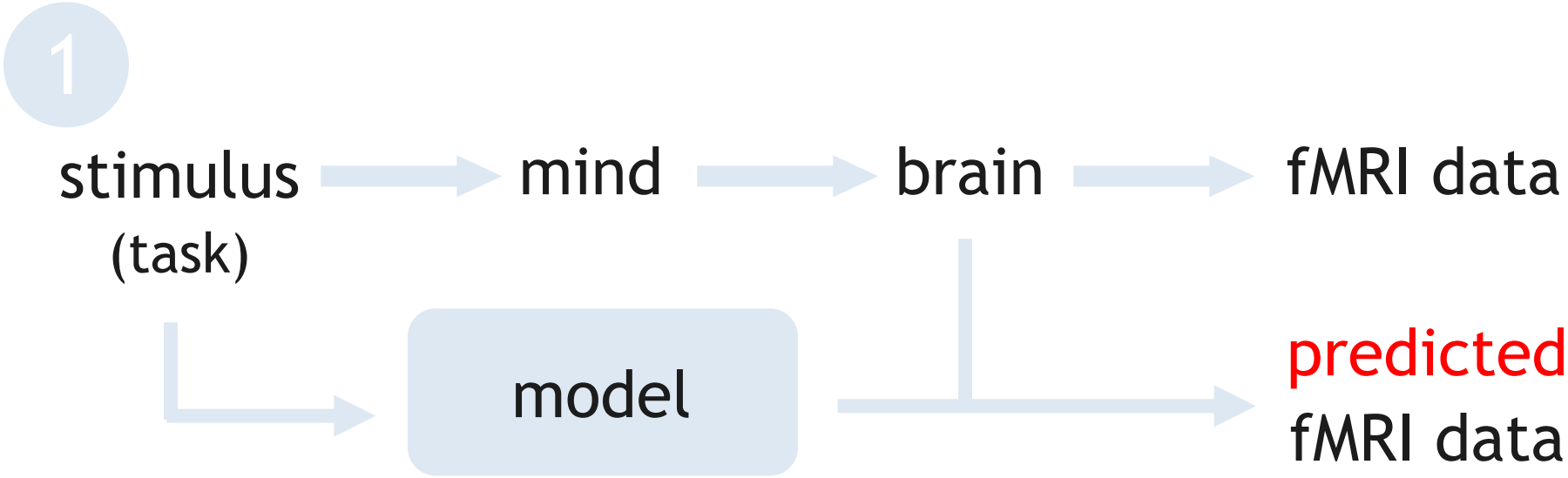
↓

what is present in the mind as the task is performed?

↓

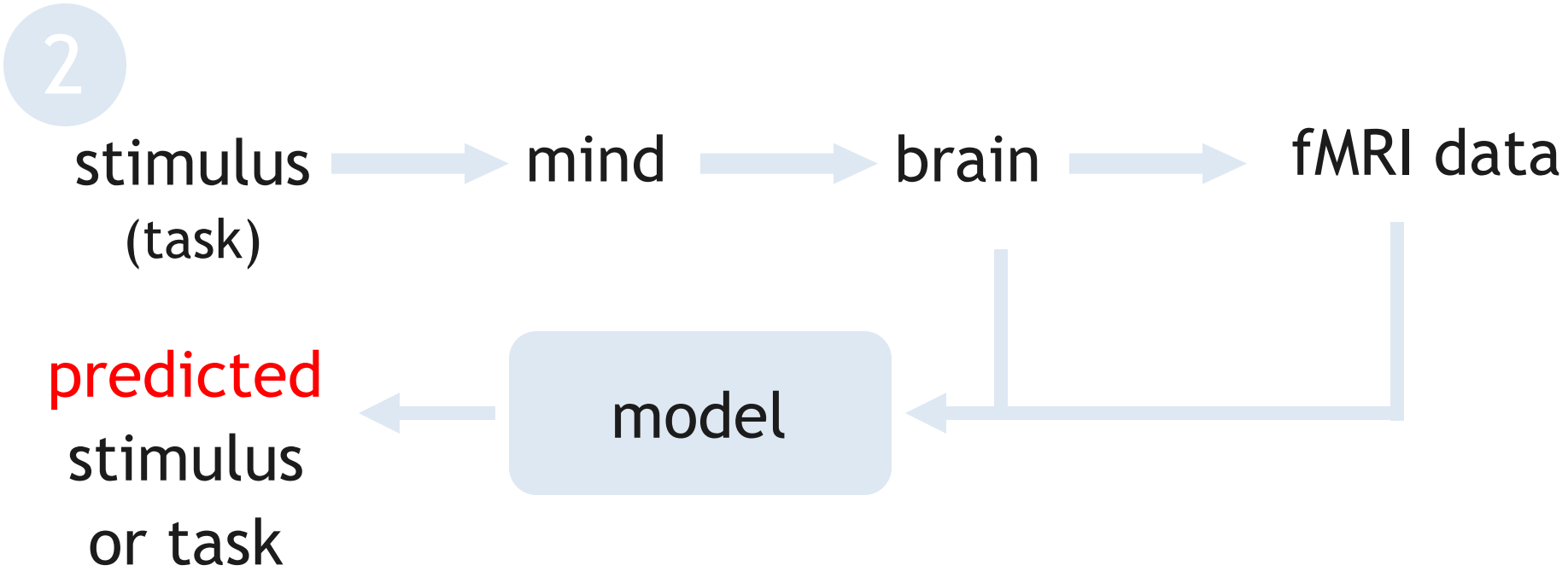
- known or constrained from behavioural experiments
- modelled mathematically or computationally
- hypothesized
- learnt elsewhere (text corpora)

how do we know it's there?



- Kay et al. 2008
- Mitchell et al. 2008

how do we know it's there?



- Thirion et al. 2006
- Miyawaki et al. 2008
- Naselaris et al. 2009
- van Gerven et al. 2010

how do we know it's there?

3

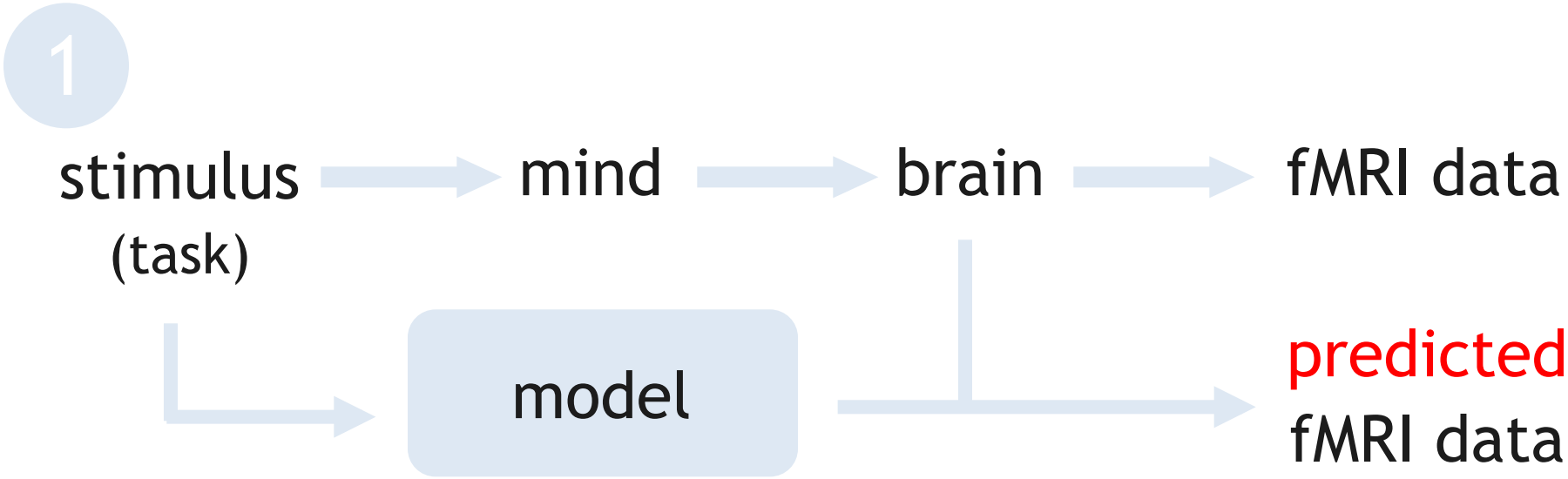
stimulus (task) → mind → brain → fMRI data

known relationships between conditions

pattern similarities or confusion matrices

- Kriegeskorte et al. 2008/2009
- Walther et al. 2009
- ...

how do we know it's there?



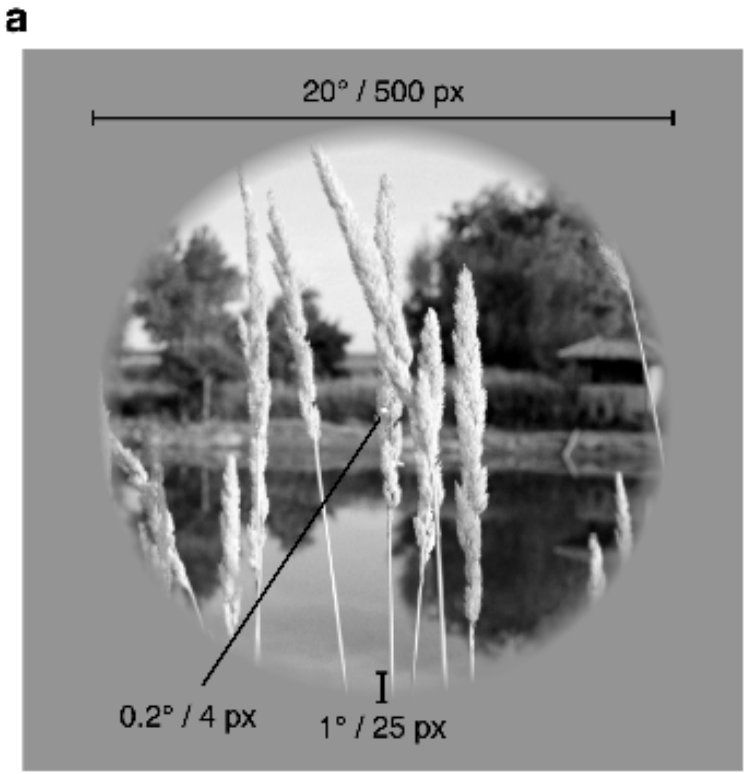
- Kay et al. 2008
- Mitchell et al. 2008

Identifying natural images from human brain activity

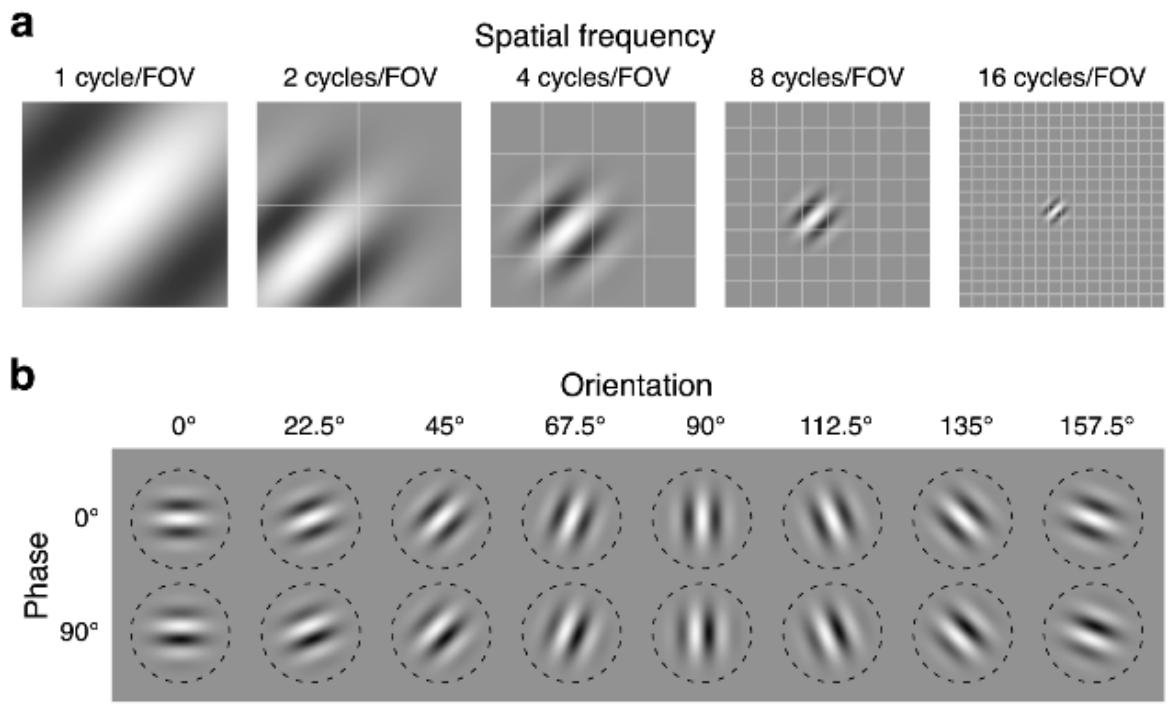
Kendrick N. Kay[1], Thomas Naselaris[2], Ryan J. Prenger[3] & Jack L. Gallant[1,2]

goal

- decode which natural image a subject is seeing (out of 1000s)
- approach:
  - assume that the image perceived is processed into features by V1/V2/V3
  - build a model of how voxels respond to the features
  - predict the fMRI response to a new image and match

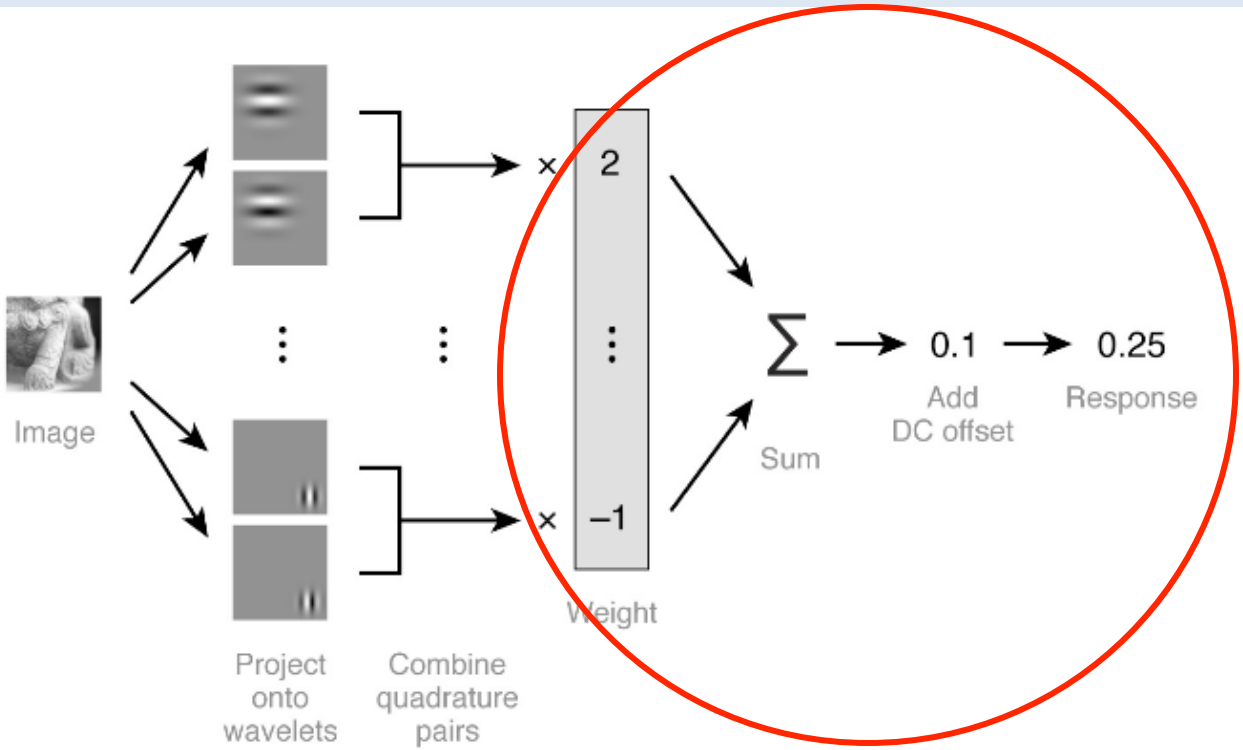


feature computation



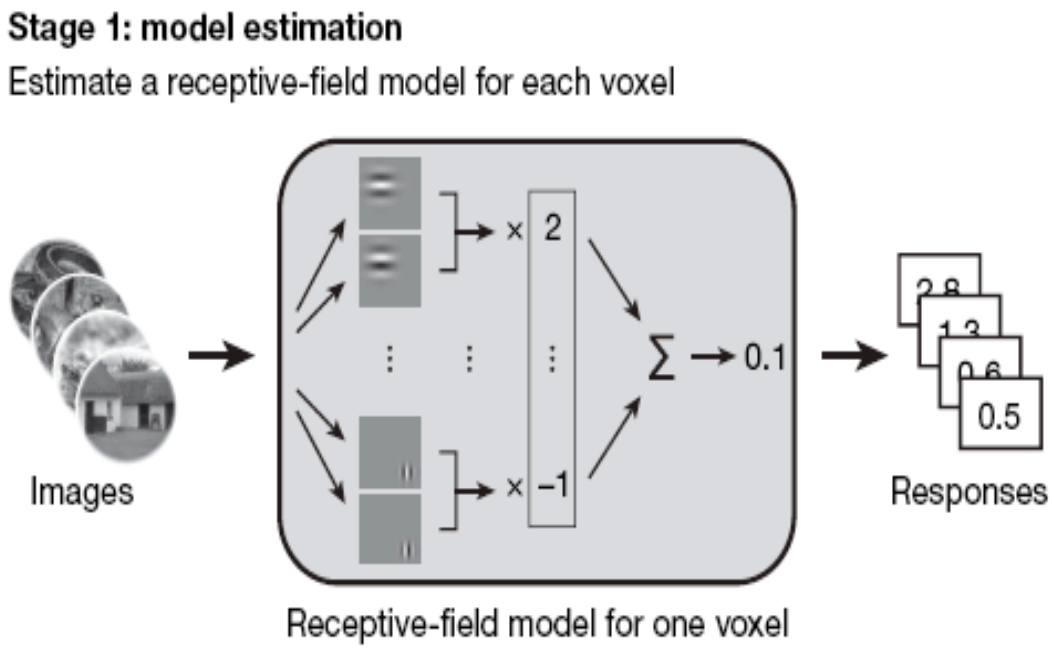
- apply filters with various scales/orientations
- the output of each filter is a feature
- deterministic function of stimulus image

feature computation + response model



- feature values are fed into a voxel response model
- voxel response model:
  linear combination of features

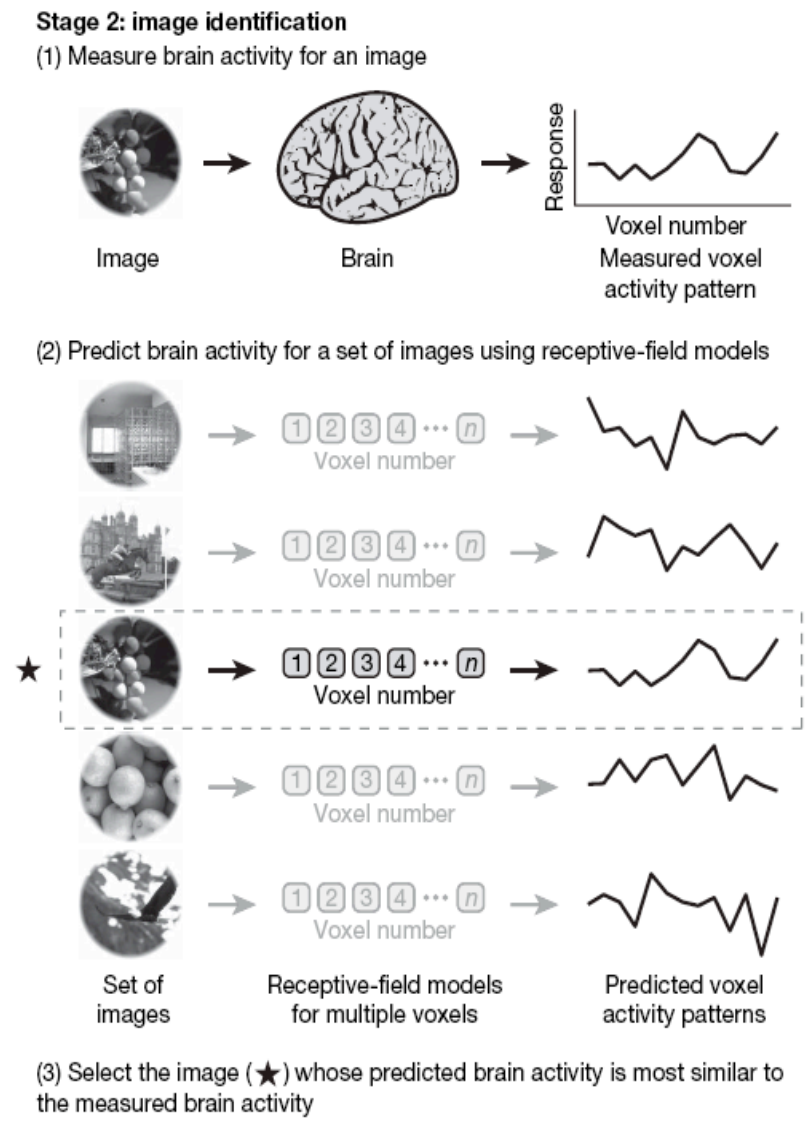
feature computation + response model



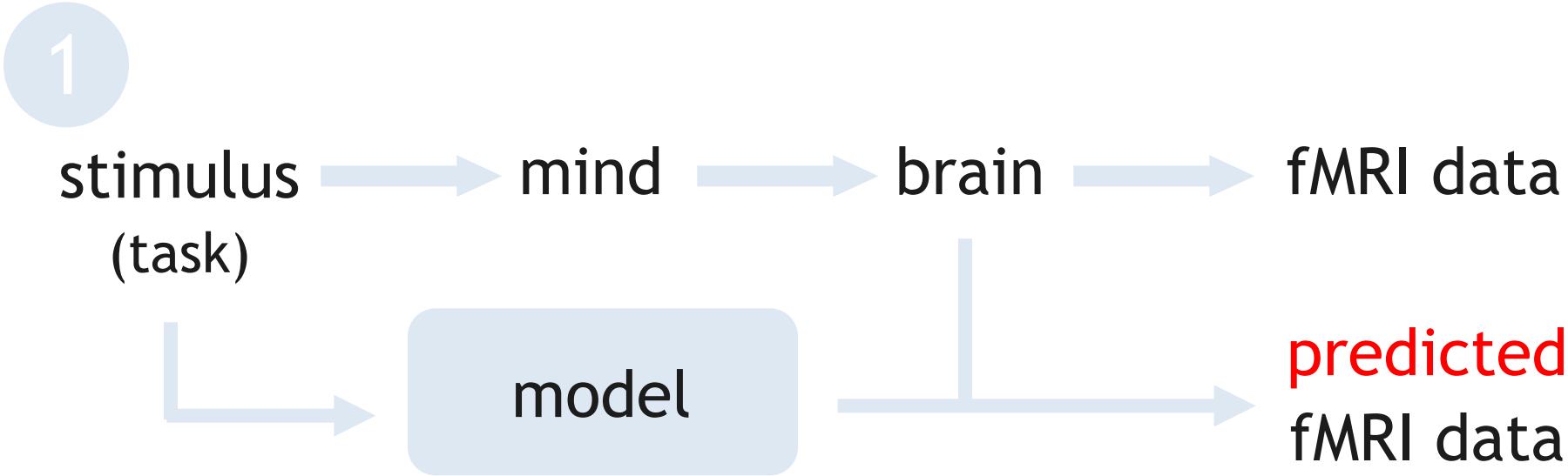
- one set of weights per voxel
- intuition: find which features a voxel responds to
- learned over training images + fMRI data

model validation

- over test stimuli not used to fit the model
- obtain fMRI data in response to stimuli
- test:
  - given natural image
  - compute features
  - predict voxel responses
  - find most similar fMRI image in test data



how do we know it's there?



- Kay et al. 2008:
  specified computation model + learnt response model
- Mitchell et al. 2008

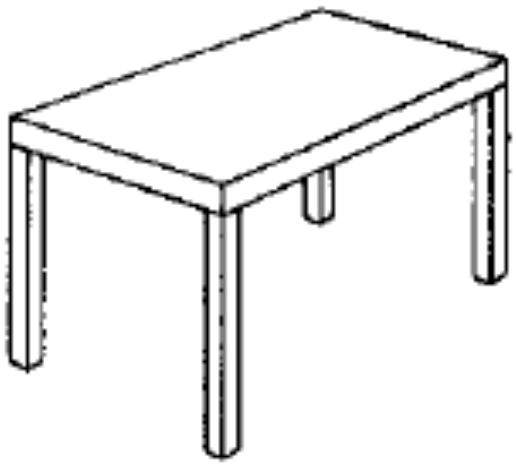
Predicting Human Brain Activity Associated with the Meanings of Nouns

Tom M. Mitchell,[1*] Svetlana V. Shinkareva,[2] Andrew Carlson,[1] Kai-Min Chang,[3,4] Vicente L. Malave,[5] Robert A. Mason,[3] Marcel Adam Just[3]

goal:

- understand how semantic information is represented in the pattern of activation

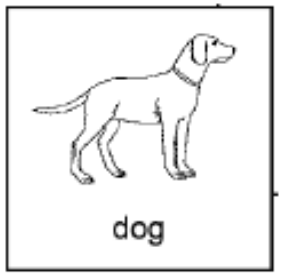
- approach
  - stimulus: word + drawing
  - subject visualizes object, thinks of properties, using it, etc
  - assume that each concept is represented in terms of features
  - learn a mapping between each feature and the voxels it affects



Table

stimuli

60 exemplars



Categories	Exemplars				
BODY PARTS	leg	arm	eye	foot	hand
FURNITURE	chair	table	bed	desk	dresser
VEHICLES	car	airplane	train	truck	bicycle
ANIMALS	horse	dog	bear	cow	cat
KITCHEN UTENSILS	glass	knife	bottle	cup	spoon
TOOLS	chisel	hammer	screwdriver	pliers	saw
BUILDINGS	apartment	barn	house	church	igloo
PART OF A BUILDING	window	door	chimney	closet	arch
CLOTHING	coat	dress	shirt	skirt	pants
INSECTS	fly	ant	bee	butterfly	beetle
VEGETABLES	lettuce	tomato	carrot	corn	celery
MAN MADE OBJECTS	refrigerator	key	telephone	watch	bell

feature extraction

- we have models of the computation being carried out by the visual cortex
- but how do we find semantic features?

feature extraction

- we have models of the computation being carried out by the visual cortex
- but how do we find semantic features?

- consider the use of the nouns naming stimuli in a very large text corpus!
- in particular, consider their co-occurrence with certain verbs...

feature extraction

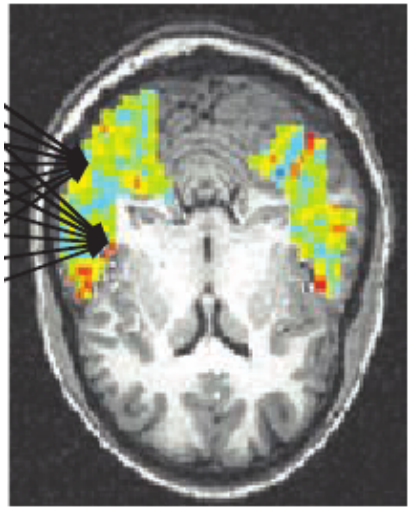
- 25 verbs reflecting sensory/motor/function aspects
  - sensory: see, hear, listen, taste, touch, smell, fear, ...
  - motor: rub, lift, run, push, move, say, eat, ...
  - abstract: fill, open, ride, approach, drive, enter, ...
- 25 features
  - co-occurrence of a stimulus noun with each of 25 verbs
  - co-occurrence is normalized to be length 1

feature extraction

- 25 verbs reflecting sensory/motor/function aspects
  - sensory: see, hear, listen, taste, touch, smell, fear, ...
  - motor: rub, lift, run, push, move, say, eat, ...
  - abstract: fill, open, ride, approach, drive, enter, ...
- 25 features
  - co-occurrence of a stimulus noun with each of 25 verbs
  - co-occurrence is normalized to be length 1
- example: “airplane”
  - 0.87, ride
  - 0.29, see
  - 0.17, near
  - 0.08, run
  - ...

predicting activation

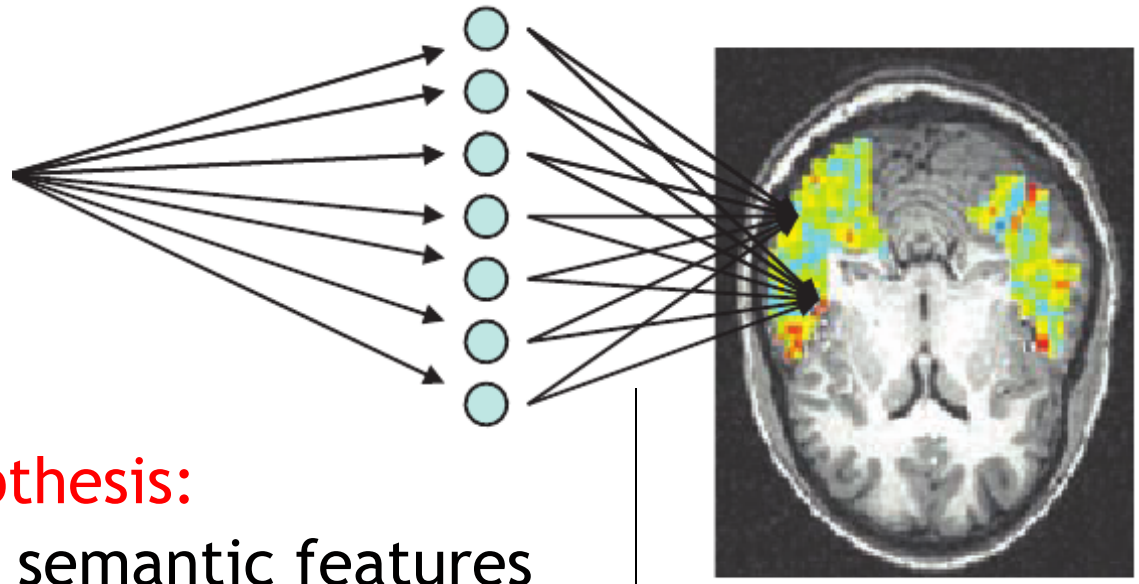
stimulus word
(+ picture)
e.g. celery



average activation
during 3 seconds

predicting activation

stimulus word
(+ picture)
e.g. celery

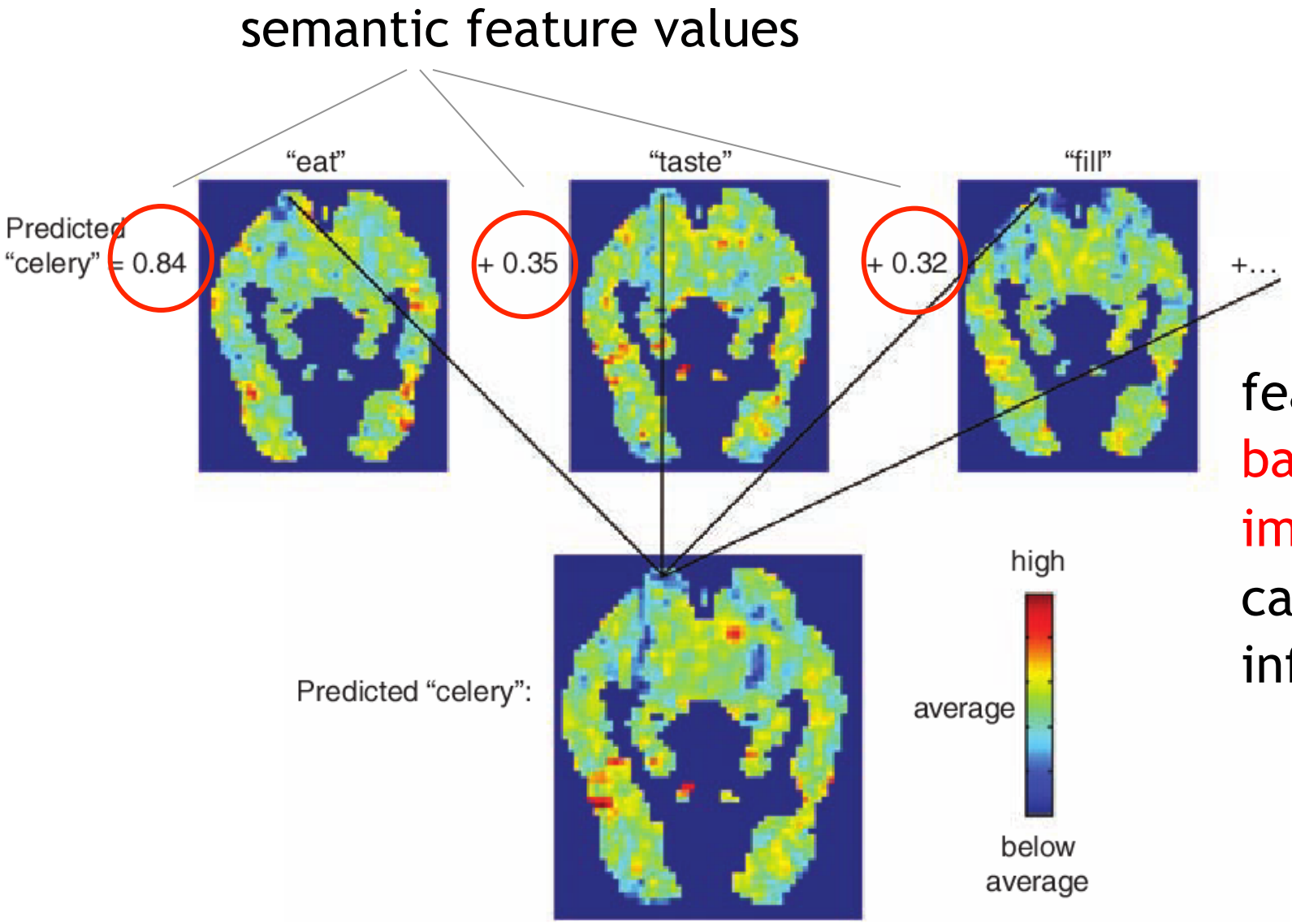


hypothesis:
verb semantic features
e.g. taste/eat/fill
word co-occurrence

average activation
during 3 seconds

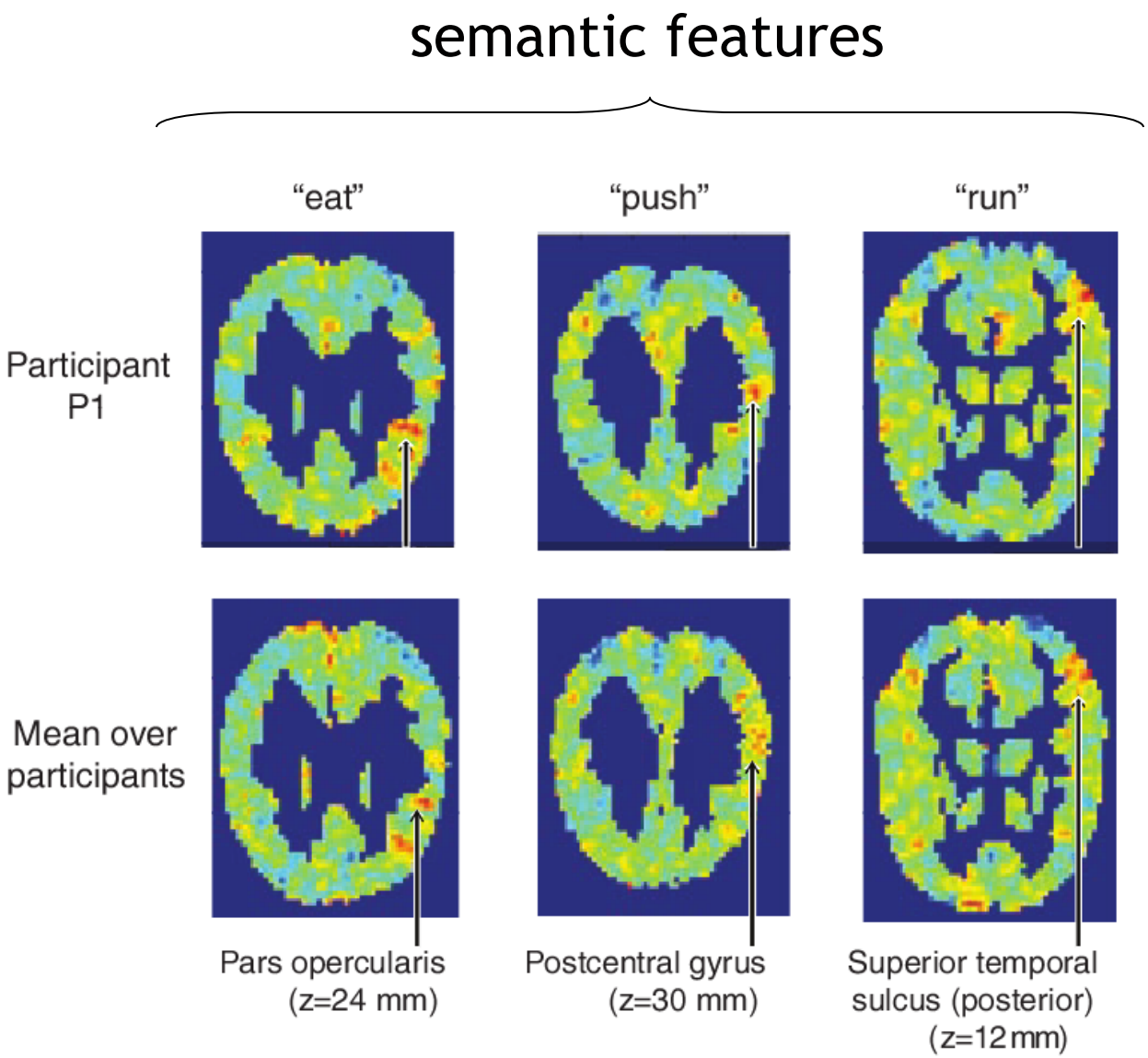
learned model:
mapping between semantic features
and their influence in activation

predicting activation



feature basis images capture influence

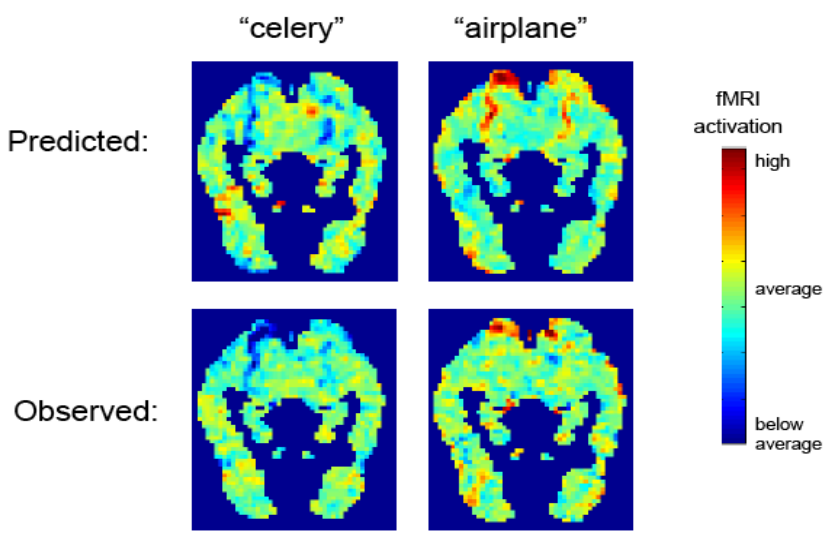
basis images



the basis images put weight in locations involved in tasks related to the verbs

evaluation

- learn basis images from 58 of 60 nouns
- use semantic features for 2 test nouns as weights for combining basis images
- result: predicted fMRI for the 2 nouns



Predicted and observed fMRI images for "celery" and "airplane" after training on 58 other words.

classification: match predicted with observed

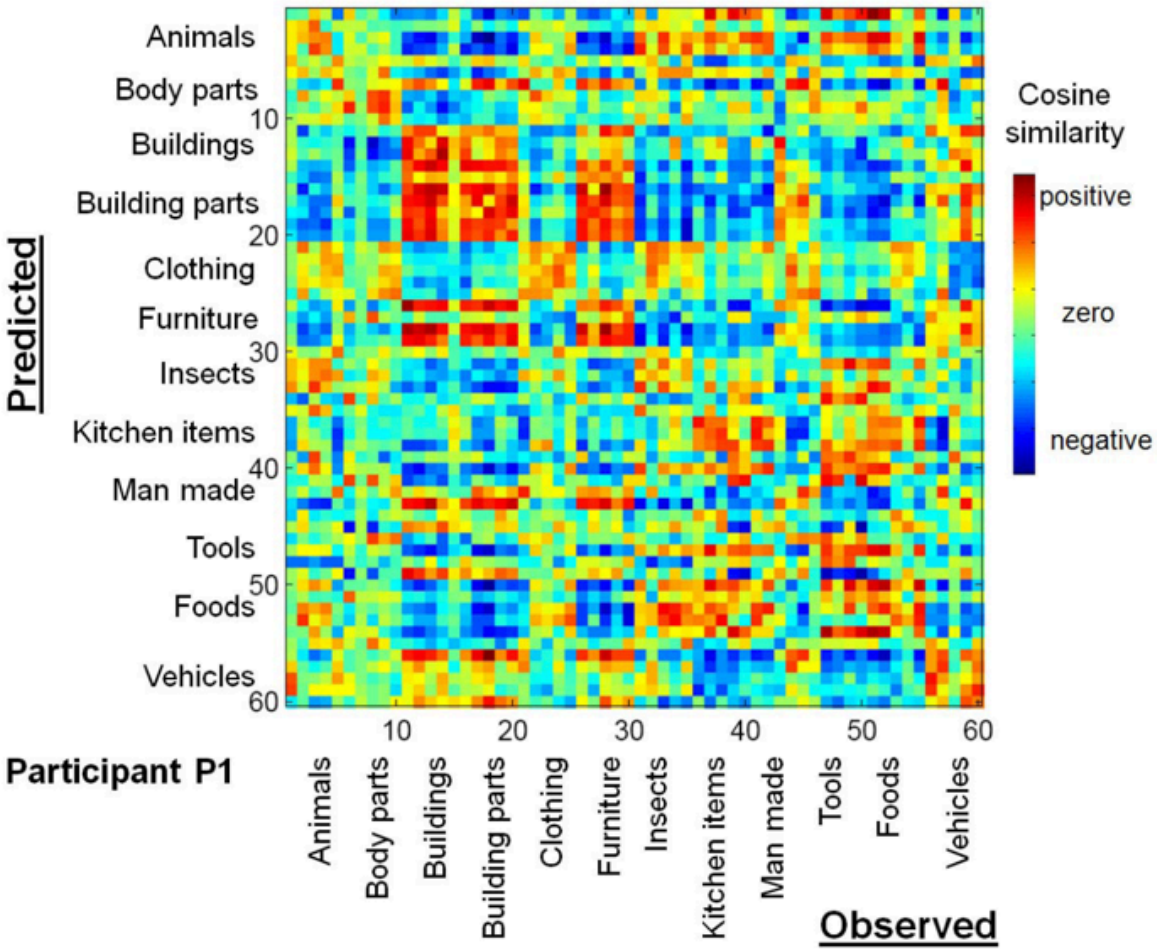
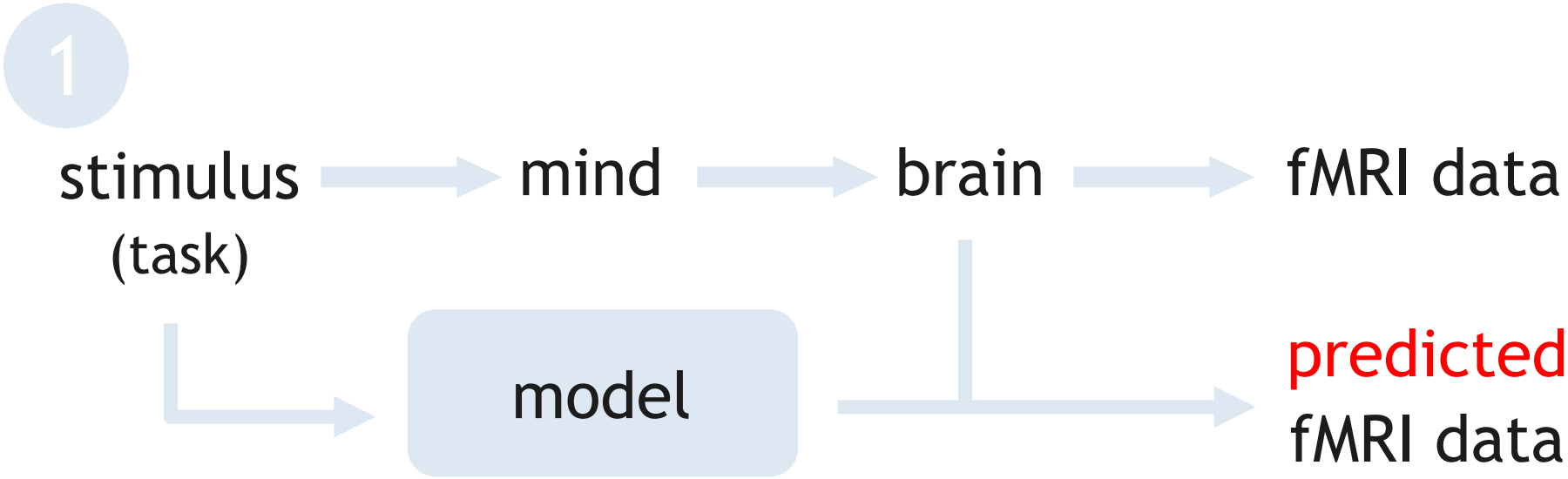


Figure S3. Cosine similarities between predicted and actual images for participant P1.

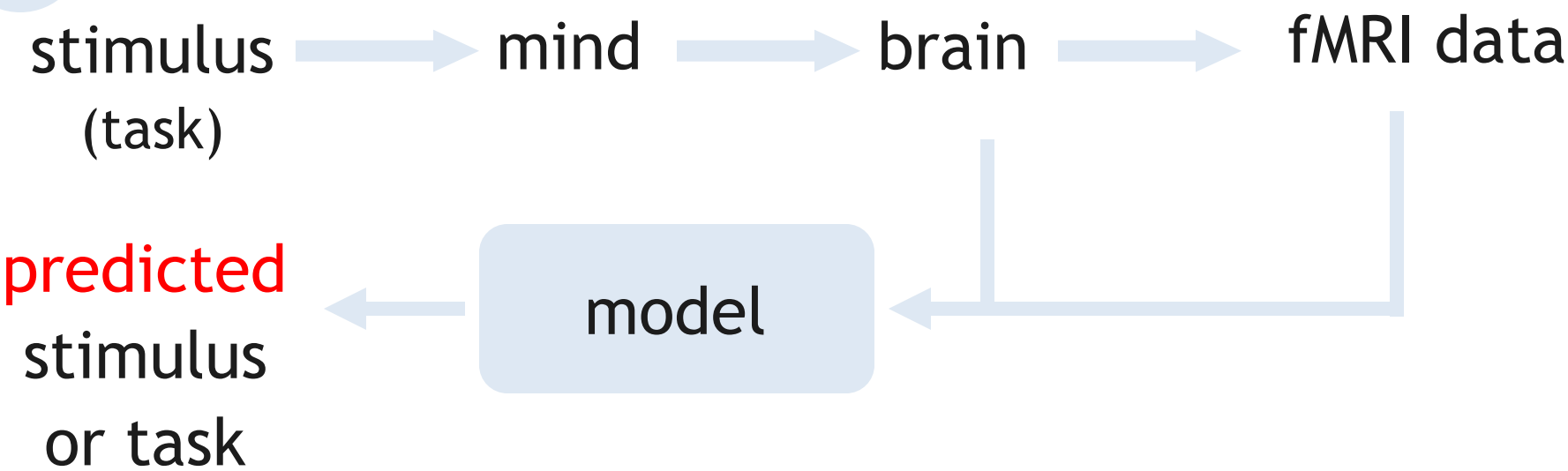
how do we know it's there?



- Kay et al. 2008:
  specified computation model + learnt response model
- Mitchell et al. 2008
  feature values from text data + learnt response model

how do we know it's there?

2



- Thirion et al. 2006
  - forward model with retinotopy task + inversion
- Miyawaki et al. 2008
- Naselaris et al. 2009
- van Gerven et al. 2010

Visual Image Reconstruction from Human Brain Activity using a Combination of Multiscale Local Image Decoders

Yoichi Miyawaki,[1,2,6] Hajime Uchida,[2,3,6] Okito Yamashita,[2] Masa-aki Sato,[2] Yusuke Morito,[4,5] Hiroki C. Tanabe,[4,5] Norihiro Sadato,[4,5] and Yukiyasu Kamitani[2,3,*]
[1]National Institute of Information and Communications Technology, Kyoto, Japan
[2]ATR Computational Neuroscience Laboratories, Kyoto, Japan
[3]Nara Institute of Science and Technology, Nara, Japan
[4]The Graduate University for Advanced Studies, Kanagawa, Japan
[5]National Institute for Physiological Sciences, Aichi, Japan
[6]These authors contributed equally to this work
*Correspondence: kmtn@atr.jp
DOI 10.1016/j.neuron.2008.11.004

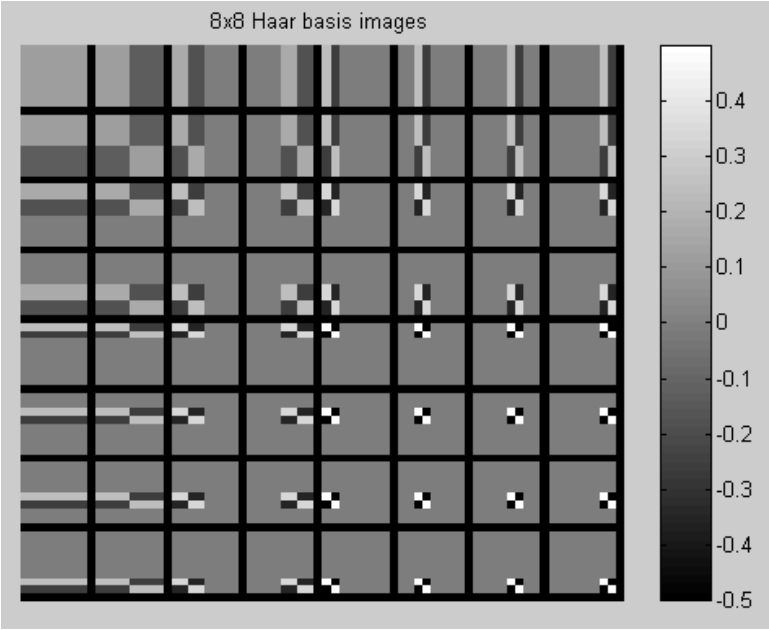
goal

- reconstruct 10 x 10 pixel binary images
- from fMRI data of subjects seeing them

- approach:
  - assume stimulus images can be represented as linear combinations of local image elements (multiple scales)
  - for each image element, learn a predictor of its state for any stimulus from corresponding fMRI data
  - reconstruct image from linear combination of predictors

38

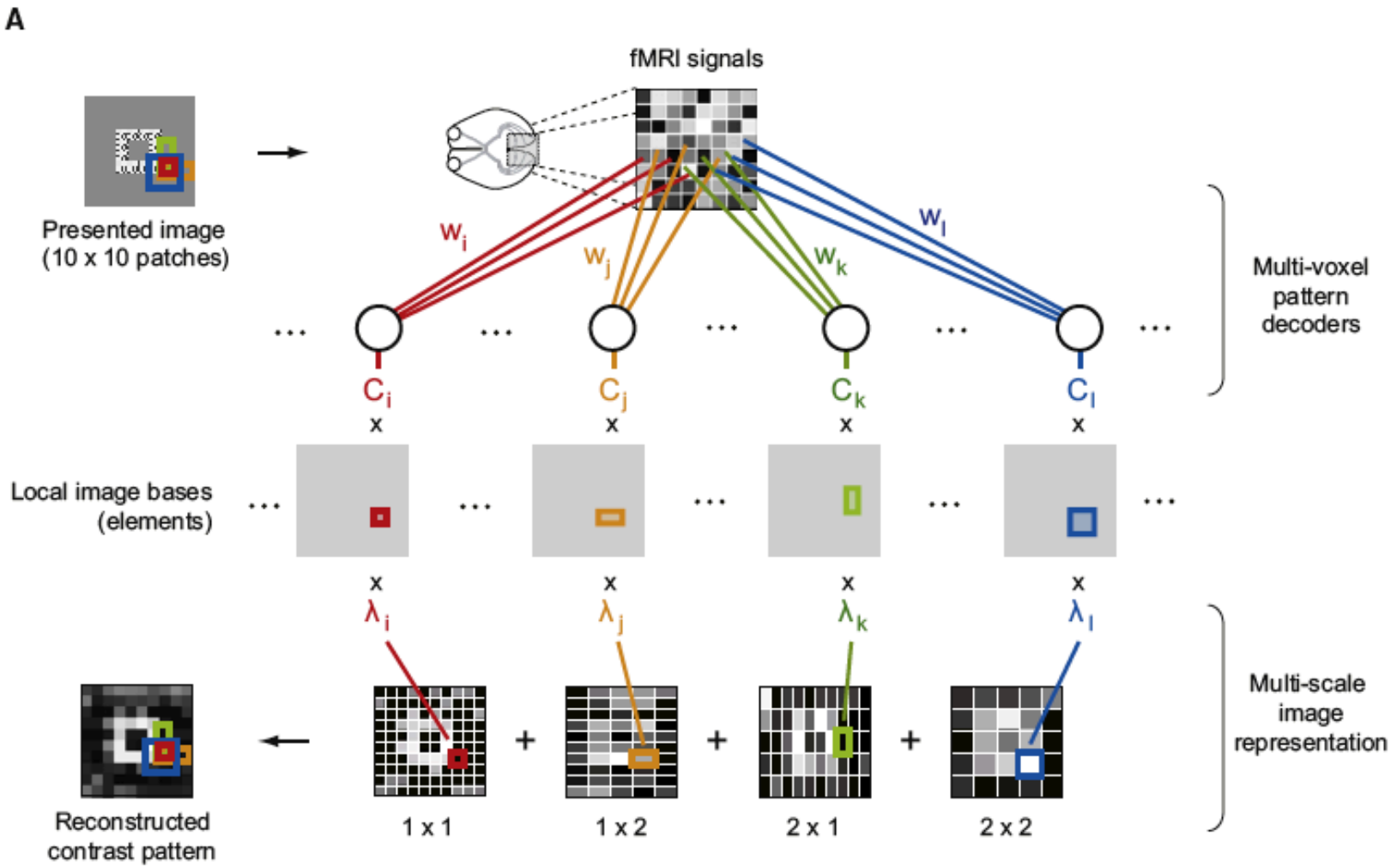
image basis



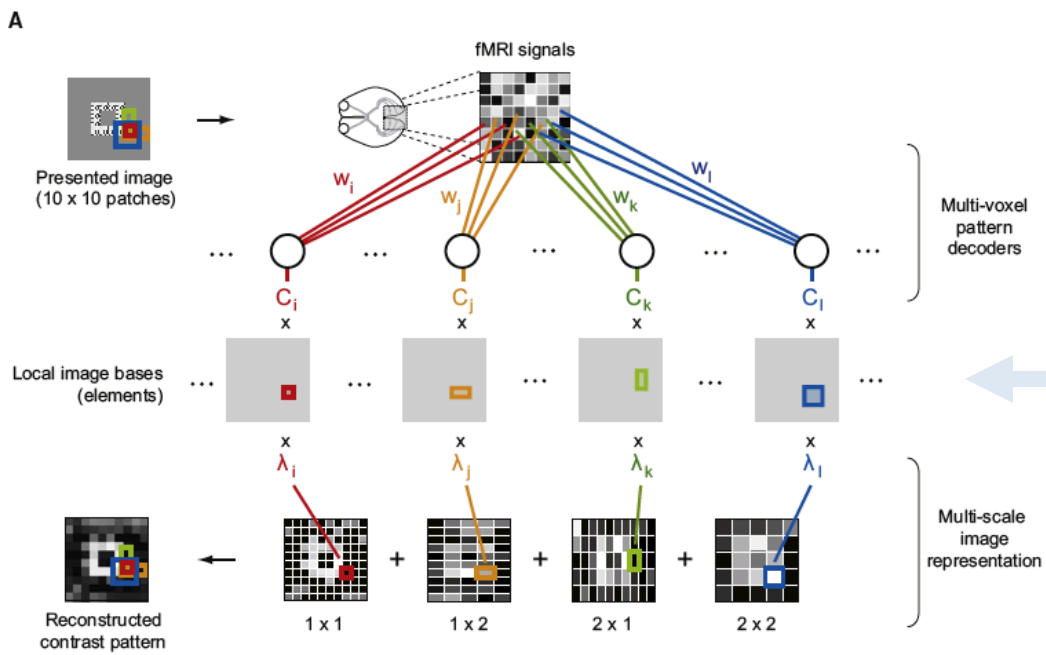
64 element
image basis

- any 8x8 binary image can be represented by a linear combination of the basis elements
- the basis is fixed
- the coefficients depend on the image represented

overall picture



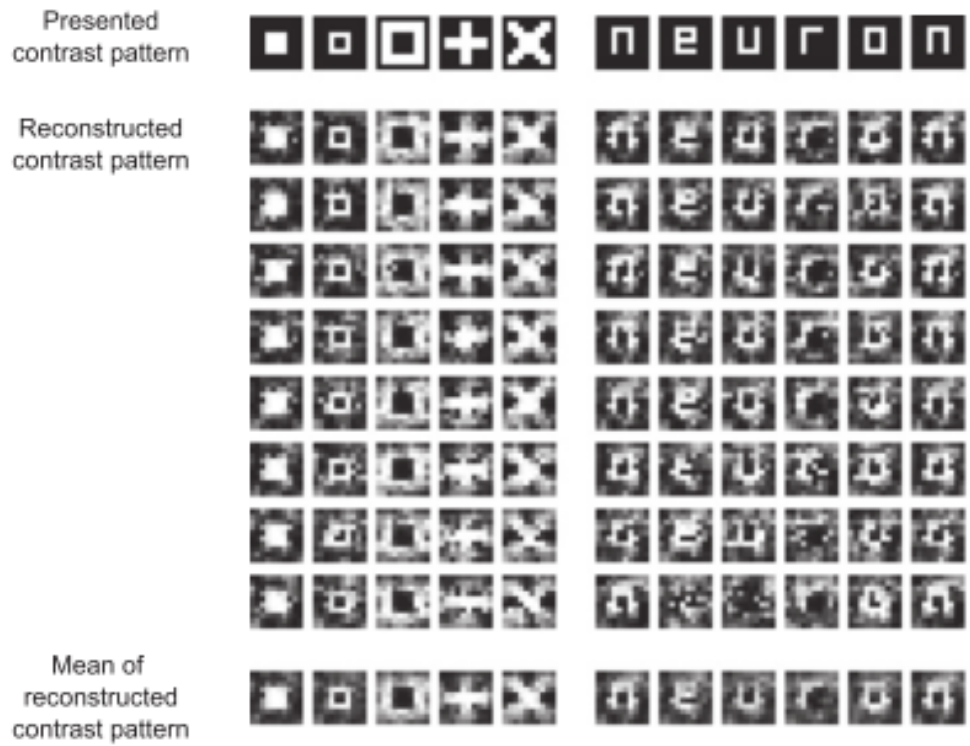
and some details



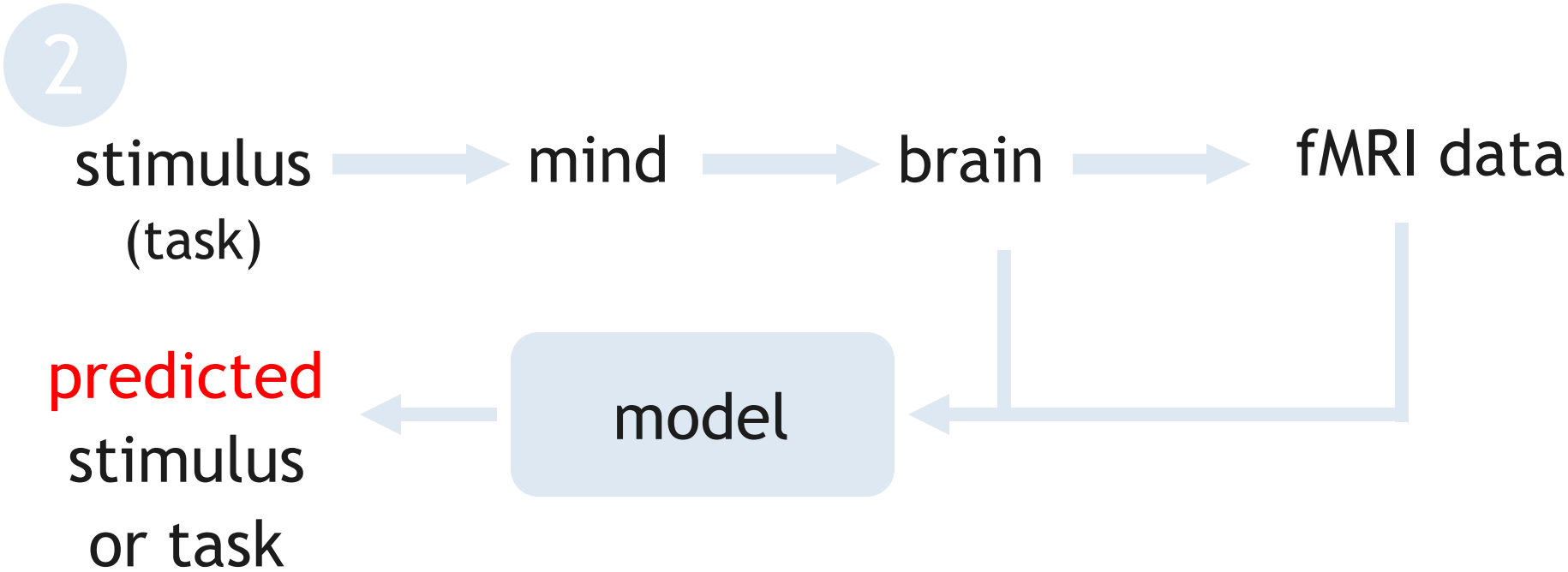
- pattern decoders work over all voxels (regularized)
- basis elements are 1x1, 1x2, 2x1 and 2x2 patches
- at every location
- output of decoders goes through a linear combination with non-negative coefficients

evaluation

- trained with fMRI responses to random images
- test reconstruction of figures, letters, other random
- note that training is such that what is learned generalizes to many stimuli
- can be used for classification



how do we know it's there?



- Miyawaki et al. 2008
  - local decoders of basis elements + linear combination
- Naselaris et al. 2009

Bayesian Reconstruction of Natural Images from Human Brain Activity

Thomas Naselaris,[1] Ryan J. Prenger,[2] Kendrick N. Kay,[3] Michael Oliver,[4] and Jack L. Gallant[1,3,4,*]
[1]Helen Wills Neuroscience Institute
[2]Department of Physics
[3]Department of Psychology
[4]Vision Science Program
University of California, Berkeley, Berkeley, CA 94720, USA
*Correspondence: gallant@berkeley.edu
DOI 10.1016/j.neuron.2009.09.006

goal

- find a natural scene that matches the stimulus
- from fMRI data of a subject
- match both visual properties and semantic content

- approach:
  - learn a model from stimulus to fMRI (same as [Kay 2008])
  - label training images with one of 23 categories
  - learn a model from category label to fMRI (AOC)

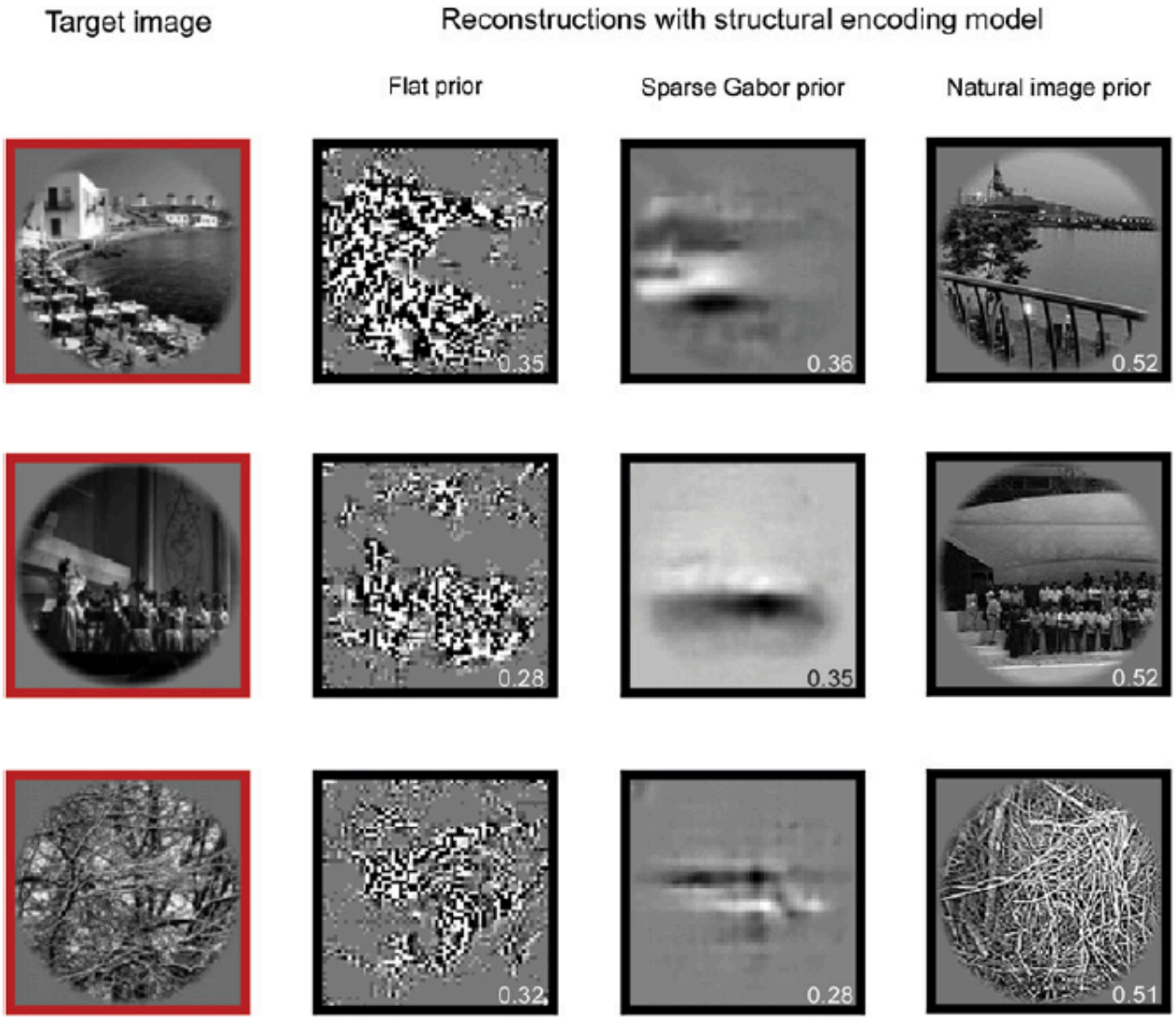
  - invert both models to find most probable stimulus
  - prior probability has a large influence

45

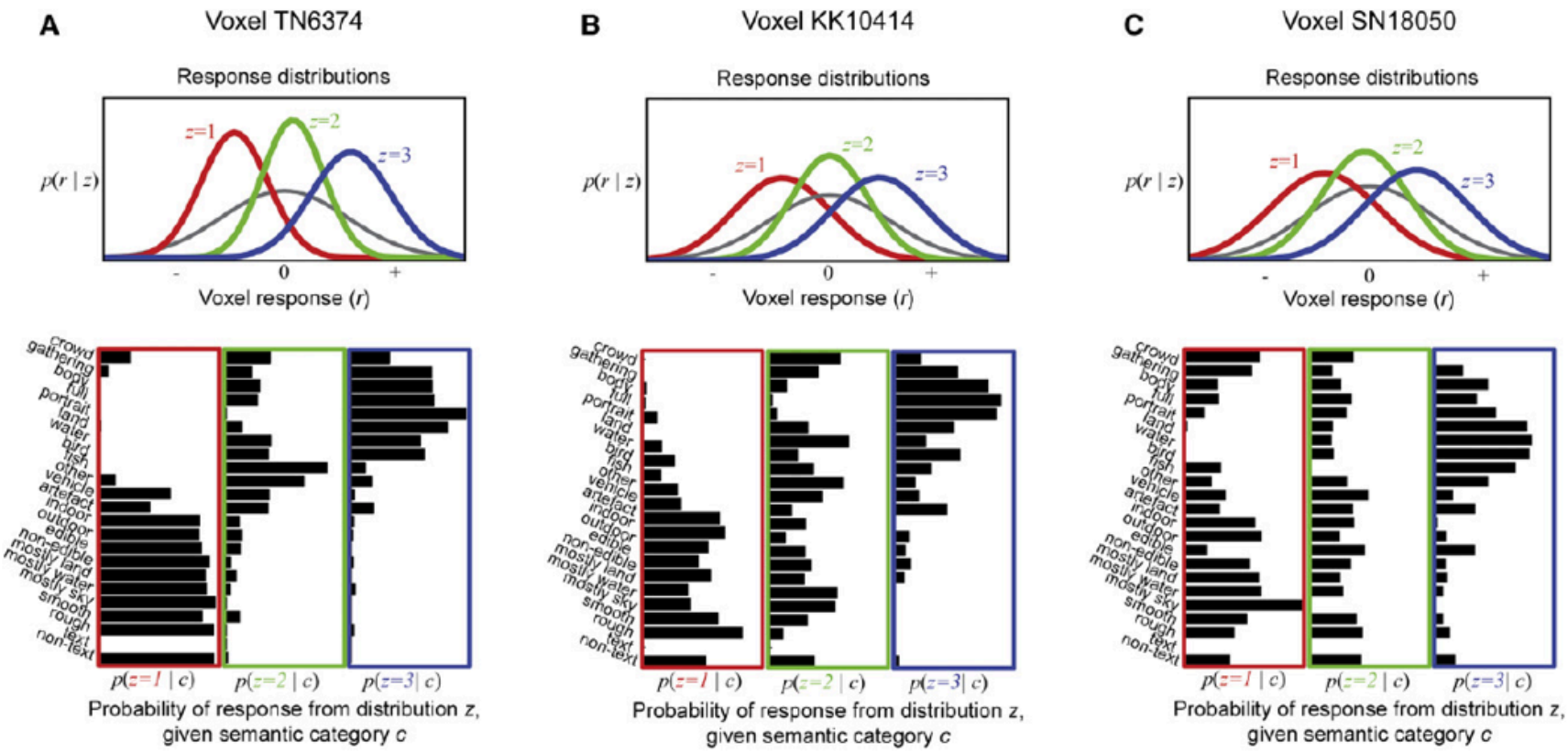
semantic model

- mostly animate
  - human
    - many (crowd/gathering)
    - few (body parts/full bodies/portrait)
  - animal
    - mammal (land/water)
    - non-mammal (bird/fish/other)
- mostly inanimate
  - man-made
    - non-building (vehicle/artifacts)
    - building (indoor/outdoor)
  - natural
    - plant (edible/non-edible)
    - non-plant (land/water/sky)
  - texture

effects of model and prior



effects of model and prior

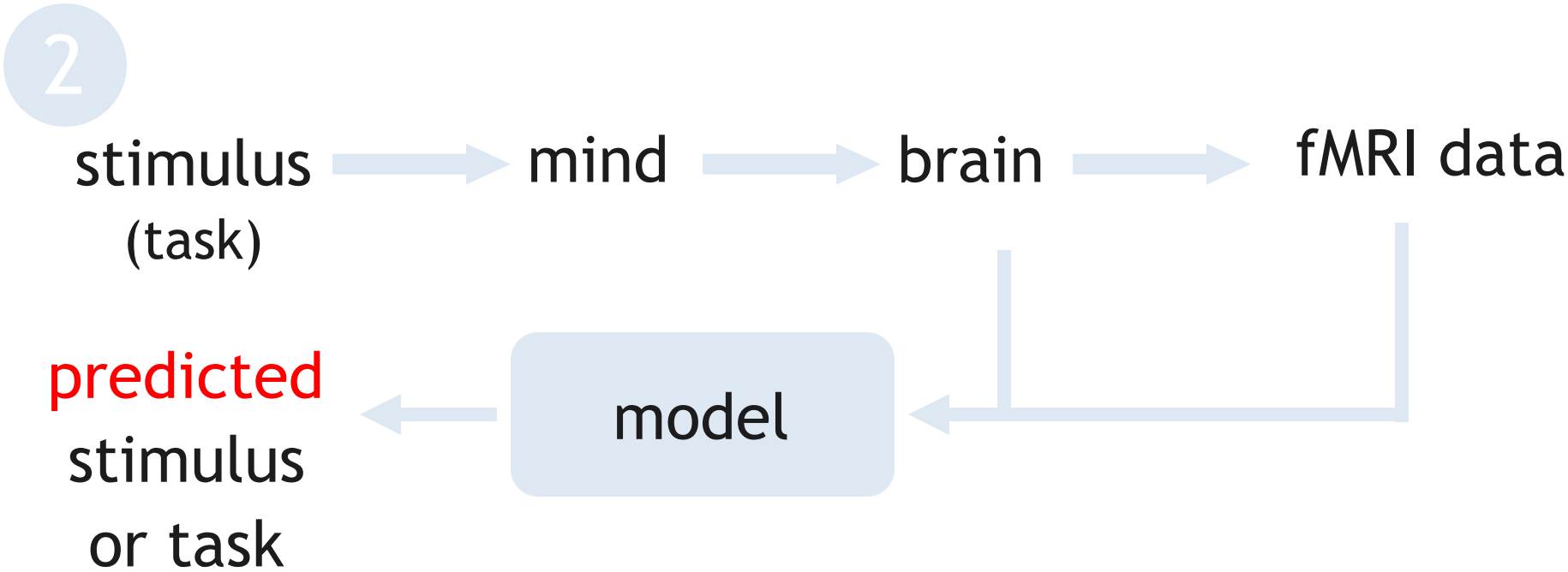

A
Voxel TN6374
Response distributions
p(r | z)
z=1
z=2
z=3
-
0
+
Voxel response (r)
crowd
gathering
body
full
portrait
land
water
bird
fish
other
vehicle
artefact
indoor
outdoor
edible
non-edible
mostly land
mostly water
mostly sky
smooth
rough
text
non-text
p(z=1 | c)
p(z=2 | c)
p(z=3 | c)
Probability of response from distribution z, given semantic category c
B
Voxel KK10414
Response distributions
p(r | z)
z=1
z=2
z=3
-
0
+
Voxel response (r)
crowd
gathering
body
full
portrait
land
water
bird
fish
other
vehicle
artefact
indoor
outdoor
edible
non-edible
mostly land
mostly water
mostly sky
smooth
rough
text
non-text
p(z=1 | c)
p(z=2 | c)
p(z=3 | c)
Probability of response from distribution z, given semantic category c
C
Voxel SN18050
Response distributions
p(r | z)
z=1
z=2
z=3
-
0
+
Voxel response (r)
crowd
gathering
body
full
portrait
land
water
bird
fish
other
vehicle
artefact
indoor
outdoor
edible
non-edible
mostly land
mostly water
mostly sky
smooth
rough
text
non-text
p(z=1 | c)
p(z=2 | c)
p(z=3 | c)
Probability of response from distribution z, given semantic category c

evaluation

- reconstruction accuracy
  - complex wavelet-domain correlation with target image

- classification accuracy
  - probability that the predicted image has the same class label as the target image
  - at multiple levels of granularity

how do we know it's there?



- Miyawaki et al. 2008
  - local decoders of basis elements + linear combination
- Naselaris et al. 2009
  - model of visual processing + semantic information

how do we know it's there?

3

stimulus (task) → mind → brain → fMRI data

known relationships between conditions

pattern similarities or confusion matrices

- Kriegeskorte et al. 2008/2009
- Walther et al. 2009
- ...

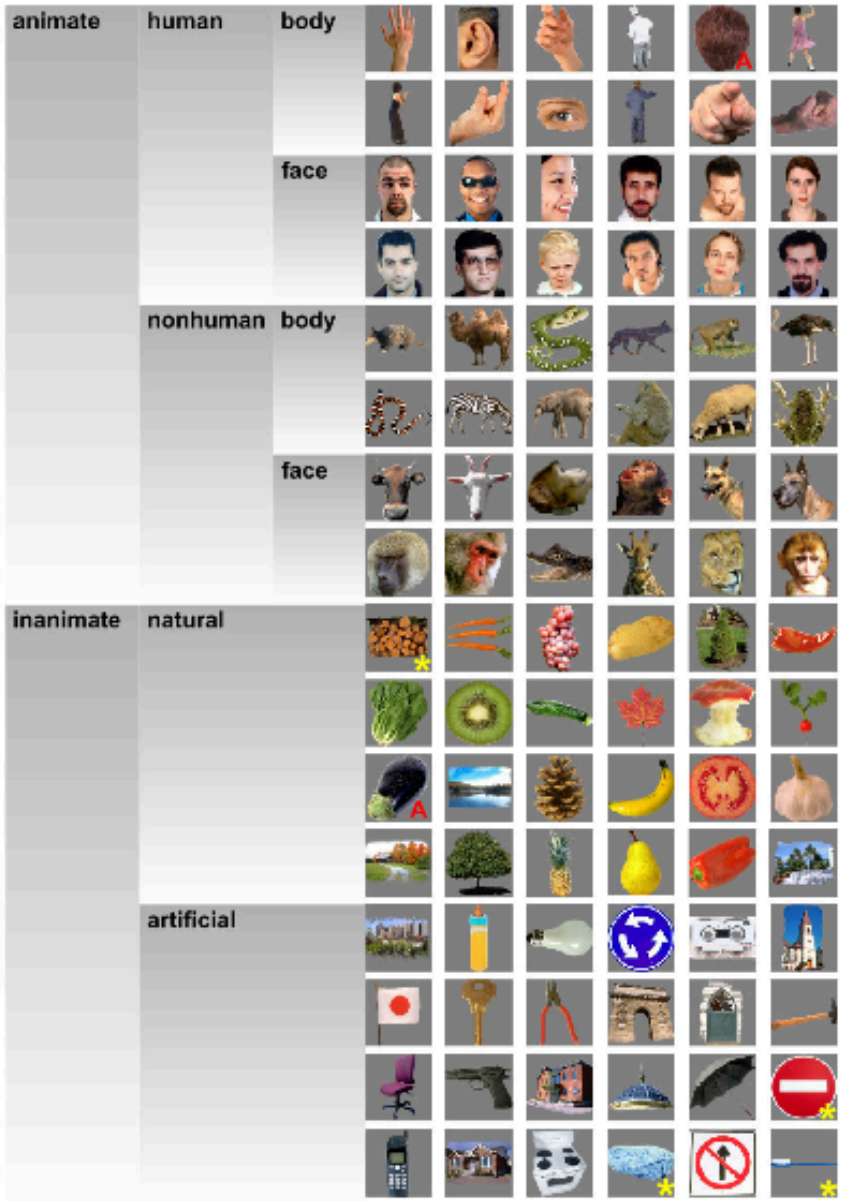
representational similarity analysis

goal:

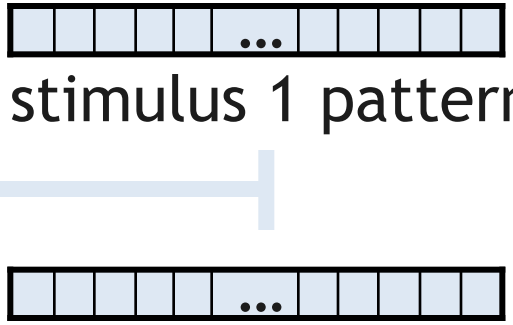
- compare fMRI patterns elicited by stimuli without assuming a priori structure
- relate similarity/distance in fMRI space to behavioural or hypothesized similarity/distance

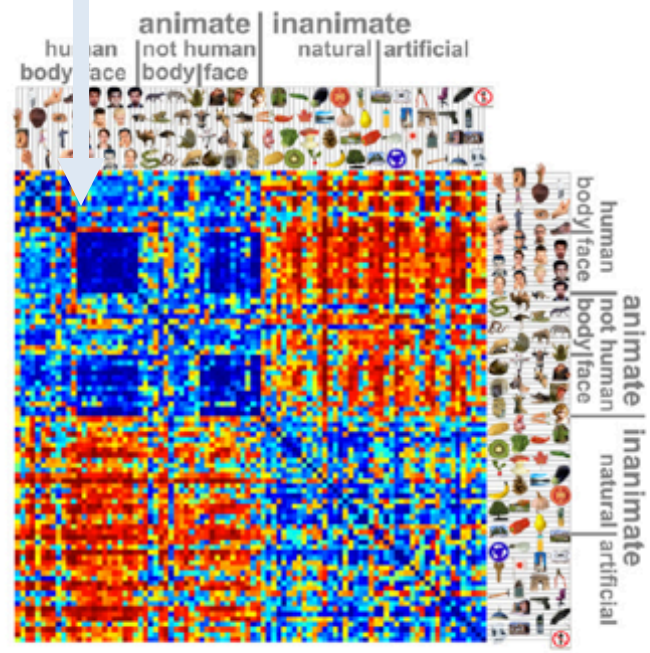
- a range of ideas more than a technique

stimuli and similarity



correlation

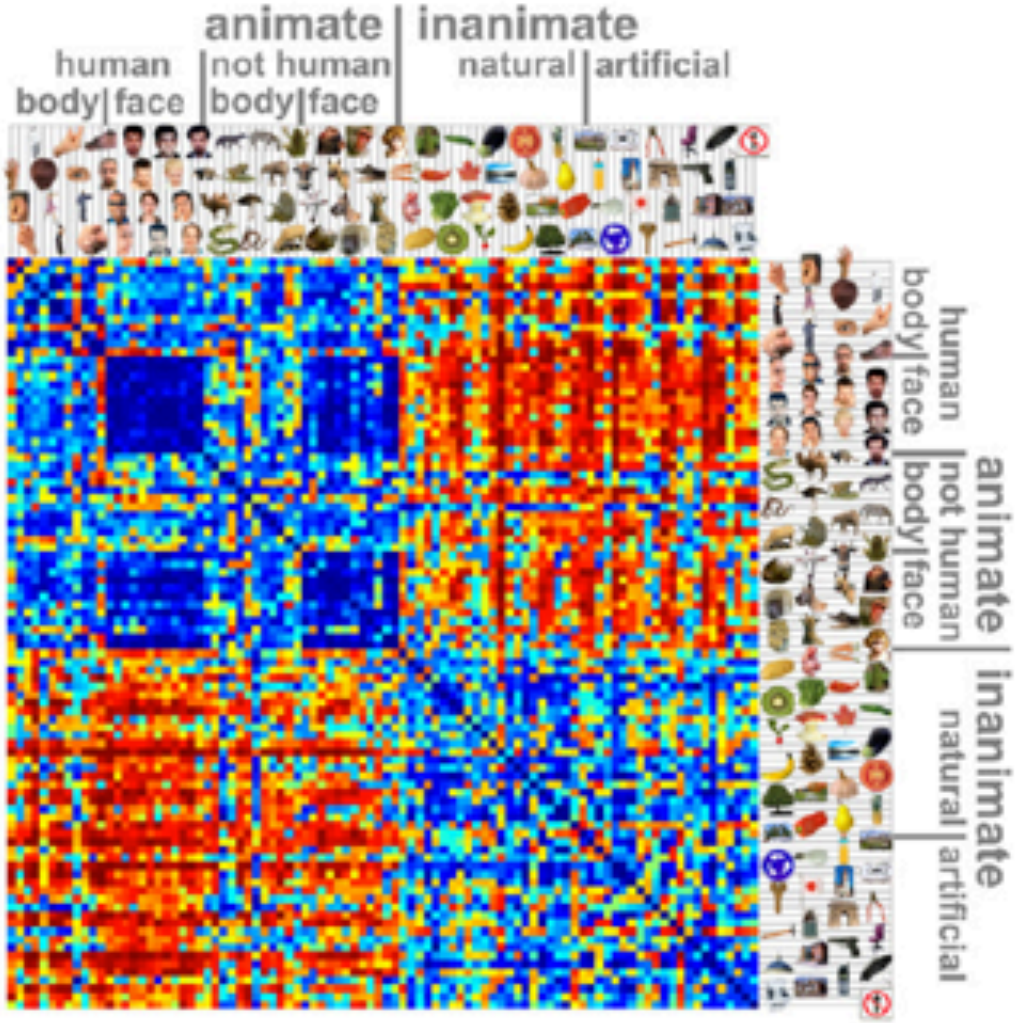




human IT

stimuli and similarity

- summarizes a lot of information
- a bit hard to make sense of



human IT

multidimensional scaling

- represent 96 items in a 2D space
- euclidean distance in 2D space corresponds (if possible) to dissimilarity in the matrix

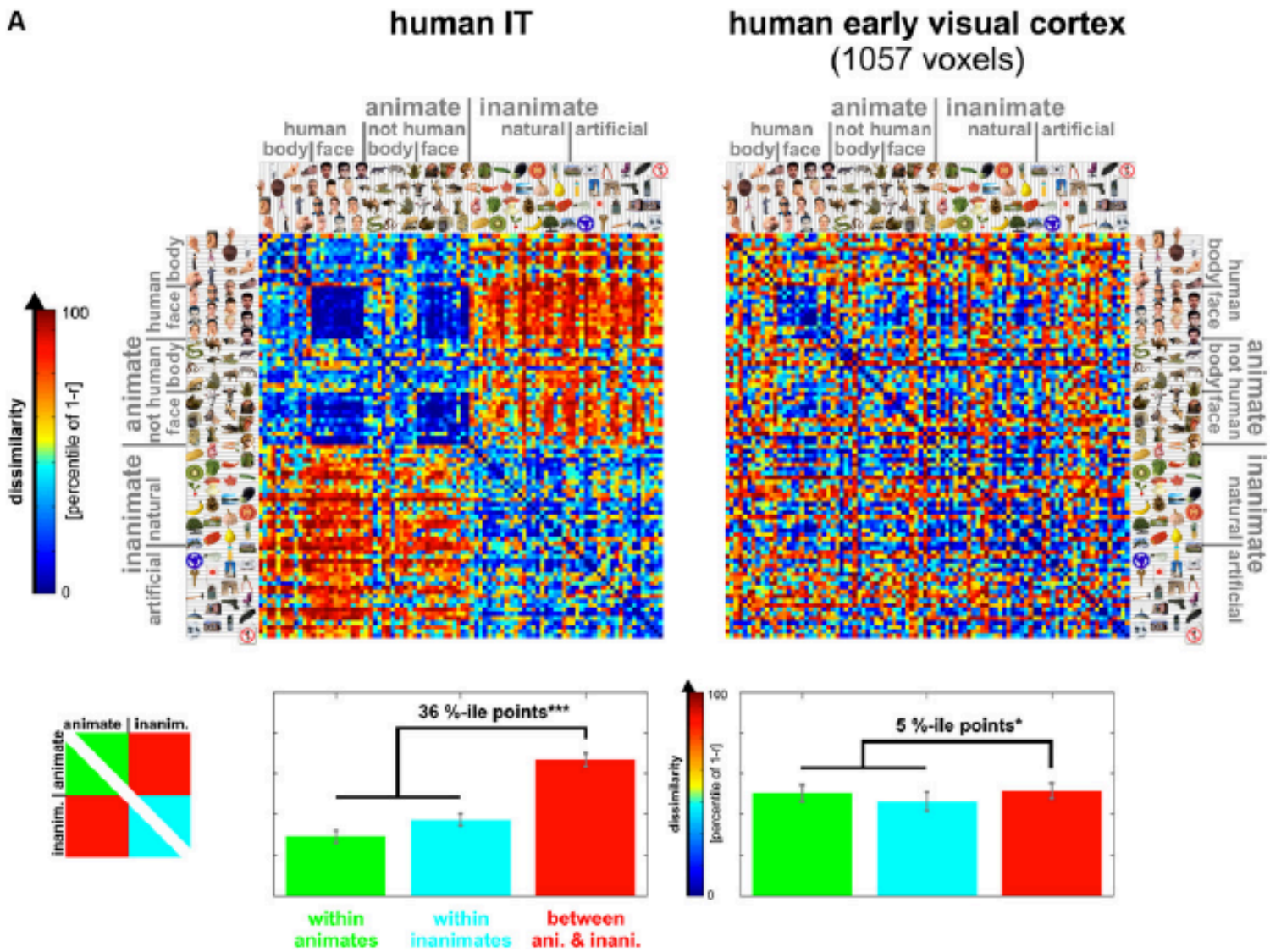
human IT

multidimensional scaling

human IT

body
face
natural obj.
artificial obj.

compare multiple areas



other aspects

- can compute matrices in different ROIs
- can look for matrices compatible with a certain hypothesis (or use it to compare competing ones)
- matrices matching behaviour, ratings, etc
- more later...

how do we know it's there?

3

stimulus (task) → mind → brain → fMRI data

known relationships between conditions

pattern similarities or confusion matrices

- Kriegeskorte et al. 2008/2009
- Walther et al. 2009
- ...

Behavioral/Systems/Cognitive

Natural Scene Categories Revealed in Distributed Patterns of Activity in the Human Brain

Dirk B. Walther,[1] Eamon Caddigan,[1,2] Li Fei-Fei,[3]* and Diane M. Beck[1,2]*

[1]Beckman Institute for Advanced Science and Technology, University of Illinois Urbana-Champaign, Urbana, Illinois, 61801-2325, [2]Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, Illinois 61820-6232, and [3]Computer Science Department, Stanford University, Stanford, California 94305-9025

relating classifiers to behaviour

goal:

- identify brain regions

  V1, FFA, LOC, PPA, RSC (retrosplenial)
- involved in rapid natural scene categorization

  beach, city, road, ...

approach:

- in each candidate region
  - train classifier to distinguish the six possibilities
  - compute confusion matrix
- contrast with other matrices...

rapid scene categorization

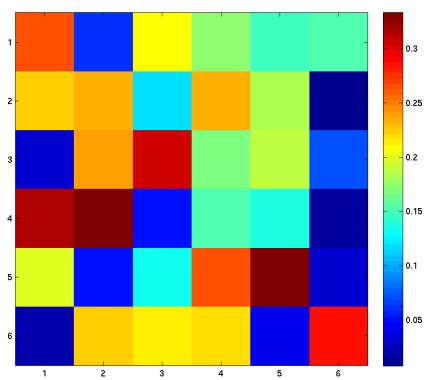
- humans are good at extracting information about a scene
- 100ms can be enough

per-ROI classifier

ROI data

ROI classifier

ROI confusion matrix



behavioural information

- collect data on confusability
- flash scenes (11-45 ms) have subjects make 6-way judgement
- behavioural confusion matrix

stimulus information

- compute how similar scenes from various categories are
- low-level visual properties
- stimulus similarity matrix

analysis

- question 1:
  which ROIs drive scene classification?
  ROIs where classifier and behavioural
  confusion matrices match

- question 2:
  which ROIs encode low-level image properties?
  ROIs where image and behavioural
  similarity matrices match

results

Table 1. Summary of main results

ROI	Decoding accuracy	Error correlation
V1	26%*	0.21
FFA	22%	0.10
LOC	24%*	0.42*
RSC	27%*	0.34 †
PPA	31%**	0.57**

Decoding accuracy is measured in percentage of blocks predicted correctly, and significance is assessed relative to chance (17%). Error correlation establishes a correlation between misclassifications (off-diagonal entries in the confusion matrices) (Figs. 2, 3) between decoding from ROIs and human behavior. Image similarity correlation correlates the image similarities matrix with the confusion matrix from fMRI decoding. The inversion effect is defined as the difference in accuracy of a decoder trained and tested with upright versus trained and tested with inverted scene presentations. PPA shows significant effects in all analyses except for the image similarity correlation. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{\dagger}p = 0.069$.

- all 5 ROIs have above chance classification
- PPA, LOC, RSC show correlation between classifier and subject errors

results

Table 1. Summary of main results

ROI	Decoding accuracy	Error correlation	Image similarity correlation
V1	26%*	0.21	0.46**
FFA	22%	0.10	0.03
LOC	24%*	0.42*	−0.22
RSC	27%*	0.34 †	−0.24
PPA	31%**	0.57**	−0.07

Decoding accuracy is measured in percentage of blocks predicted correctly, and significance is assessed relative to chance (17%). Error correlation establishes a correlation between misclassifications (off-diagonal entries in the confusion matrices) (Figs. 2, 3) between decoding from ROIs and human behavior. Image similarity correlation correlates the image similarities matrix with the confusion matrix from fMRI decoding. The inversion effect is defined as the difference in accuracy of a decoder trained and tested with upright versus trained and tested with inverted scene presentations. PPA shows significant effects in all analyses except for the image similarity correlation. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{\dagger}p = 0.069$.

- low-level similarity correlates with behavioural confusion in V1 but not elsewhere

how do we know it's there?

3

stimulus (task) → mind → brain → fMRI data

known relationships between conditions

pattern similarities or confusion matrices

- Kriegeskorte et al. 2008/2009
- Walther et al. 2009
- ...

where can it come from?

stimulus (task) → mind → brain → fMRI data

↓

what is present in the mind as the task is performed?

↓

- known or constrained from behavioural experiments
- modelled mathematically or computationally
- hypothesized
- learnt elsewhere (text corpora)

take home points

- prediction goes beyond classification
- similarity/distance are useful too

take home points

- prediction goes beyond classification
- similarity/distance are useful too
- data-driven prediction has limits
  - what we can feed the prediction mechanism
  - what kind of bias it has
  - what we can dissect from it

take home points

- prediction goes beyond classification
- similarity/distance are useful too
- data-driven prediction has limits
  - what we can feed the prediction mechanism
  - what kind of bias it has
  - what we can dissect from it
- many other things to use in the mechanism
  - known or constrained from behavioural experiments
  - modelled mathematically or computationally
  - hypothesized
  - learnt elsewhere (text corpora)

Thank you!

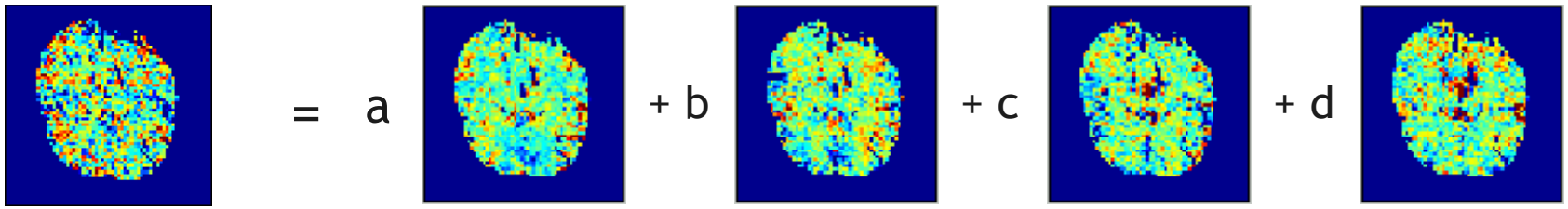
questions?

a mathematical perspective for the advanced methods

Francisco Pereira

Botvinick Lab
Princeton Neuroscience Institute
Princeton University

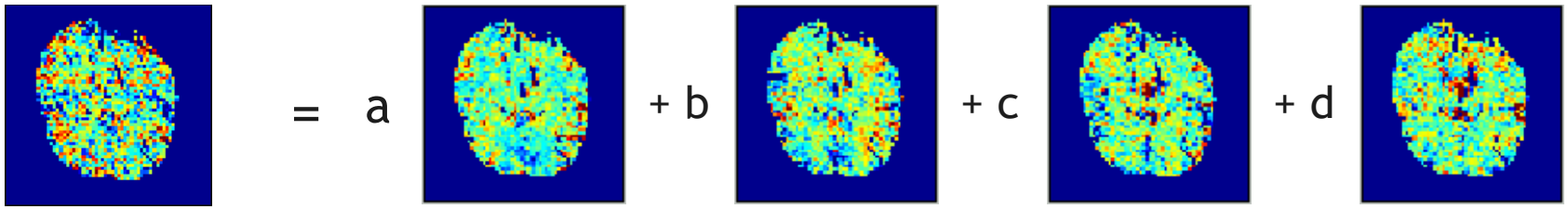
low-dimensional spatial decompositions



example

(a,b,c,d) in a basis of (eigen)images
are coordinates

low-dimensional spatial decompositions



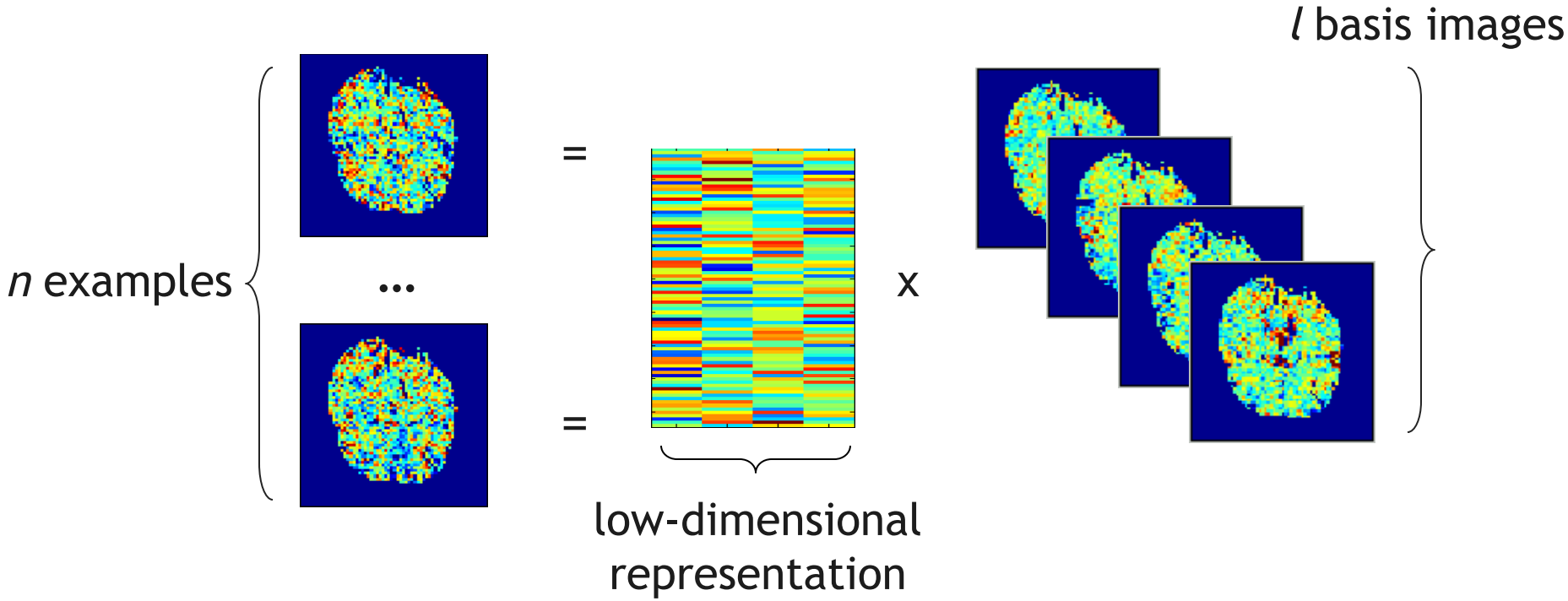
example

(a,b,c,d)
are coordinates

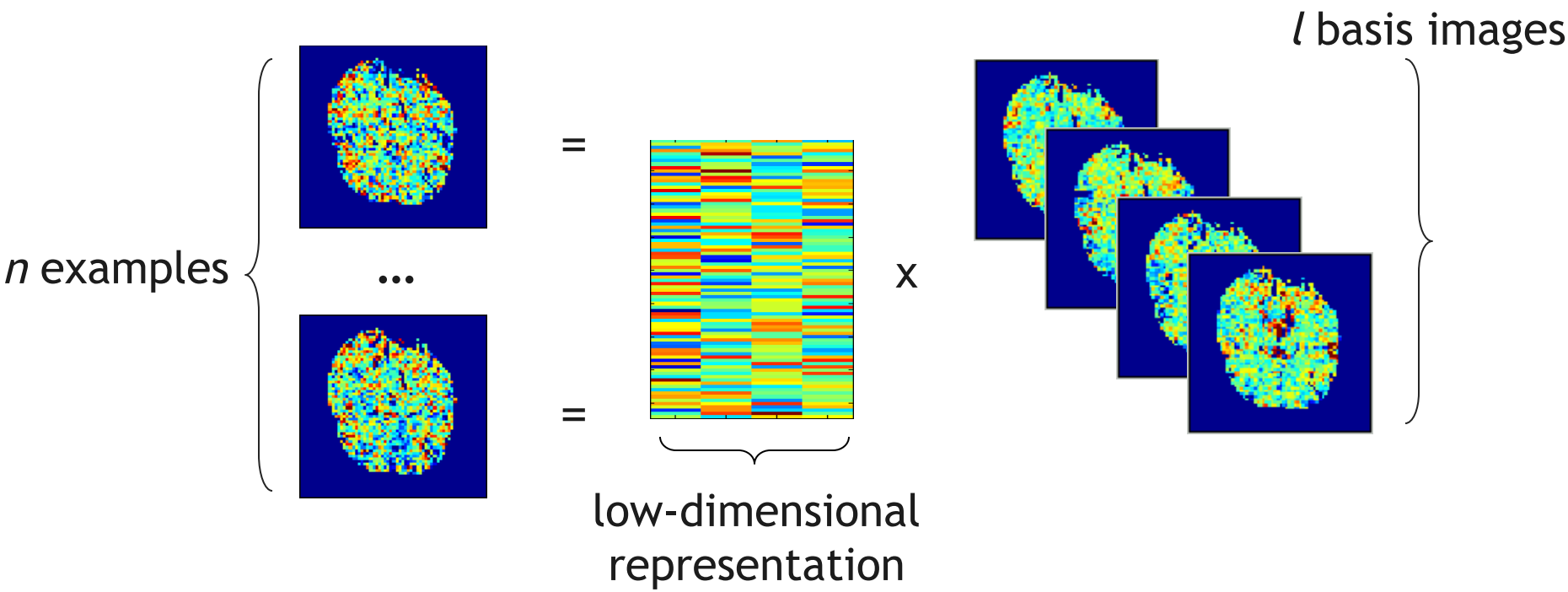
in a basis of (eigen)images

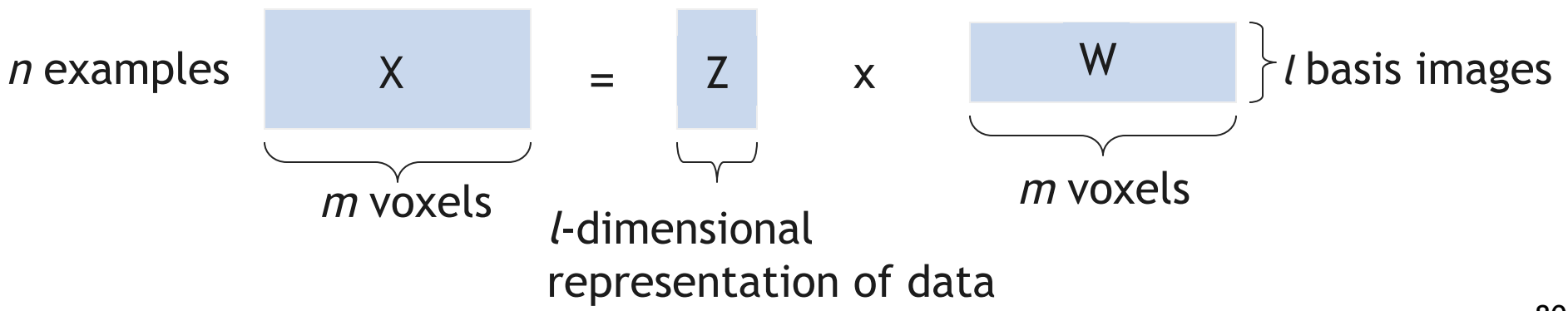
(a,b,c,d)
can also be seen as
feature values
in a low-dimensional
feature representation

low-dimensional spatial decompositions



low-dimensional spatial decompositions

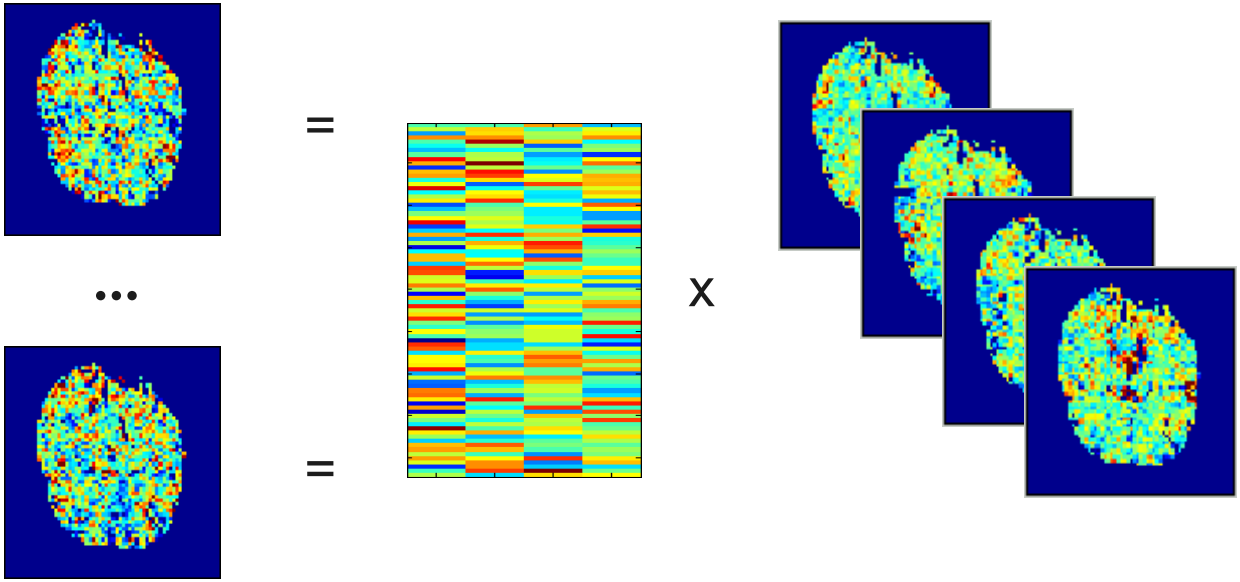




low-dimensional spatial decompositions

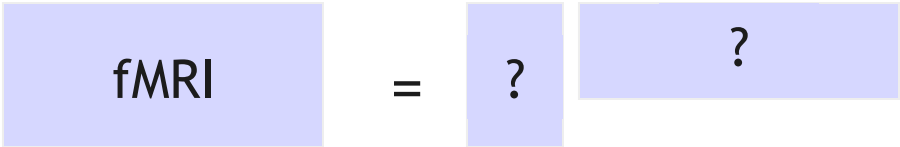
Why express an image in terms of a basis of images?

- capture spatial patterns of activity over many voxels
- coordinates give a succinct representation of the data



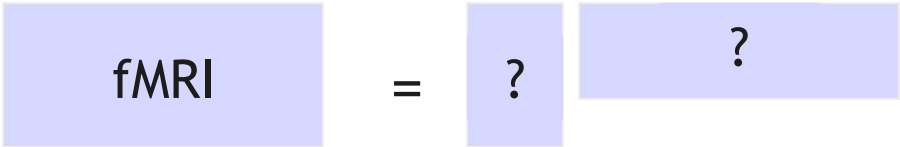
features

- from data alone: learn Z and basis
  - any matrix factorization
  - SVD, ICA, NMF,...
  - regularization matters

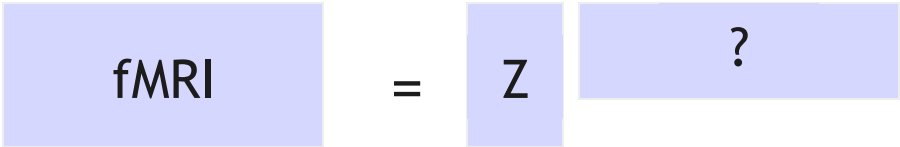


features

- from data alone: learn Z and basis
  - any matrix factorization
  - SVD, ICA, NMF,...
  - regularization matters

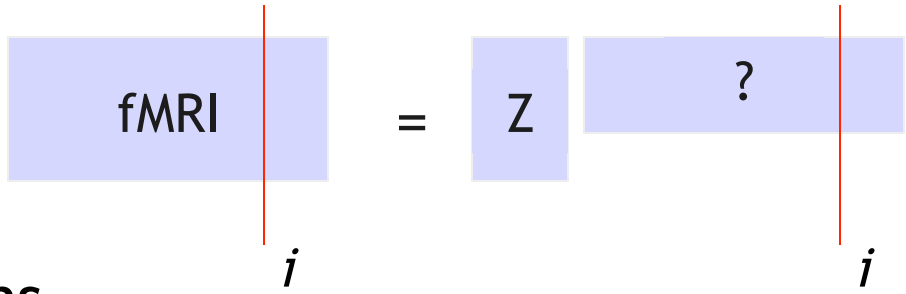


- hypothesized
  - behavioural experiments
  - mathematical model



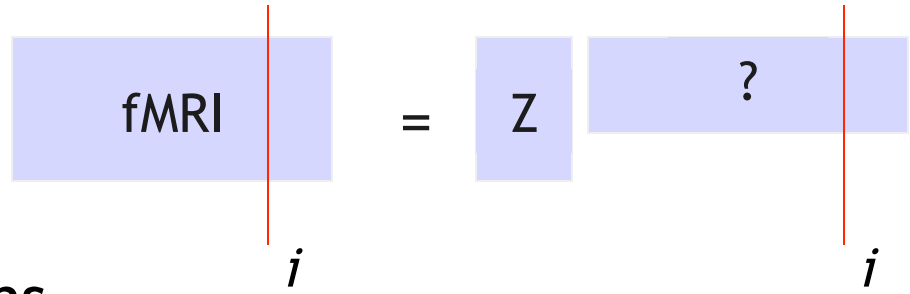
- learned from text

two problems

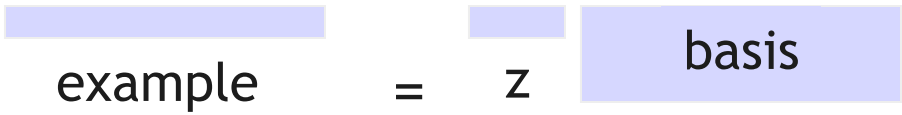


- given Z, learn basis
  - #voxels regression problems
  - predict voxel *i* from Z using column *i* of basis matrix as regression coefficients

two problems



- given Z, learn basis
  - #voxels regression problems
  - predict voxel *i* from Z using column *i* of basis matrix as regression coefficients



- given a new example X, find its basis coordinates
  - 1 regression problem
  - predict x vector from basis rows using z as regression coefficients

issues

- if Z columns are orthogonal, finding basis is simple
- if basis images are orthogonal, finding z is simple

issues

- if Z columns are orthogonal, finding basis is simple
- if basis images are orthogonal, finding z is simple

- if they are not, we need regularization
  - make things positive
  - make things have low L2 (small) or L1 (sparse) norms

issues

- if Z columns are orthogonal, finding basis is simple
- if basis images are orthogonal, finding z is simple

- if they are not, we need regularization
  - make things positive
  - make things have low L2 (small) or L1 (sparse) norms

- there might also be constraints
  - e.g. entries of z must be positive and add to 1

a suggestion

- previous problems are likely convex
- CVX
  - a solver for a subset of convex programming
  - uses MATLAB notation to express the problem
  - many functions in the objective and constraints
  - http://cvxr.com/cvx
  - also available in Python

a suggestion

- previous problems are likely convex
- CVX
  - a solver for a subset of convex programming
  - uses MATLAB notation to express the problem
  - many functions in the objective and constraints
  - http://cvxr.com/cvx
  - also available in Python

- if not convex,
  - consider dividing into subproblems that are and solve them in alternation
  - Carl Rasmussen's minimize.m

Thank you!

questions?

assumptions were made II

- stimulus differences lead to fMRI differences
  - very task dependent

- same task produces train and test examples

assumptions were made II

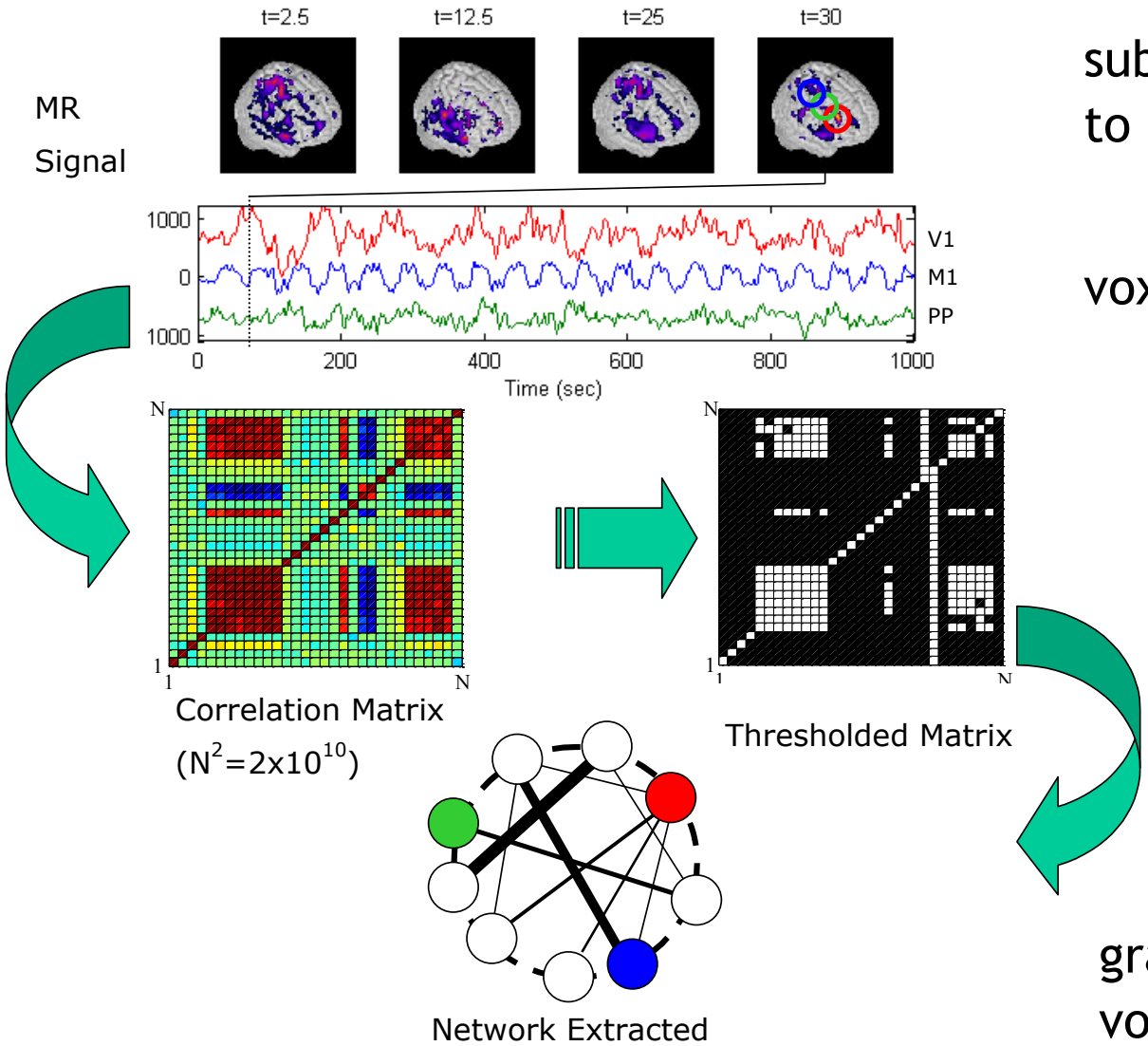
- stimulus differences lead to fMRI differences
  - very task dependent
- same task produces train and test examples
- reproducibility:
  - subject behaviour from trial to trial is the same
  - same behaviour leads to same neural response
  - same behaviour/response lead to same fMRI signal
  - patterns elicited are stable across runs (or sessions)

4 case studies in relaxing assumptions

- features are functions of voxels/interactions
  - auditory processing in schizophrenia
- endogenous events
  - memory context reinstatement
- detecting activation at multiple time scales

- predictions of many simultaneous events
  - virtual reality environment

auditory processing in schizophrenia



subjects attend to sentences

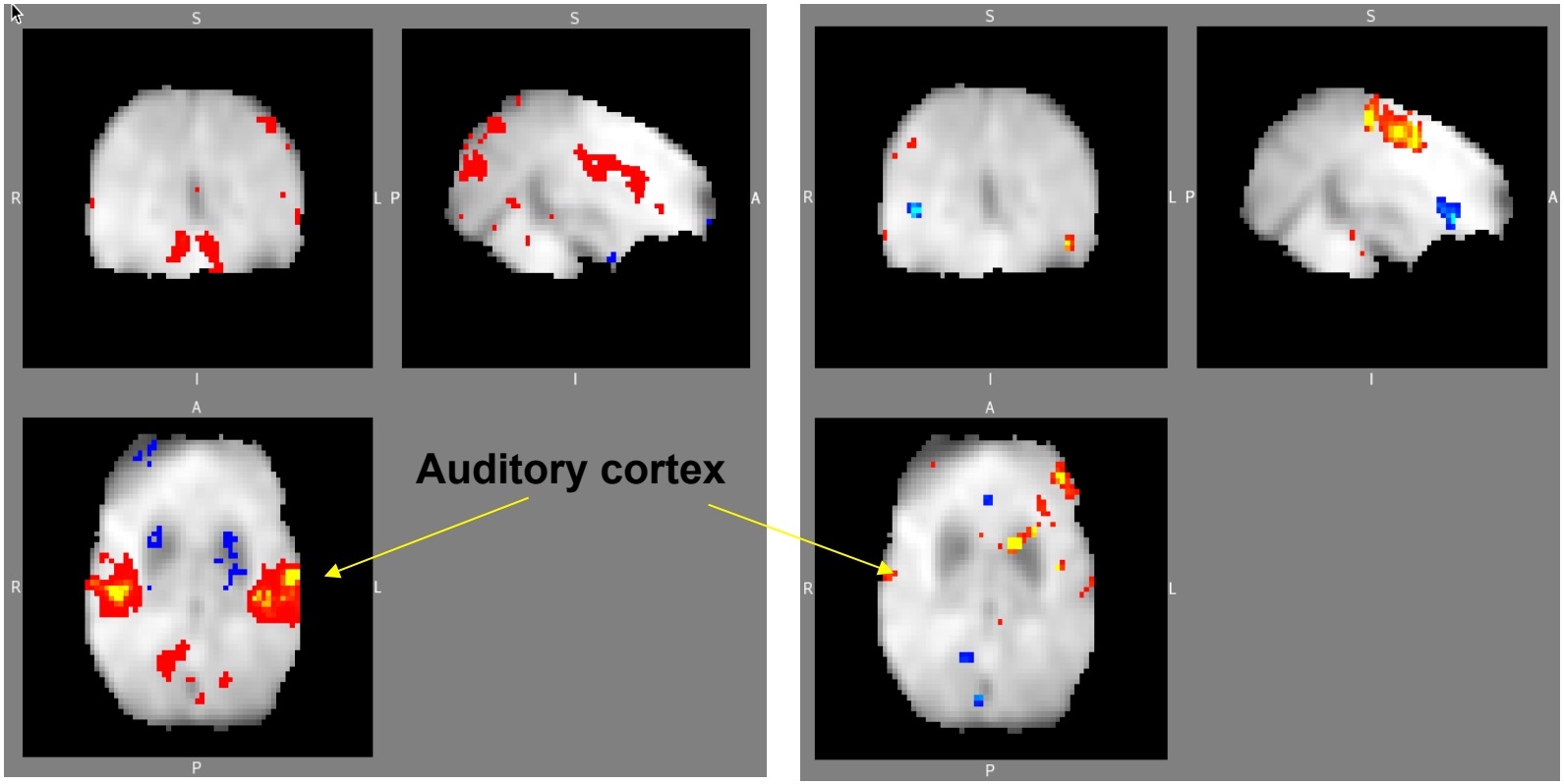
voxel time series

graph connecting voxels with other correlated voxels

[Cecchi/Rish]

auditory processing in schizophrenia

- graph can be turned into per-voxel measures (e.g. degree)
- each map contrasts normals vs patients
- examples with voxel degree features can be classified



Degree Map
(voxel degree)

Statistical Parametric Map
(voxel response to task)

[Cecchi/Rish]

auditory processing in schizophrenia

key ideas:

- we can consider features other than voxel values
- functions of voxels can integrate over time
- there are different types of information present

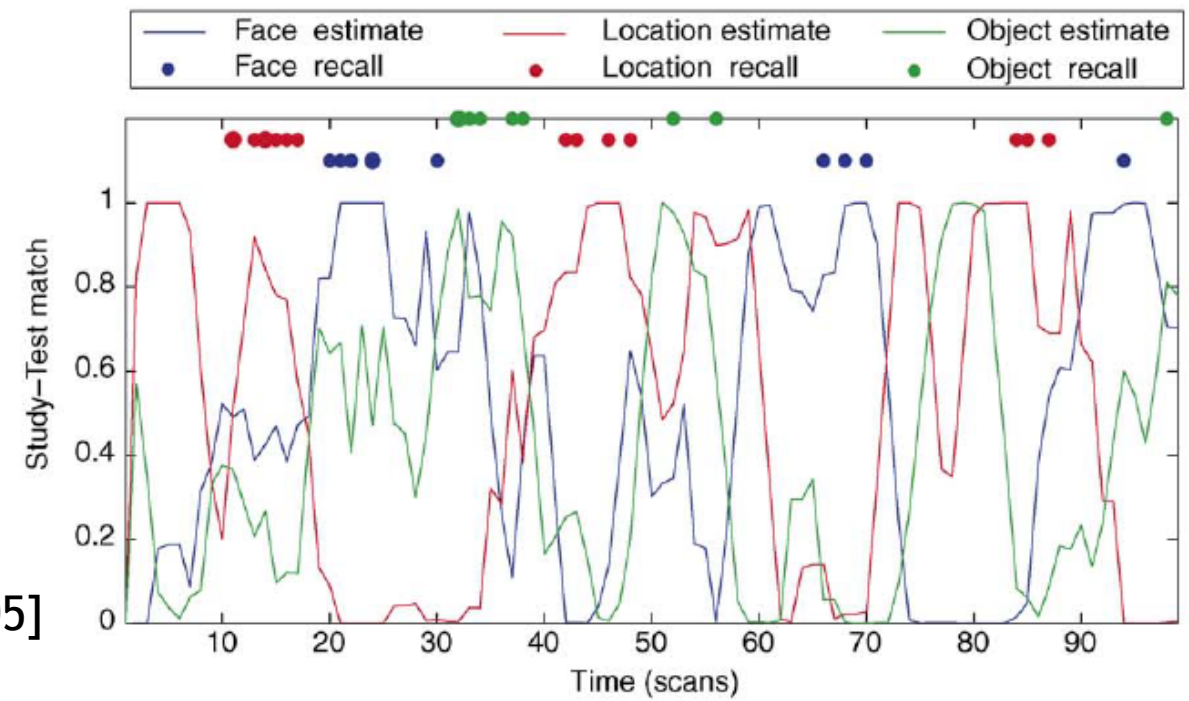
detecting memory context reinstatement

- items from 3 categories: faces, locations, objects
- classifier trained during study of the grouped items

detecting memory context reinstatement

- items from 3 categories: faces, locations, objects
- classifier trained during study of the grouped items
- detect category reinstatement during free recall (test)

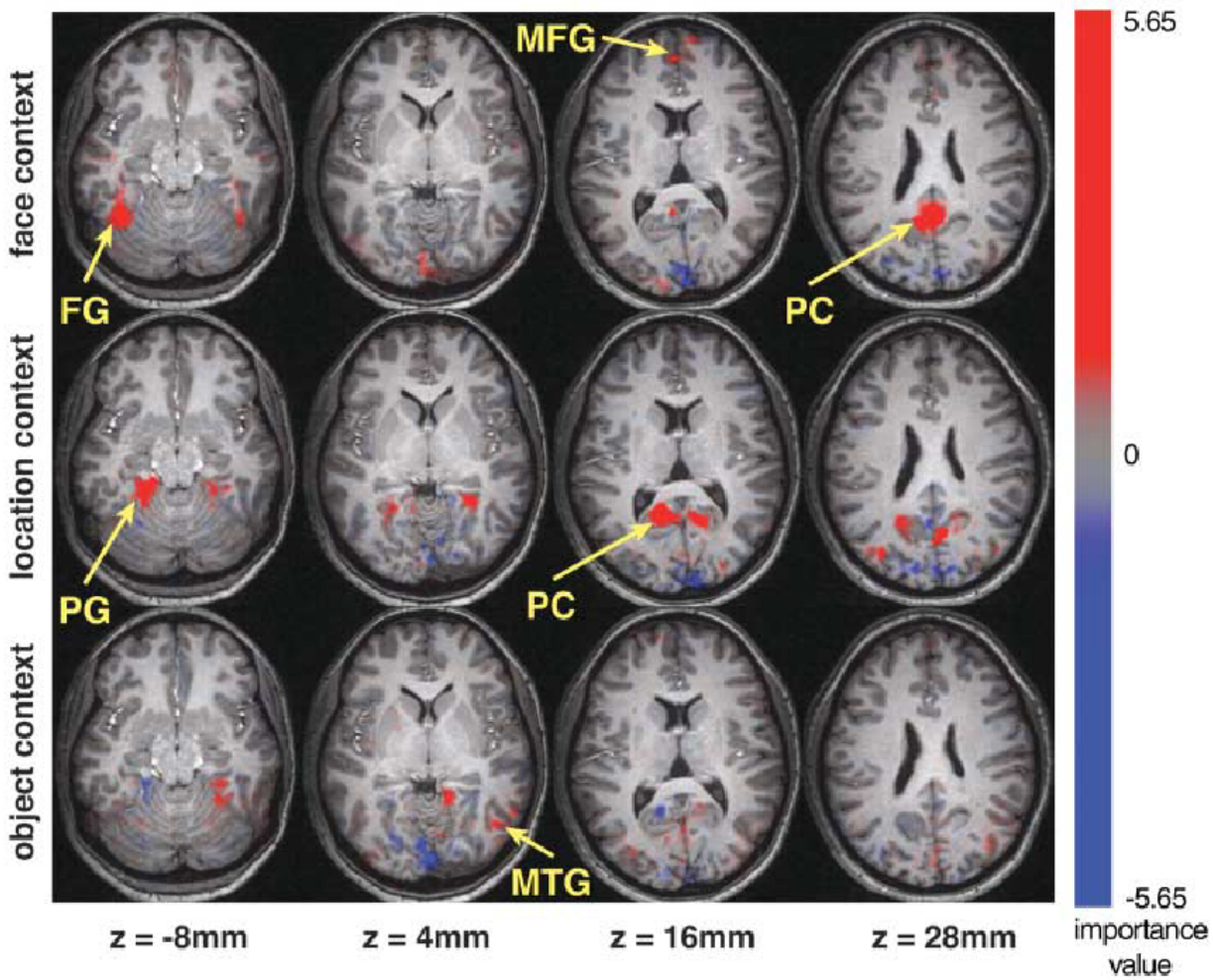
faces | locations | objects



[Polyn et al 2005]

detecting memory context reinstatement

voxel influence on reinstatement estimate



faces

locations

objects

[Polyn et al 2005]

detecting memory context retrieval

key ideas:

- trained classifiers may be used as detectors
- test examples can be temporally overlapping
- it is feasible to decode endogenous events

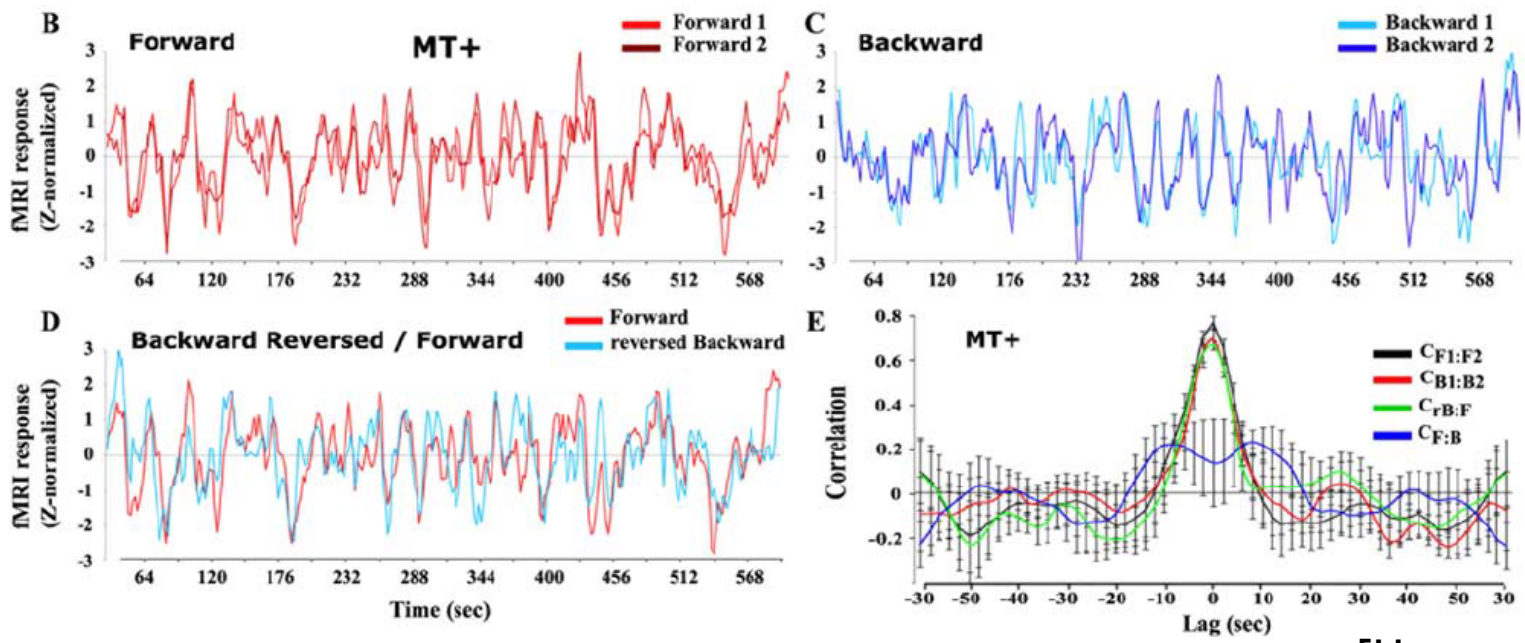
detecting multiple processing time scales

subjects watch silent movie, forward and backward

detecting multiple processing time scales

subjects watch silent movie, forward and backward

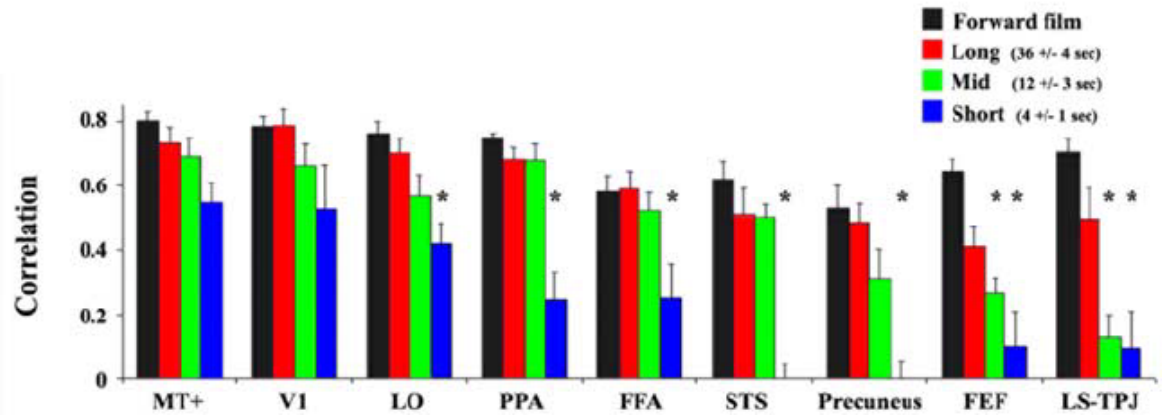
for certain areas, flow of time is irrelevant



[Hasson et al 2008]

detecting multiple processing time scales

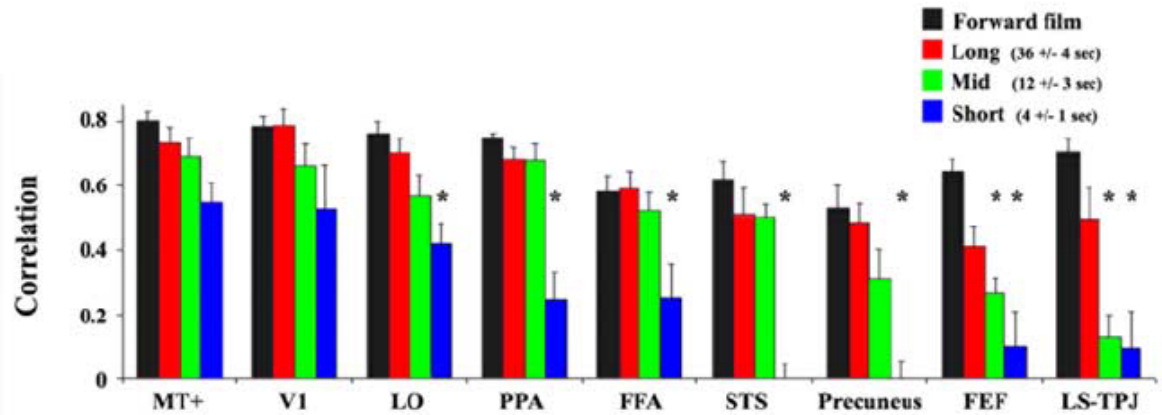
- subjects watch film, scrambled at 3 time scales
- areas exhibit different scale sensitivity

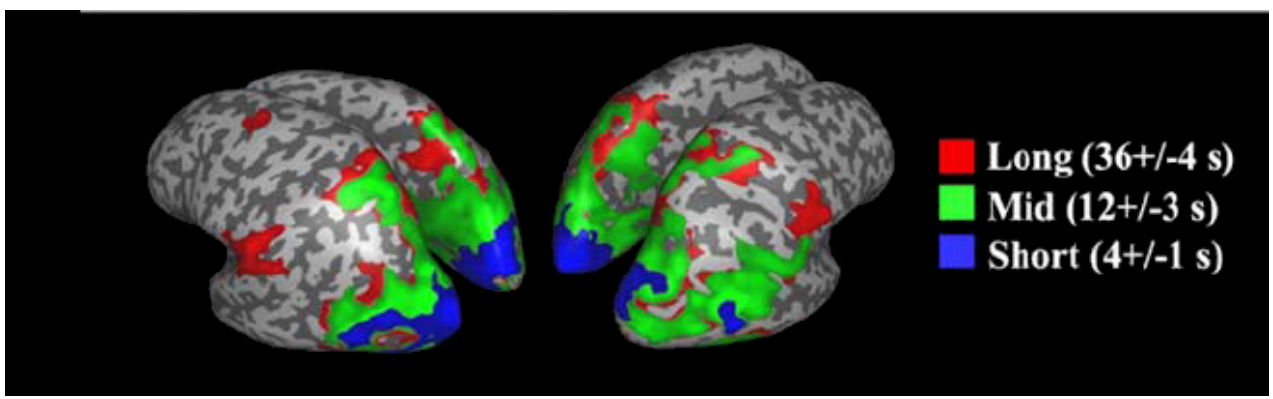


[Hasson et al 2008]

detecting multiple processing time scales

- subjects watch film, scrambled at 3 time scales
- areas exhibit different scale sensitivity





[Hasson et al 2008]

detecting multiple processing time scales

key ideas:

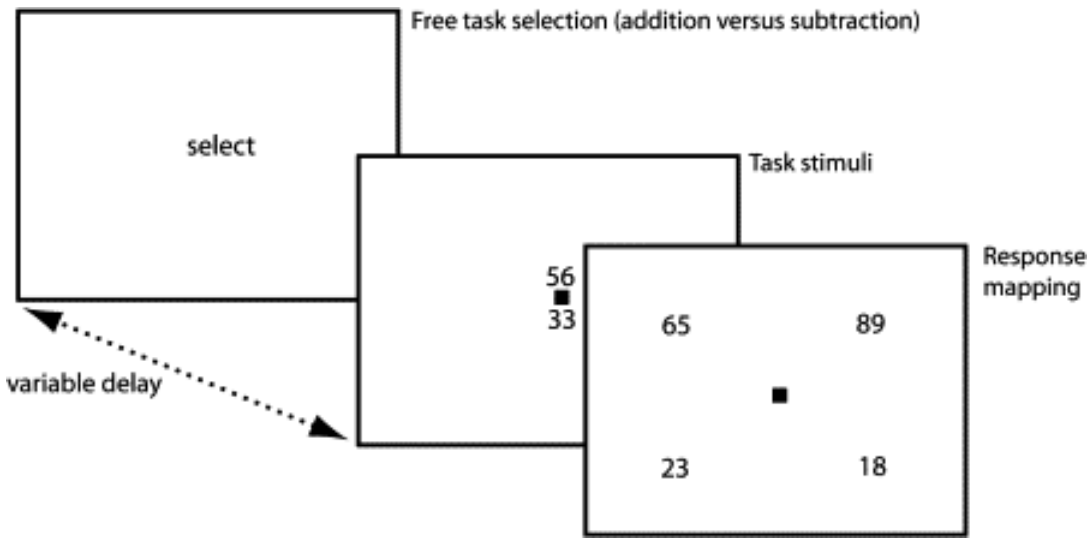
- experiment design tailored to analysis method
- activation happens at multiple scales, generally not all controlled for

predicting everything

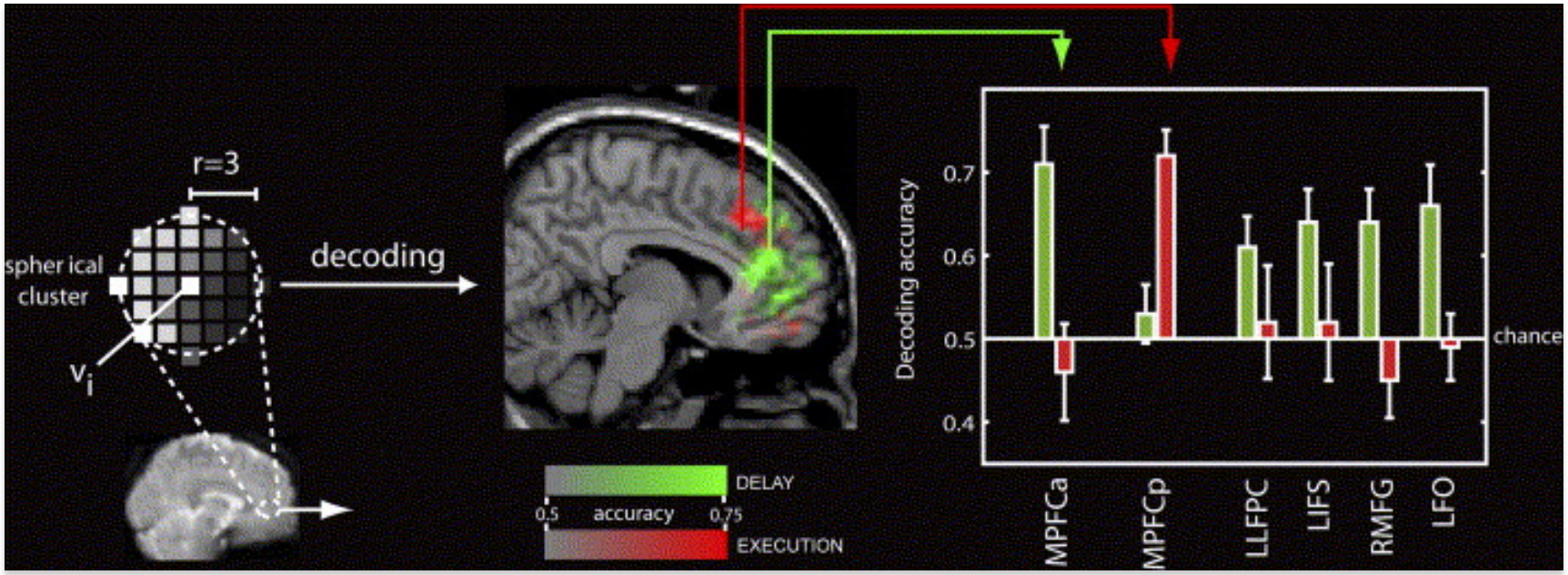
- subjects scanned in a “field anthropology” VR
- subjects have goals directing their actions
- subjects react to what happens to them

predicting everything

- subjects scanned in a “field anthropology” VR
- subjects have goals directing their actions
- subjects react to what happens to them
- prediction examples:
  - doing a particular task
  - picking up an object
  - having a dog in the picture
  - emotional state
  - ...



Train on 7 runs, test on 8th



Haynes et al., 2007, *Current Biology*

Polyn et al.

- Train classifier to detect processes associated with face/location/object study processing
- Use these study classifiers to detect reinstatement during recall test