Using neuroimaging to infer mental states: A guided tour through the minefield

Russell Poldrack

Department of Psychology
Stanford University

Can neuroimaging tell us anything about the mind?

Can neuroimaging tell us anything about the mind?

Max Coltheart

“No amount of knowledge about the hardware of a computer will tell you anything serious about the nature of the software that the computer runs. In the same way, no facts about the activity of the brain could be used to confirm or refute some information-processing model of cognition." (Coltheart, 2004, p. 22)

The New York Times

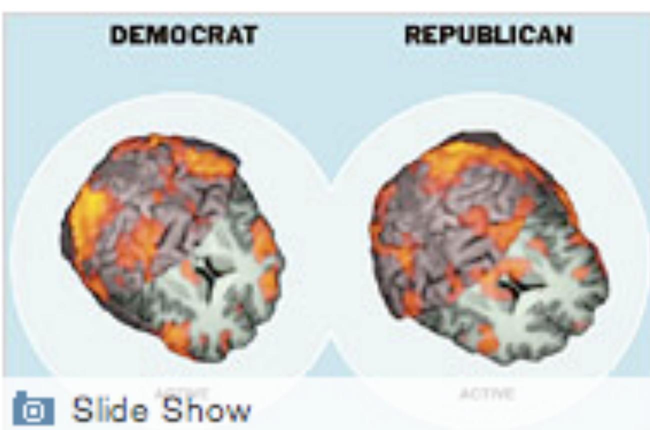
OP-ED CONTRIBUTORS

This Is Your Brain on Politics

Published: November 11, 2007

This article was written by Marco Iacoboni, Joshua Freedman and Jonas Kaplan of the University of California, Los Angeles, Semel Institute for Neuroscience; Kathleen Hall Jamieson of the Annenberg Public Policy Center at the University of Pennsylvania; and Tom Freedman, Bill Knapp and Kathryn Fitzgerald of FKF Applied Research.

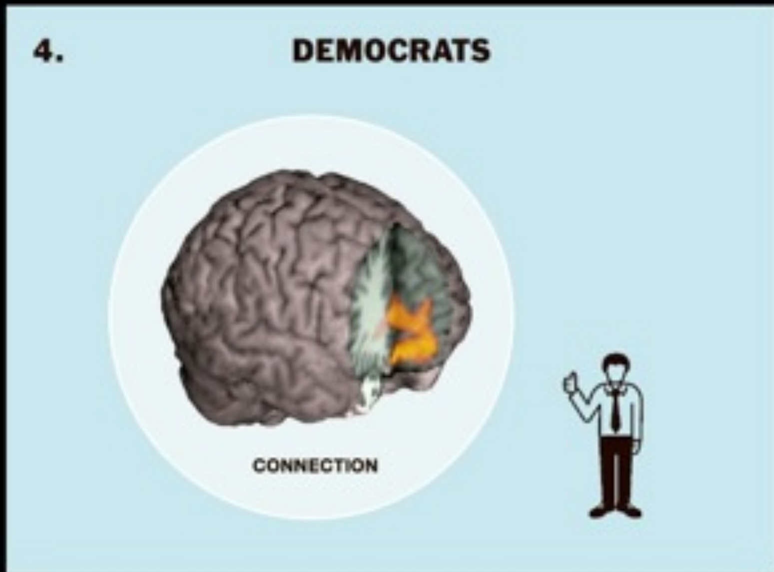
Multimedia



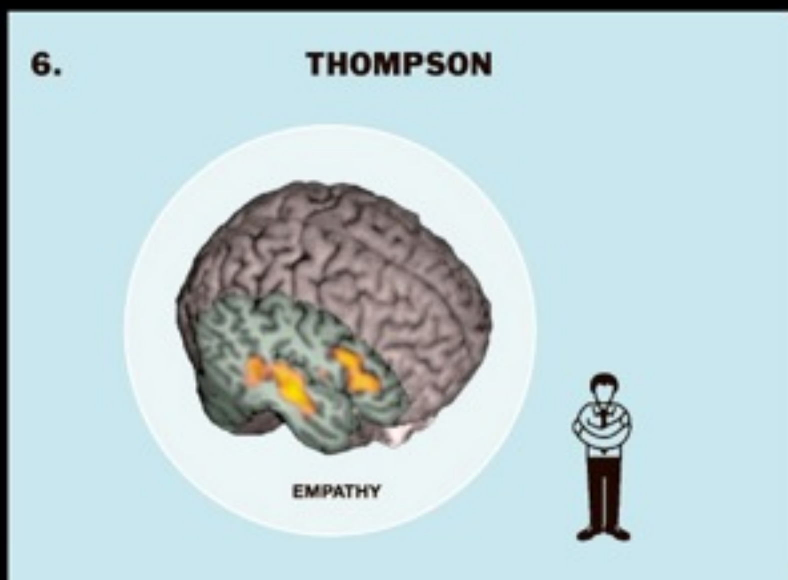
This Is Your Brain on Politics

IN anticipation of the 2008 presidential election, we used functional magnetic resonance imaging to watch the brains of a group of swing voters as they responded to the leading presidential candidates. Our results reveal some voter impressions on which this election may well turn.

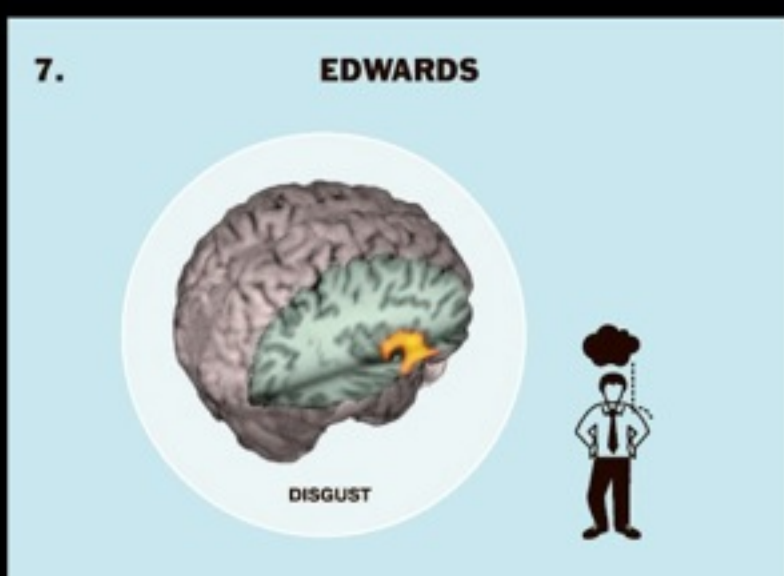
Our 20 subjects — registered voters who stated that they were open to choosing a candidate from either party next November — included 10 men and 10 women. In late summer, we asked them to answer a list of questions about their political preferences, then observed their brain activity for nearly an hour in the scanner at the Ahmanson Lovelace Brain Mapping Center at the University of California, Los Angeles. Afterward, each subject filled out a second questionnaire.



"In response to images of Democratic candidates, men exhibited activity in the medial orbital prefrontal cortex, indicating emotional connection and positive feelings."



"Images of Fred Thompson led to increased activity in the inferior frontal cortex, a brain structure associated with empathy."



"Subjects who had an unfavorable view of John Edwards responded to pictures of him with feelings of disgust, evidenced by increased activity in the insula, a brain area associated with negative emotions."

Do you really love your iPhone?

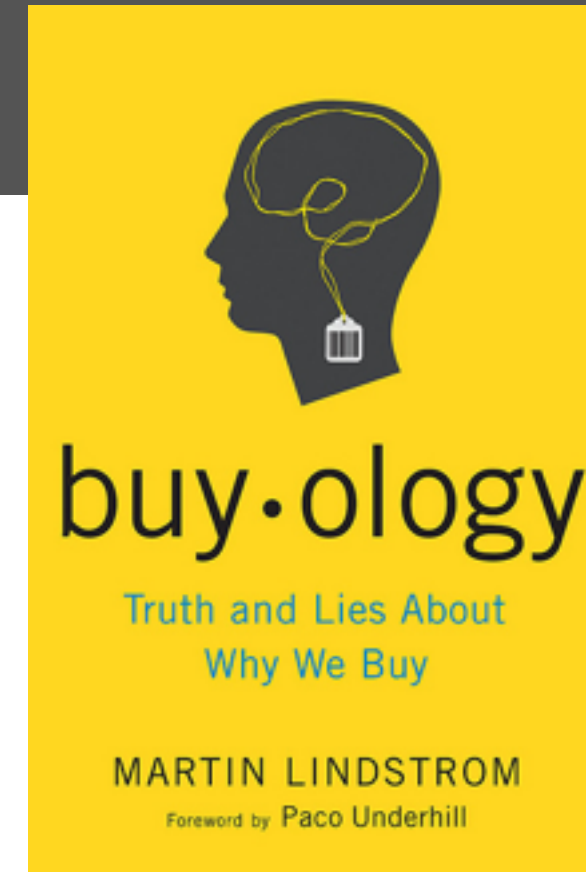
The New York Times

The Opinion Pages

OP-ED CONTRIBUTOR

You Love Your iPhone. Literally.

By MARTIN LINDSTROM
Published: September 30, 2011



- “Earlier this year, I carried out an fMRI experiment to find out whether iPhones were really, truly addictive, no less so than alcohol, cocaine, shopping or video games. In conjunction with the San Diego-based firm MindSign Neuromarketing, I enlisted eight men and eight women between the ages of 18 and 25. Our 16 subjects were exposed separately to audio and to video of a ringing and vibrating iPhone…most striking of all was the flurry of activation in the insular cortex of the brain, which is associated with feelings of love and compassion. The subjects’ brains responded to the sound of their phones as they would respond to the presence or proximity of a girlfriend, boyfriend or family member. In short, the subjects didn’t demonstrate the classic brain-based signs of addiction. Instead, they loved their iPhones.

To the Editor:

"You Love Your iPhone. Literally," by Martin Lindstrom (Op-Ed, Oct. 1), purports to show, using brain imaging, that our attachment to digital devices reflects not addiction but instead the same kind of emotion that we feel for human loved ones.

However, the evidence the writer presents does not show this.

The brain region that he points to as being "associated with feelings of love and compassion" (the insular cortex) is active in as many as one-third of all brain imaging studies.

Further, in studies of decision making the insular cortex is more often associated with negative than positive emotions.

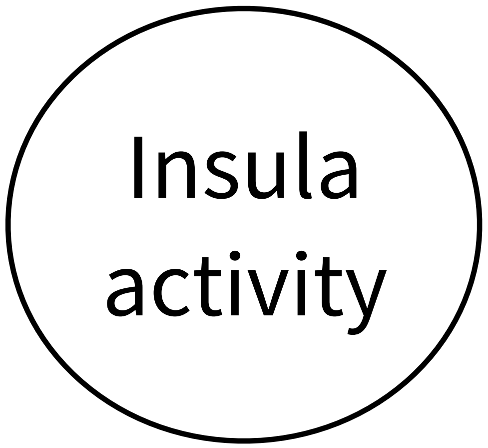
The kind of reasoning that Mr. Lindstrom uses is well known to be flawed, because there is rarely a one-to-one mapping between any brain region and a single mental state; insular cortex activity could reflect one or more of several psychological processes.

We find it surprising that The Times would publish claims like this that lack scientific validity.

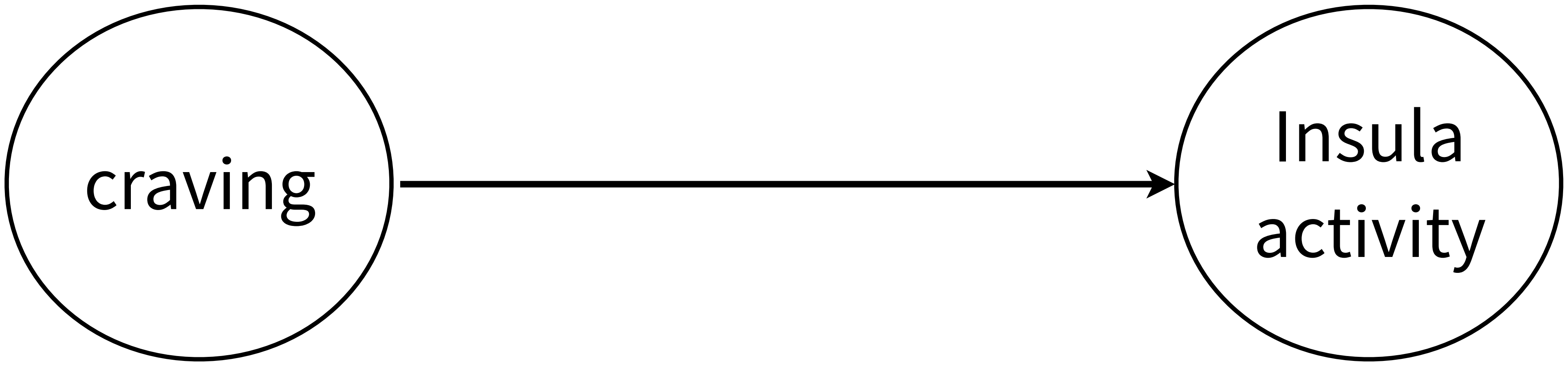
RUSSELL POLDRACK
Austin, Tex., Oct. 3, 2011

The writer is a professor of psychology and neurobiology at the University of Texas at Austin. His letter was signed by 44 other neuroscientists.

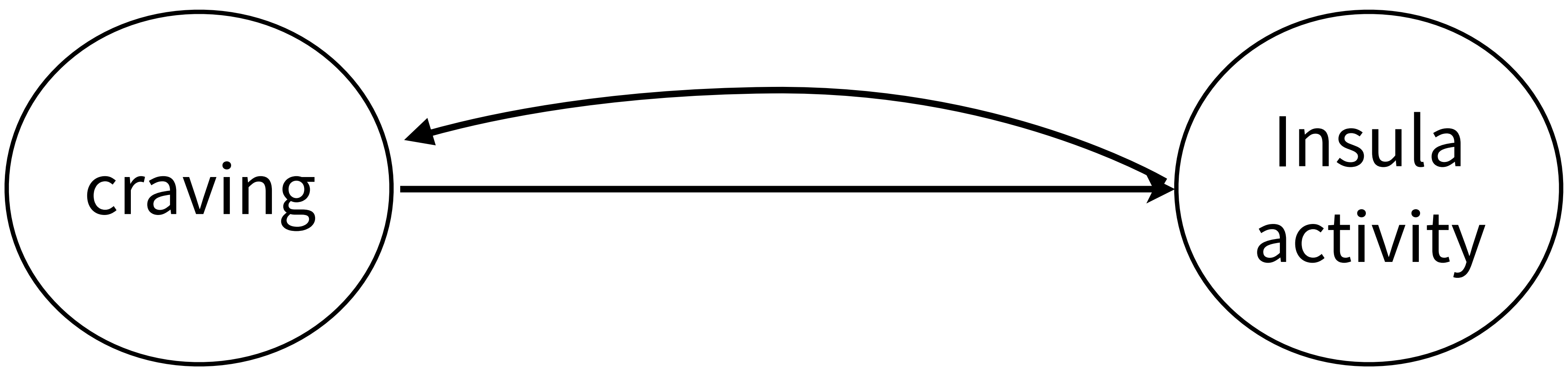
Does reverse inference work?



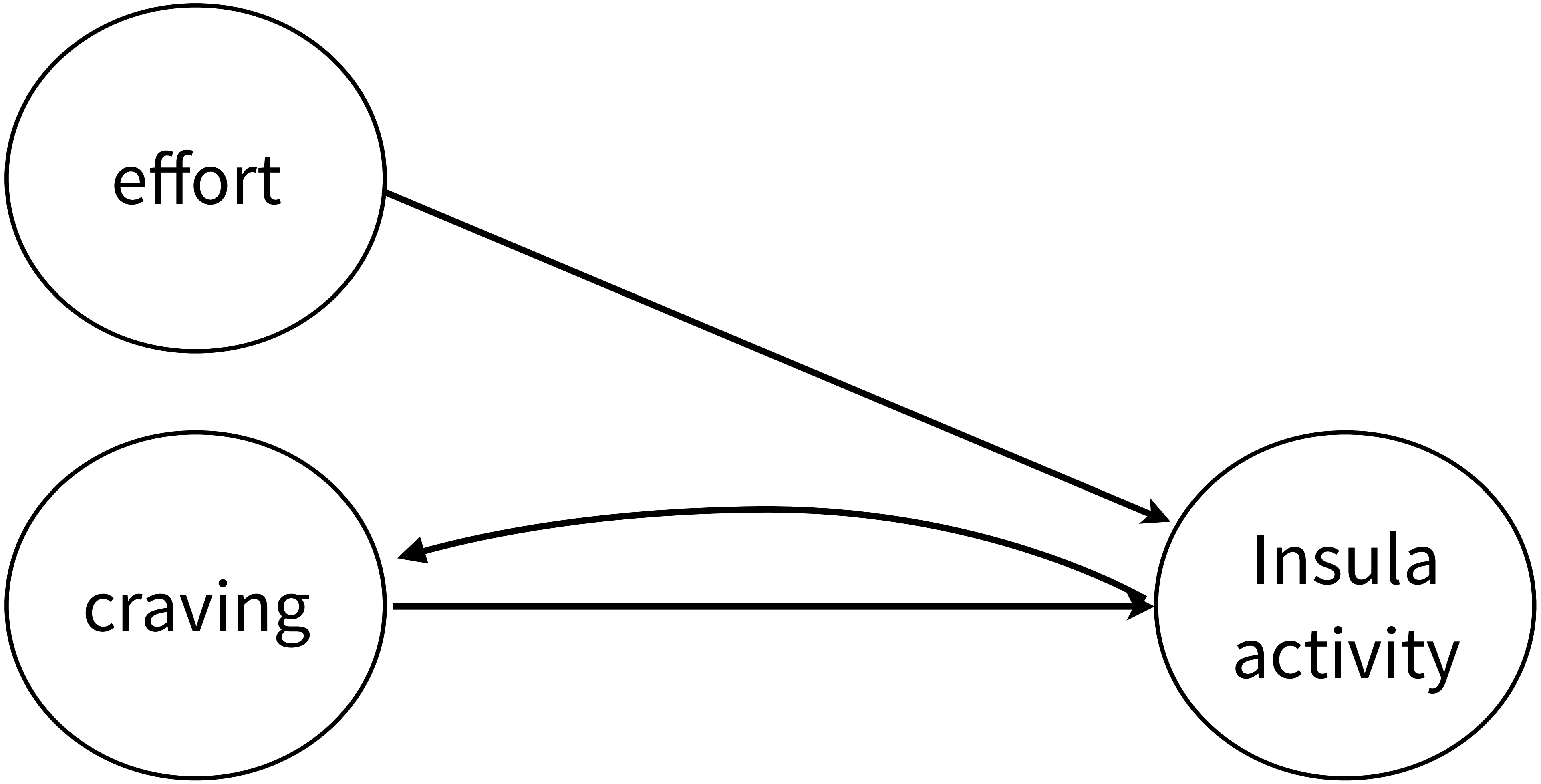
Does reverse inference work?



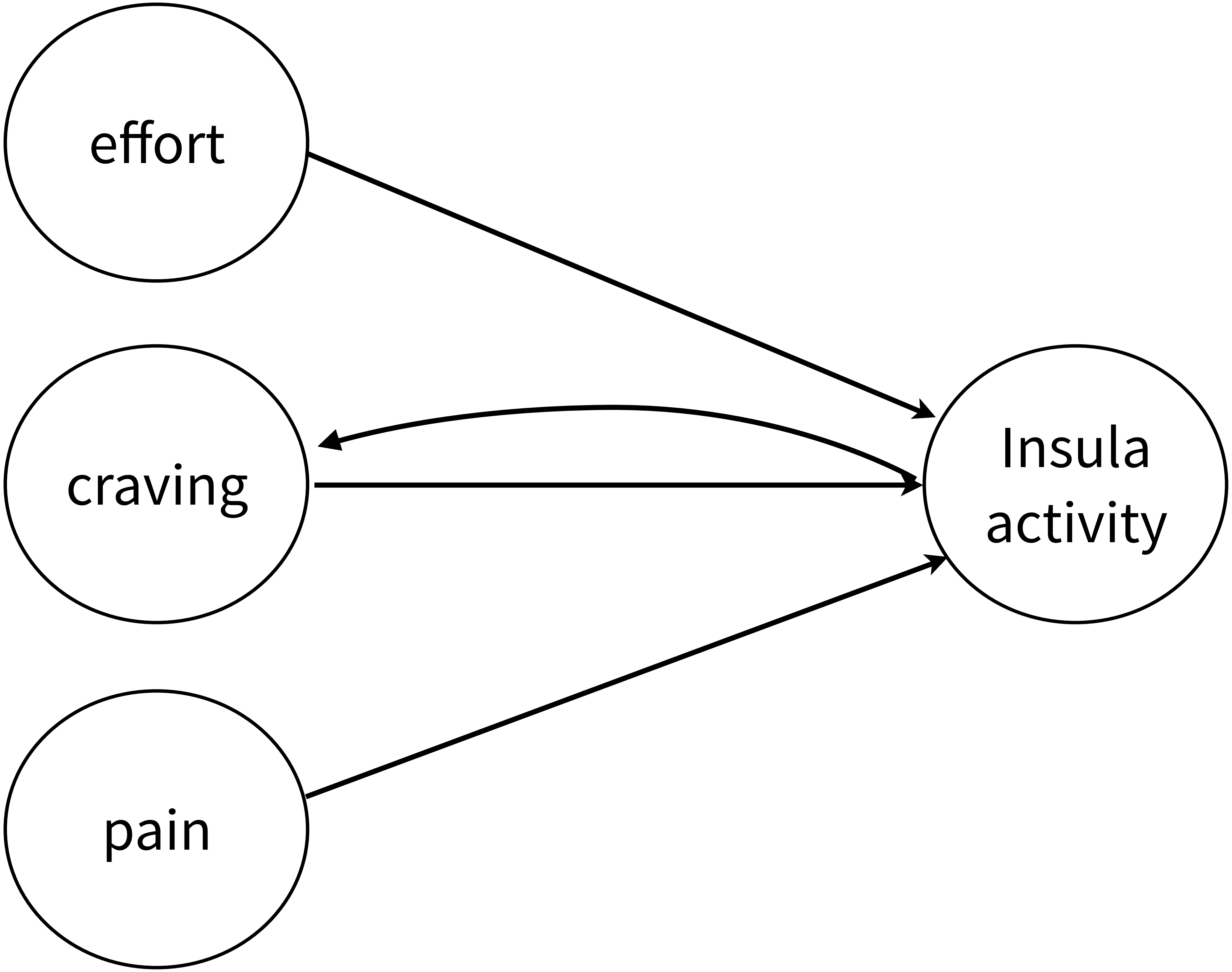
Does reverse inference work?



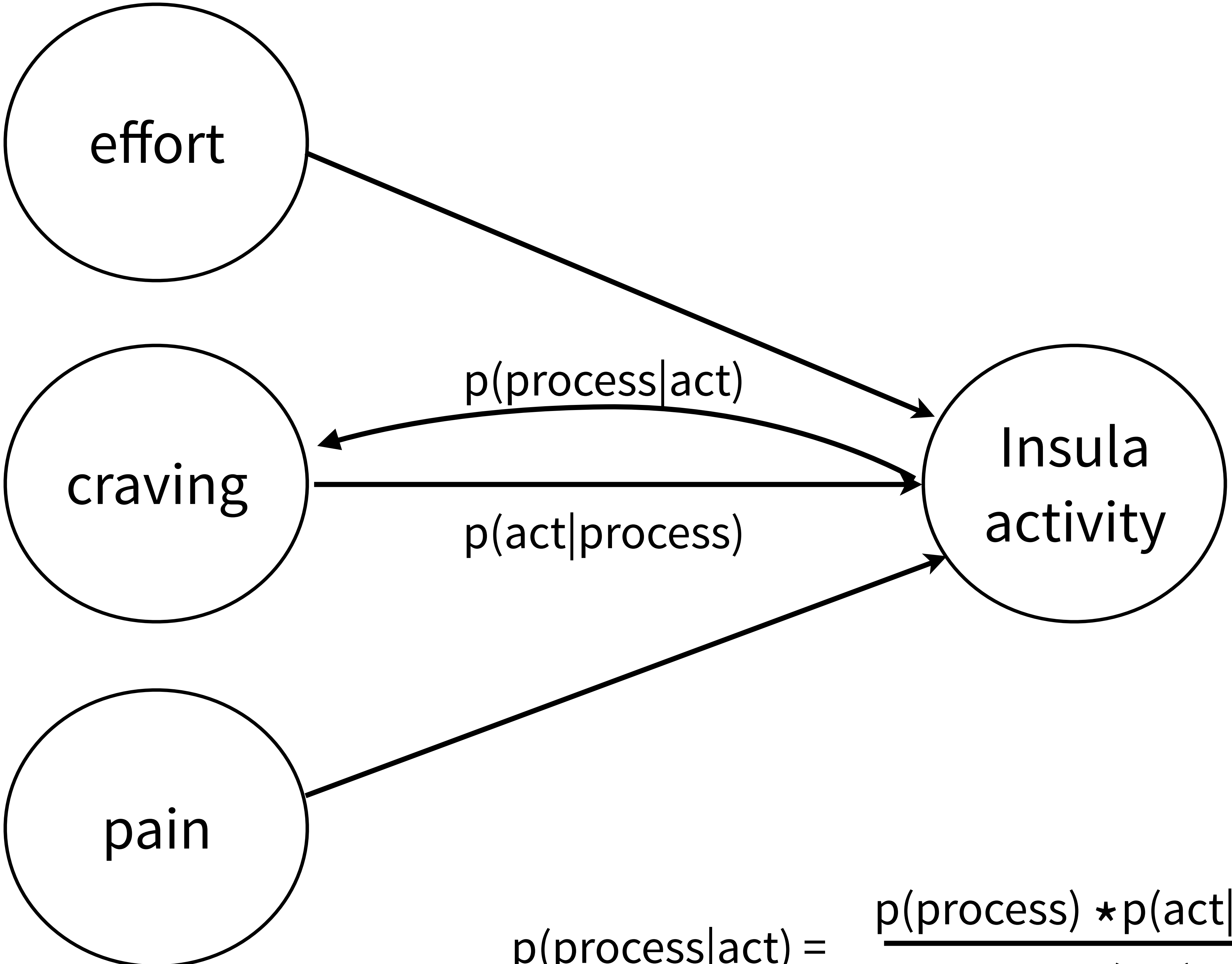
Does reverse inference work?



Does reverse inference work?

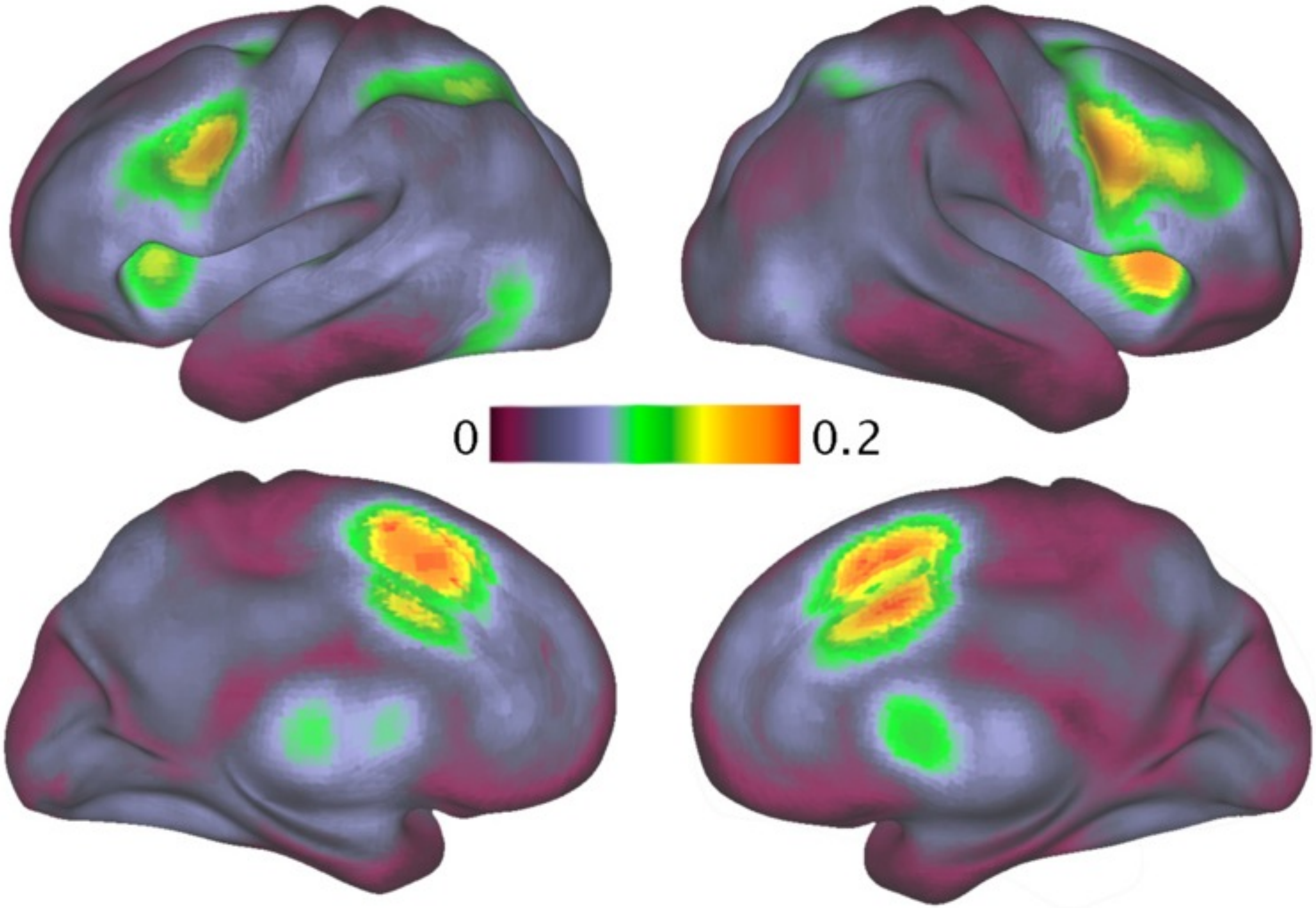


Does reverse inference work?



$$p(\text{process}|\text{act}) = \frac{p(\text{process}) * p(\text{act}|\text{process})}{p(\text{act})}$$

Insula activation is weakly selective



Some voxels active in more than 20% of studies

Yarkoni et al., 2011

Reverse inference

- Informal reverse inference provides relatively weak evidence

Opinion TRENDS in Cognitive Sciences Vol.10 No.2 February 2006

Can cognitive processes be inferred from neuroimaging data?

Russell A. Poldrack

Department of Psychology and Brain Research Institute, UCLA, Los Angeles, CA 90095-1563, USA

There is much interest currently in using functional neuroimaging techniques to understand better the nature of cognition. One particular practice that has become common is 'reverse inference', by which the engagement of a particular cognitive process is inferred from the activation of a particular brain region. Such inferences are not deductively valid, but can still provide some information. Using a Bayesian analysis of the BrainMap neuroimaging database, I characterize the amount of additional evidence in favor of the engagement of a cognitive process that can be offered by a reverse inference. Its usefulness is particularly limited by the selectivity of activation in the region of interest. I argue that cognitive neuroscientists should be circumspect in the use of reverse inference, particularly when selectivity of the region in question cannot be established or is known to be weak.

Introduction

Functional neuroimaging techniques such as functional magnetic resonance imaging (fMRI) provide a measure of local brain activity in response to cognitive tasks undertaken during scanning. These data allow the cognitive neuroscientist to infer something about the role of particular brain regions in cognitive function. However, there is increasing use of neuroimaging data to make the opposite inference; that is, to infer the engagement of particular cognitive functions based on activation in particular brain regions. My goal here is to analyze this practice, known as 'reverse inference', and to characterize some limitations on the effectiveness of this strategy. The companion paper in this issue by Henson [1] discusses a complementary strategy for using neuroimaging to distinguish competing cognitive theories.

The goal of cognitive psychology is to understand the underlying mental architecture that supports cognitive functions. To this end, cognitive psychologists examine the effects of task manipulations on behavioral variables, such as response time or accuracy, and use these data to test models of cognitive function. However, it is often not possible to determine on the basis of behavioral variables alone whether a particular cognitive process is engaged, or whether a particular theory of cognitive architecture is correct; for example, there are well-known examples of theoretical indeterminacy based on behavioral data [2]. If neuroimaging were able to provide information regarding what cognitive processes were engaged in performance of a particular task, cognitive psychologists would have gained a powerful new tool. Researchers outside cognitive psychology are also sometimes interested in using neuroimaging to determine the engagement of particular cognitive processes. For example, philosophers might wish to know the degree to which emotion versus deliberative reasoning plays a role in moral judgments [3].

Inference in neuroimaging

The usual kind of inference that is drawn from neuroimaging data is of the form '*if cognitive process X is engaged, then brain area Z is active*'. Perusal of the discussion sections of a few fMRI articles will quickly reveal, however, an epidemic of reasoning taking the following form:

(1) In the present study, when task comparison A was presented, brain area Z was active.
(2) In other studies, when cognitive process X was putatively engaged, then brain area Z was active.
(3) Thus, the activity of area Z in the present study demonstrates engagement of cognitive process X by task comparison A.

This is a 'reverse inference', in that it reasons backwards from the presence of brain activation to the engagement of a particular cognitive function.

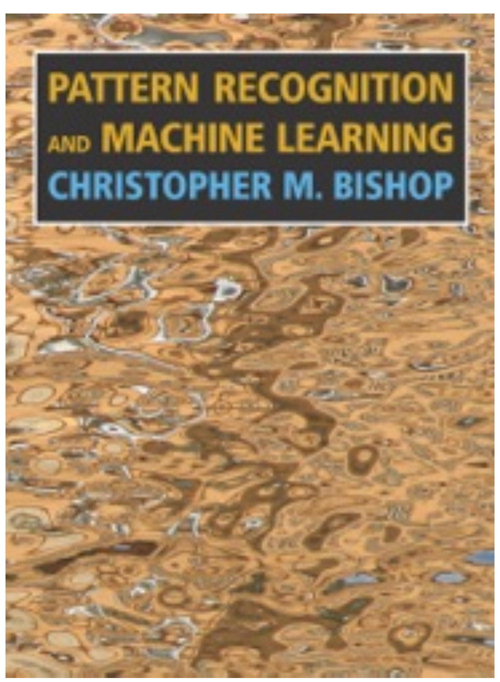
In many cases the use of reverse inference is informal; the presence of unexpected activation in a particular region is explained by reference to other studies that found activation in the same region. However, in some studies the reverse inference is a central feature. In one study [4], subjects were scanned using PET while they performed an economic exchange task in which they had the chance to punish those who defected. Activation was observed in the dorsal striatum when participants subjected defectors to effective punishment; this activation was inferred to reflect the rewarding properties of altruistic punishment. Similarly, a study using fMRI in rats [5] compared activity during pup suckling versus cocaine administration. Greater activity in the dorsal and ventral striatum during suckling compared with cocaine administration led the authors to conclude that 'pup suckling is more rewarding than cocaine' (p. 149). In each of these studies, a cognitive process ('reward') was inferred from activation in a particular brain system (the striatum). Nearly every

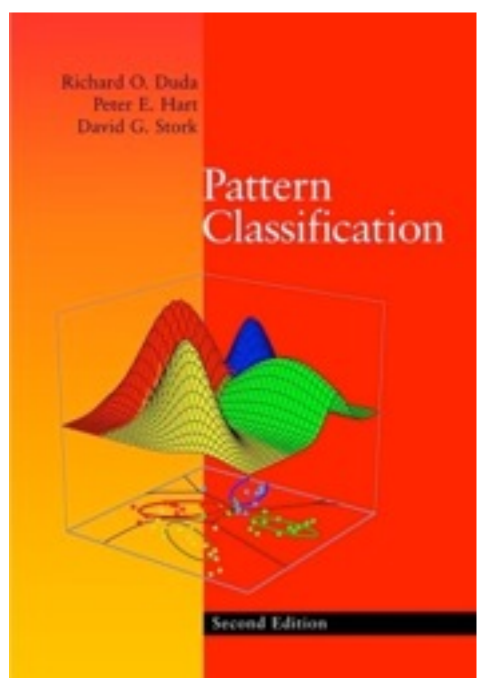
Corresponding author: Poldrack, R.A. (poldrack@ucla.edu).
Available online 6 January 2006

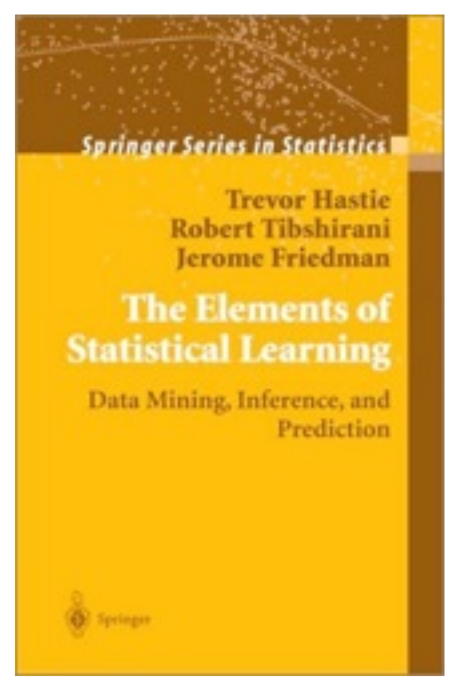
TICS, 2006

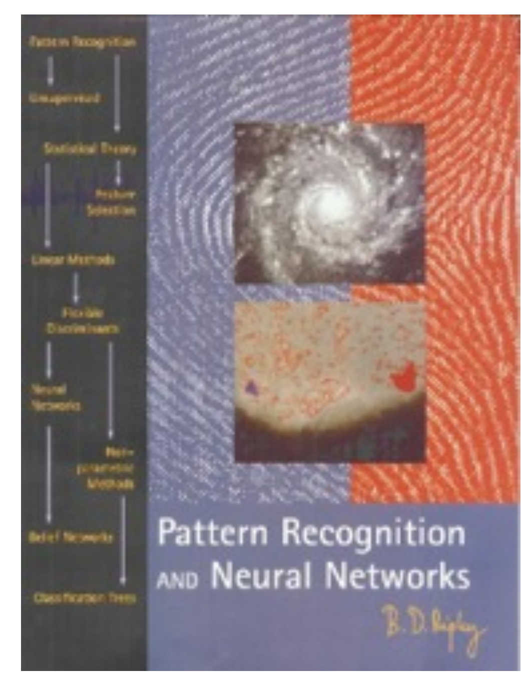
Formalizing reverse inference

- How can we more formally test the predictive ability of fMRI?
- Answer: statistical methods for prediction
  - Machine learning/statistical learning/pattern recognition

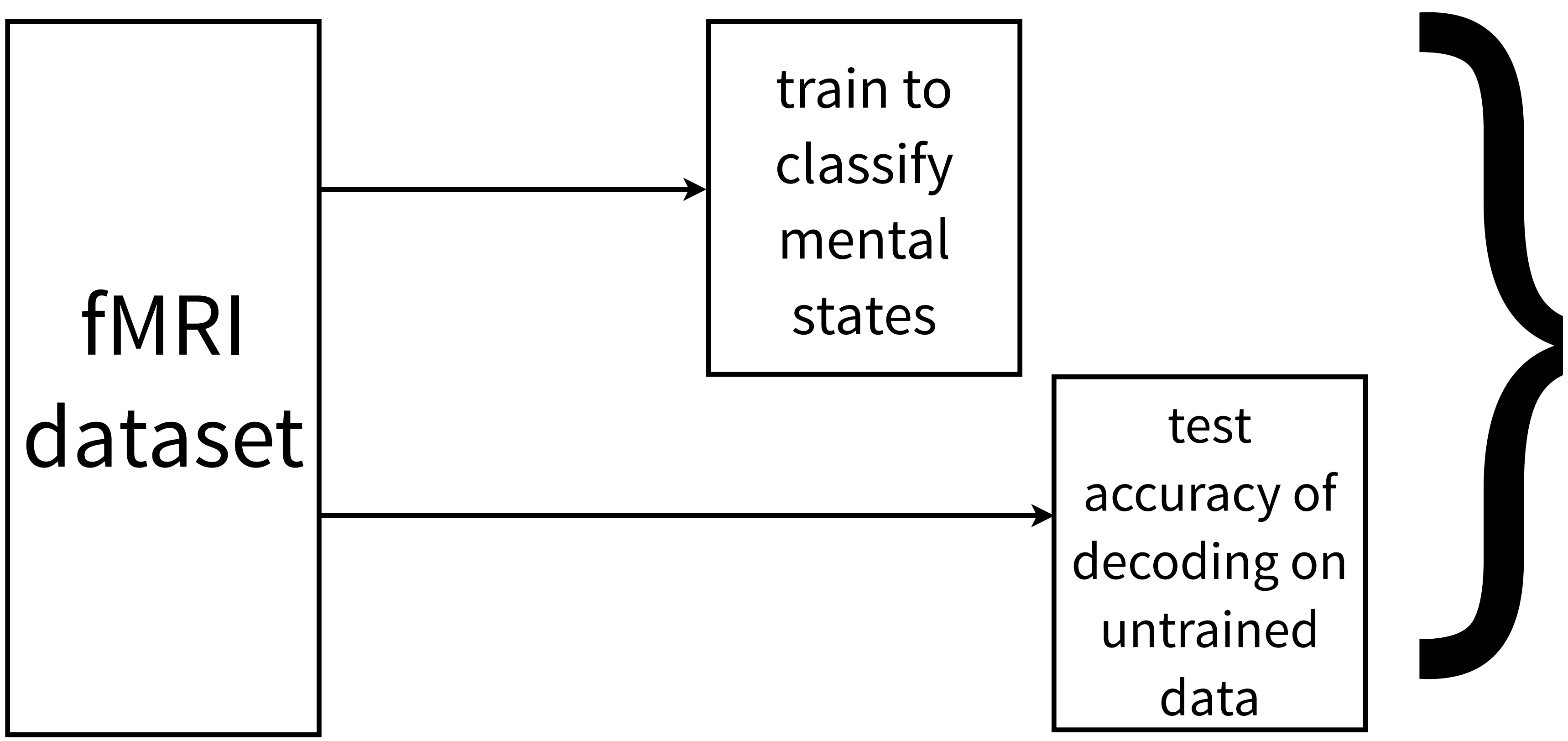






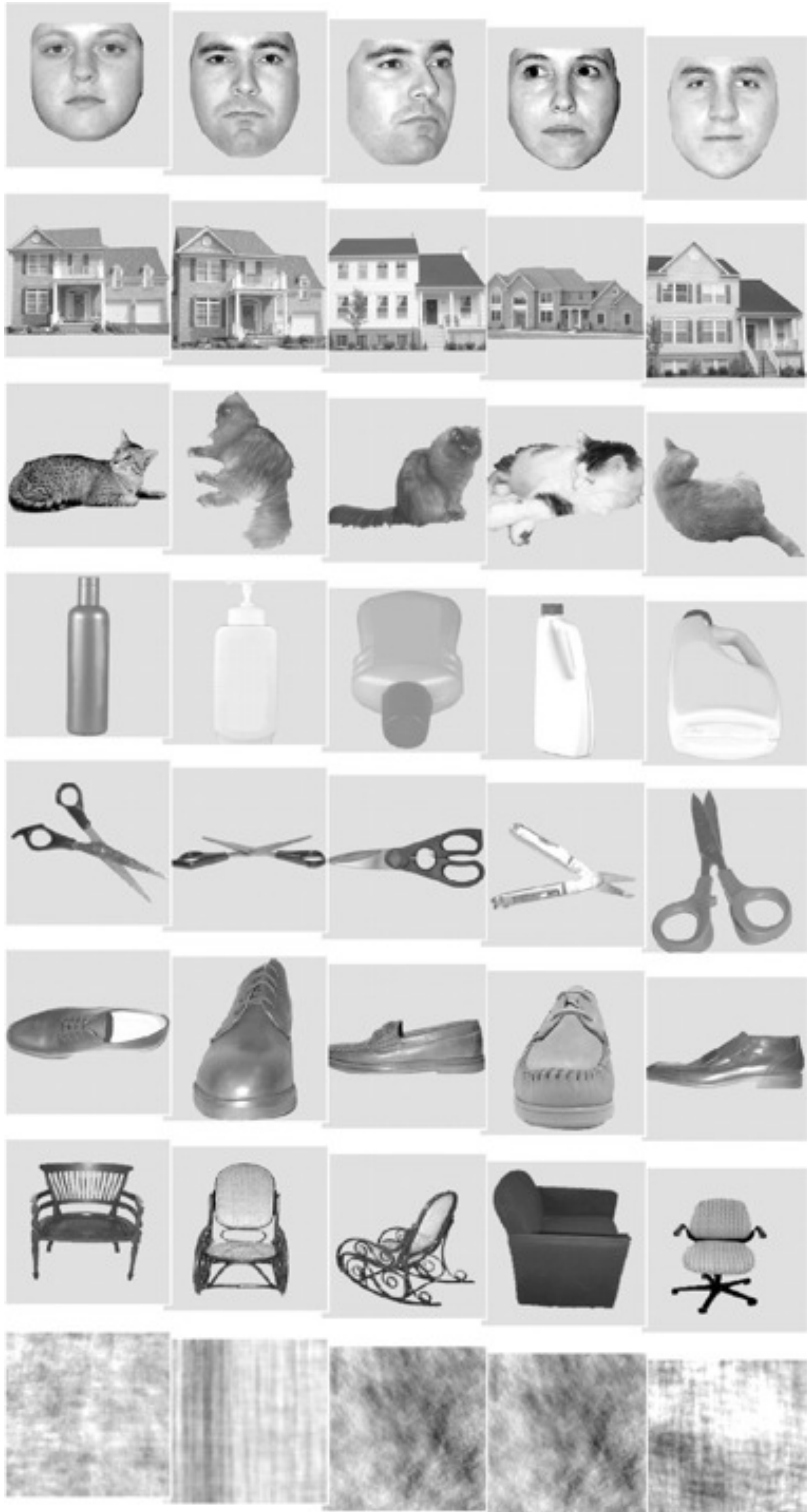


Decoding mental states using machine learning

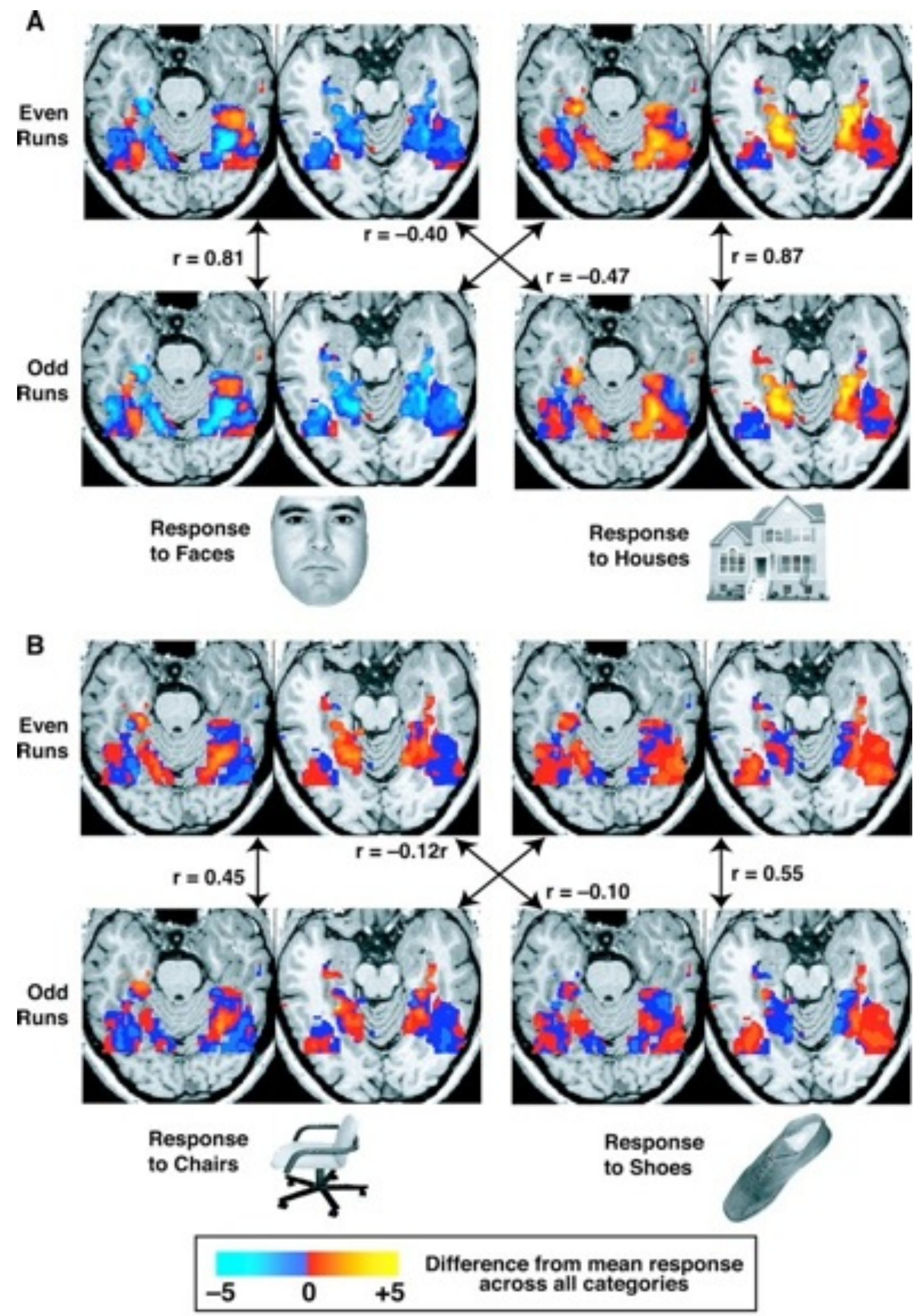


Cross-validation:

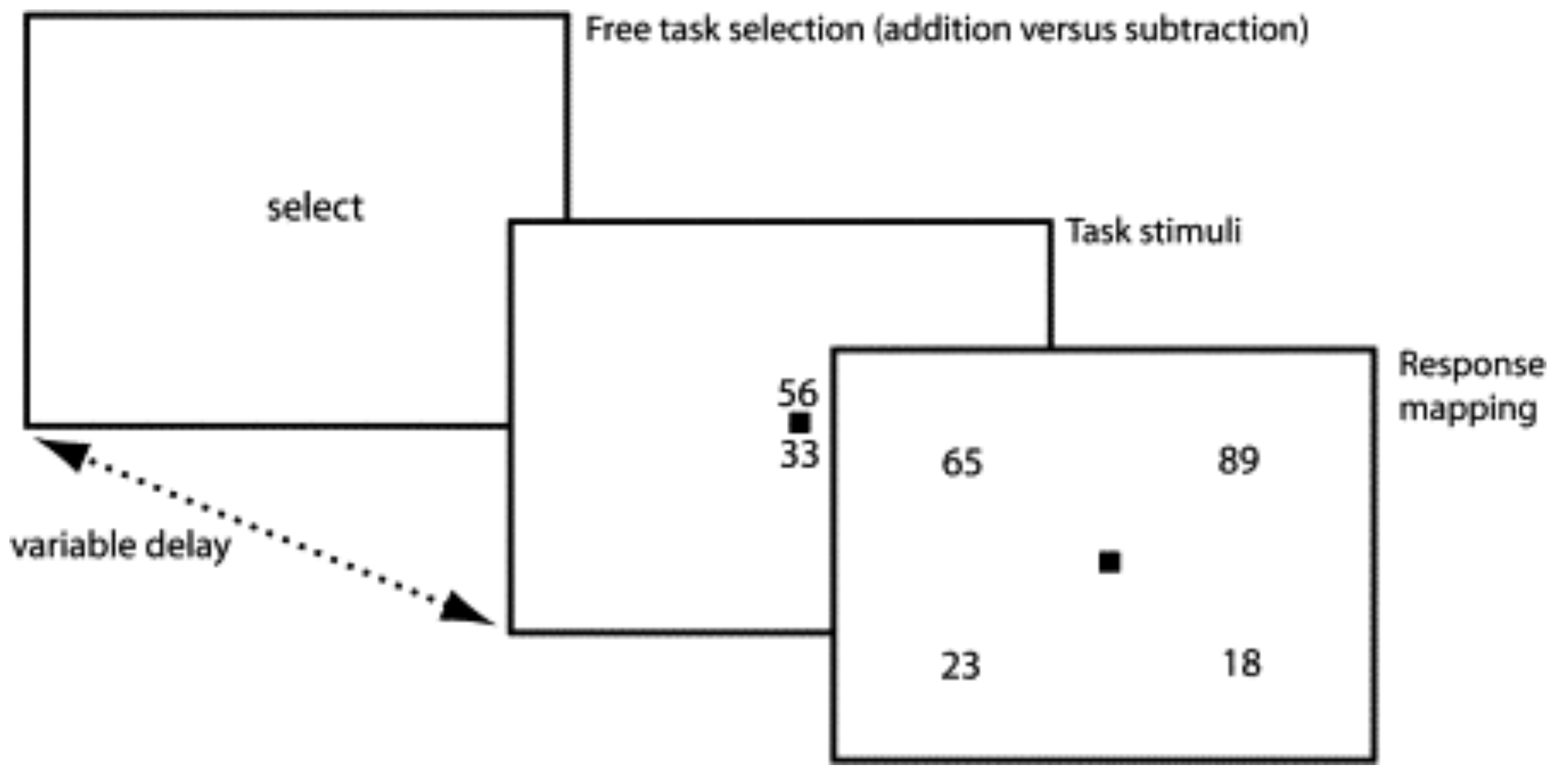
- Train for each split of size k
- Compute average predictive accuracy on left-out data



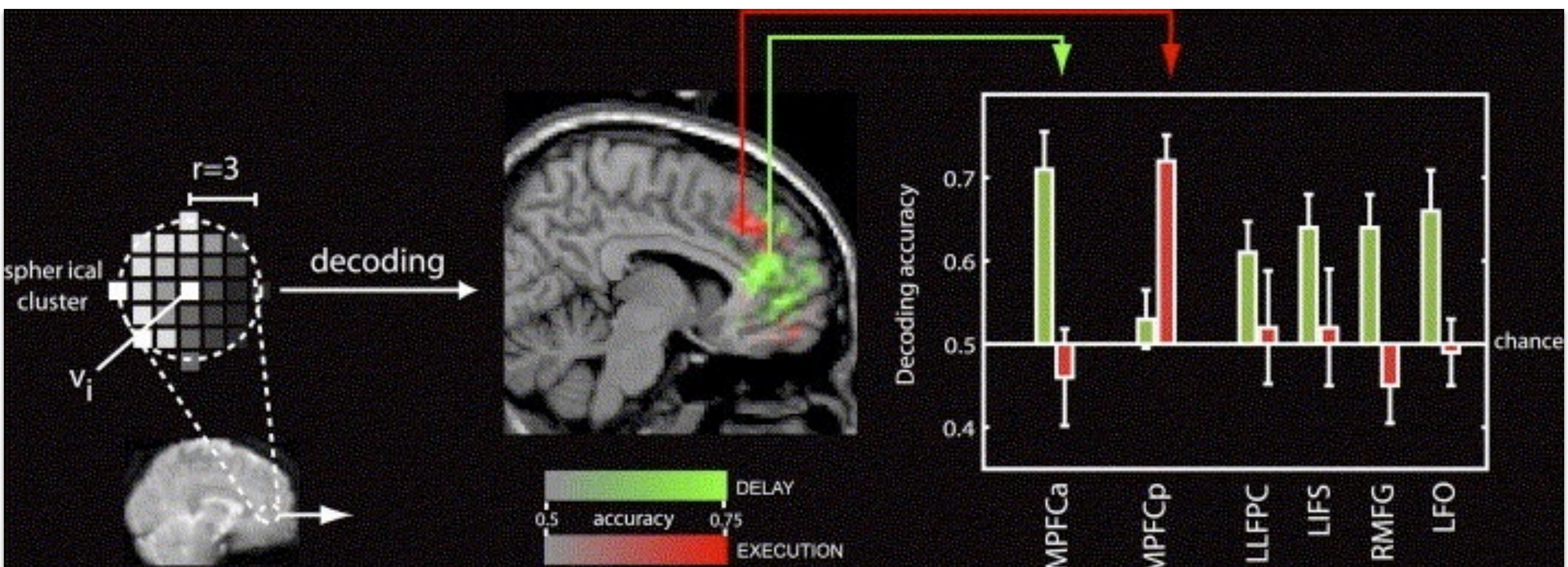
96% correct classification



Haxby et al., 2001, *Science*



Train on 7 runs, test on 8th



Haynes et al., 2007, *Current Biology*

Mind-reading machine knows what you see

15:26 25 April 2005
NewScientist.com news service

SCIENTIFIC AMERICAN.COM

April 25, 2005

Brain Scans Helps Scientists "Read" Minds

BBC NEWS

OPEN The News in 2 minutes

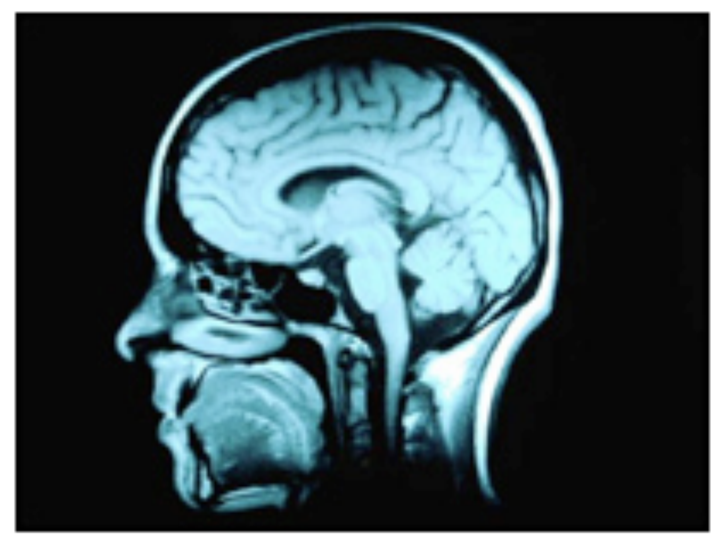
News Front Page
Africa
Americas
Asia-Pacific
Europe
Middle East
South Asia
UK
Business
Health
Medical notes
Science/Nature
Technology
Entertainment

Last Updated: Monday, 25 April, 2005, 00:05 GMT 01:05 UK

E-mail this to a friend   Printable version

Brain scan 'sees hidden thoughts'

Scientists say they can read a person's unconscious thoughts using a simple brain scan.



The scan picks up subliminal thought activity

Functional MRI scans plot brain activity by looking at brain blood flow and are already used by researchers.

A team at University College London found with fMRI they could tell what a person was thinking deep down even when the individual was unaware themselves.

“60 Minutes”, January 4, 2009

“60 Minutes”, January 4, 2009

“60 Minutes”, January 4, 2009

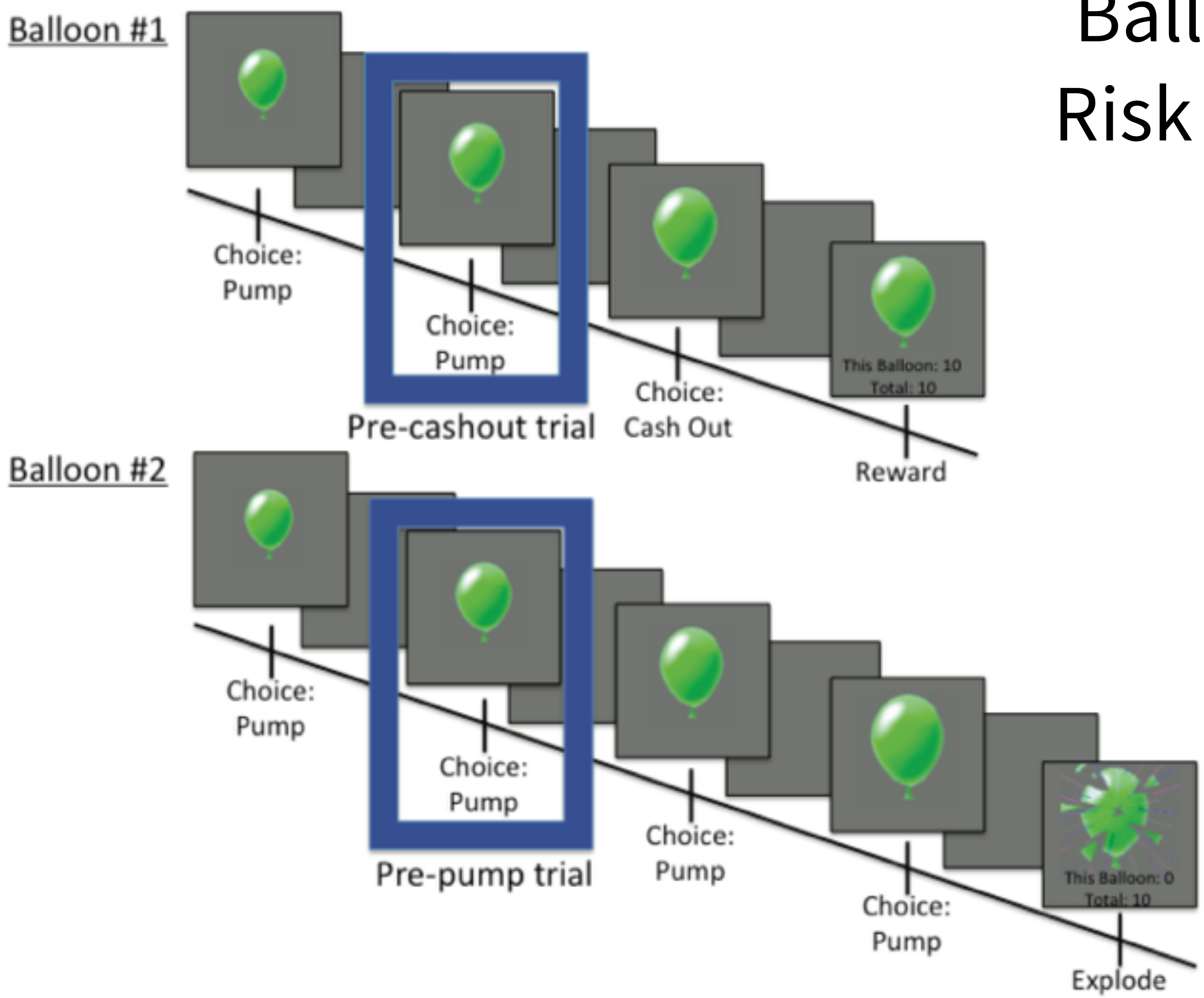
“It's tough to make predictions, especially about the future.” - Yogi Berra

Predicting mental states across people

- Existing work has primarily examined ability to predict mental states using a classifier trained on data from the same person
- For many applications of interest, such training data would not exist for the individual being tested
- Can we accurately generalize to new individuals?

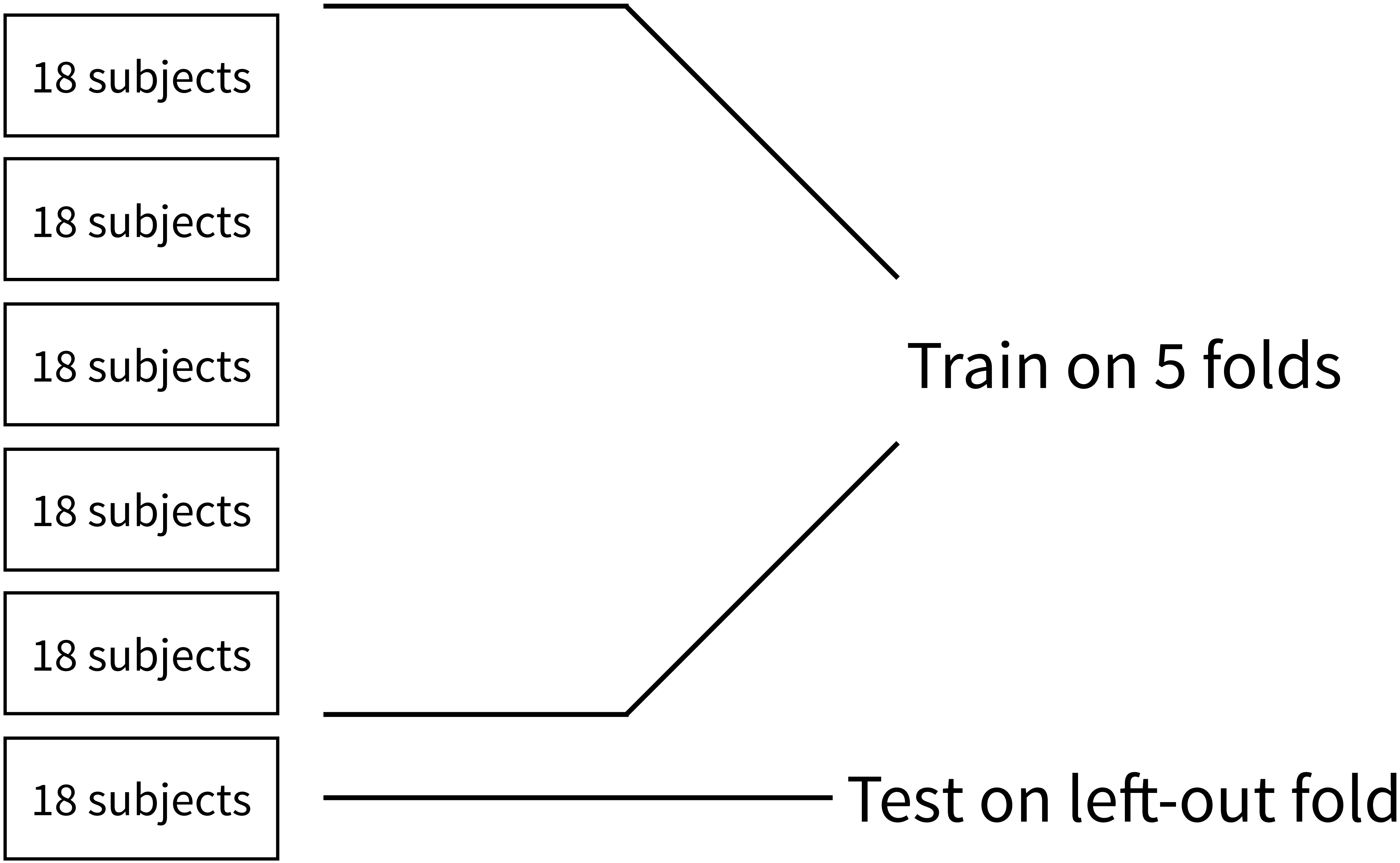
Predicting risky decisions

Balloon Analog Risk Task (BART)



Helfinstein et al, 2014, PNAS

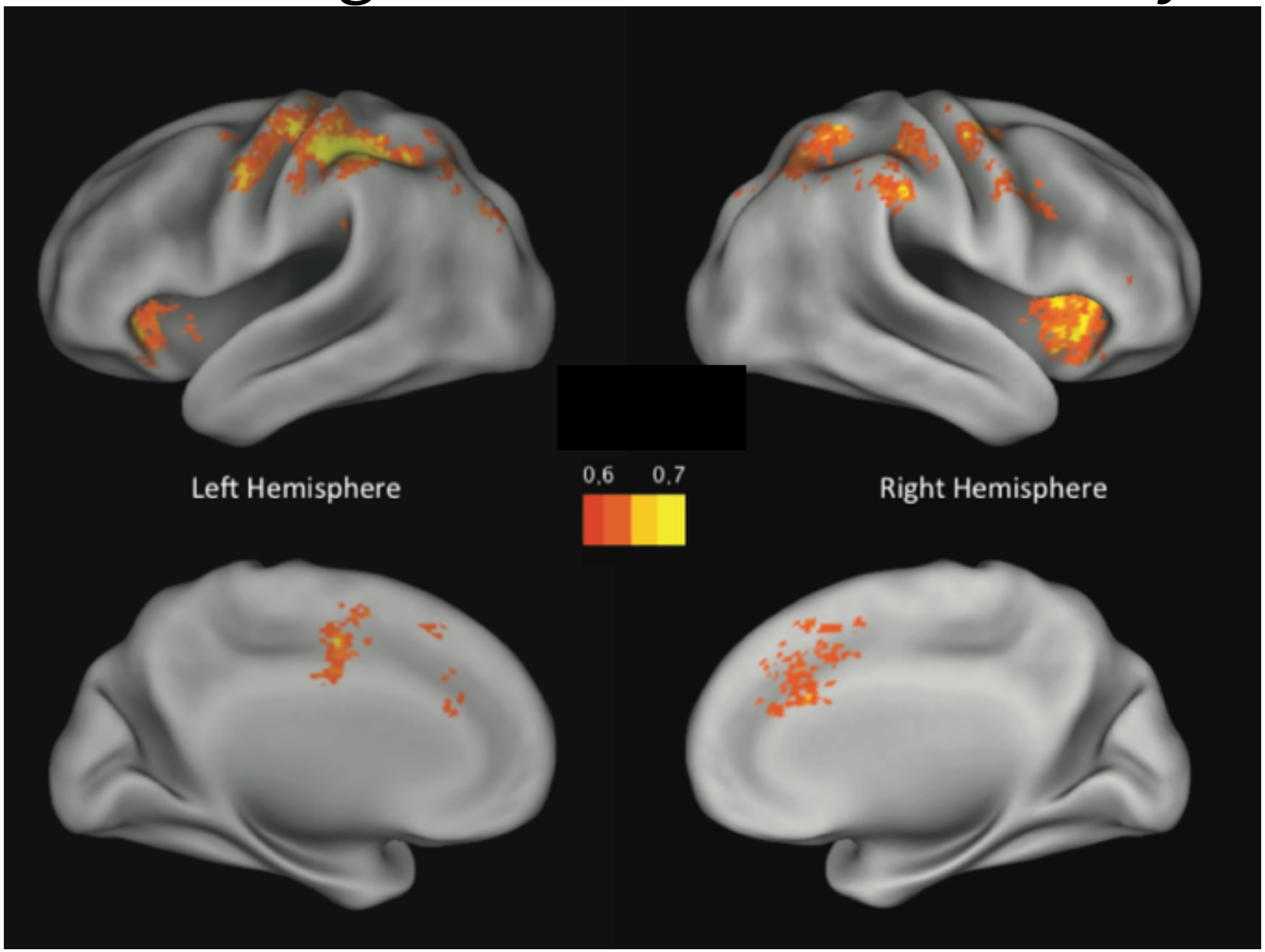
Crossvalidation across subjects



Randomly assign to folds 50 times and average results

Helfinstein et al, 2014, PNAS

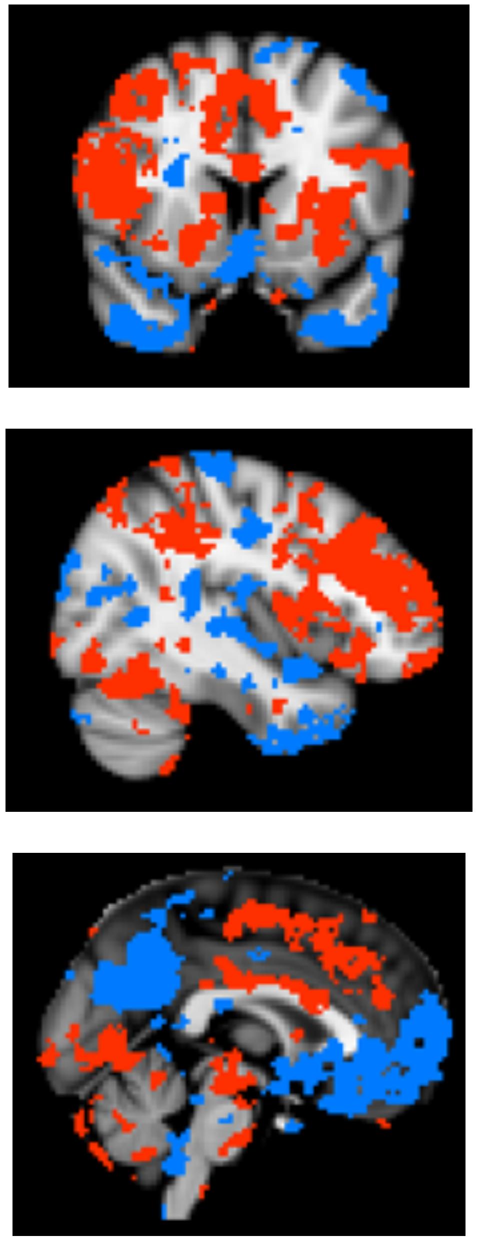
Searchlight classification accuracy

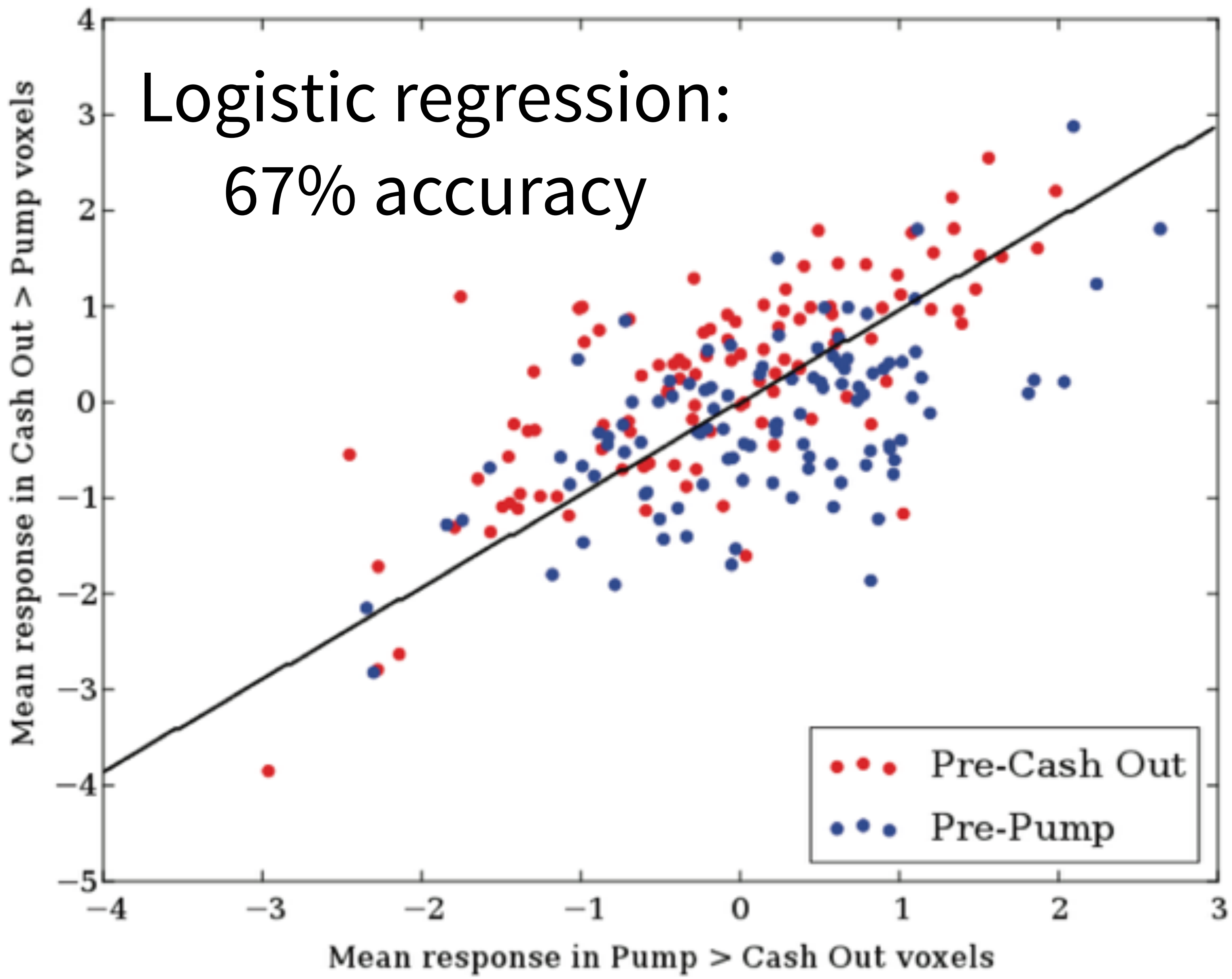


Whole-brain classification:
72%
p<0.002 under null hypothesis
(by randomization)

Helfinstein et al, 2014, PNAS

Classifying based on activity balance





Blue: Pre-Pump > Pre-Cashout
Red: Pre-Cashout > Pre-Pump

Helfinstein et al, 2014, PNAS

Decoding different mental states

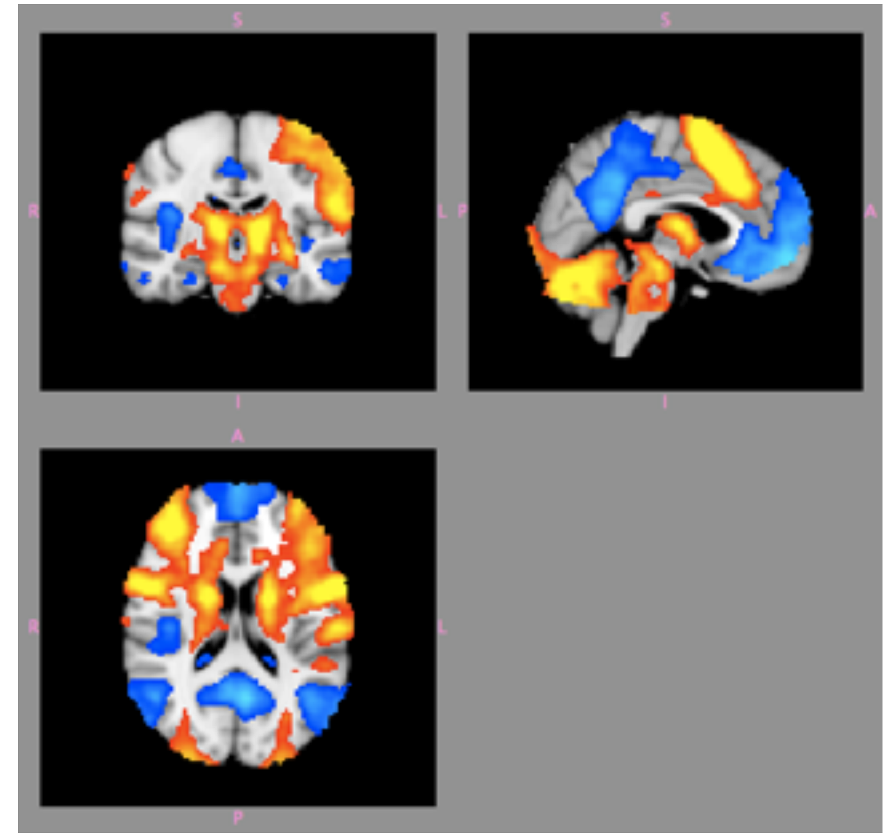
- Can we predict what task a subject was performing, using a classifier trained on other people?
  - 8 tasks, 130 subjects total

Task #	Task description	# subjects	Design type
1	Risky decision making (Balloon analog risk task) (Stover et al., in preparation)	16	Event-related
2	Probabilistic classification (no feedback) (Aron et al., unpublished)	20	Event-related
3	Rhyme judgments on pseudowords (Xue et al., unpublished)	13	Event-related
4	Working memory (tone counting) (Foerde et al., 2006)	17	Event-related
5	50/50 gain-loss gamble decisions (Tom et al., 2007)	16	Blocked
6	Living/nonliving decision on mirror-reversed words (Poldrack et al., unpublished)	14	Blocked
7	Reading pseudowords aloud (Xue et al., submitted)	19	Event-related
8	Response inhibition (successful stopping) (Aron & Poldrack, 2006)	15	Event-related

Poldrack, Halchenko, & Hanson, 2009, *Psychological Science*

Analysis	**Crossvalidated accuracy**	**# of voxels included**
Union of all in-mask voxels across subjects (one-vs-one)	74%	417,231
Intersection of in-mask voxels across subjects (one-vs-many)	80.8%	214,940
Positively activated voxels only (across all 130 subjects, t > 3, p<.002) (one-vs-many)	74.6%	83,825
Deactivated voxels only (t < -3, p<.002) (one-vs-many)	50.8%	23,736

Accuracy above 18.5% is significant at p<.05 by randomization



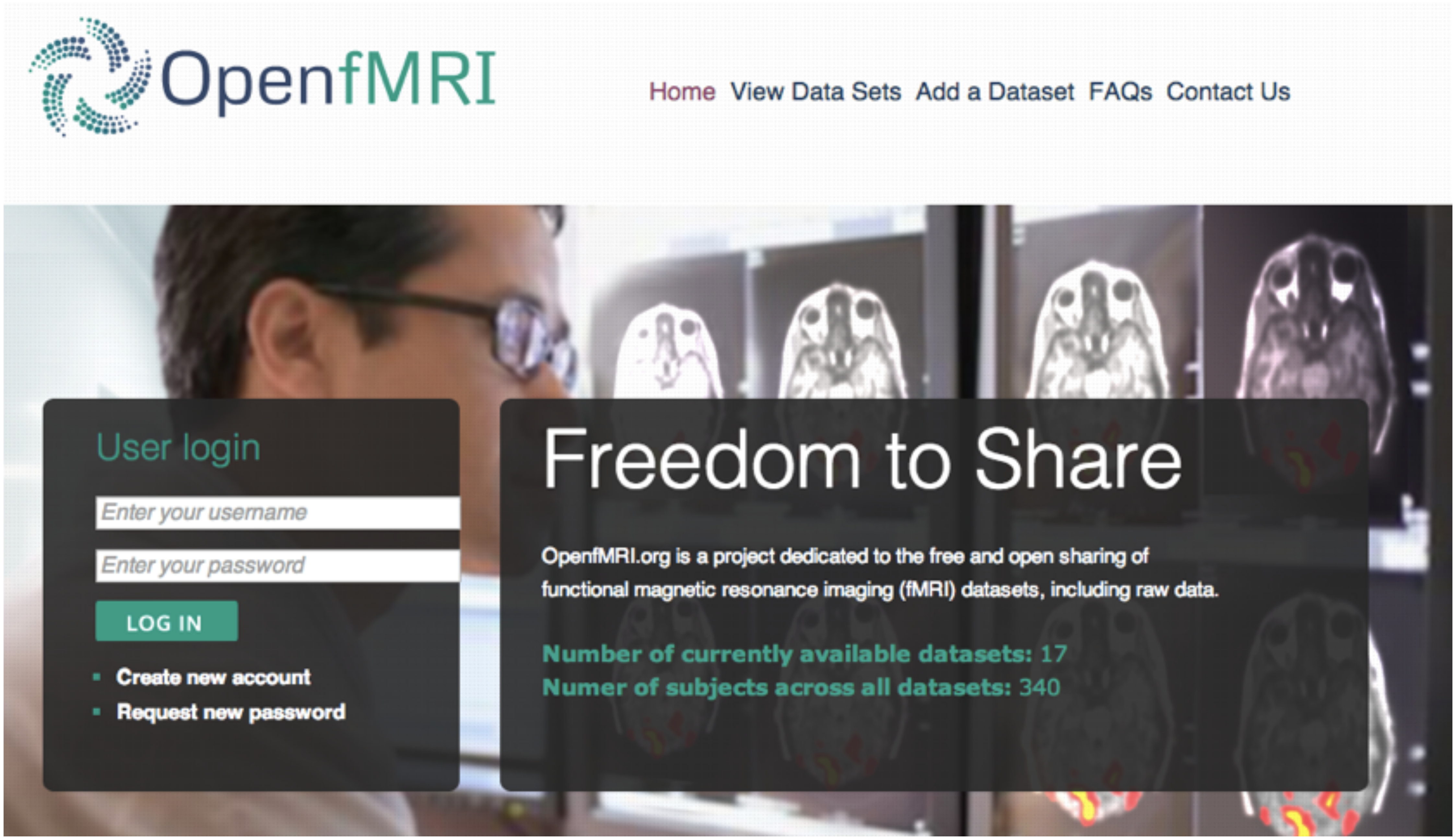
Task chosen by classfier

True task

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
Task 1	**87.5**	6.0	0.0	0.0	6.0	0.0	0.0	0.0
Task 2	0.0	**90.0**	0.0	0.0	0.0	0.0	5.0	5.0
Task 3	8.0	23.0	**61.5**	0.0	0.0	8.0	0.0	0.0
Task 4	0.0	0.0	0.0	**82.4**	0.0	0.0	0.0	18.0
Task 5	0.0	38.0	0.0	0.0	**43.8**	18.2	0.0	0.0
Task 6	0.0	28.0	0.0	0.0	0.0	**71.4**	0.0	0.0
Task 7	0.0	11.0	0.0	0.0	0.0	0.0	**84.0**	5.0
Task 8	0.0	0.0	7.0	0.0	0.0	0.0	27.0	**63.0**

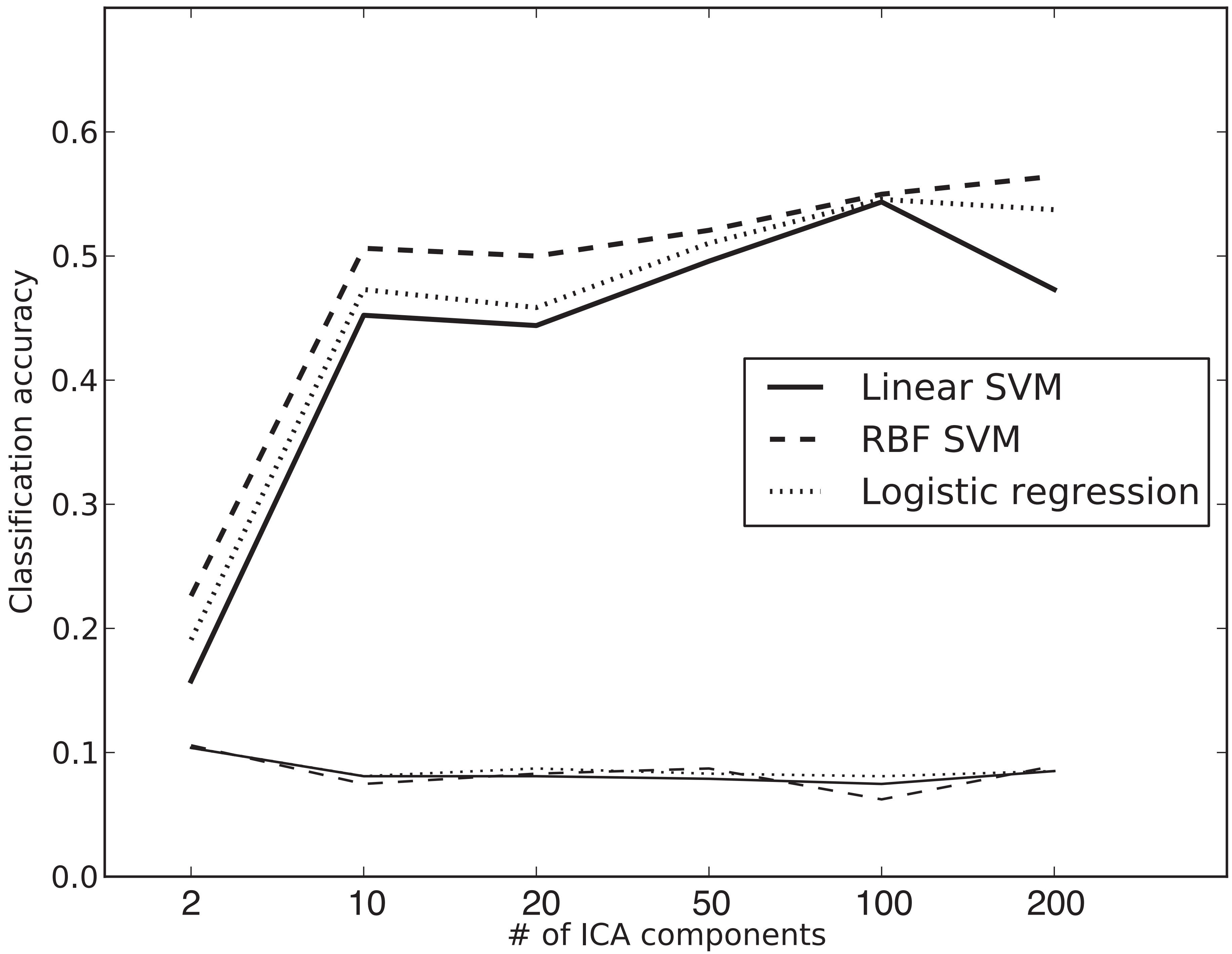
Poldrack, Halchenko, & Hanson, 2009, *Psychological Science*

Larger-scale decoding



26 tasks, 482 images from 338 subjects

Classification results

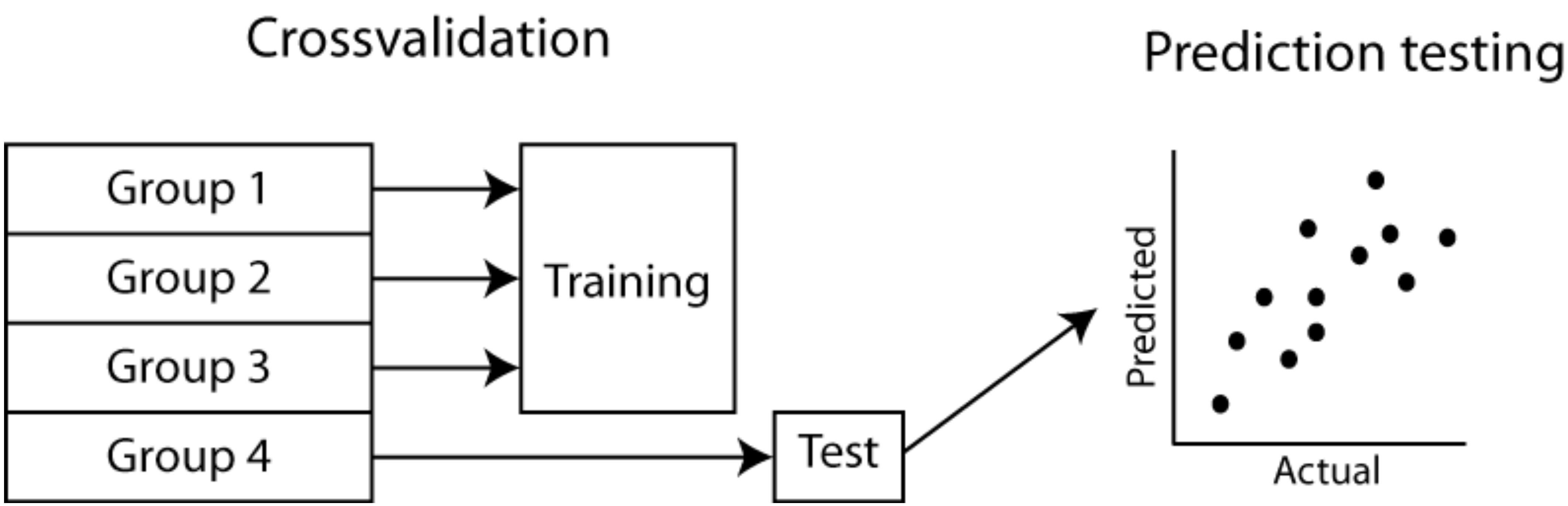


Whole-brain:
47% accuracy

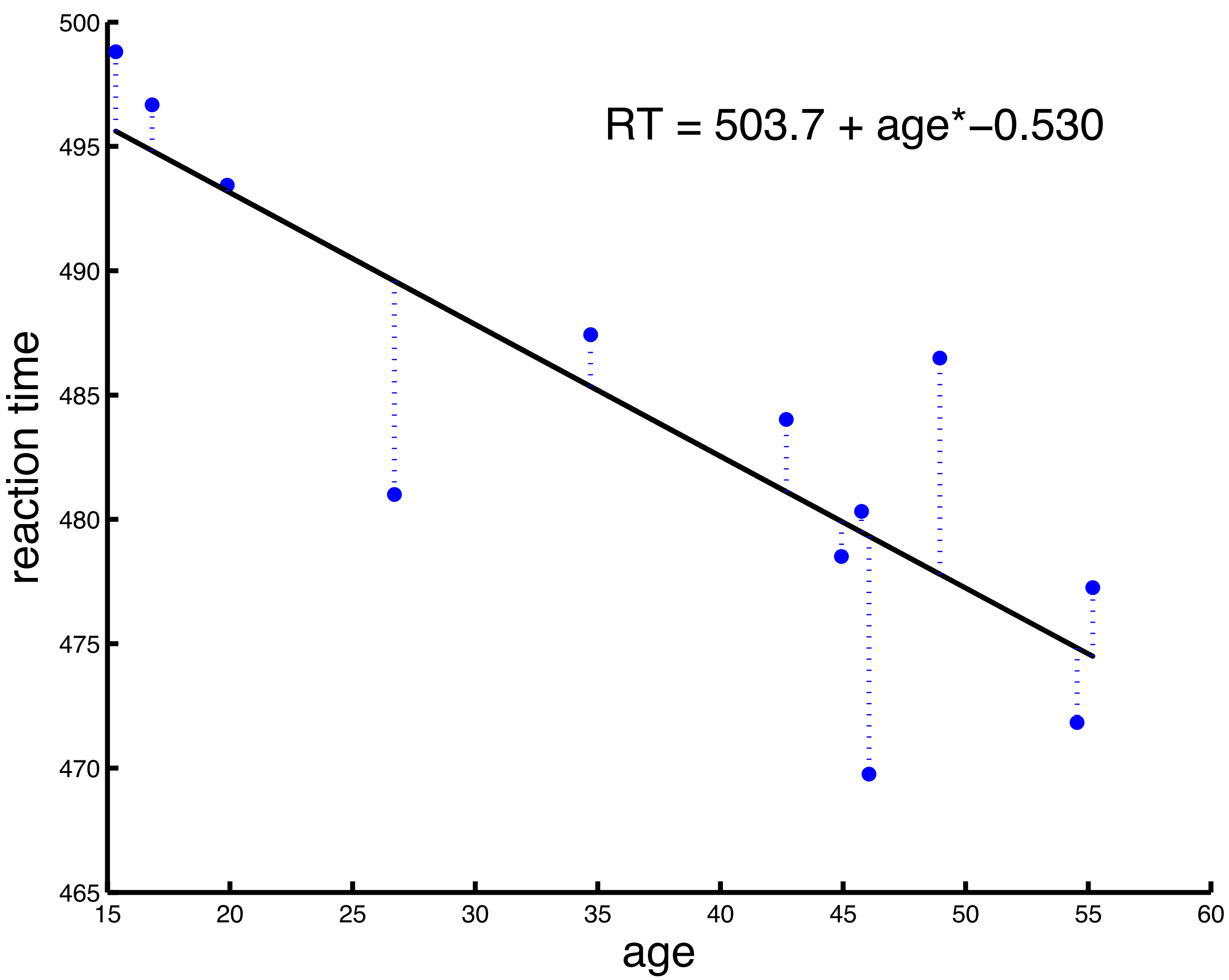
Poldrack et al., 2013, *Frontiers in Neuroinformatics*

Predicting individual differences from fMRI

- In neuroscience, correlations are often colloquially described as "prediction", but true prediction requires generalization to new samples
- The ability to predict quantitative variables for new individuals can be tested using crossvalidation

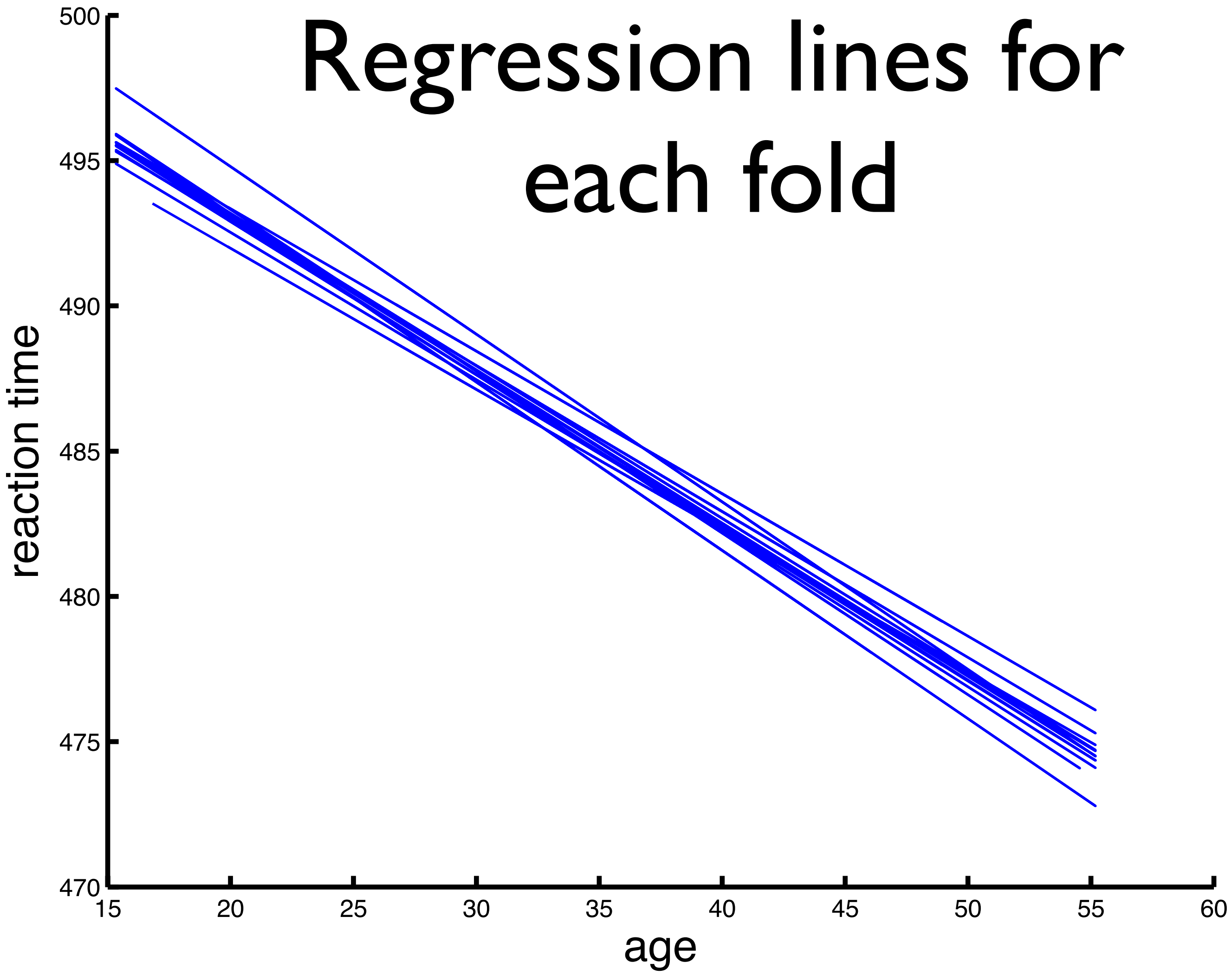


Leave-one-out crossvalidation



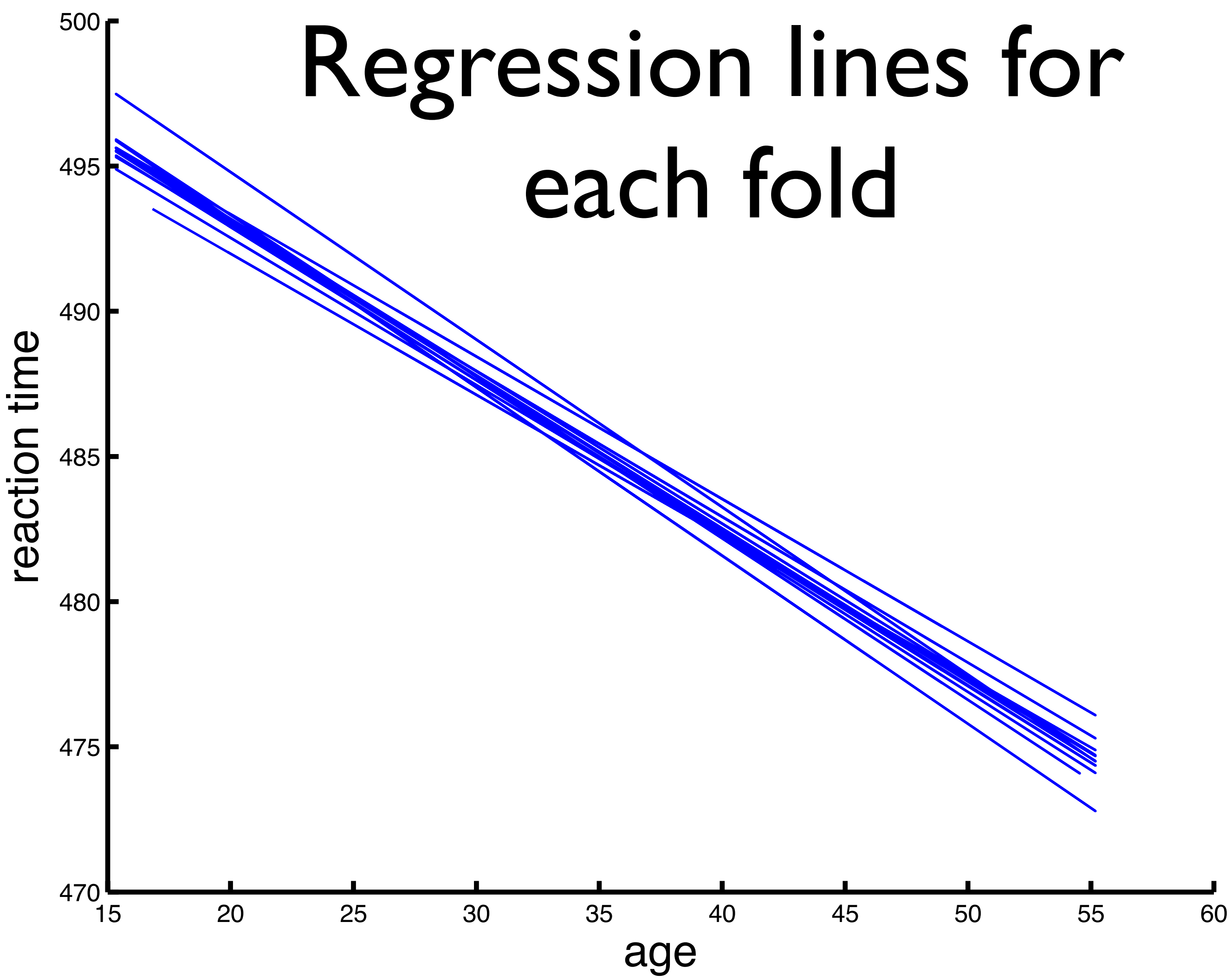
full-sample $R^2 = 0.694$

Leave-one-out crossvalidation



full-sample $R^2 = 0.694$

Leave-one-out crossvalidation

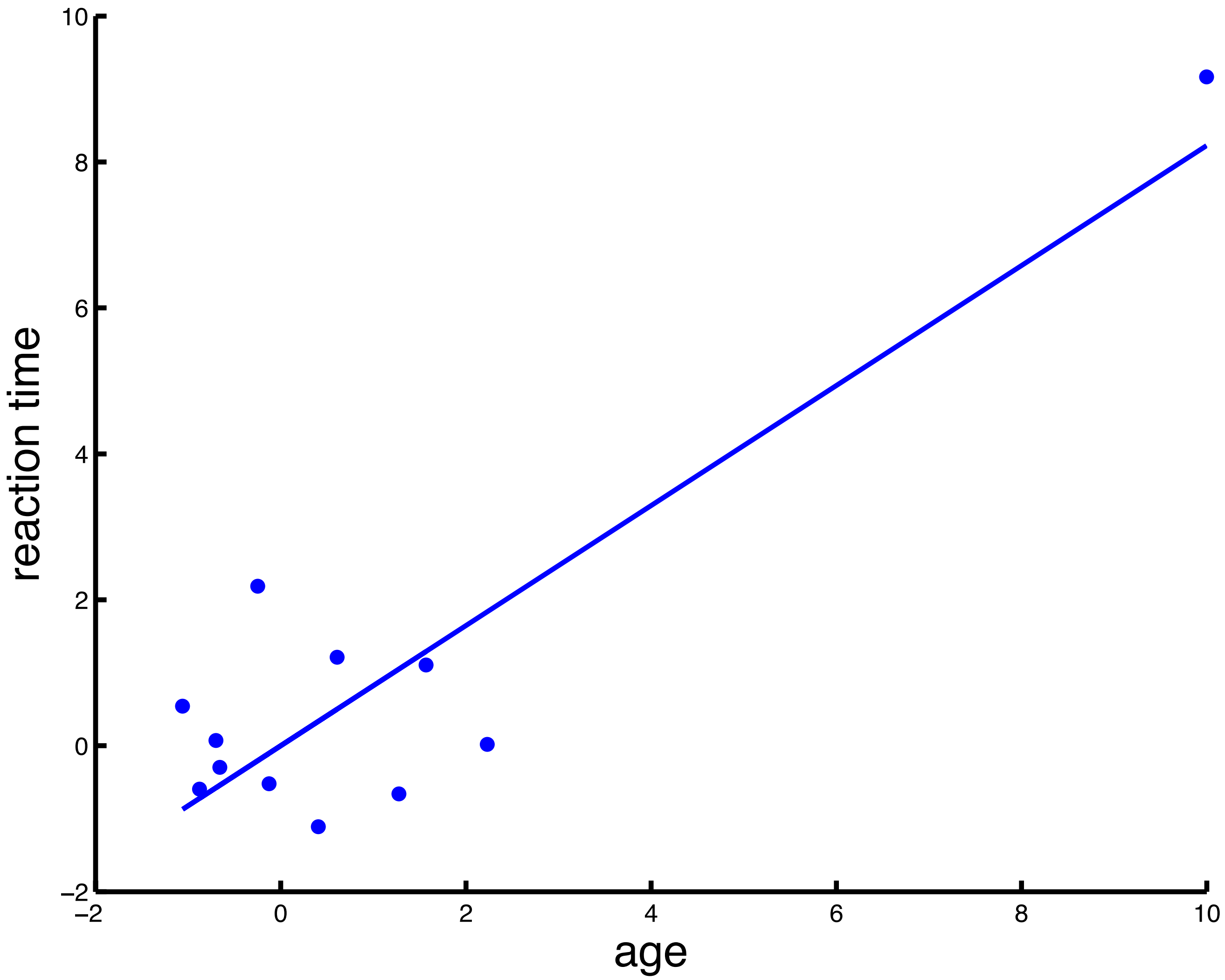


full-sample R^2 = 0.694

LOO CV R^2 = 0.586

mean new sample R^2 =0.591

Correlation ≠ Prediction



- 11 datapoints sampled from normal distribution (no correlation)
- one outlier
- full-sample R^2 =0.785
- LOO CV R^2 = 0.025

- Highlights importance of examining the raw data!

Proceedings of the National Academy of Sciences of the United States of America

PNAS

Neuroprediction of future rearrest

Eyal Aharoni[a,b,1,2], Gina M. Vincent[c], Carla L. Harenski[a], Vince D. Calhoun[a,d], Walter Sinnott-Armstrong[e], Michael S. Gazzaniga[f], and Kent A. Kiehl[a,b,2]

Author Affiliations

Edited by Robert Desimone, Massachusetts Institute of Technology, Cambridge, MA, and approved February 27, 2013 (received for review November 7, 2012)

Current Issue
E-mail Alerts
Subscribe
RSS

Published online before print March 27, 2013, doi: 10.1073/pnas.1219302110
PNAS March 27, 2013

"The present analysis shows that hemodynamic activity within the brain prospectively predicted rearrest in an offender sample."

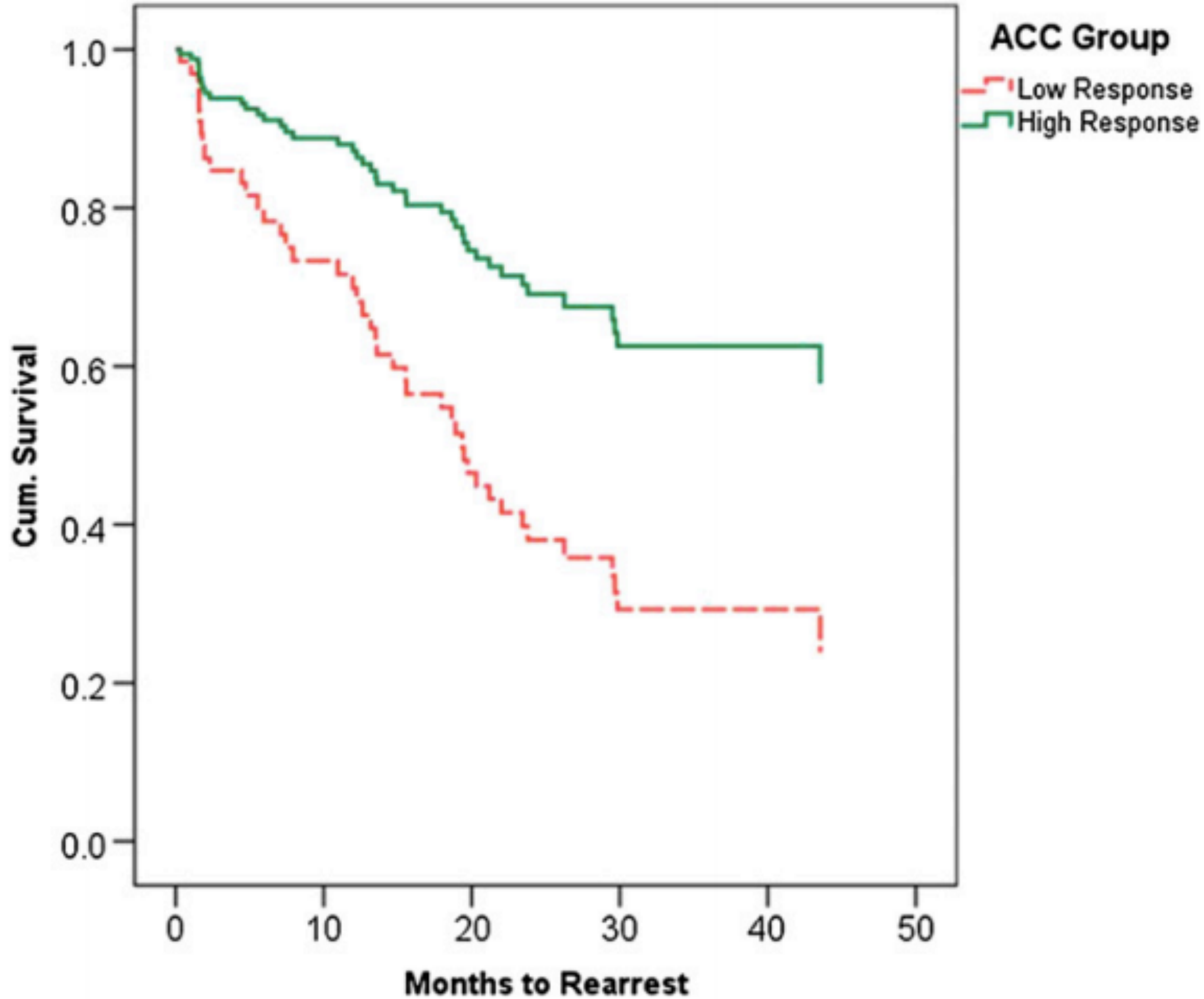
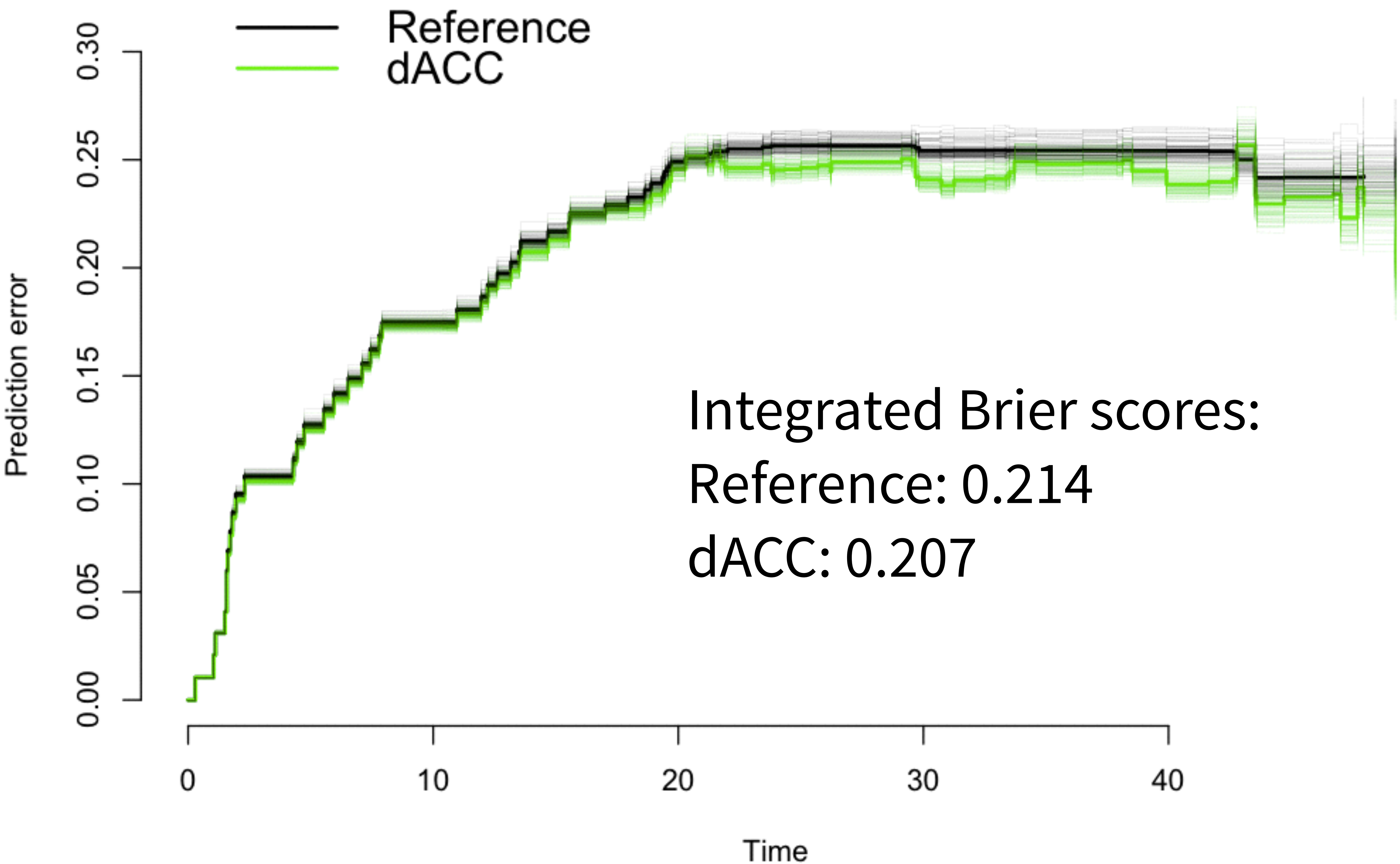


Fig. 1. Cox survival function showing proportional rearrest survival rates of high (solid green) vs. low (dashed red) ACC response groups for any crime over a 4-y period. Results of this median split analysis were equivalent to that of the parametric model: bootstrapped $B = 0.96$; SE $= 0.40$; $P < 0.01$; 95% CI, 0.29–1.84. The mean survival times to rearrest for the low and high ACC activity groups were 25.27 (2.80) mo and 32.42 (2.73) mo, respectively. The overall probabilities of rearrest were 60% for the low ACC group and 46% for the high ACC group.

Aharoni et al., 2013

Prediction error using crossvalidation

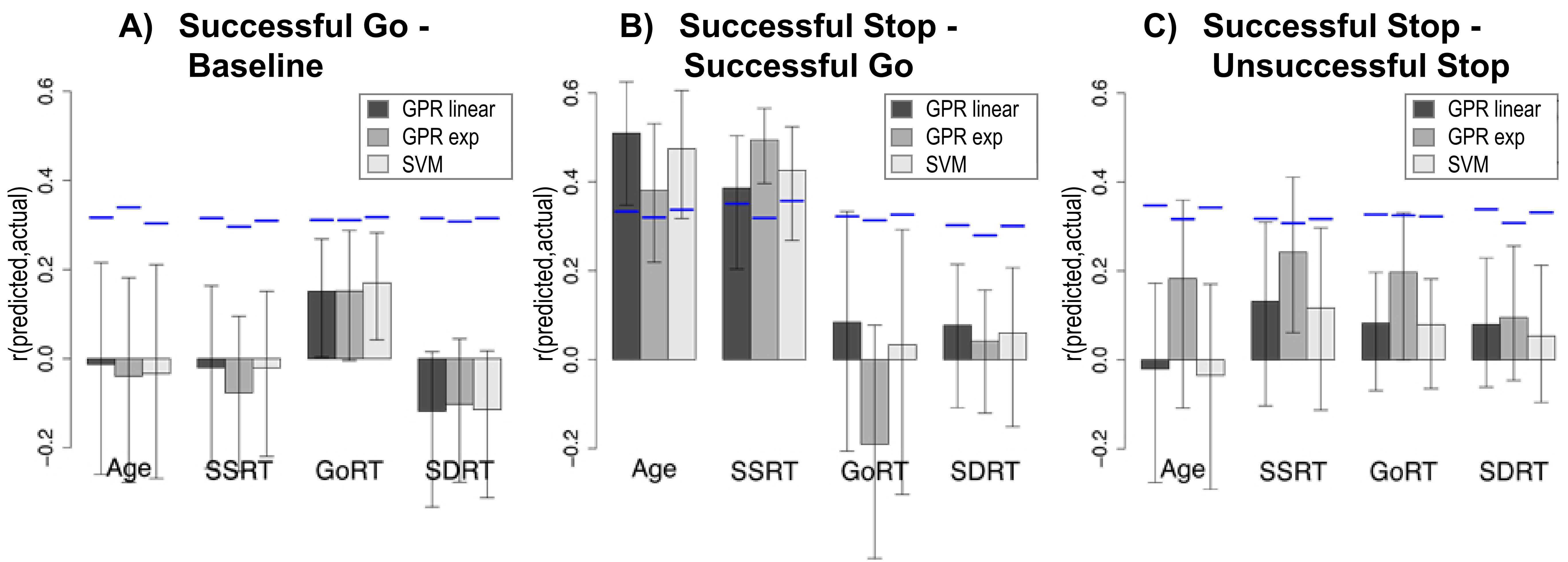


Integrated Brier scores:
Reference: 0.214
dACC: 0.207

http://www.russpoldrack.org/2013/04/how-well-can-we-predict-future-criminal.html

https://github.com/poldrack/criminalprediction

Predicting individual differences from fMRI



p<.05 (by randomization)

Variables tested:

Age: subject's age

SSRT: stop signal reaction time

GoRT: go reaction time

SDRT: std. dev. of go reaction time

Cohen et al., 2010, *Frontiers in Human Neuroscience*

Meta-analytic decoding

- All of the results to this point were based on fMRI data from individual subjects
- Can we push this even further?
  - Can we use meta-analytic data from papers?

Activation locations

- Brain activity is reported in (somewhat) standardized format

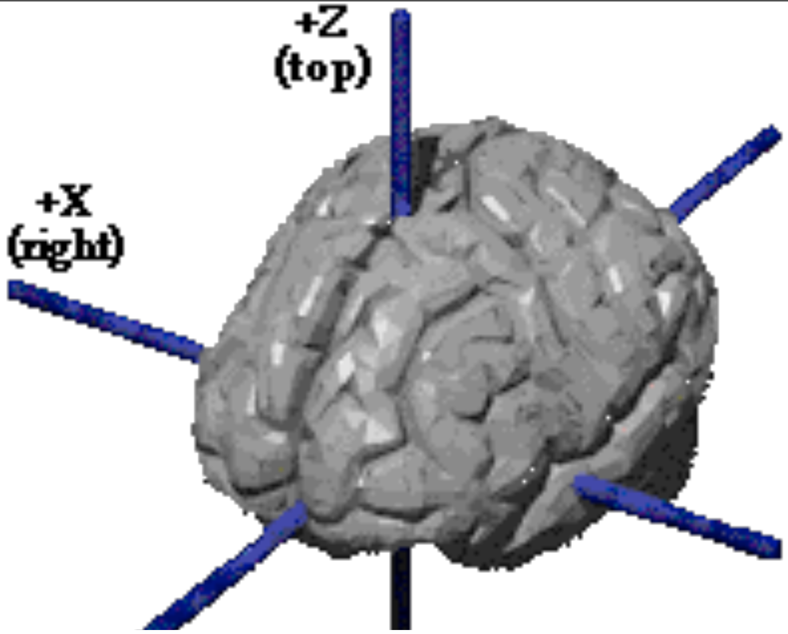


Table 1
Regions that showed a condition × time interaction in the ANOVA analysis

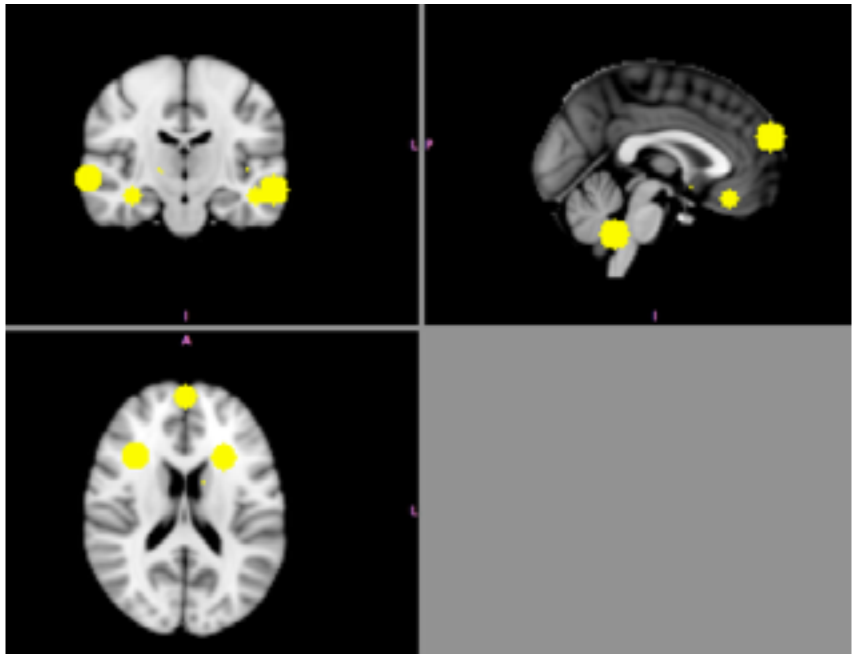
No.	Region	Hemisphere	BA	x	y	z	mm³
1	Middle/superior temporal gyrus	L	21/22/37	−52	−54	9	13257
2	Inferior frontal gyrus	L	45/46/9	−49	26	6	2781
3	Posterior cerebellum	L		−19	−79	−38	2214
4	Dorsomedial PFC	L	9/8	−11	42	47	3051
5	Left anterior PFC	L	10	−37	49	15	2025
6	Inferior parietal cortex	L	40/7	−42	−58	47	3132
7	Dorsal premotor cortex	L	6	−43	0	50	1485
8	Lingual gyrus	L	17	−10	−95	−2	378
9	Middle /superior temporal gyrus	R	21/22/37	52	−40	5	16470
10	Inferior frontal gyrus	R	45/46	51	28	6	2241
11	Posterior cerebellum	R		23	−78	−34	2808
12	Dorsomedial PFC	R	9	5	53	29	405
13	Right anterior PFC	R	10	38	42	21	5022
14	Inferior parietal cortex	R	40/7	42	−53	48	9963
15	Superior frontal gyrus	R	6/8	10	28	60	297
16	Anterior cingulate cortex	M	32	0	26	35	5076
17	Posterior cingulate cortex	M	23/31/7	0	−35	31	9612
18	Precuneus	M	7/19	1	−76	36	10044

Creating meta-analytic brain maps

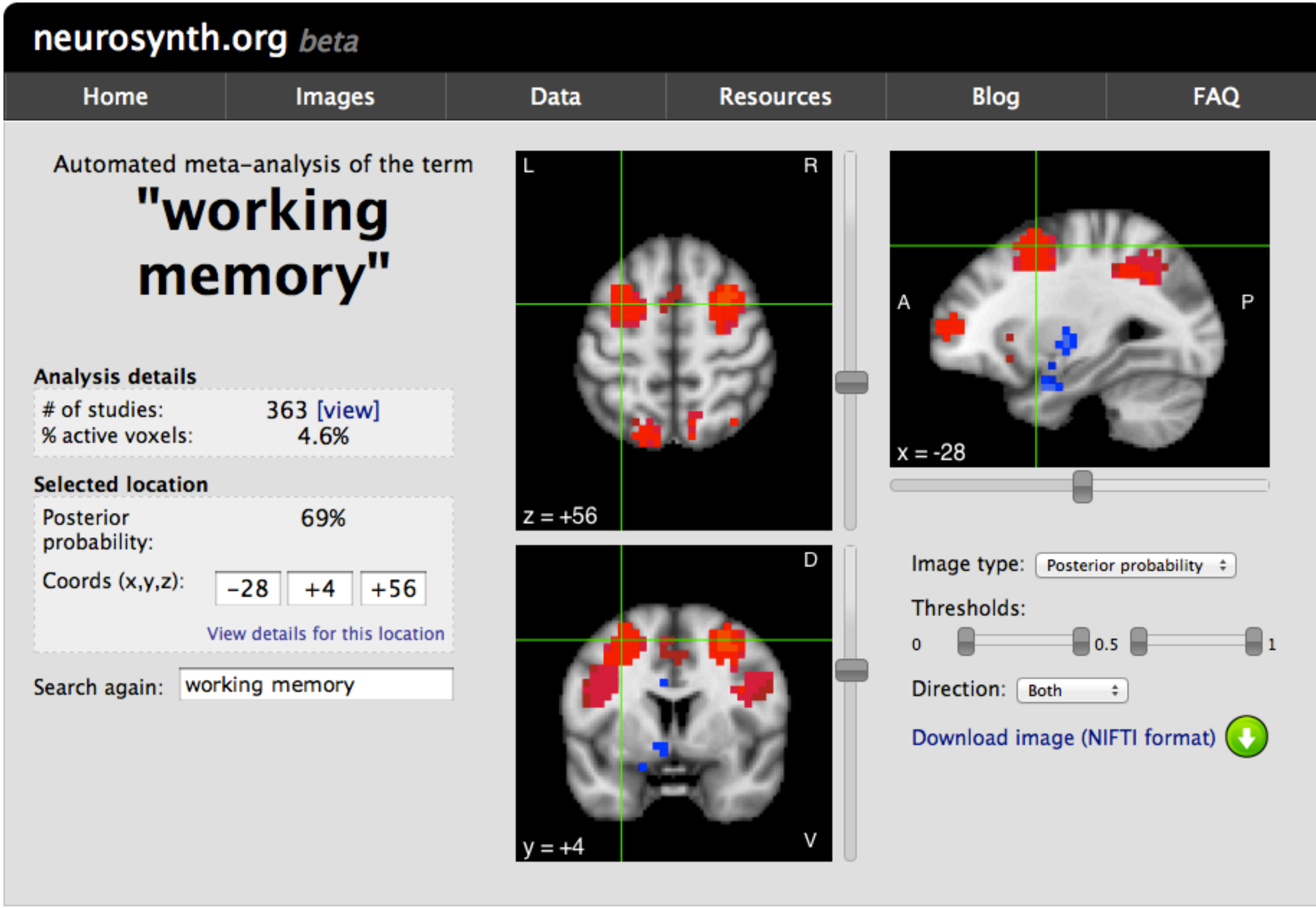
- Automated Coordinate Extraction (Yarkoni et al, 2011, *Nature Methods*)
  - Automatically extracts activation tables from fMRI papers for 17 journals
  - Current database has 4,393 papers (with full text)
  - Good accuracy
    - 84% sensitivity, 97% specificity against SumsDB manual database
- Meta-analytic maps created for each paper
  - 10mm sphere placed at each focus

X	Y	Z
12	57	-6
33	21	15
24	15	60
42	6	51
24	-3	57

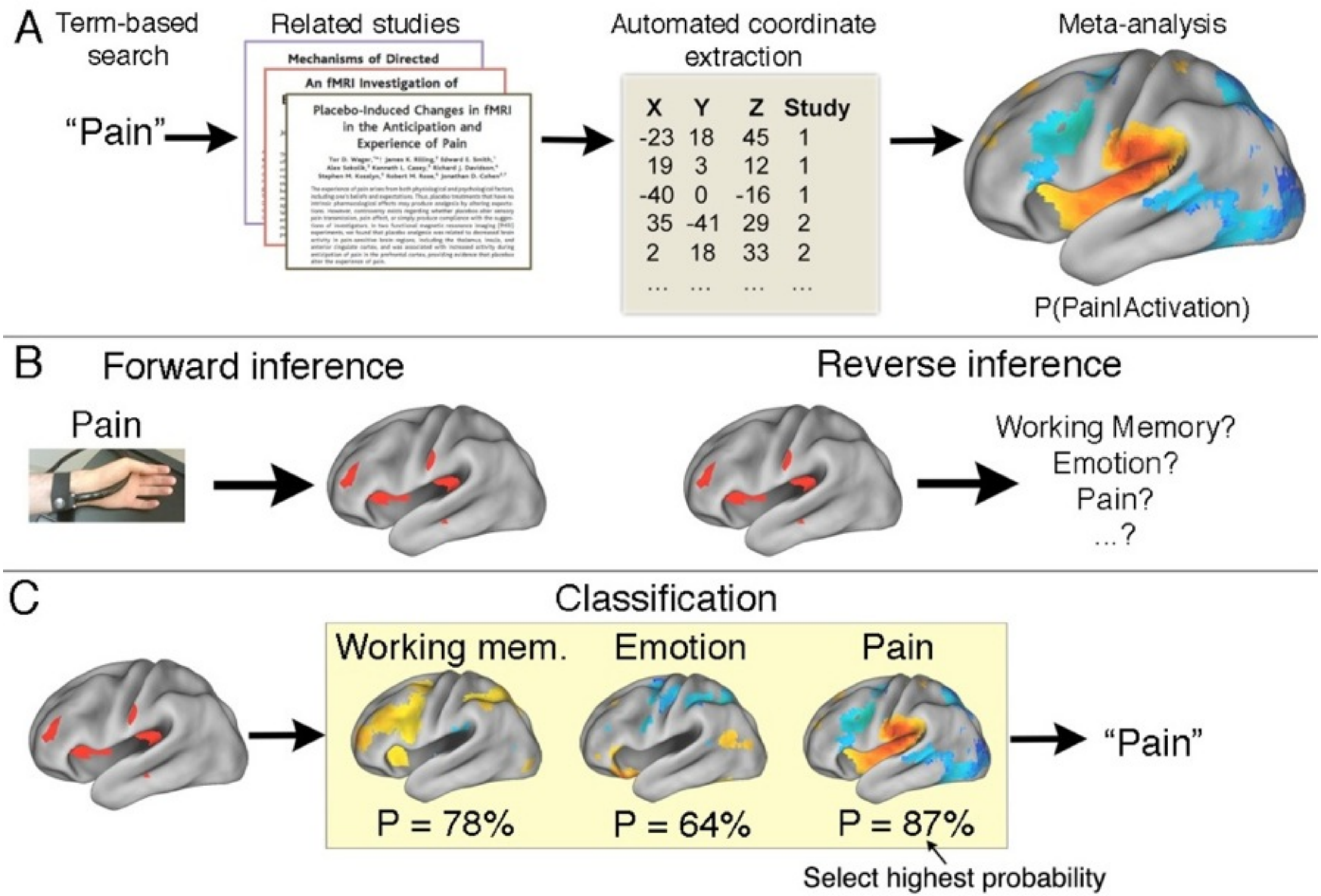
Automated coordinate extraction →



Neurosynth.org

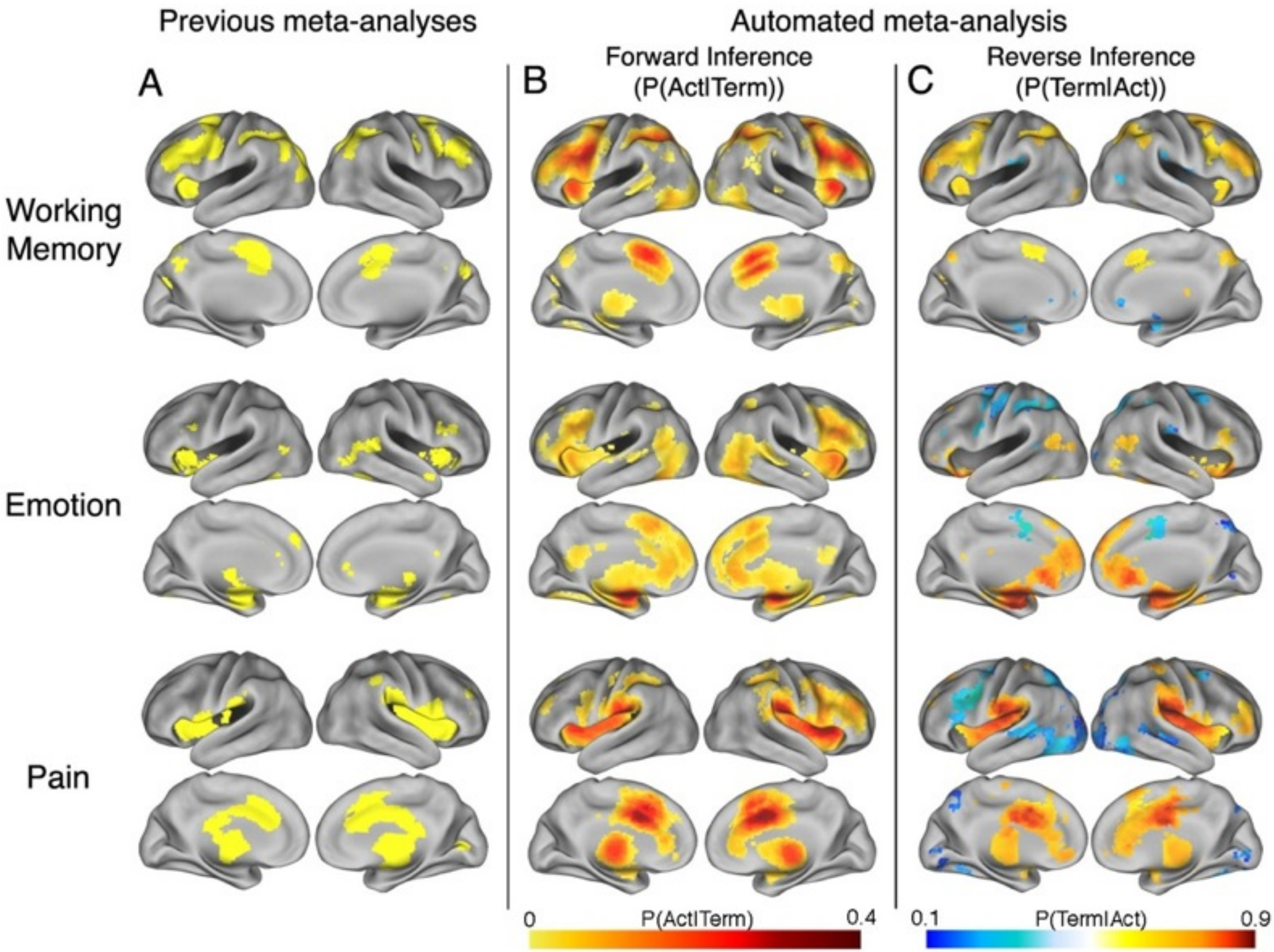


Automated meta-analysis



Yarkoni et al., 2011, *Nature Methods*

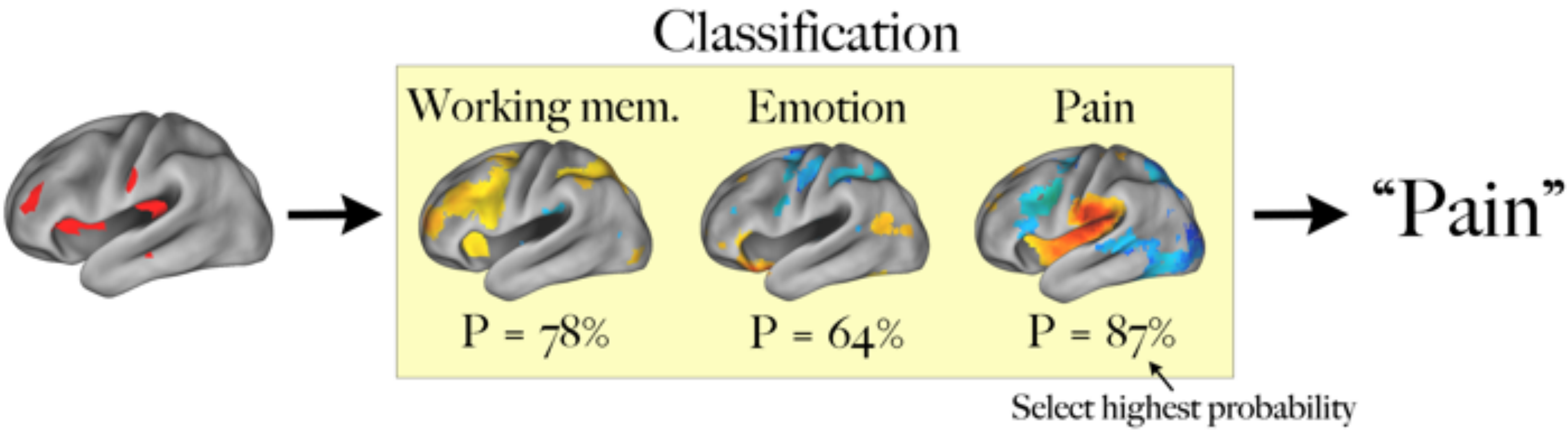
Automated meta-analysis



Yarkoni et al., 2011, *Nature Methods*

Classification of cognitive states

- Given 2+ terms, can determine which is most likely given the data
- Naive Bayes classifier: assumes that all features (voxels) are independent; selects the most probable class
- Can apply this to any activation map—studies, individual subjects, etc.

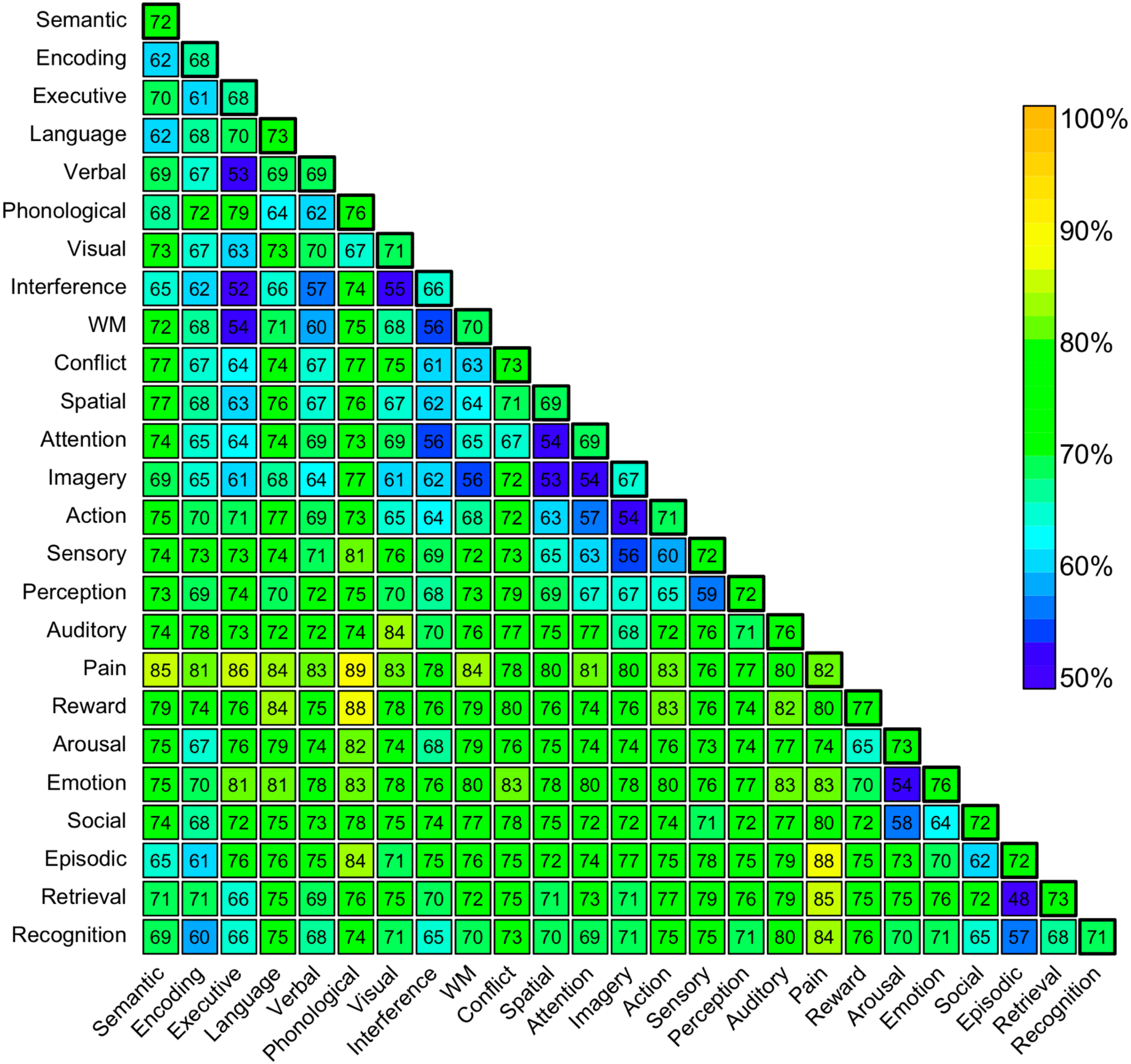


Yarkoni et al, 2011, *Nature Methods*

Classification of new studies

- Cross-validated classification of all studies in database
- Select 25 high-frequency terms
- Pairwise classification: how well can we distinguish between each pair of terms?

Yarkoni et al, 2011, *Nature Methods*

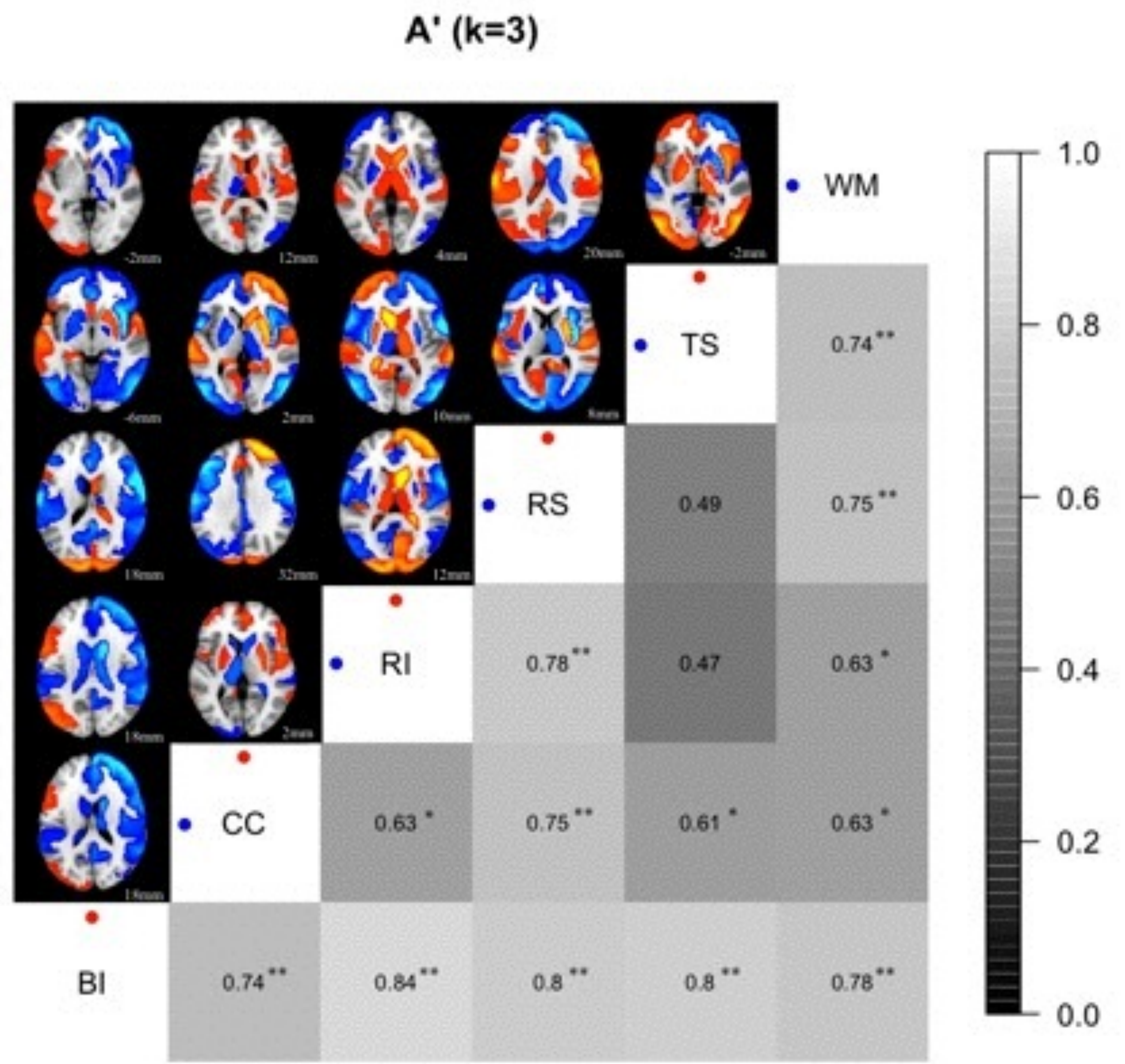


	Semantic	Encoding	Executive	Language	Verbal	Phonological	Visual	Interference	WM	Conflict	Spatial	Attention	Imagery	Action	Sensory	Perception	Auditory	Pain	Reward	Arousal	Emotion	Social	Episodic	Retrieval	Recognition
Semantic	72																								
Encoding	62	68																							
Executive	70	61	68																						
Language	62	68	70	73																					
Verbal	69	67	53	69	69																				
Phonological	68	72	79	64	62	76																			
Visual	73	67	63	73	70	67	71																		
Interference	65	62	52	66	57	74	55	66																	
WM	72	68	54	71	60	75	68	56	70																
Conflict	77	67	64	74	67	77	75	61	63	73															
Spatial	77	68	63	76	67	76	67	62	64	71	69														
Attention	74	65	64	74	69	73	69	56	65	67	54	69													
Imagery	69	65	61	68	64	77	61	62	56	72	53	54	67												
Action	75	70	71	77	69	73	65	64	68	72	63	57	54	71											
Sensory	74	73	73	74	71	81	76	69	72	73	65	63	56	60	72										
Perception	73	69	74	70	72	75	70	68	73	79	69	67	67	65	59	72									
Auditory	74	78	73	72	72	74	84	70	76	77	75	77	68	72	76	71	76								
Pain	85	81	86	84	83	89	83	78	84	78	80	81	80	83	76	77	80	82							
Reward	79	74	76	84	75	88	78	76	79	80	76	74	76	83	76	74	82	80	77						
Arousal	75	67	76	79	74	82	74	68	79	76	75	74	74	76	73	74	77	74	65	73					
Emotion	75	70	81	81	78	83	78	76	80	83	78	80	78	80	76	77	83	83	70	54	76				
Social	74	68	72	75	73	78	75	74	77	78	75	72	72	74	71	72	77	80	72	58	64	72			
Episodic	65	61	76	76	75	84	71	75	76	75	72	74	77	75	78	75	79	88	75	73	70	62	72		
Retrieval	71	71	66	75	69	76	75	70	72	75	71	73	71	77	79	76	79	85	75	75	76	72	48	73	
Recognition	69	60	66	75	68	74	71	65	70	73	70	69	71	75	75	71	80	84	76	70	71	65	57	68	71

Yarkoni et al, 2011, *Nature Methods*

Using classification to understand mental structure

WM: working memory
TS: Task switching
RS: Response selection
RI: Response inhibition
CC: Cognitive control
BI: Bilingual language



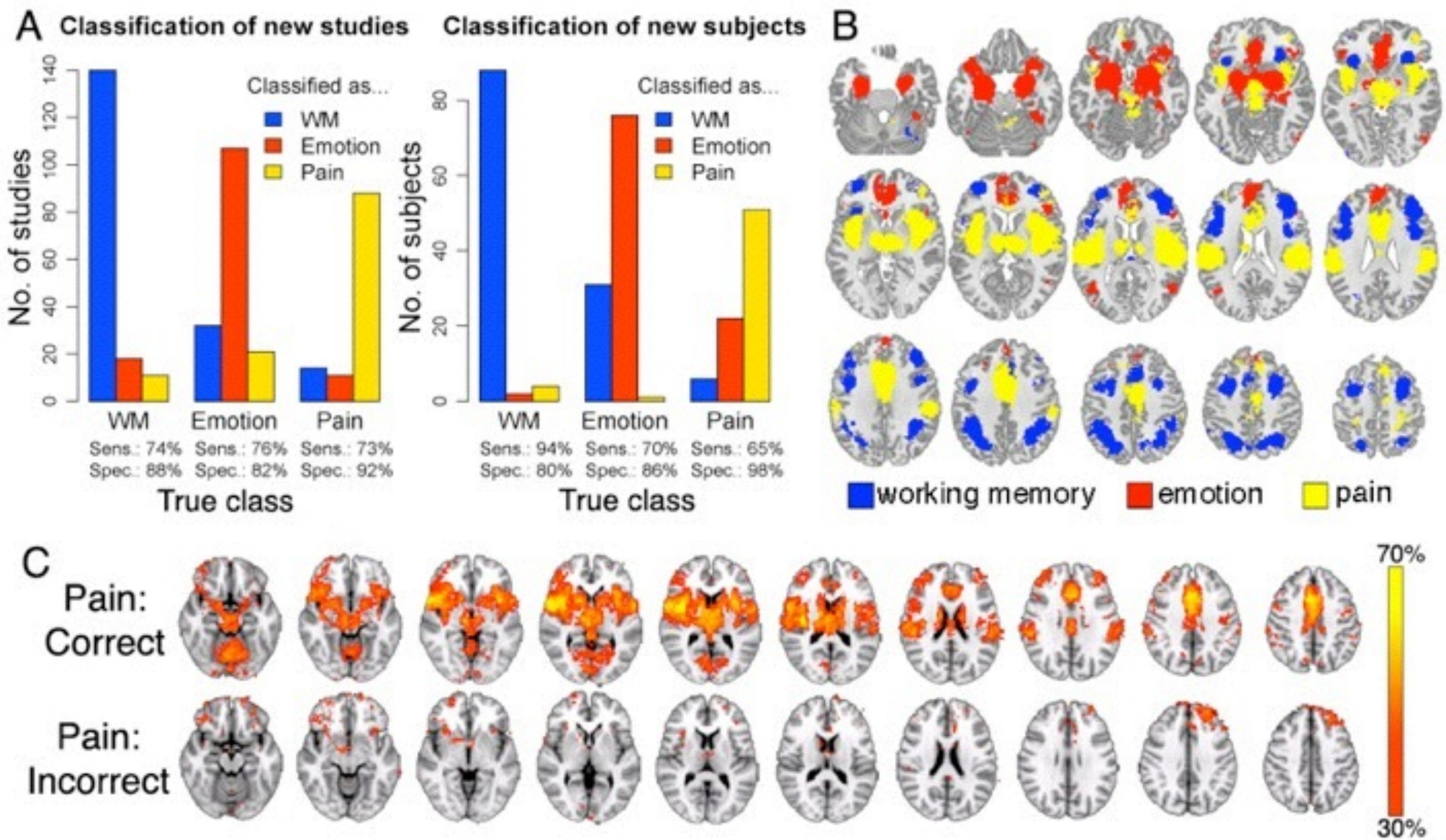
Lenartowicz et al, 2010, *Topics in Cognitive Science*

What about individual subjects?

- Can we identify cognitive states in individual (new) subjects?
- Difficult, because:
  - No opportunity for training
  - Data is of a fundamentally different type
- Tested in samples of subjects from working memory, emotion, and pain studies
  - Can we predict source study type?

Yarkoni et al, 2011, *Nature Methods*

Classifying individual subjects



Yarkoni et al, 2011, *Nature Methods*

Automating reverse inference

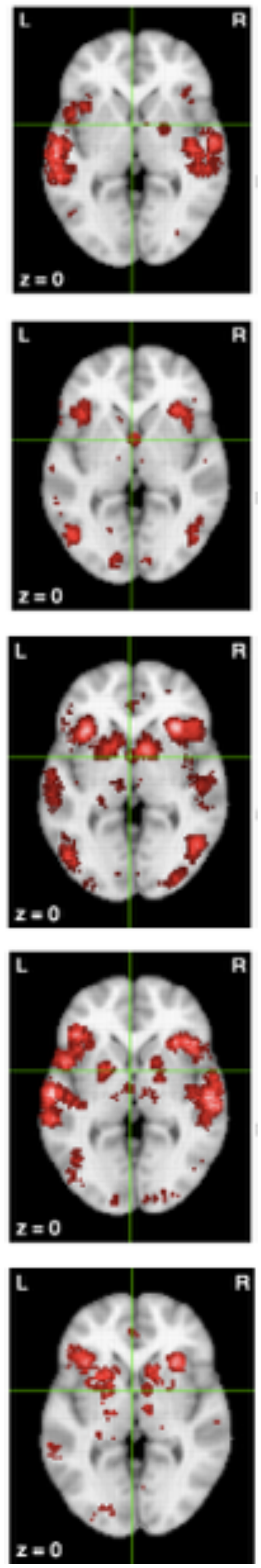


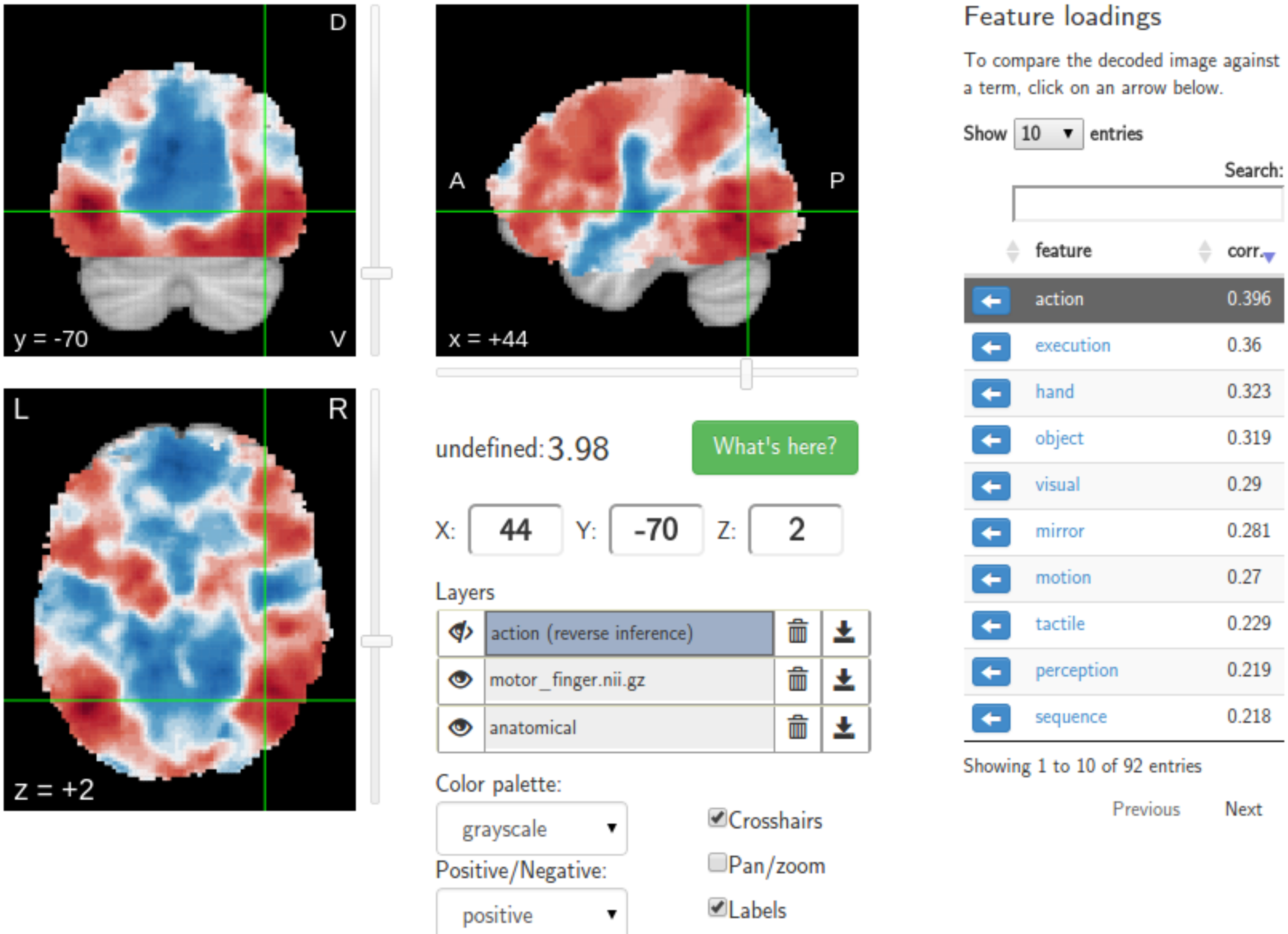
Table 2. Pearson correlations between searchlight classification map and NeuroSynth term-based reverse inference activation maps

Term	Correlation (*r*)
Control	0.1451
Working	0.1159
Numerical	0.1157
Letter	0.1081
Attention	0.1062
Correct	0.1060
Cue	0.0995
Preparatory	0.0970
Load	0.0959
Hand	0.0924

The 10 most highly correlated terms are listed. From Yarktoni et al. (26).

Helfinstein et al, 2014, PNAS

Neurovault + Neurosynth = automated reverse inference



Gorgolewski et al., submitted

Summary: Decoding mental states

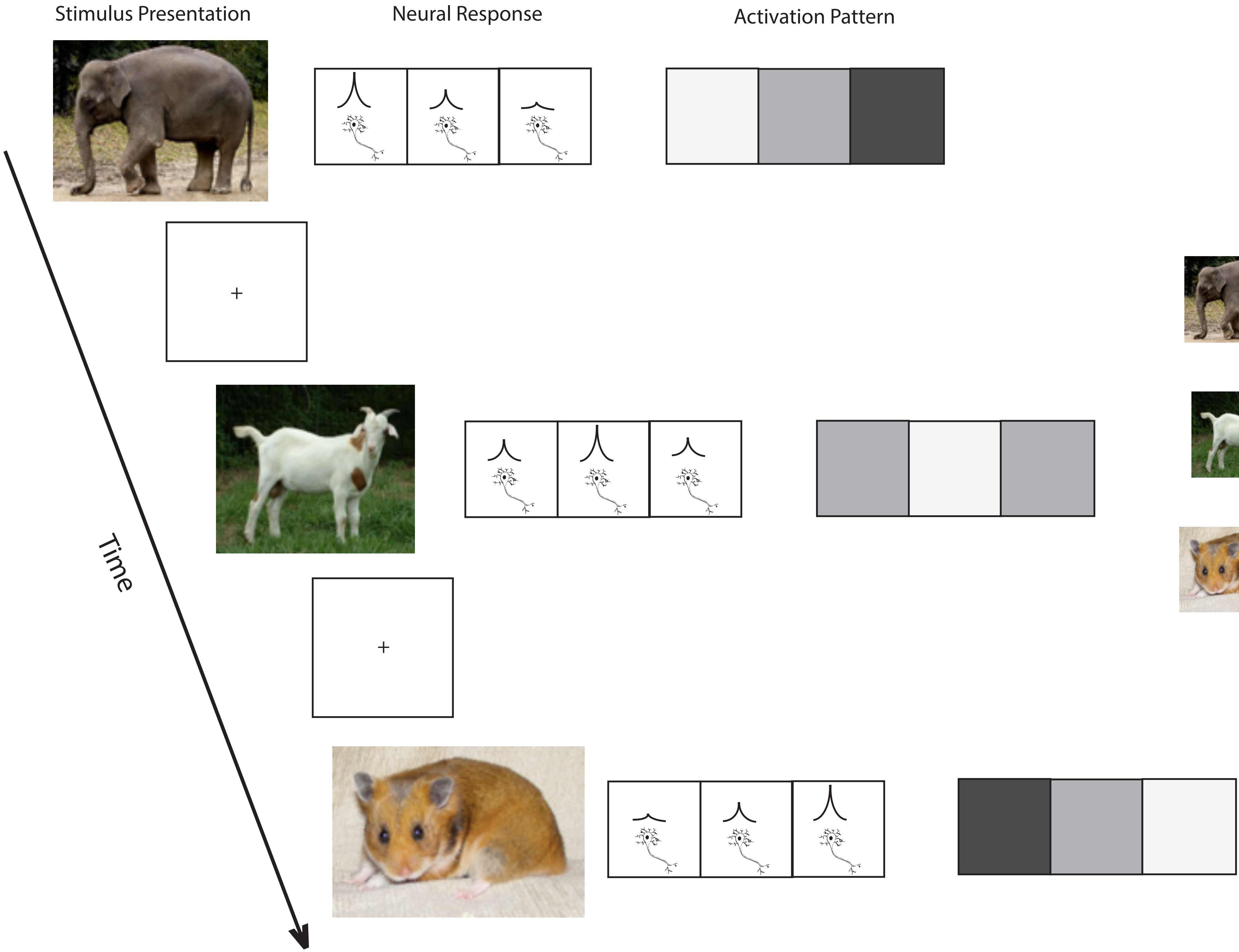
- We can decode mental states across individuals
- This can provide insights into the similarity space of mental processes
  - And ultimately inform our ontology of mental processes

Decoding representational structure using fMRI

- Psychological theories rarely make clear predictions about activation
- But they often make predictions about similarity relations between stimuli

- We can test those against neuroimaging data
  - In principle we don't even have to care *where* the effects happen in the brain

Representational similarity analysis

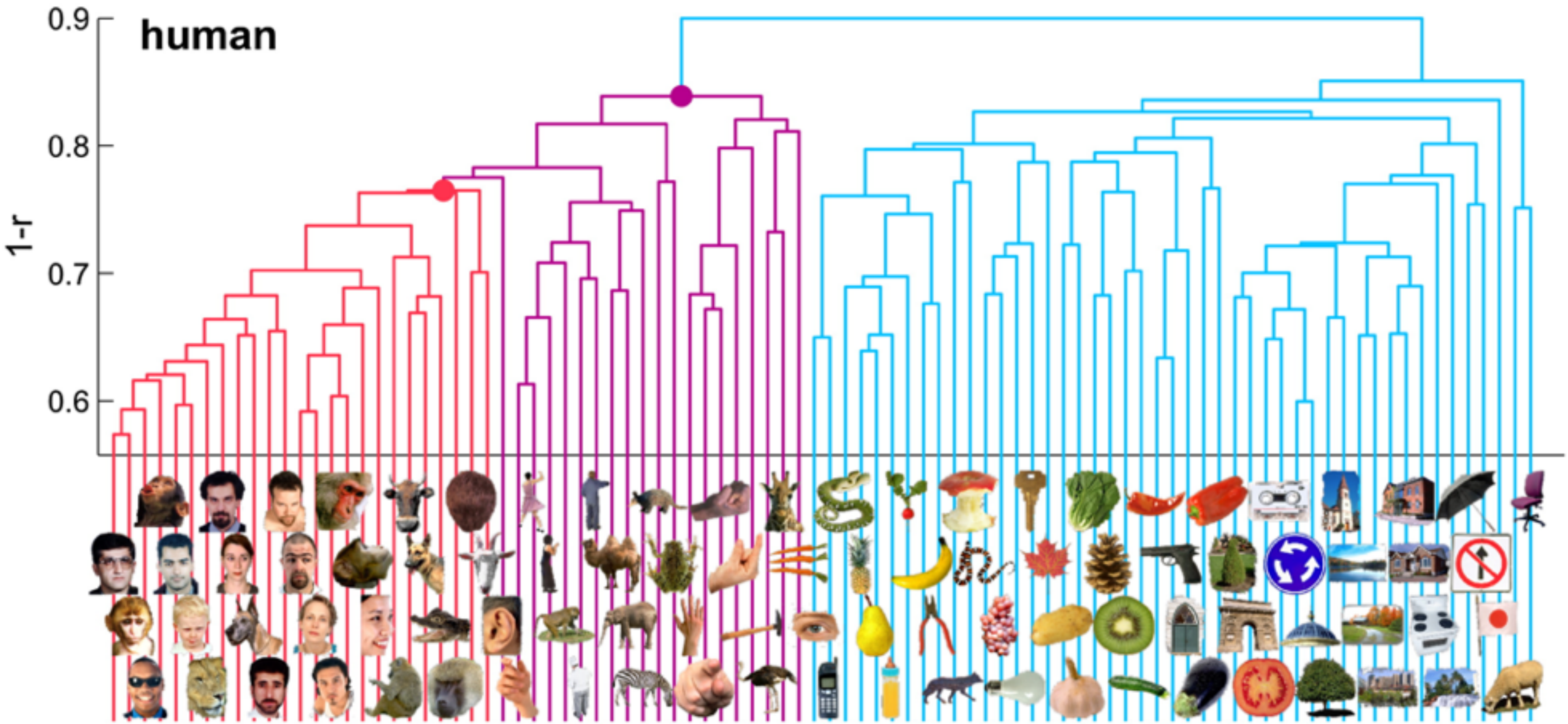


Similarity matrix

1	med	low
	1	med
		1

Davis & Poldrack, 2013

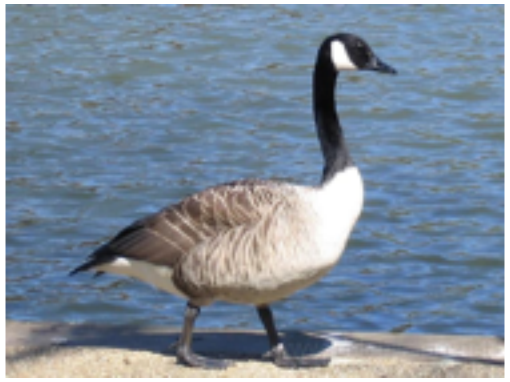
Representational analysis using fMRI



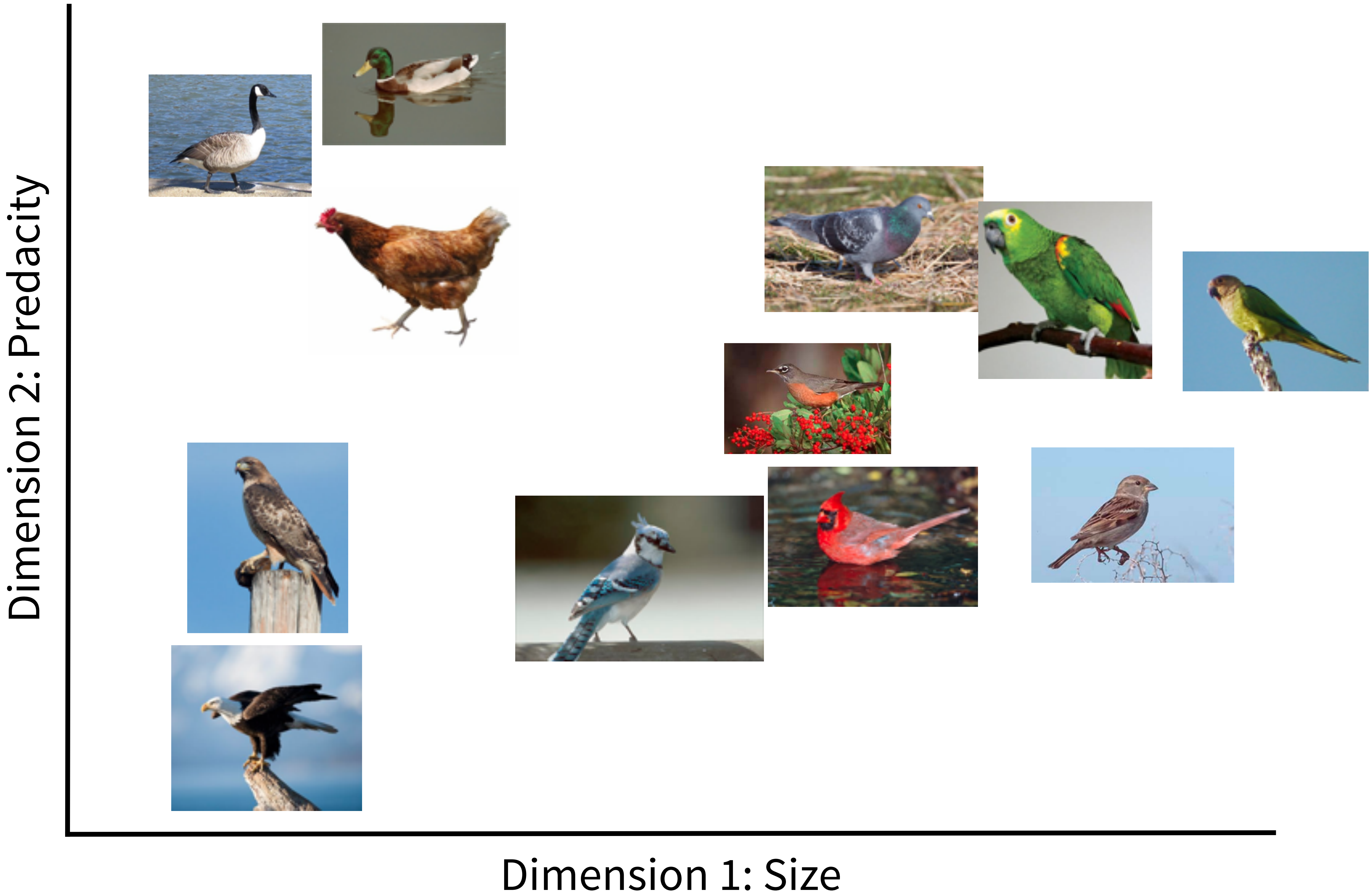
Kriegeskorte et al., 2008

Typicality

- Some birds are “birdier” than others


Typicality



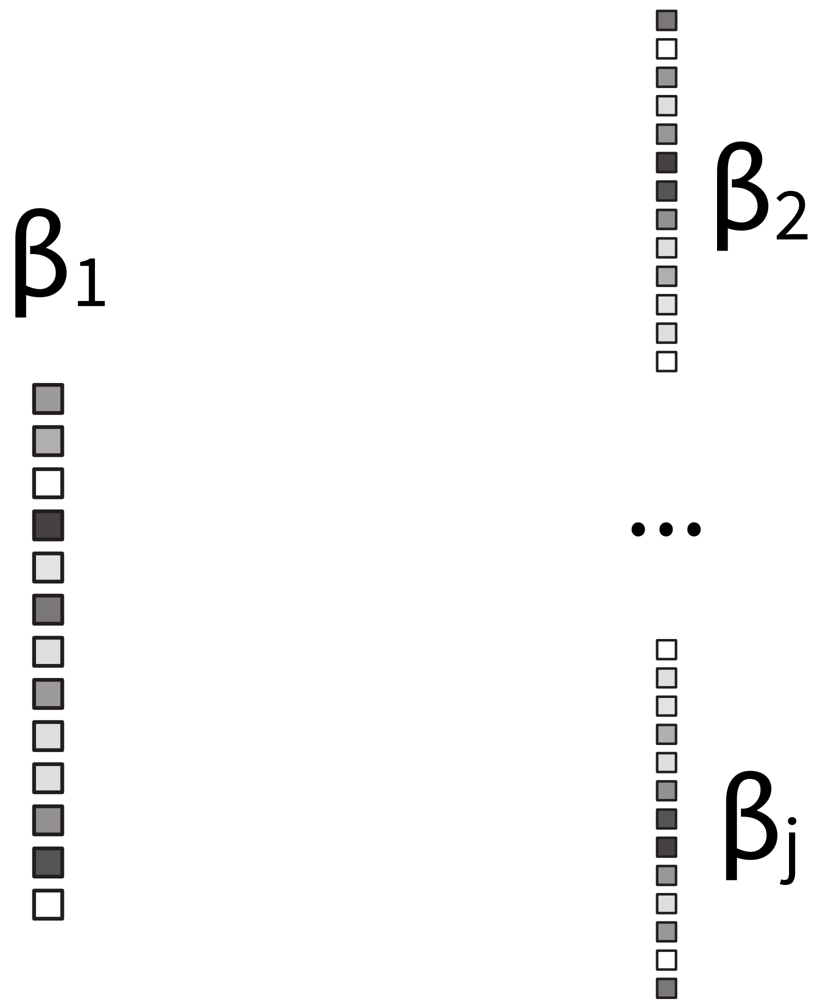
After Smith, Shoben, & Rips (1974) Photos via http://www.birdphotography.com/

Typicality is highly unconstrained

- May reflect:
  - Average similarity to other category members
  - Similarity to idealized members
    - Caricature effects

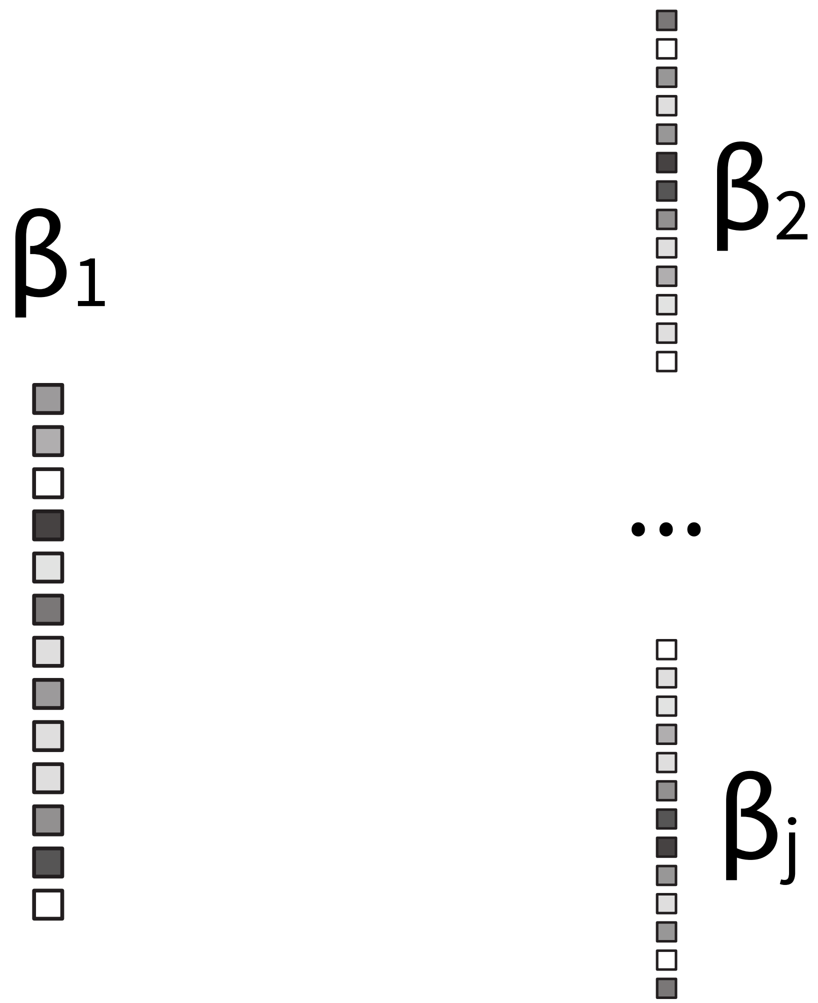
- Can we find a neural signature that is related to psychological typicality?

Computing neural typicality



Davis & Poldrack, 2013, *Cerebral Cortex*

Computing neural typicality

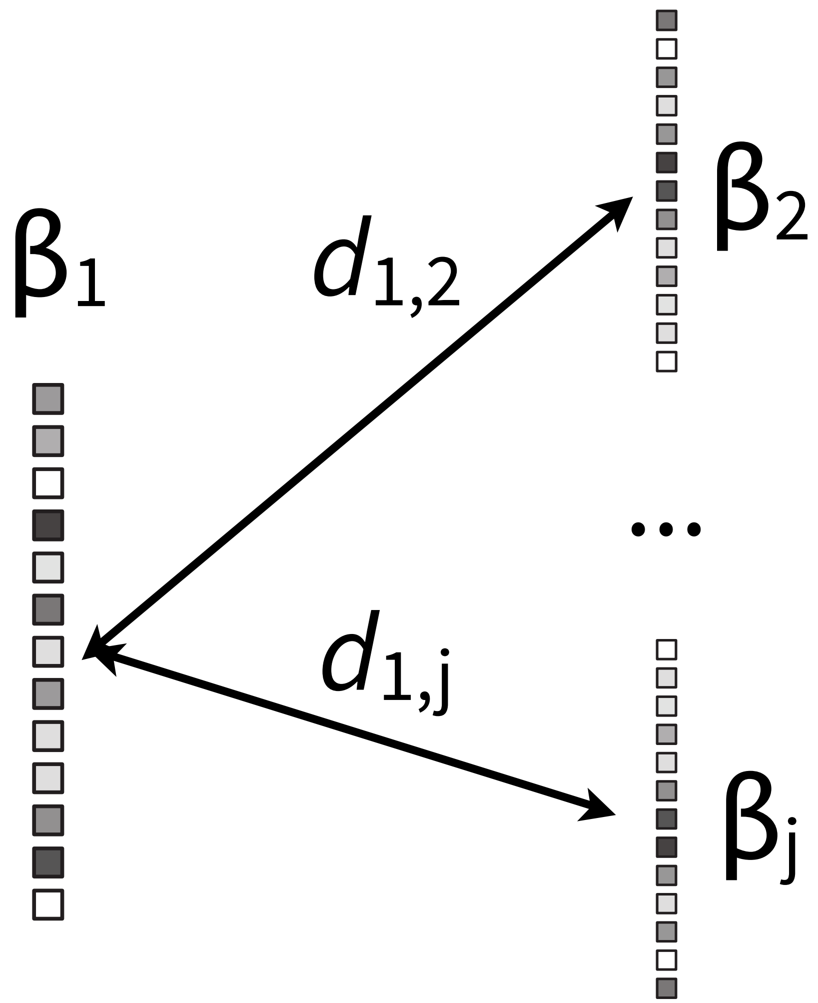


DISTANCE IS DEFINED AS THE CORRELATION DISTANCE BETWEEN TWO BETA-SERIES ACTIVATION PATTERNS

$$d_{ij} = \left[1 - corr(\beta_i, \beta_j)\right]/2$$

Davis & Poldrack, 2013, *Cerebral Cortex*

Computing neural typicality

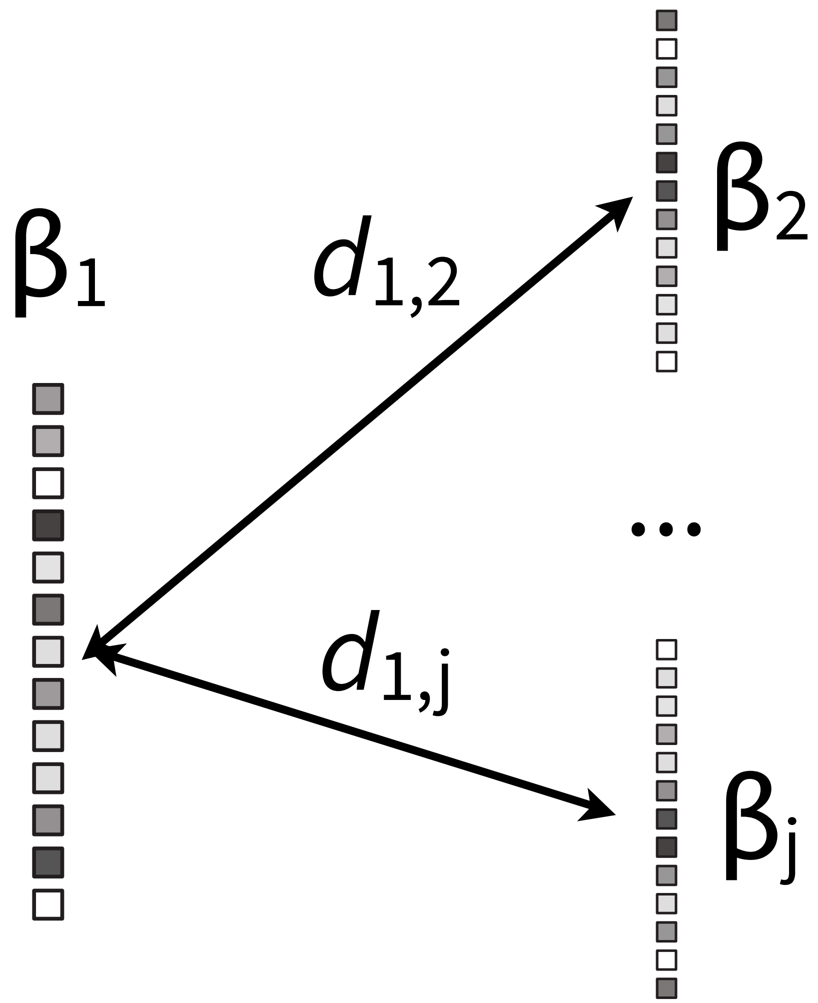


DISTANCE IS DEFINED AS THE CORRELATION DISTANCE BETWEEN TWO BETA-SERIES ACTIVATION PATTERNS

$$d_{ij} = \left[1 - corr(\beta_i, \beta_j)\right]/2$$

Davis & Poldrack, 2013, *Cerebral Cortex*

Computing neural typicality



DISTANCE IS DEFINED AS THE CORRELATION DISTANCE BETWEEN TWO BETA-SERIES ACTIVATION PATTERNS

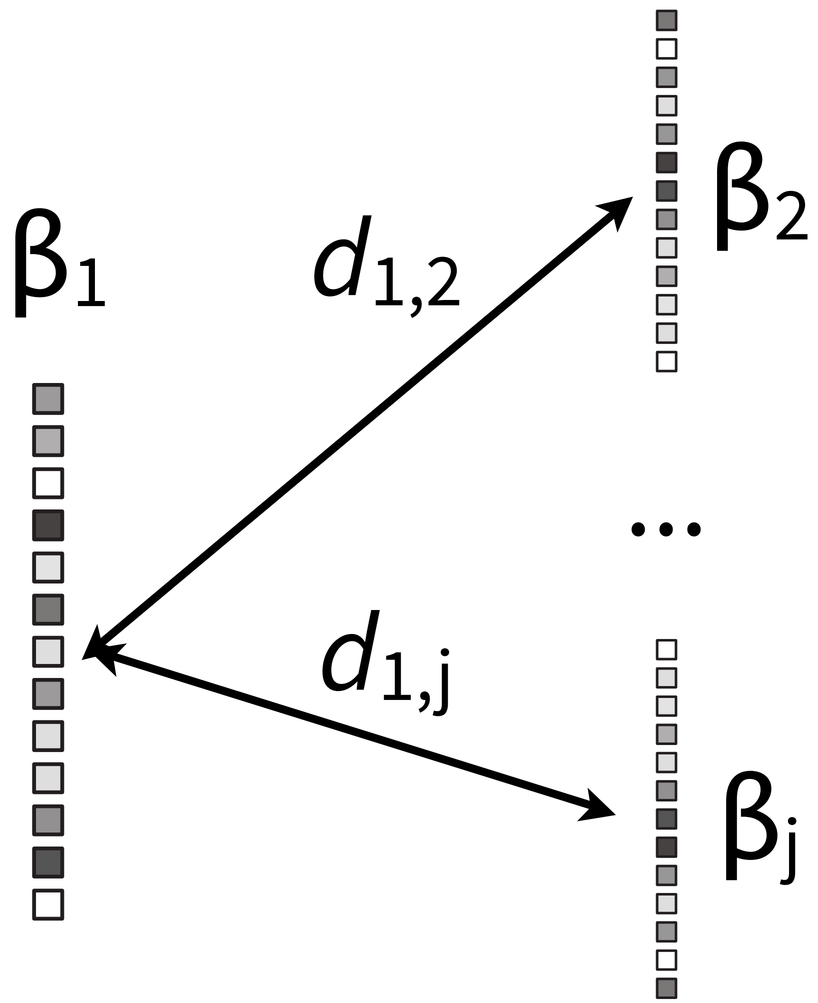
$$d_{ij} = \left[1 - corr(\beta_i, \beta_j)\right]/2$$

SIMILARITY IS AN EXPONENTIAL FUNCTION OF THE DISTANCE BETWEEN TWO REPRESENTATIONS

$$s_{ij} = \exp(-d_{ij})$$

Davis & Poldrack, 2013, *Cerebral Cortex*

Computing neural typicality



DISTANCE IS DEFINED AS THE CORRELATION DISTANCE BETWEEN TWO BETA-SERIES ACTIVATION PATTERNS

$$d_{ij} = \left[1 - corr(\beta_i, \beta_j)\right]/2$$

SIMILARITY IS AN EXPONENTIAL FUNCTION OF THE DISTANCE BETWEEN TWO REPRESENTATIONS

$$s_{ij} = \exp(-d_{ij})$$

$$typ(i \mid J) = \sum_{j \in J} S_{ij}$$

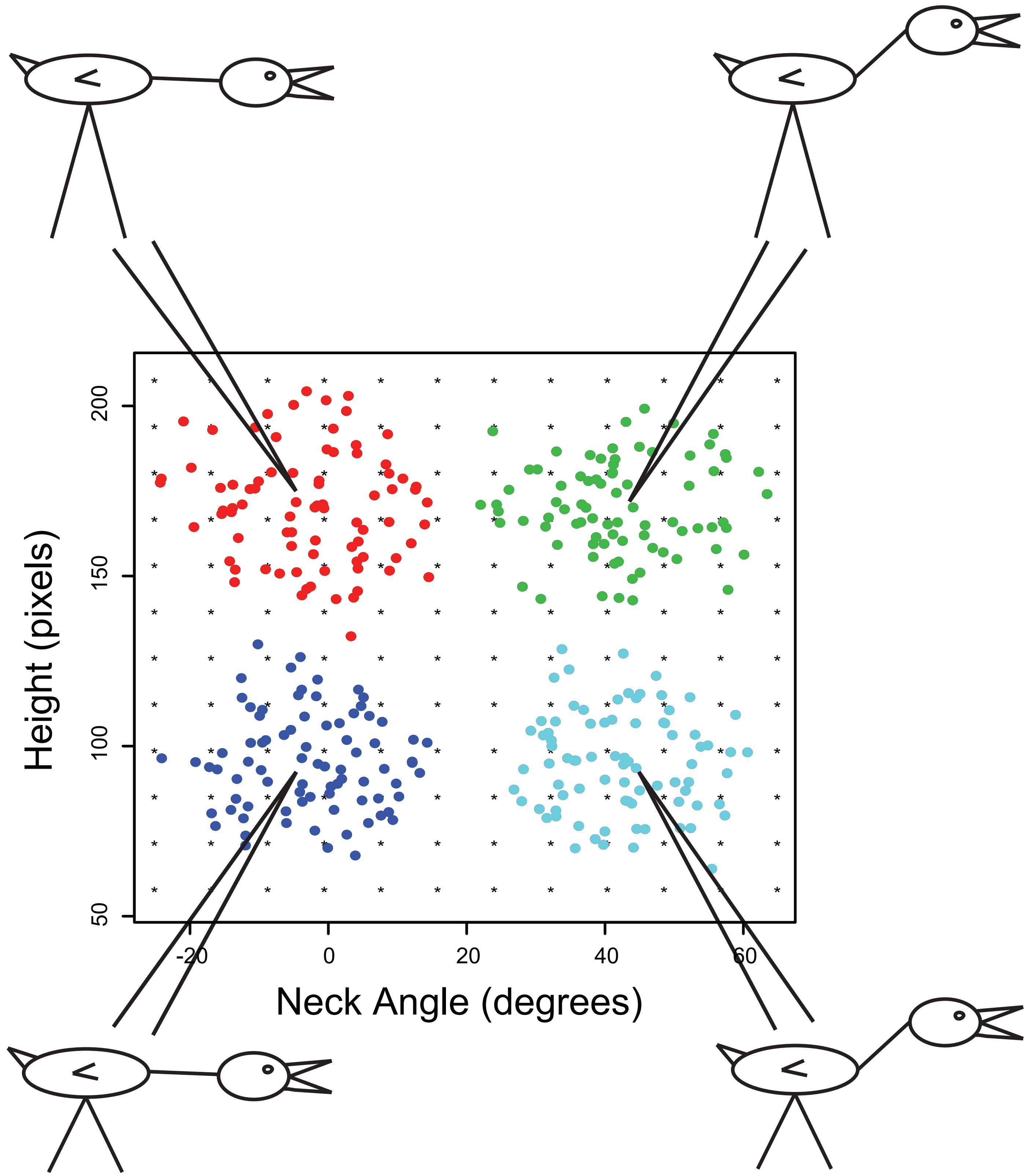
TYPICALITY IS BASED ON THE SUM OF SIMILARITIES BETWEEN REPRESENTATIONS OF AN OBJECT AND OTHER CATEGORY MEMBERS

Davis & Poldrack, 2013, *Cerebral Cortex*

Testing the measure

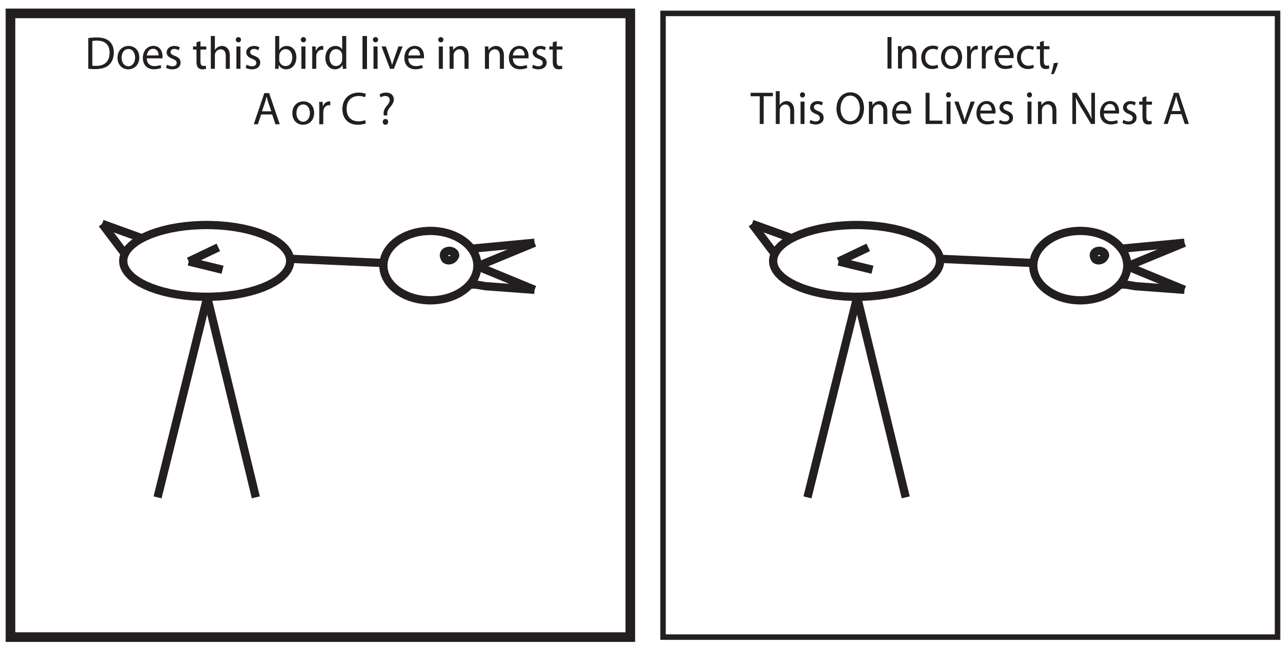
- Is neural typicality associated with subjective typicality ratings?
- Used a task in which subjective typicality and physical feature resemblance are dissociated

Category Structure (Davis & Love, 2010)

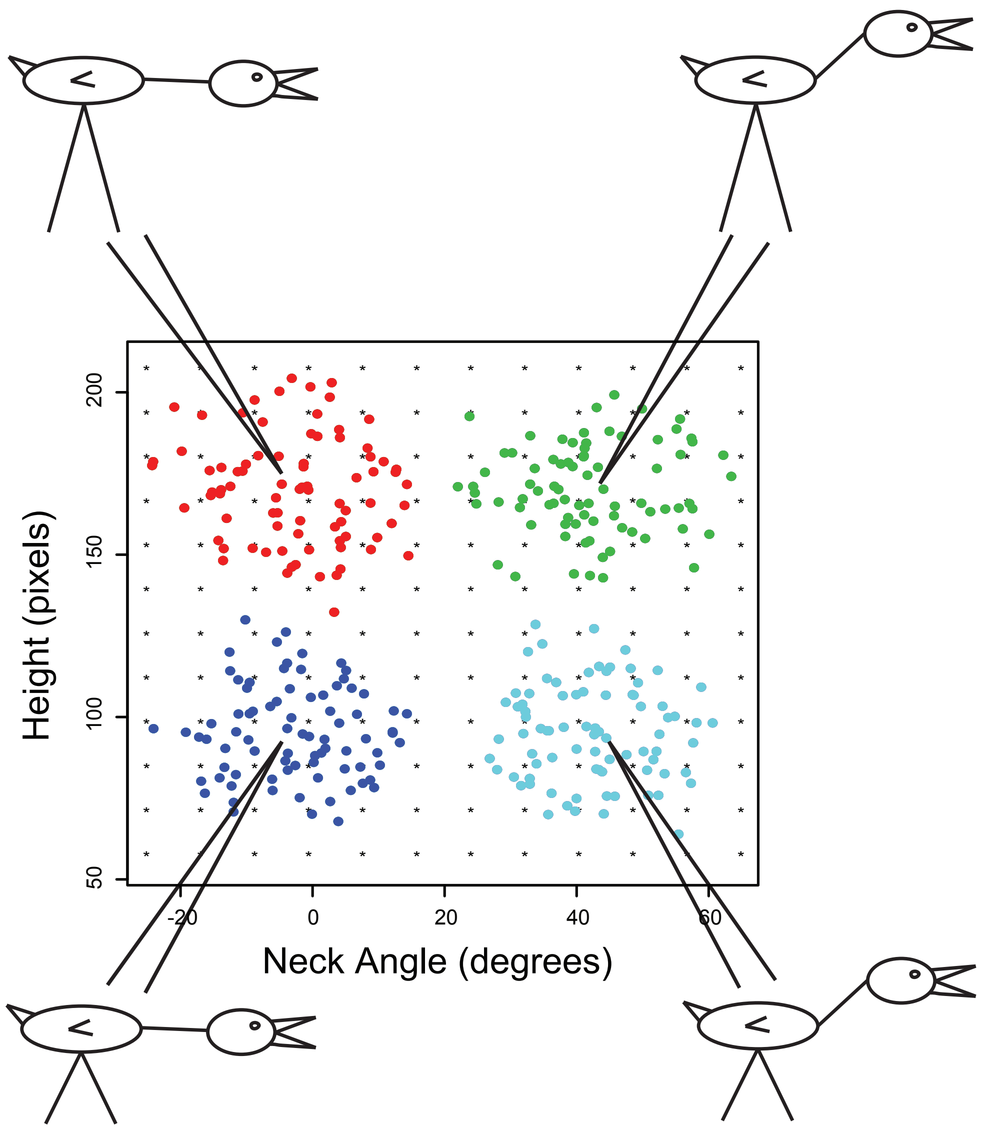


Davis & Poldrack, 2013, *Cerebral Cortex*

Task: Learning phase



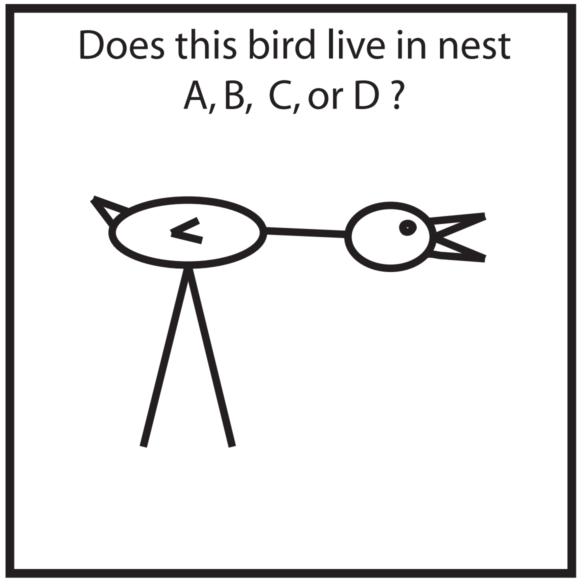
1.5 seconds

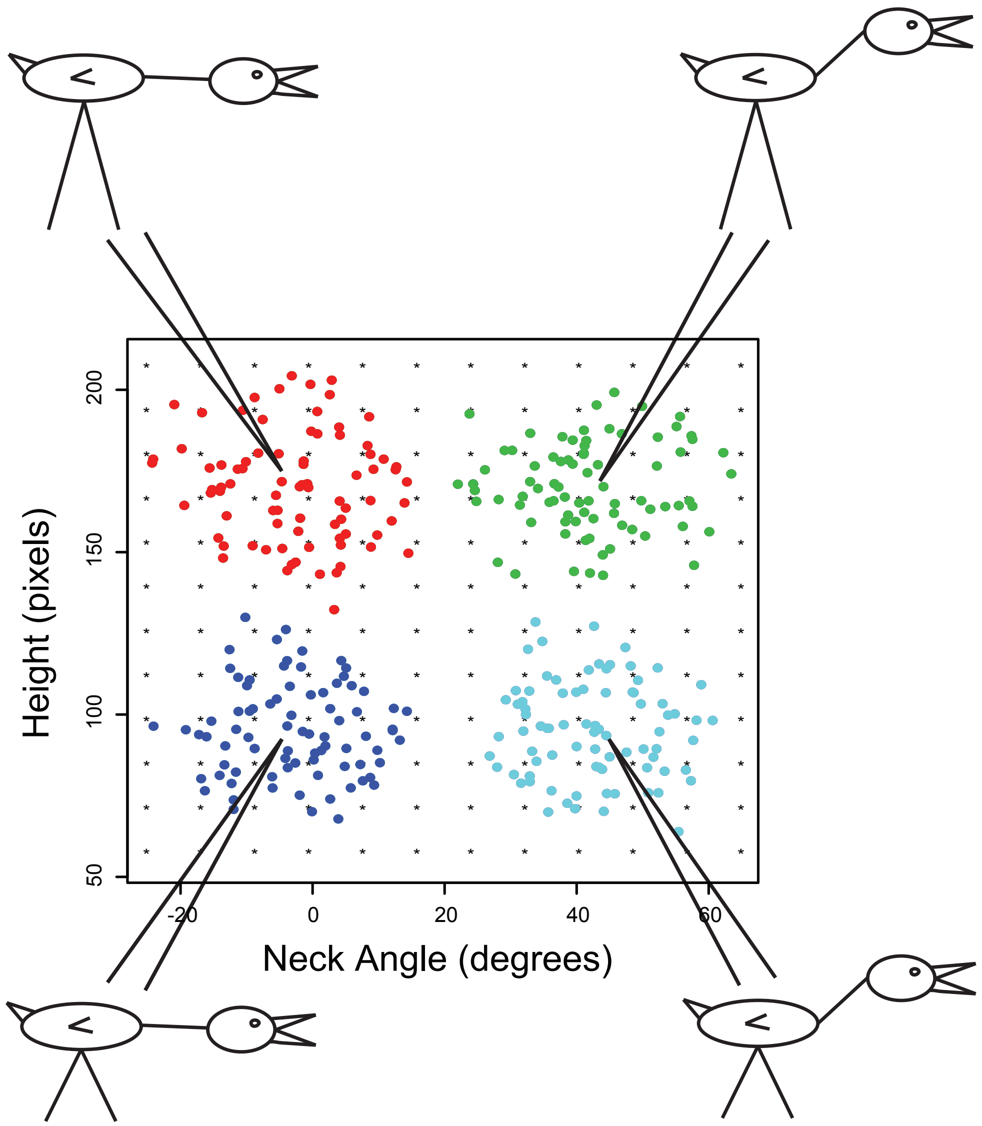


performed during
structural imaging

Davis & Poldrack, 2013, *Cerebral Cortex*

Task: Categorization judgment

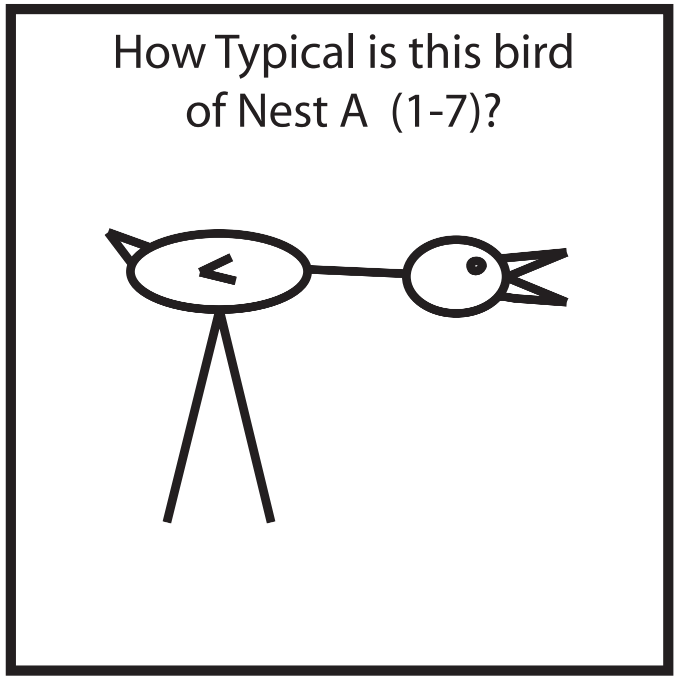


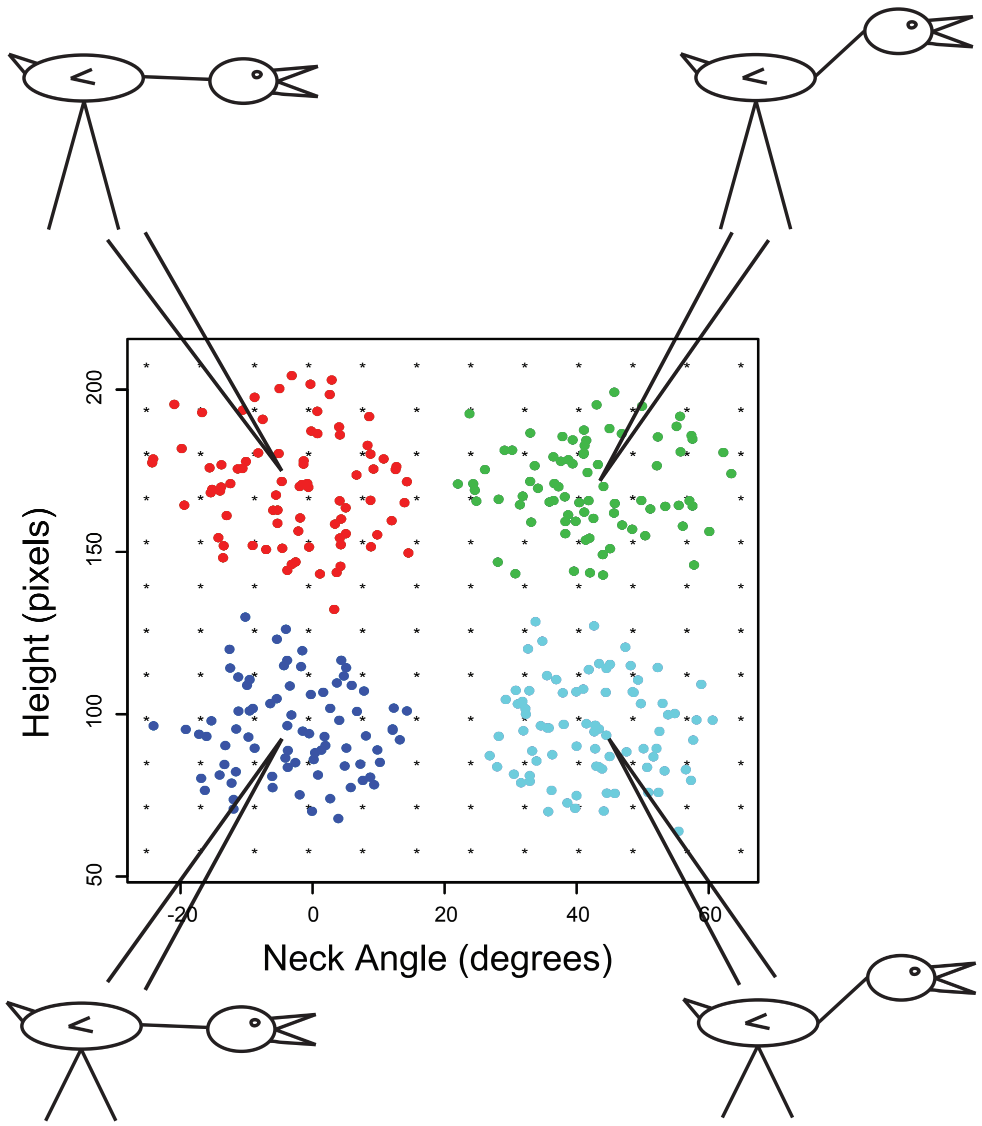


performed while scanning

Davis & Poldrack, 2013, *Cerebral Cortex*

Task: Typicality judgment

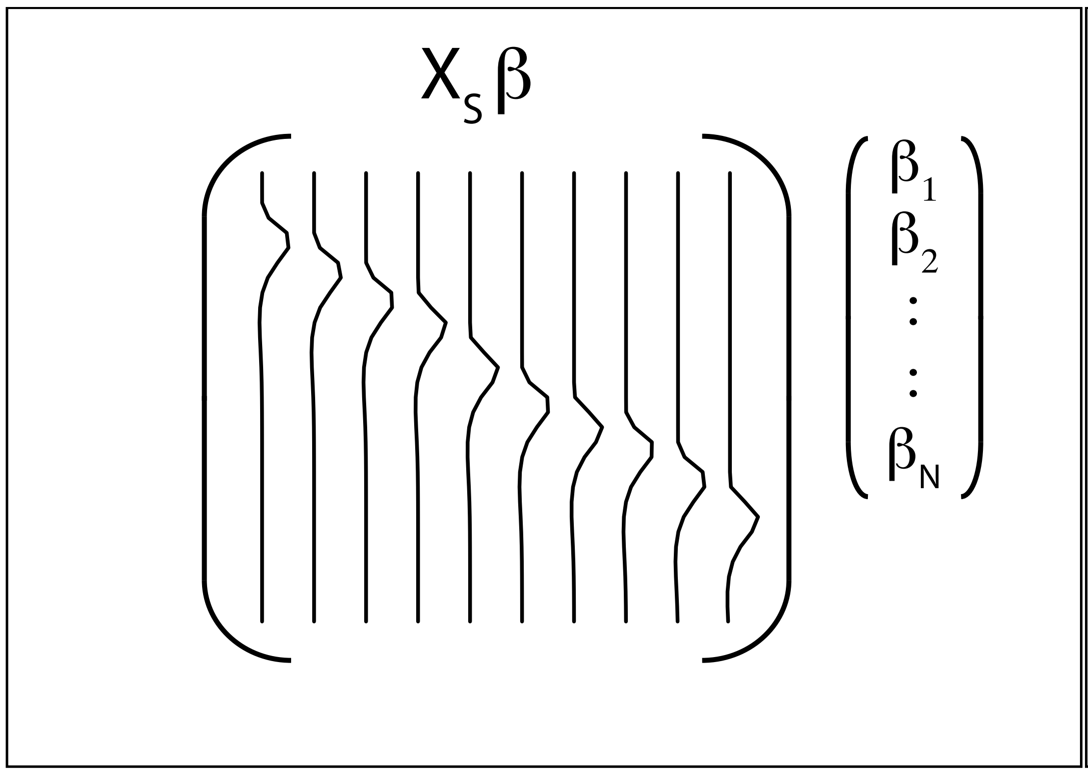




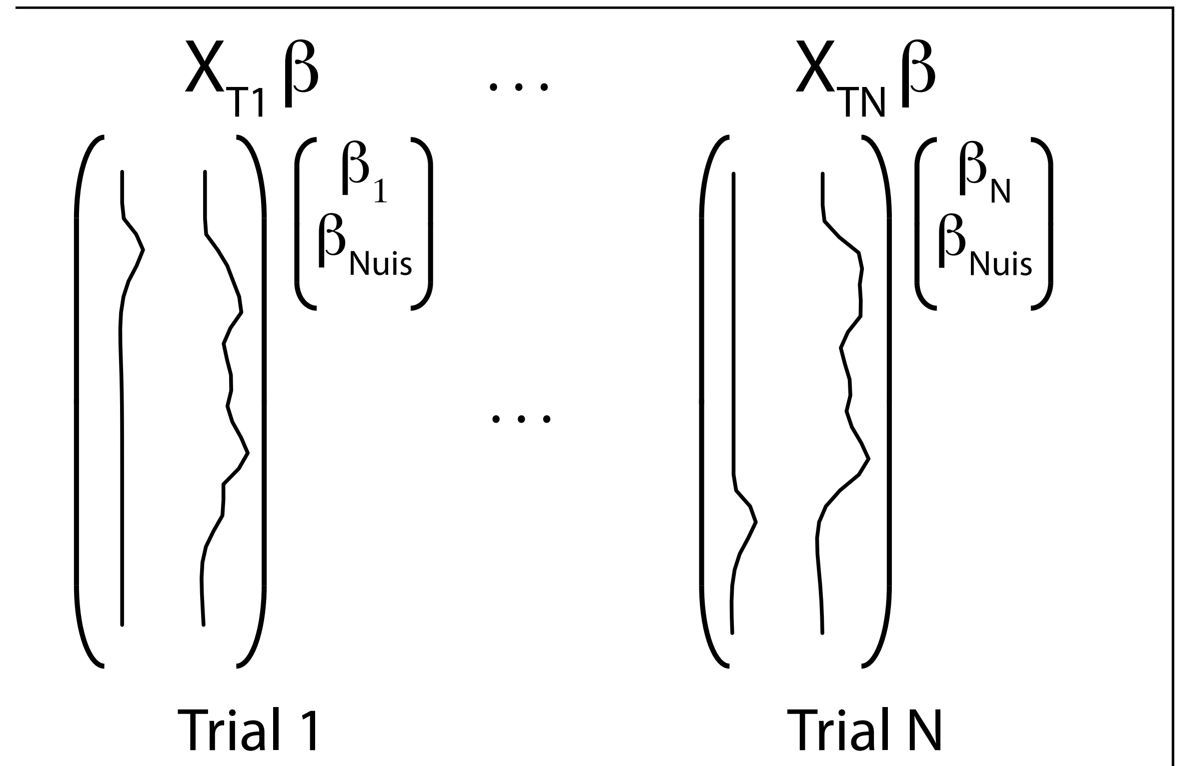
performed outside scanner

Davis & Poldrack, 2013, *Cerebral Cortex*

Single-trial fMRI response estimation



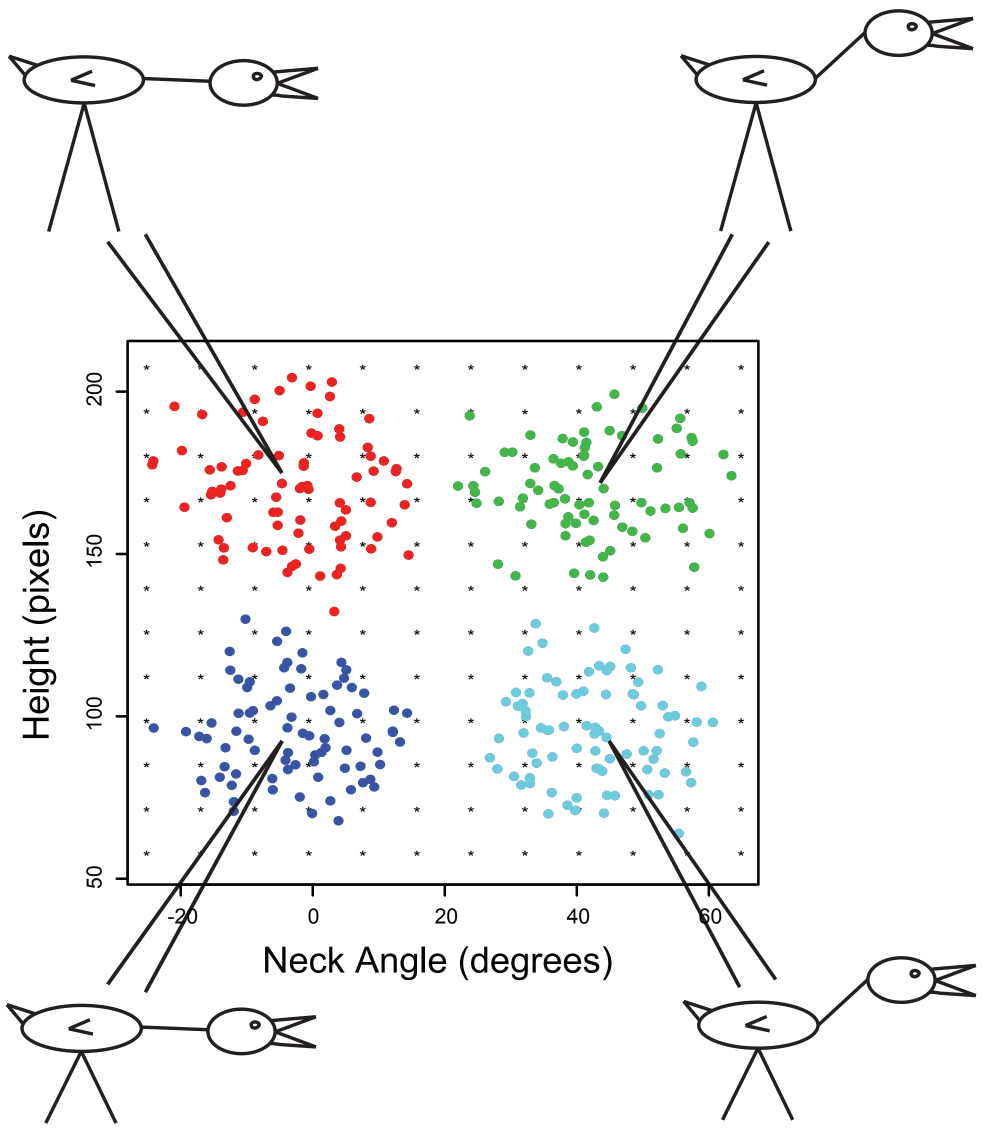
beta-series



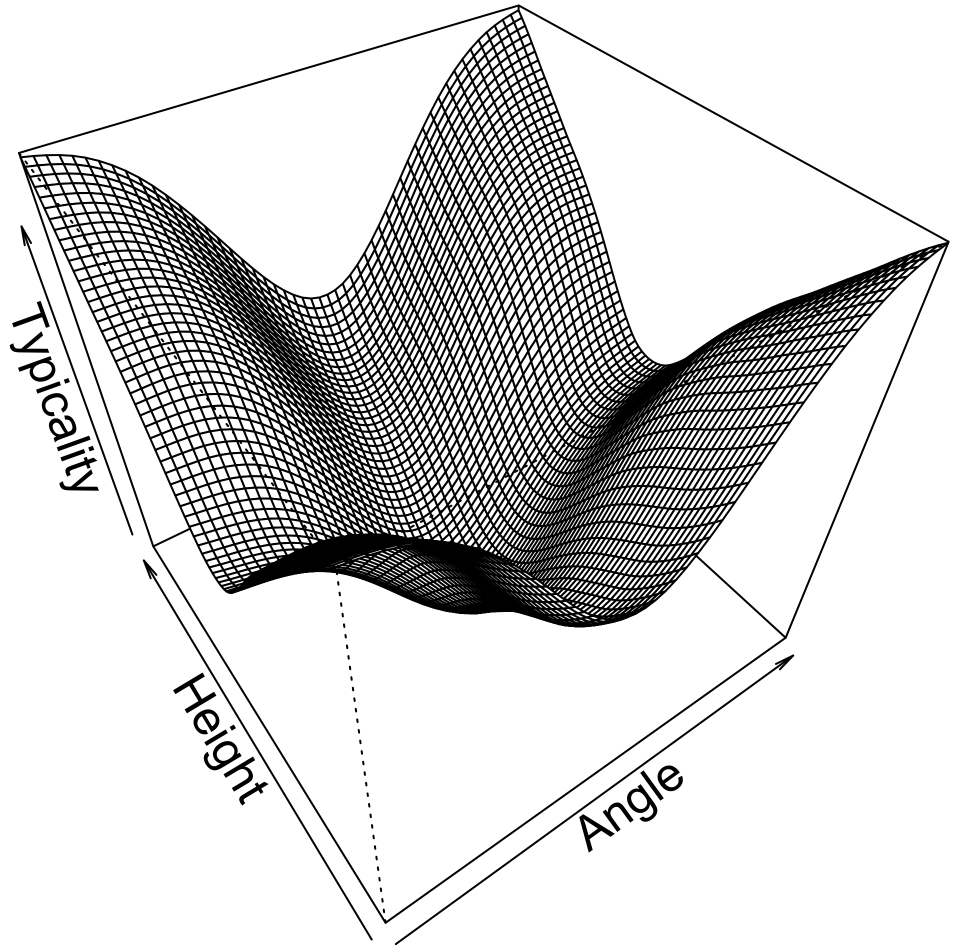
LS-single
(Mumford et al., 2012)

Design matrices for single-trial estimation

Idealized stimuli are judged most typical



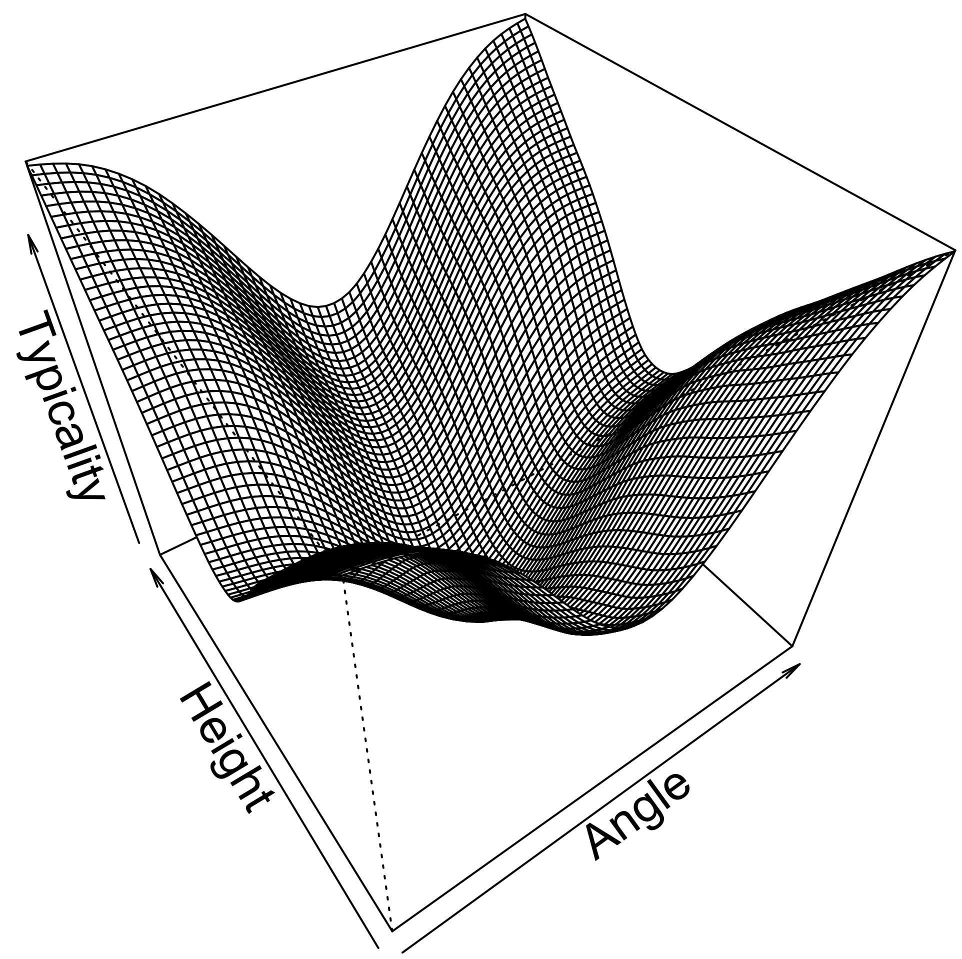
GAM fit to typicality ratings



Davis & Poldrack, 2013, *Cerebral Cortex*

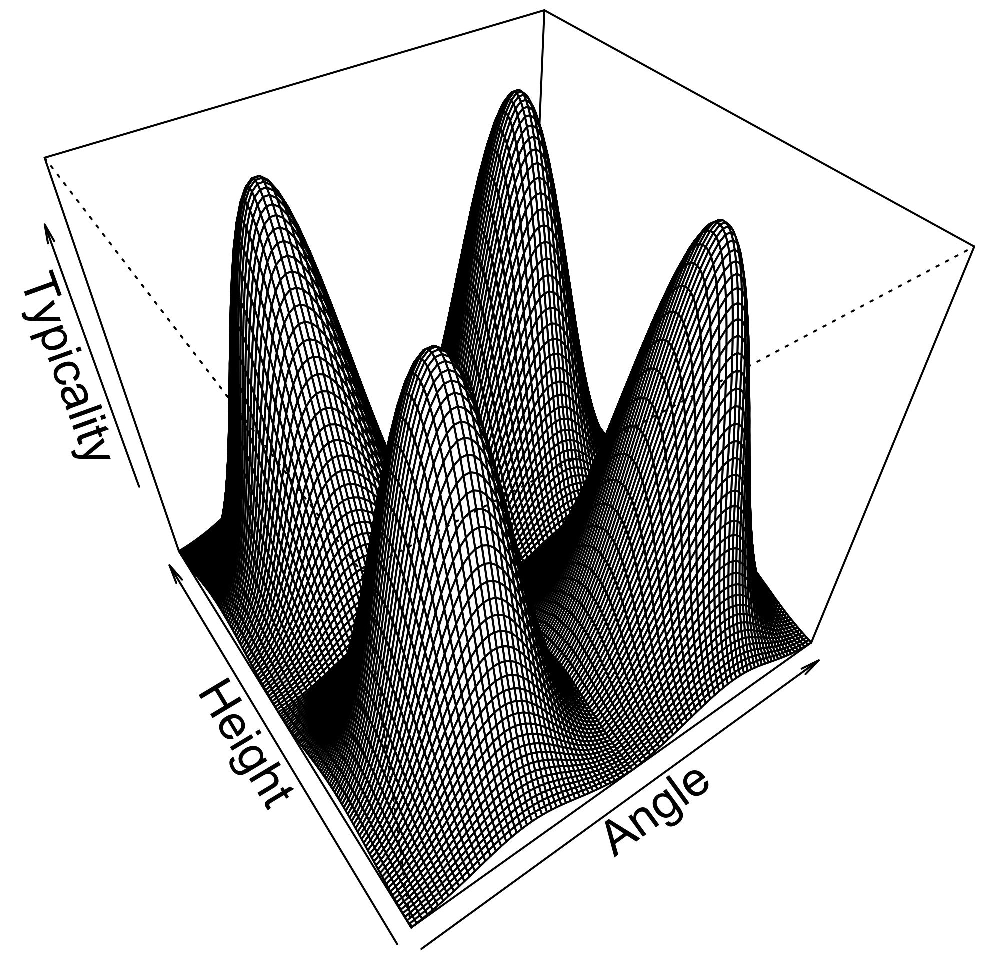
Alternative Predictions

- Neural similarity space will reflect subjective typicality



Obtained from behavior

- Neural similarity space will reflect physical typicality (likelihood given category)

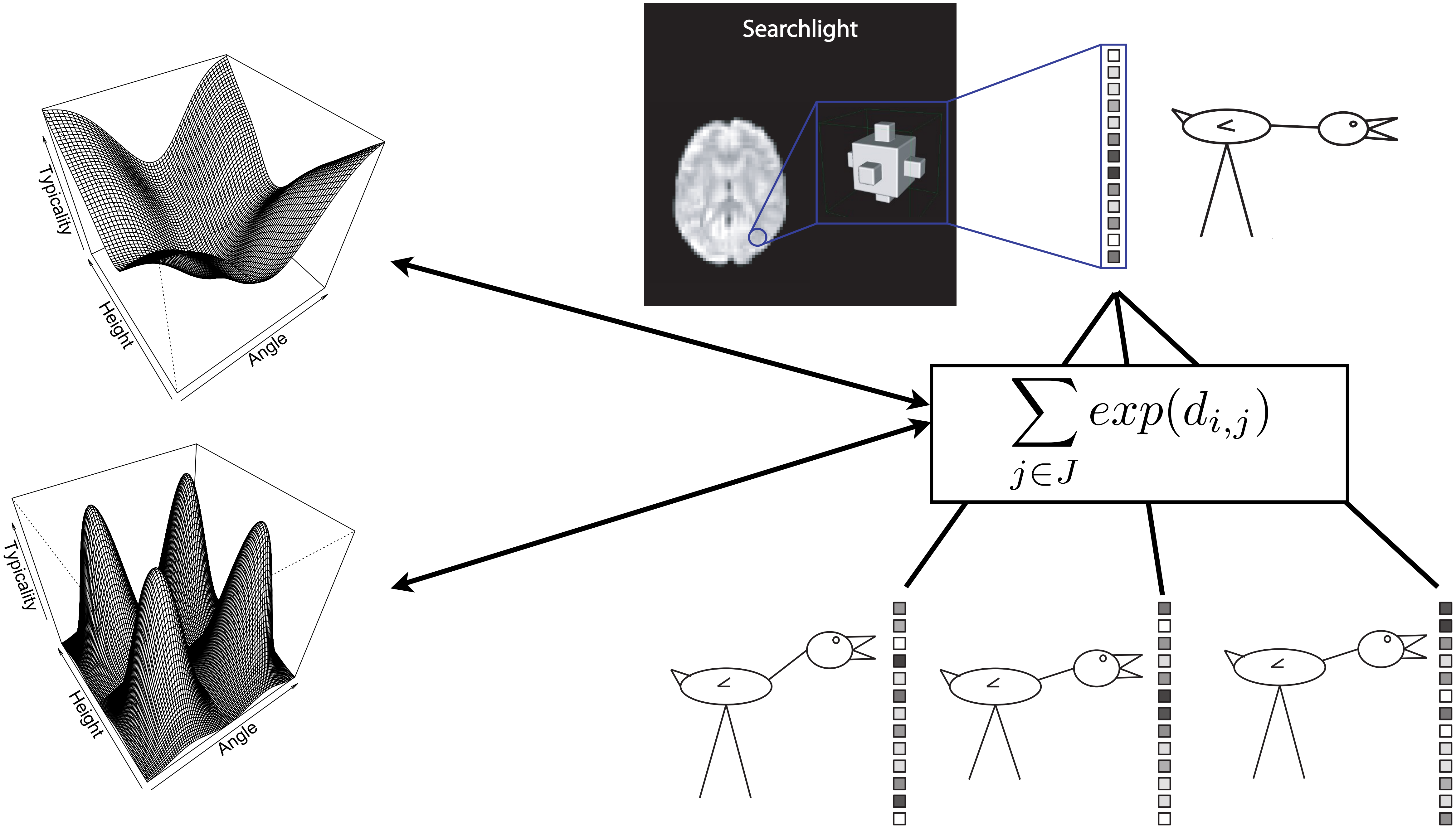


Obtained using GCM

Davis & Poldrack, 2013, *Cerebral Cortex*

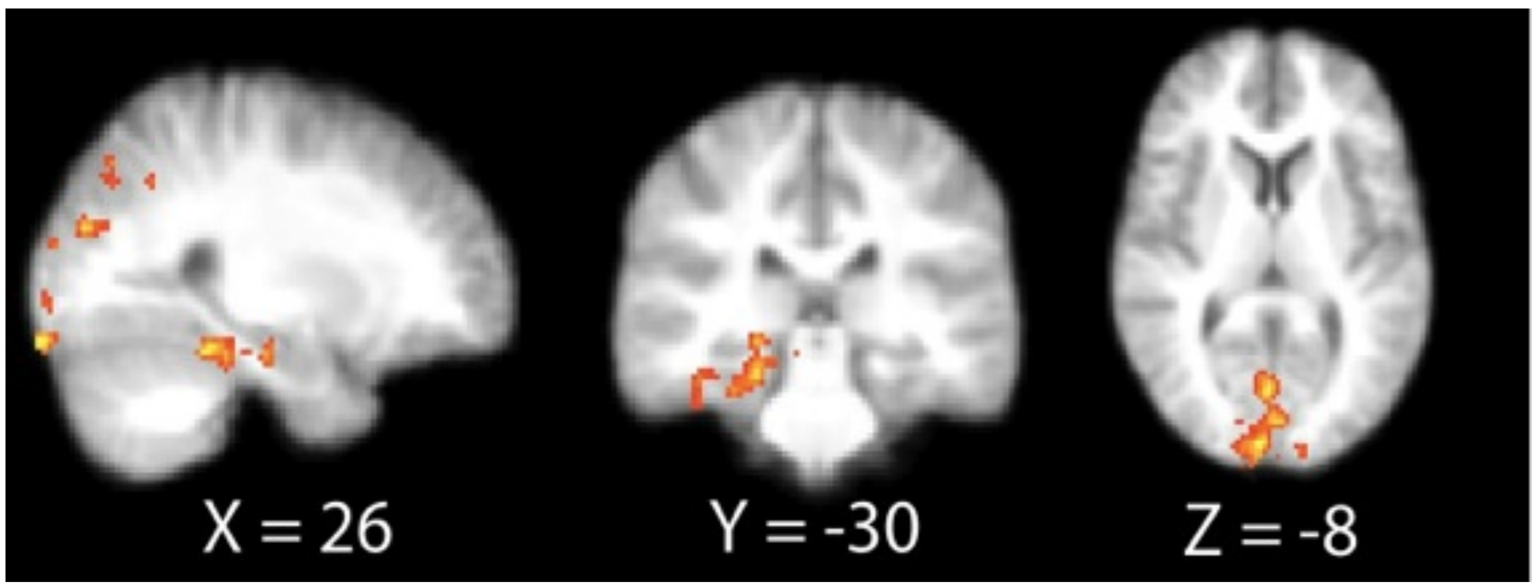
Applying the neural typicality measure

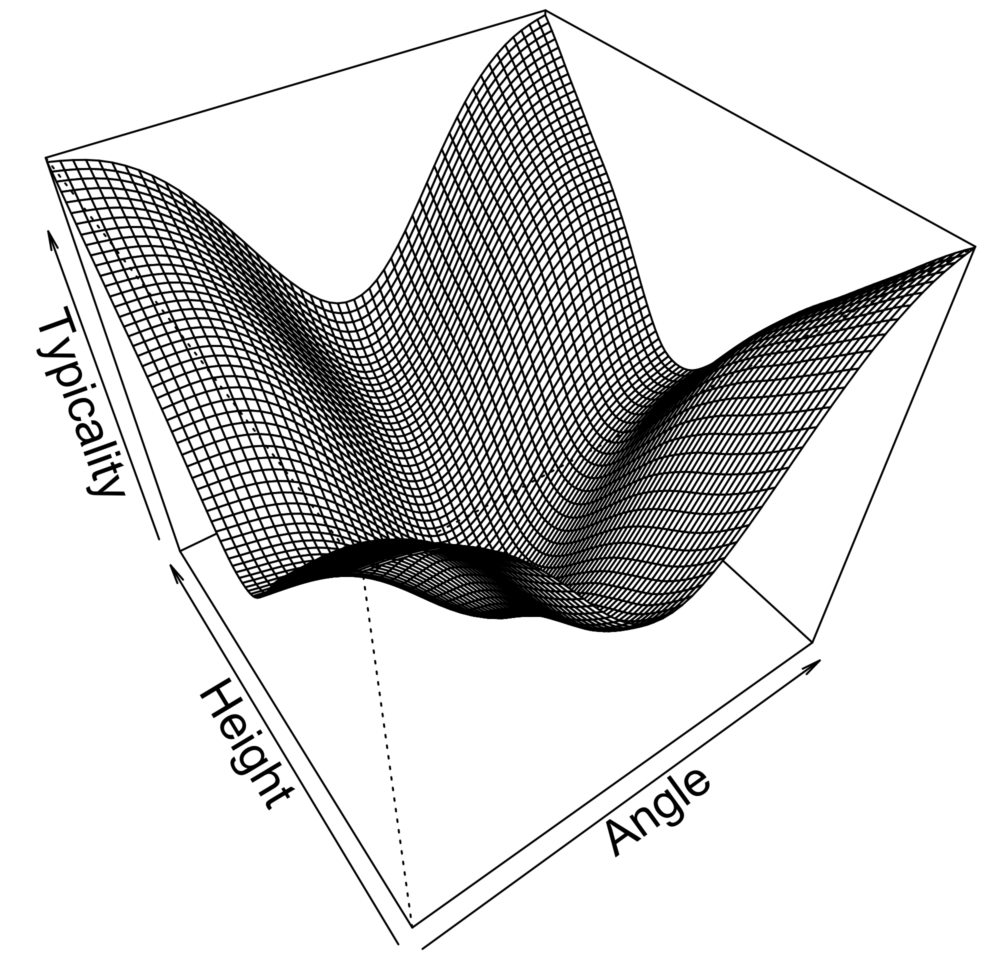
- Correlate neural typicality with psychological and physical predictions



Neural typicality

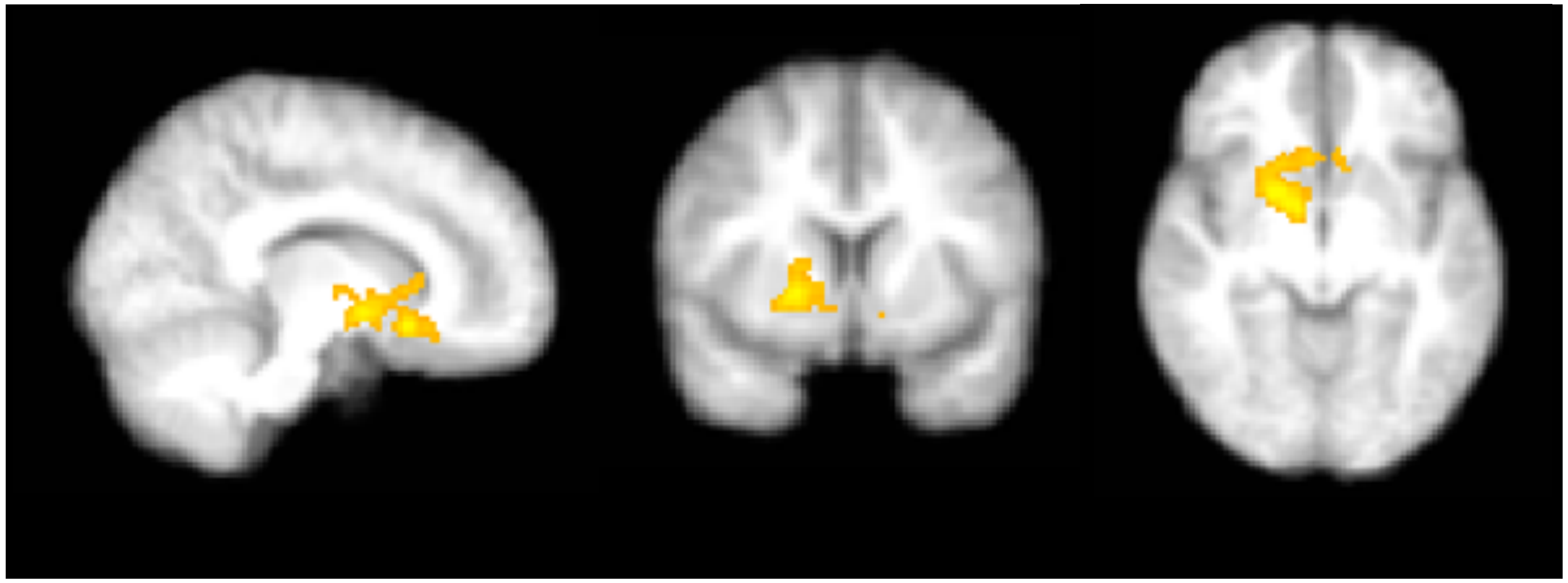
Regions in which neural typicality and psychological typicality are correlated



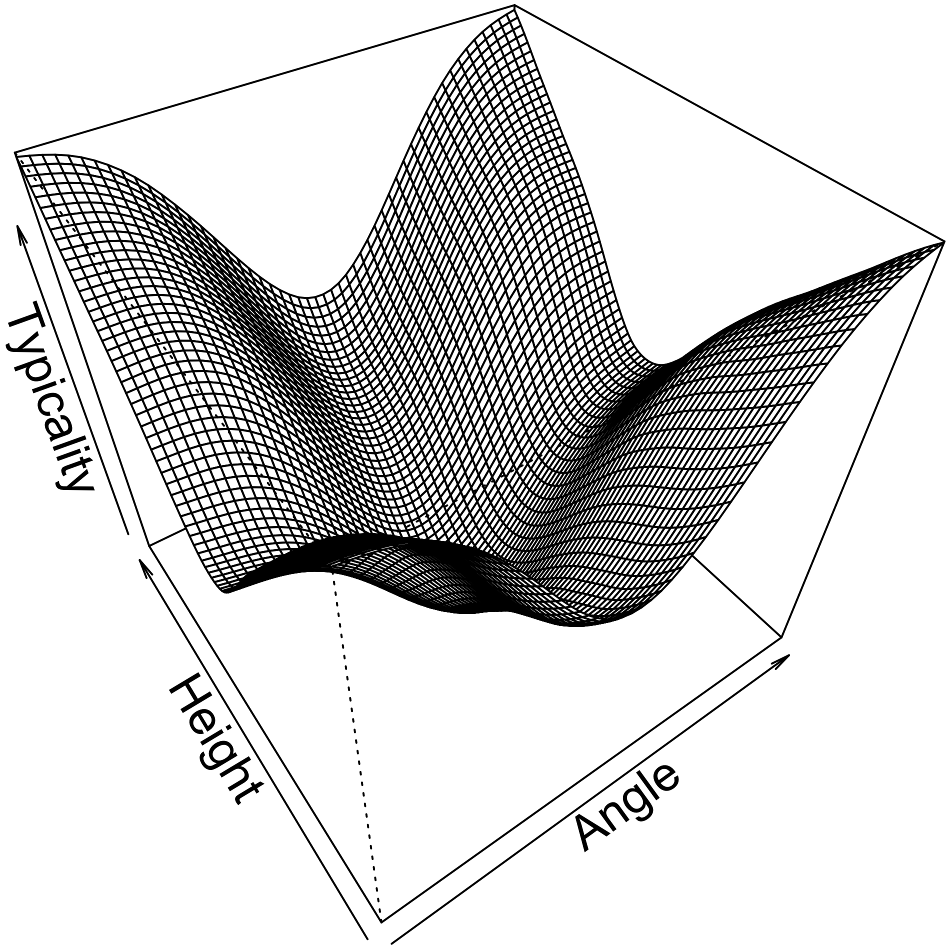


Davis & Poldrack, 2013, *Cerebral Cortex*

Univariate typicality effect



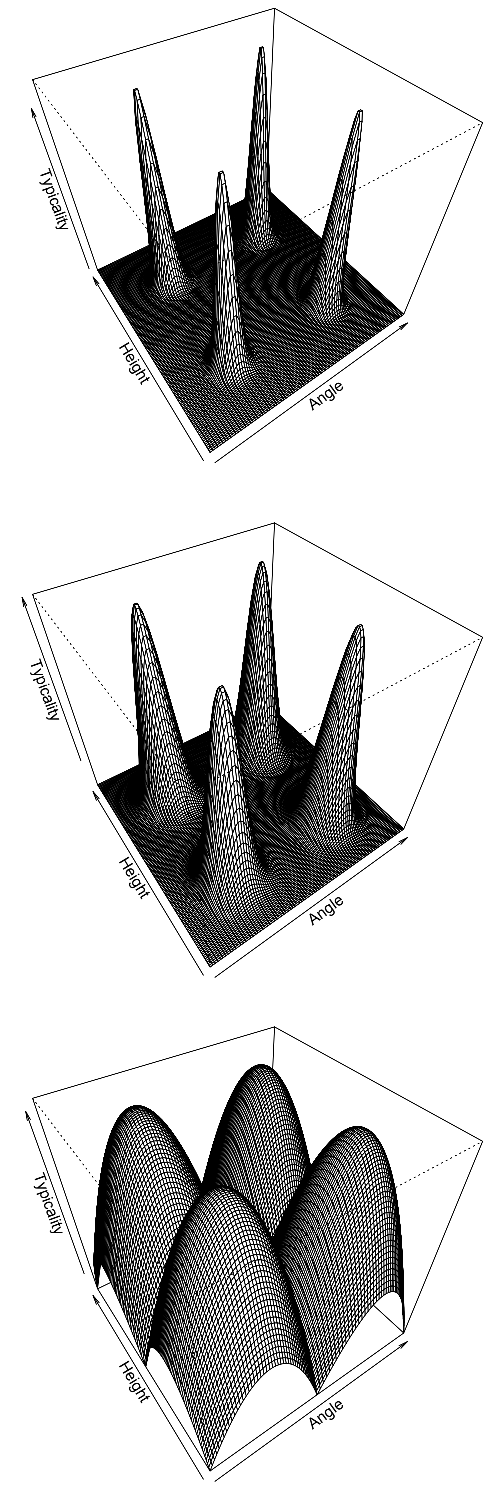
Regions in which univariate activation
and psychological typicality are correlated



Davis & Poldrack, 2013, *Cerebral Cortex*

Physical similarity

- Physical similarity predictions obtained using GCM
    - Examined across multiple levels of variance
- There were no regions in which neural typicality or activation reflected physical typicality/

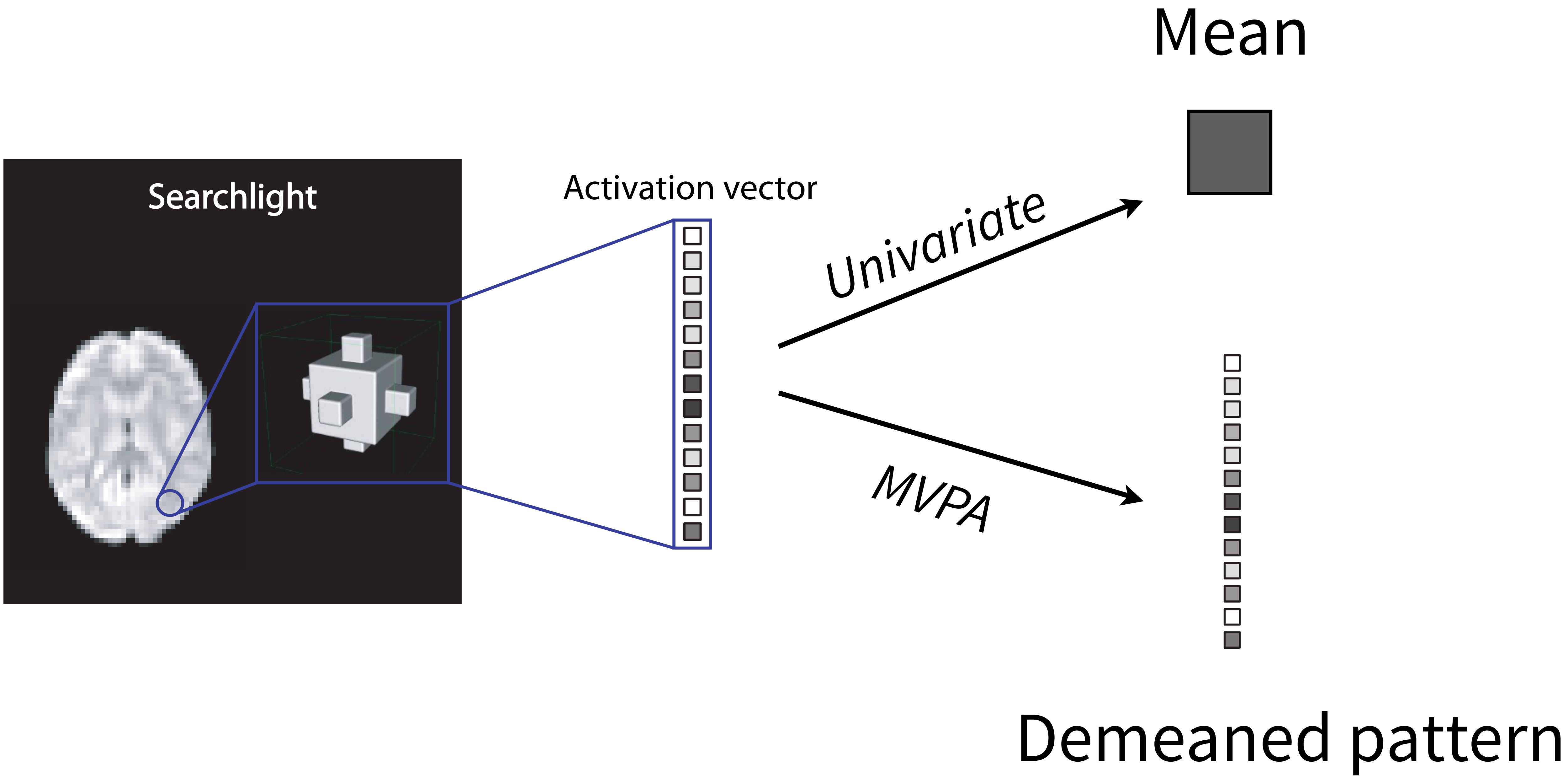


Davis & Poldrack, 2013, *Cerebral Cortex*

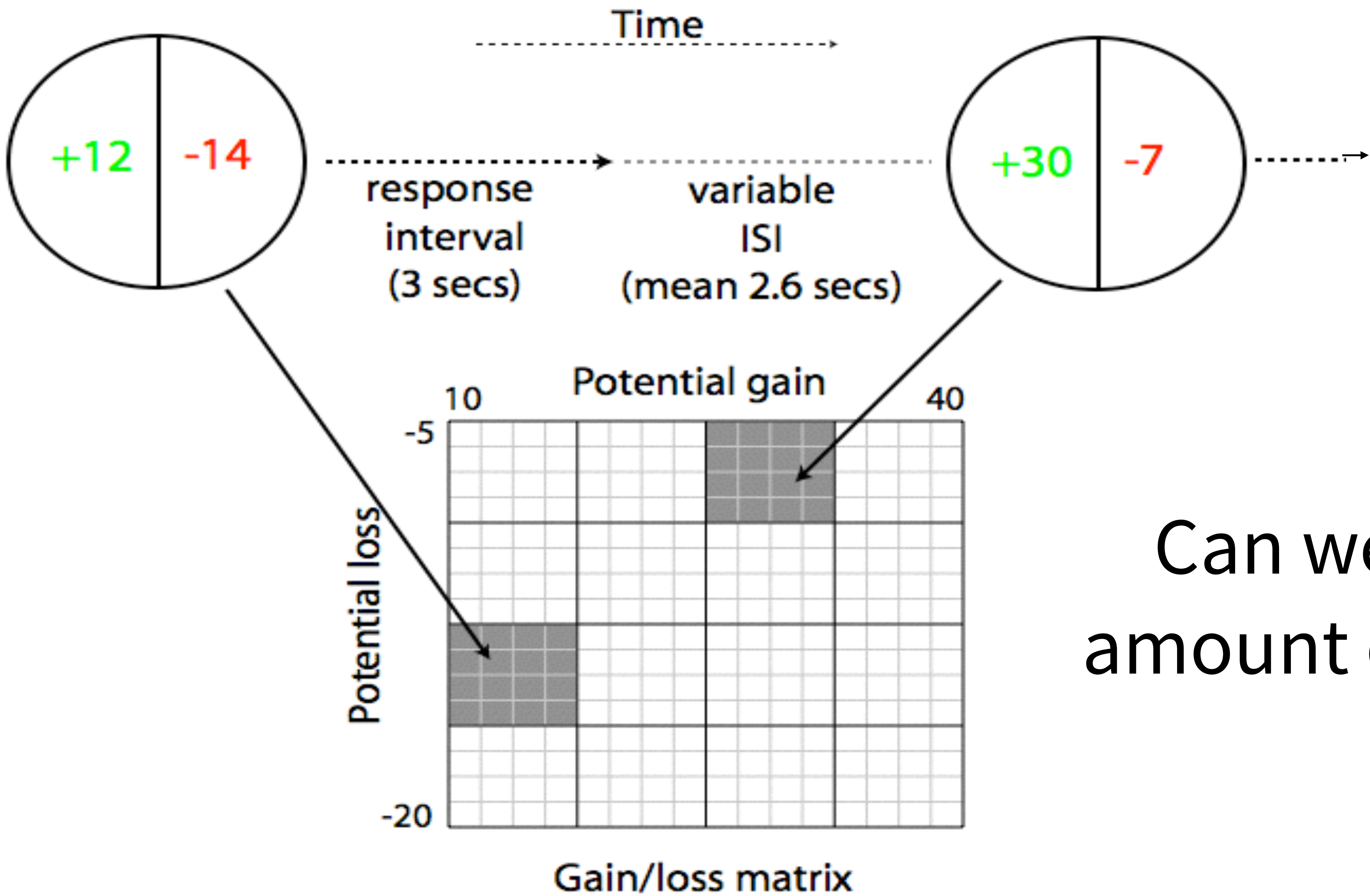
Implications

- Activation patterns are isomorphic to mental representations
- Subjective typicality reflected in neural typicality
  - Univariate analysis would have told a very different story

How does MVPA compare to univariate analysis?



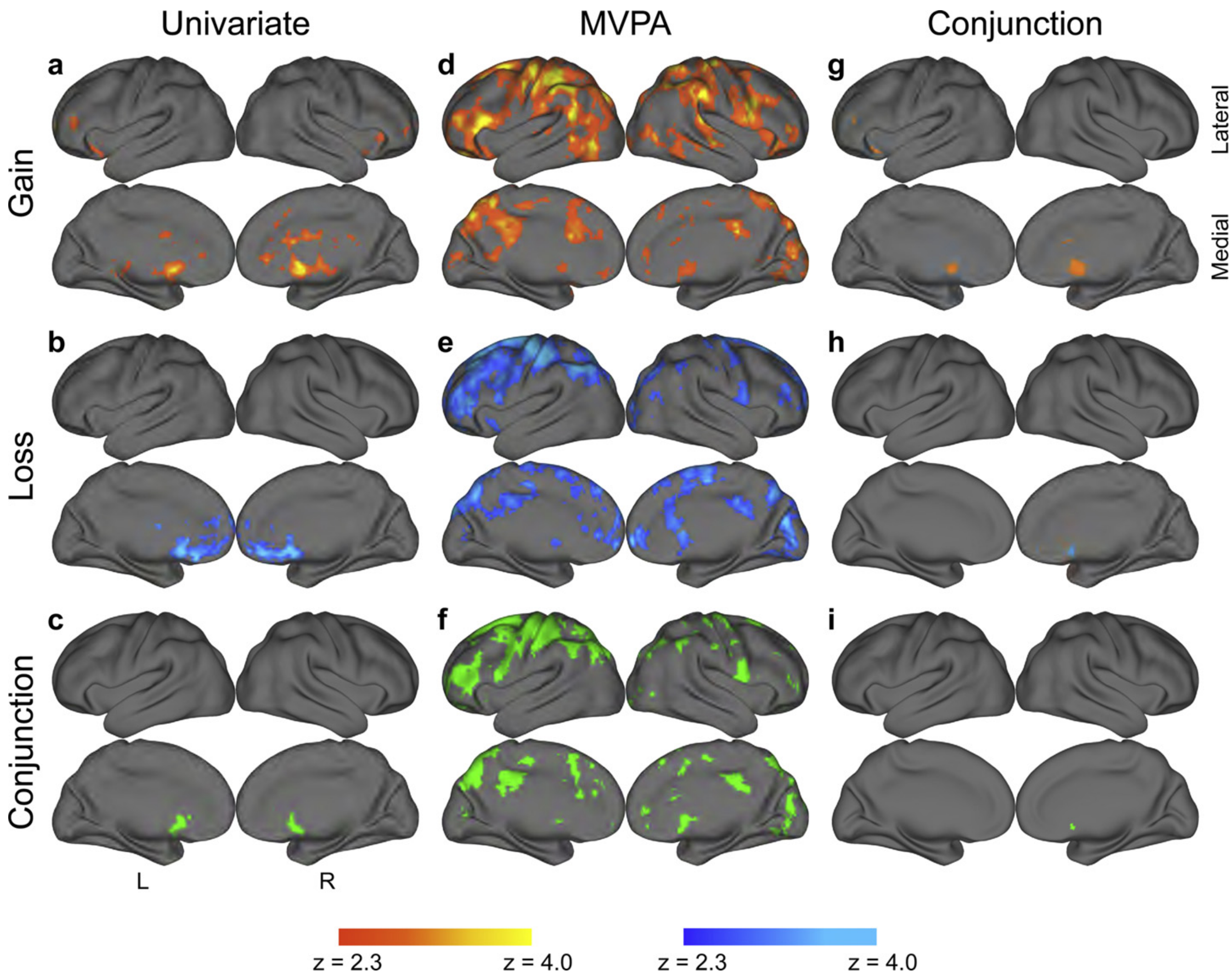
Decoding continuous variables



Can we decode the amount of gain or loss?

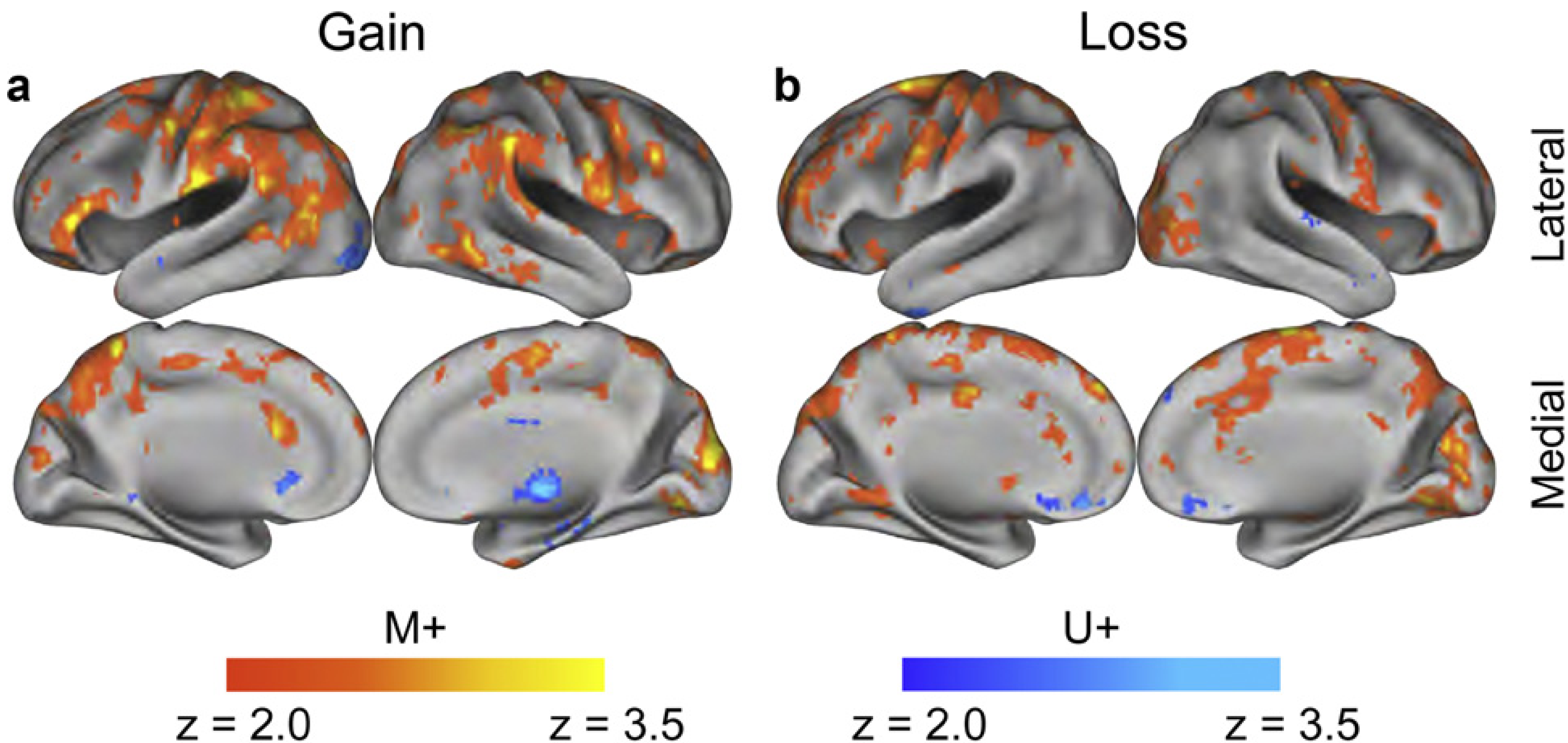
Tom et al., 2007, *Science*

The opportunistic nature of MVPA



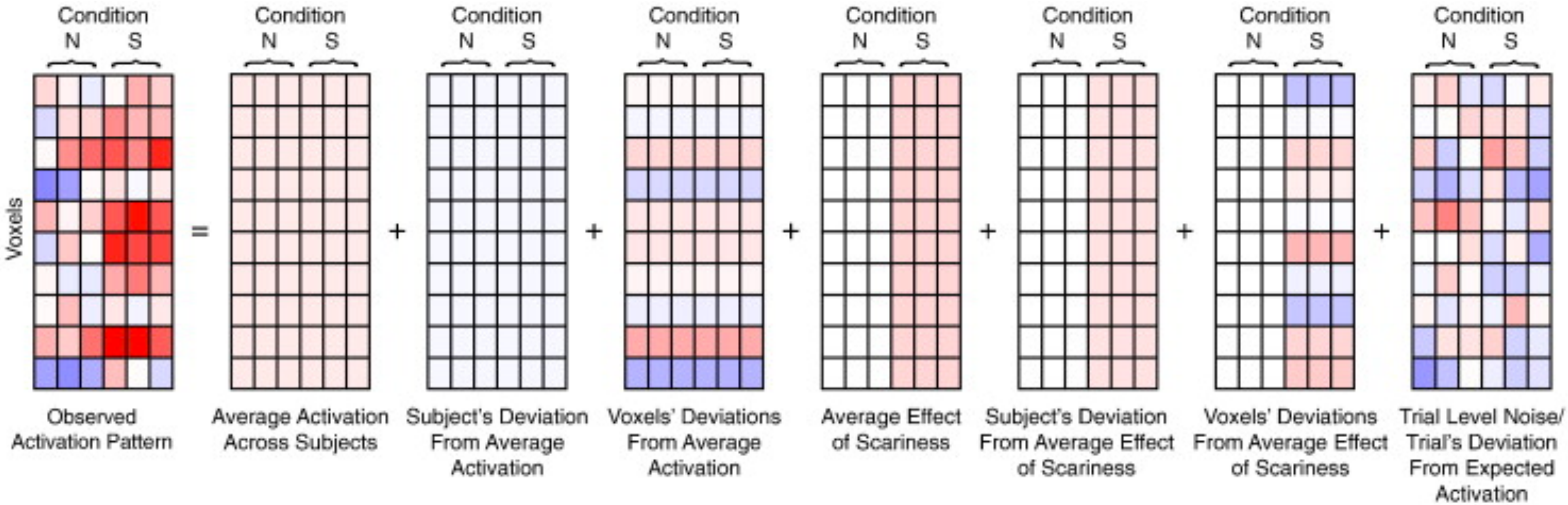
Jimura & Poldrack, 2011, *Neuropsychologia*

Differential sensitivity of MVPA



Jimura & Poldrack, 2011, *Neuropsychologia*

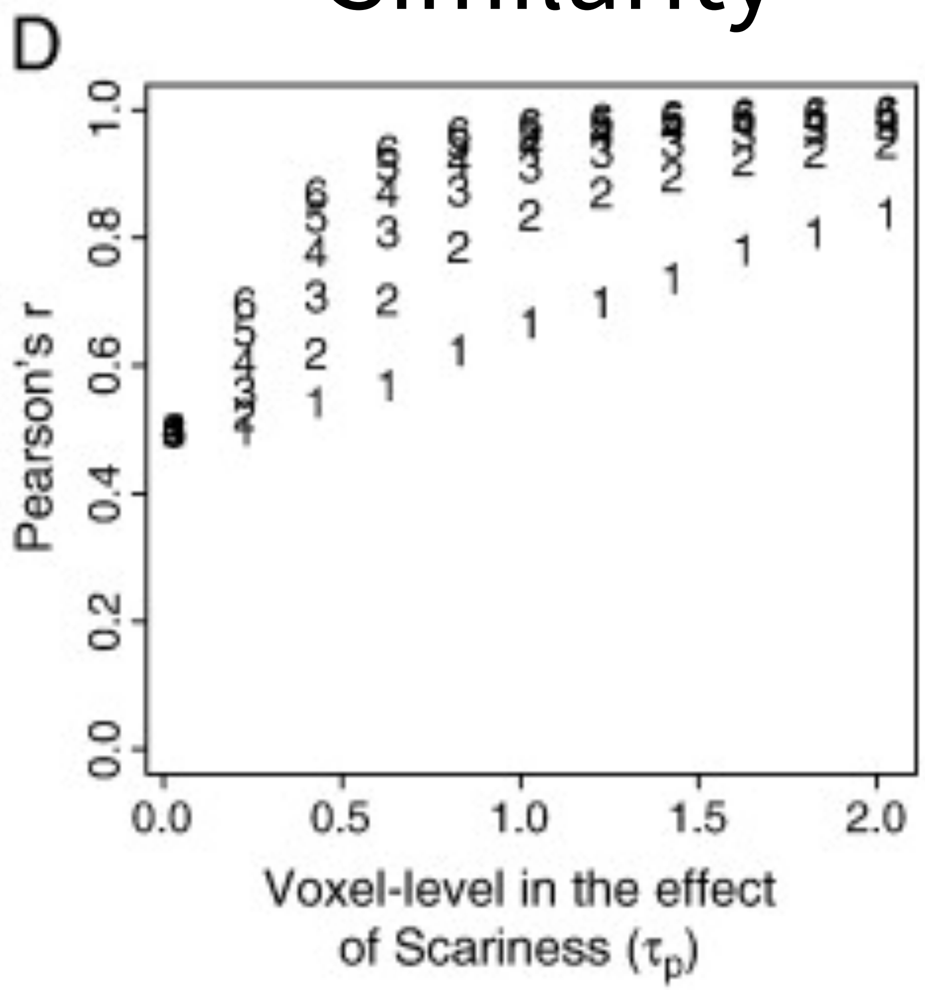
MVPA is sensitive to univariate patterns



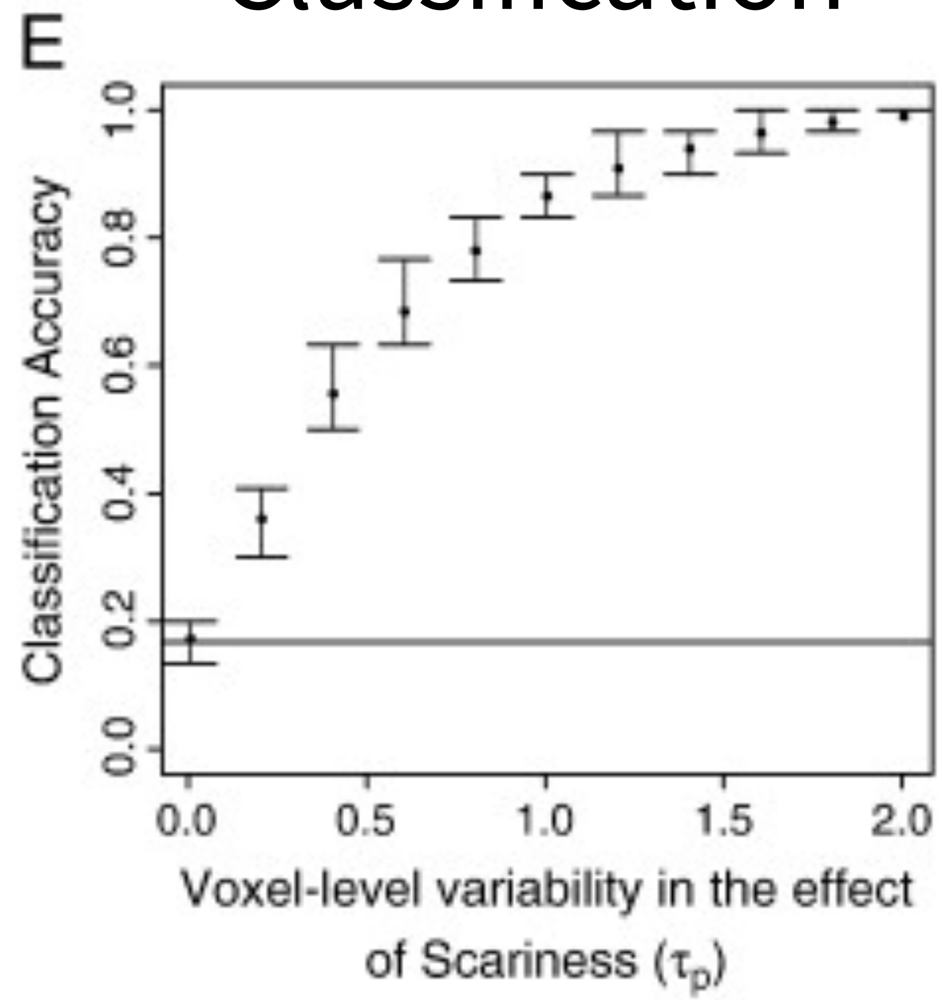
Davis, Laroque et al., 2014, *Neuroimage*

Voxelwise variability in univariate signals can drive MVPA

Similarity



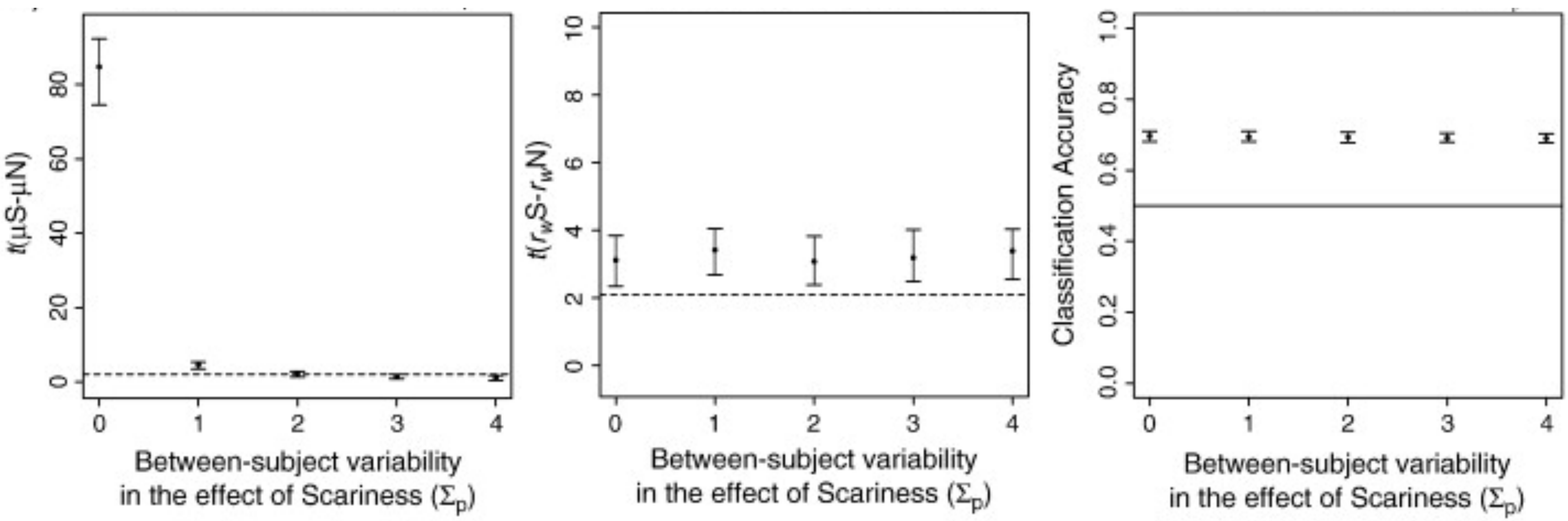
Classification



Davis, Laroque et al., 2014, *Neuroimage*

MVPA is less sensitive to between-subject variability

Univariate | Similarity | Classification



Davis, Laroque et al., 2014, *Neuroimage*

Some things we have learned

- If classification results look too good, you have most likely done something wrong
- Always confirm results by randomizing data from the very beginning
    - run many times to get null distribution, make sure it's actually at chance
- Crossvalidation with regression is very tricky (don't use LOO)
- Differences between univariate and multivariate analyses can't be easily interpreted (Davis et al., 2014, *Neuroimage*)
- Trials orders must be separately randomized for each subject (Mumford et al., 2014, *Neuroimage*)

Conclusions

- Neuroimaging data CAN provide evidence relevant to psychological questions
    - But informal reverse inference is not the way!
- Machine learning methods provide the means to decode and predict mental states from neuroimaging data
- Multivariate analyses can establish isomorphisms between neural and mental representations

Acknowledgments

Stanford
Sanmi Koyejo
Chris Filo Gorgolewski
Karen LaRoque
Anthony Wagner

UT Austin
Jeanette Mumford
Sarah Helfinstein
Tom Schonberg
Tyler Davis
Tal Yarkoni

UCLA
Jessica Cohen
Robert Bilder
Eliza Congdon
Eydie London
Tyrone Cannon
Nelson Freimer

Rutgers
Stephen J. Hanson
Yaroslav Halchenko

Princeton
Ken Norman

NIMH
National Institute of Mental Health

James S. McDonnell Foundation

Department of the Navy
ONR
Science & Technology

Data sets and code will be made available at www.openfmri.org