Inferring mental states from neuroimaging data

Russell Poldrack

Departments of Psychology
and Neurobiology

Imaging Research Center
University of Texas at Austin

The New York Times

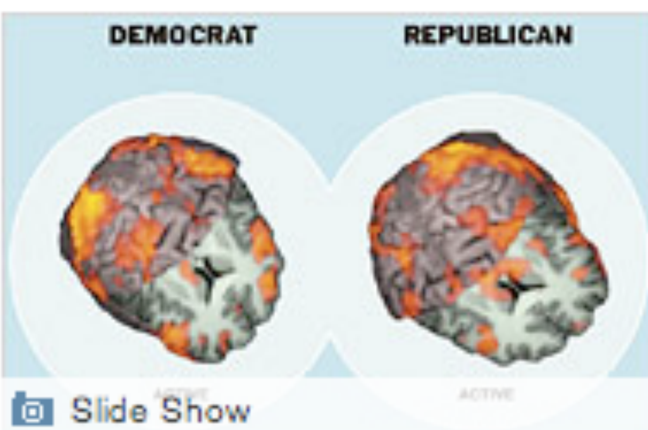
OP-ED CONTRIBUTORS

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.

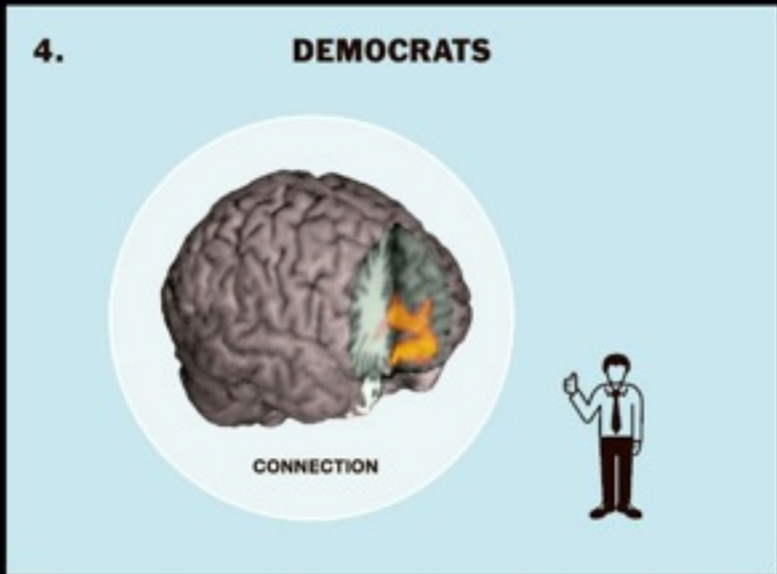
Multimedia



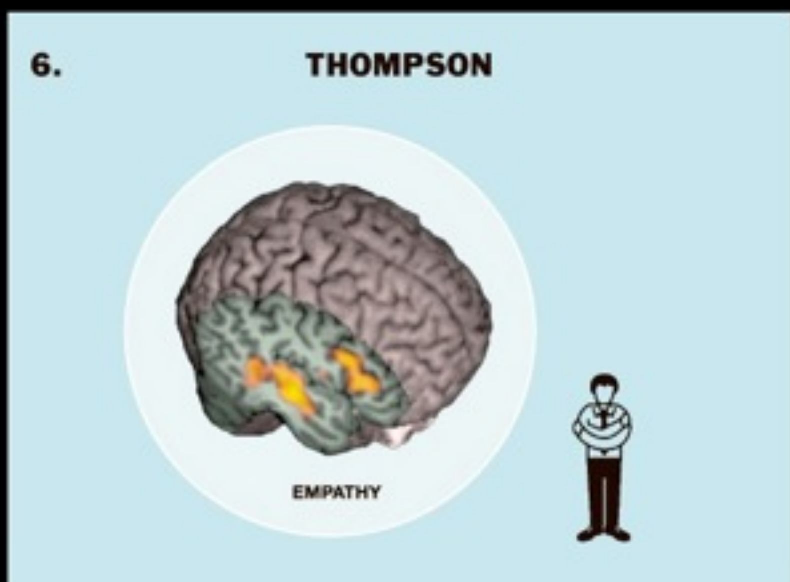
This Is Your Brain on Politics

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.

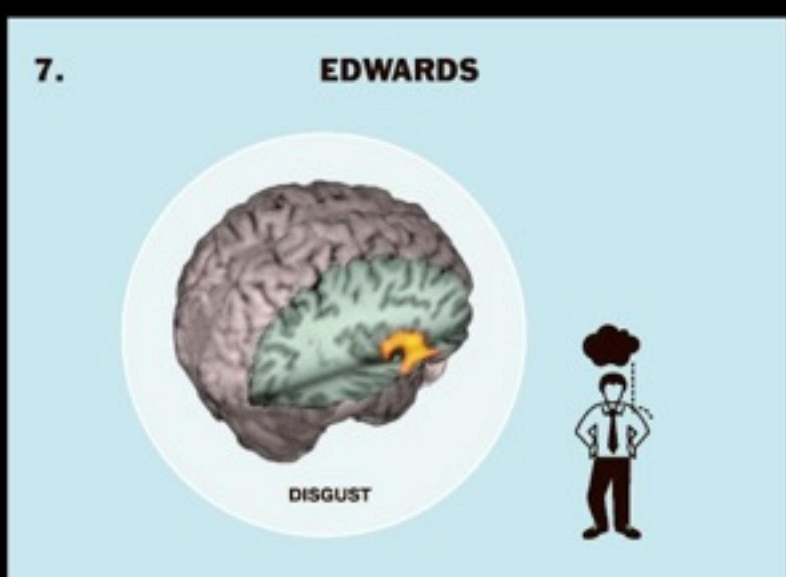
Our 20 subjects — registered voters who stated that they were open to choosing a candidate from either party next November — included 10 men and 10 women. In late summer, we asked them to answer a list of questions about their political preferences, then observed their brain activity for nearly an hour 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."

LETTER; Politics and the Brain

Published: November 14, 2007

To the Editor:

''This Is Your Brain on Politics'' (Op-Ed, Nov. 11) used the results of a brain imaging study to draw conclusions about the current state of the American electorate. The article claimed that it is possible to directly read the minds of potential voters by looking at their brain activity while they viewed presidential candidates.

For example, activity in the amygdala in response to viewing one candidate was argued to reflect ''anxiety'' about the candidate, whereas activity in other areas was argued to indicate ''feeling connected.'' While such reasoning appears compelling on its face, it is scientifically unfounded.

As cognitive neuroscientists who use the same brain imaging technology, we know that it is not possible to definitively determine whether a person is anxious or feeling connected simply by looking at activity in a particular brain region. This is so because brain regions are typically engaged by many mental states, and thus a one-to-one mapping between a brain region and a mental state is not possible.

For example, rather than simply providing a brain marker of anxiety levels, as the article assumed, we know that the amygdala is activated by arousal and positive emotions as well. Such problems of interpretation with brain imaging studies can be avoided only by careful experimental design, and, as with any scientific data, the peer review process is critical to understanding whether the data are sound or based on faulty methodology.

Unfortunately, the results reported in the article were apparently not peer-reviewed, nor was sufficient detail provided to evaluate the conclusions.

As cognitive neuroscientists, we are very excited about the potential use of brain imaging techniques to better understand the psychology of political decisions. But we are distressed by the publication of research in the press that has not undergone peer review, and that uses flawed reasoning to draw unfounded conclusions about topics as important as the presidential election.

Adam Aron, Ph.D., University of California, San Diego
David Badre, Ph.D., Brown University
Matthew Brett, M.D., University of Cambridge
John Cacioppo, Ph.D., University of Chicago
Chris Chambers, Ph.D., University College London
Roshan Cools, Ph.D., Radboud University, Netherlands
Steve Engel, Ph.D., University of Minnesota
Mark D'Esposito, M.D., University of California, Berkeley
Chris Frith, Ph.D., University College London
Eddie Harmon-Jones, Ph.D., Texas A&M University
John Jonides, Ph.D., University of Michigan
Brian Knutson, Ph.D., Stanford University
Liz Phelps, Ph.D., New York University
Russell Poldrack, Ph.D., University of California, Los Angeles
Tor Wager, Ph.D., Columbia University
Anthony Wagner, Ph.D., Stanford University
Piotr Winkielman, Ph.D., University of California, San Diego

Do you really love your iPhone?

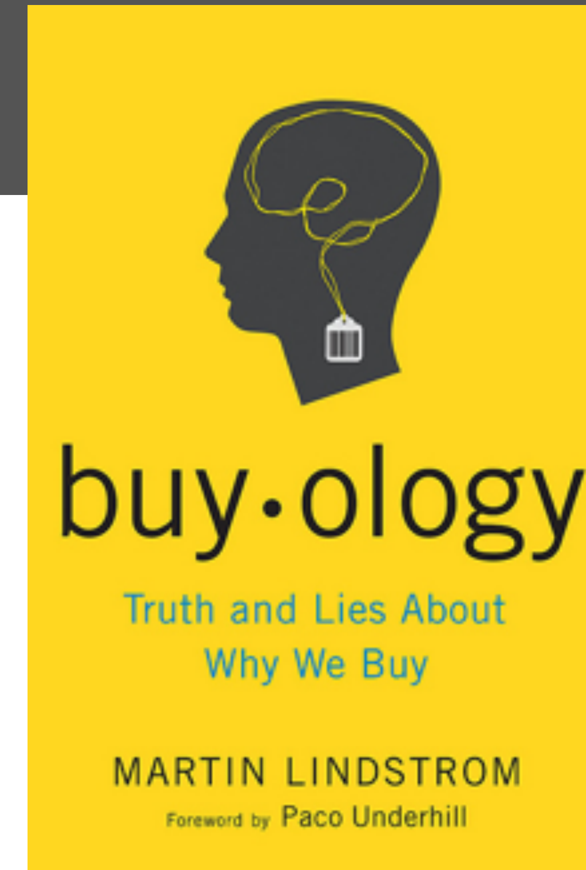
The New York Times

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.

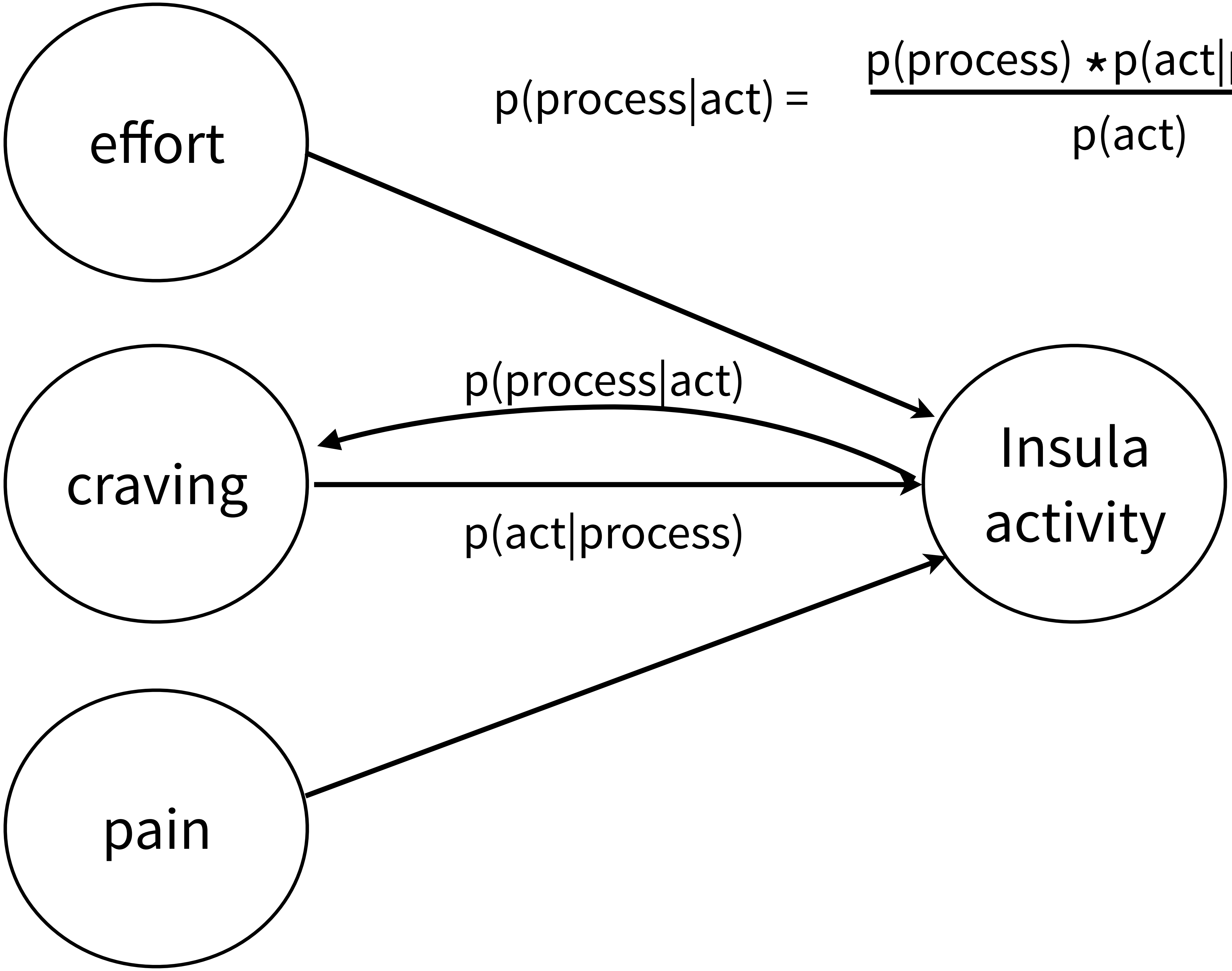
We find it surprising that The Times would publish claims like this that lack scientific validity.

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

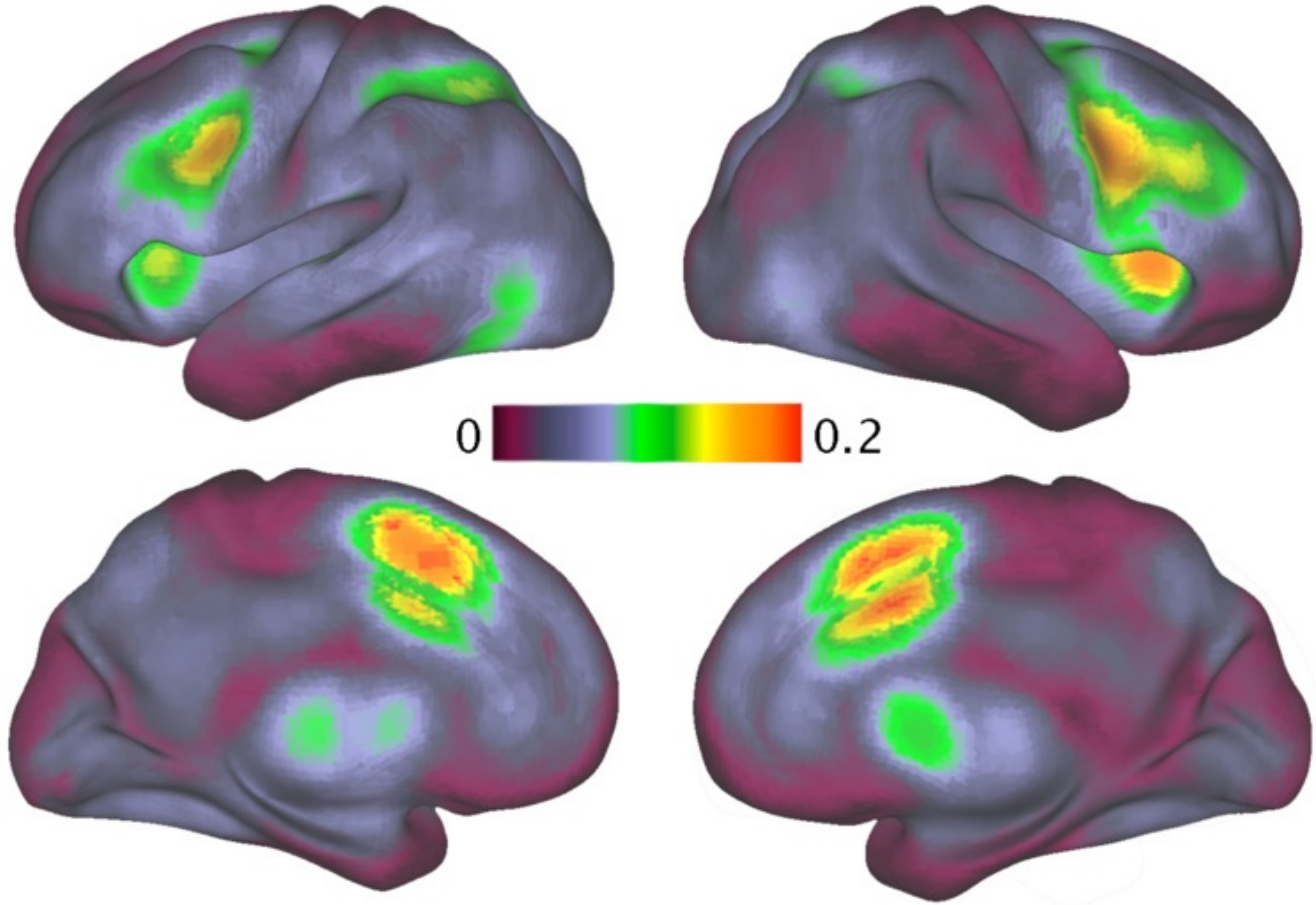
The writer is a professor of psychology and neurobiology at the University of Texas at Austin. His letter was signed by 44 other neuroscientists.

Does reverse inference work?

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



Insula activation is weakly selective



Some voxels active in as many of 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.

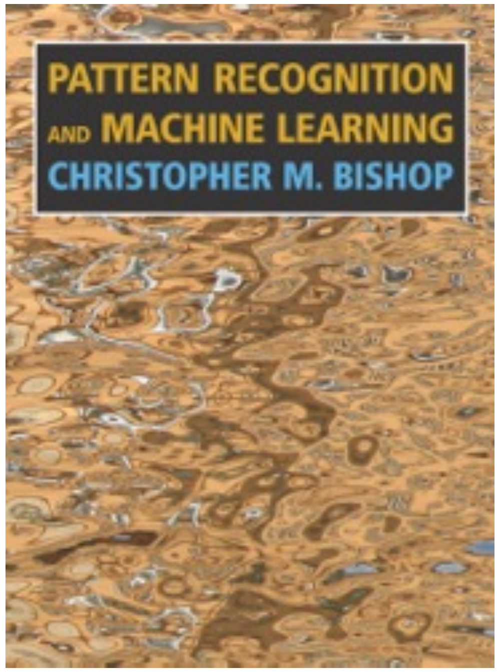
In many cases the use of reverse inference is informal; the presence of unexpected activation in a particular region is explained by reference to other studies that found activation in the same region. However, in some studies the reverse inference is a central feature. In one study [4], subjects were 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

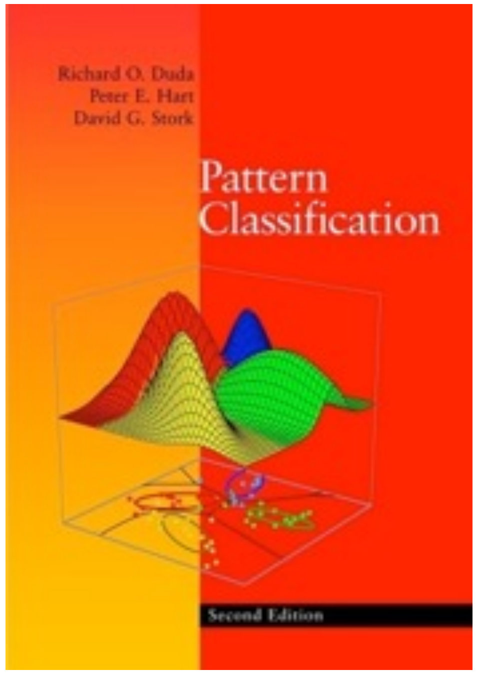
Corresponding author: Poldrack, R.A. (poldrack@ucla.edu).
Available online 6 January 2006

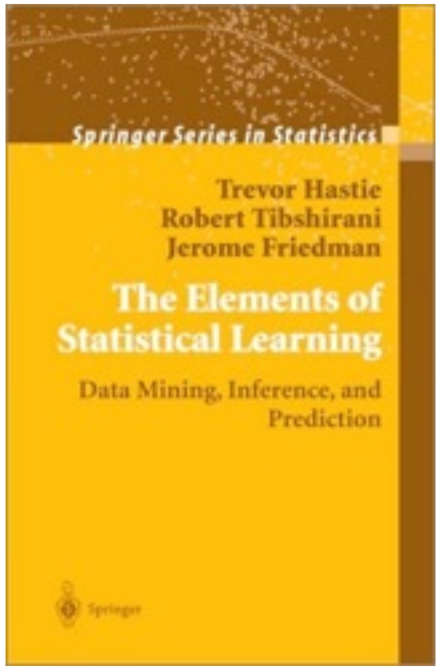
TICS, 2006

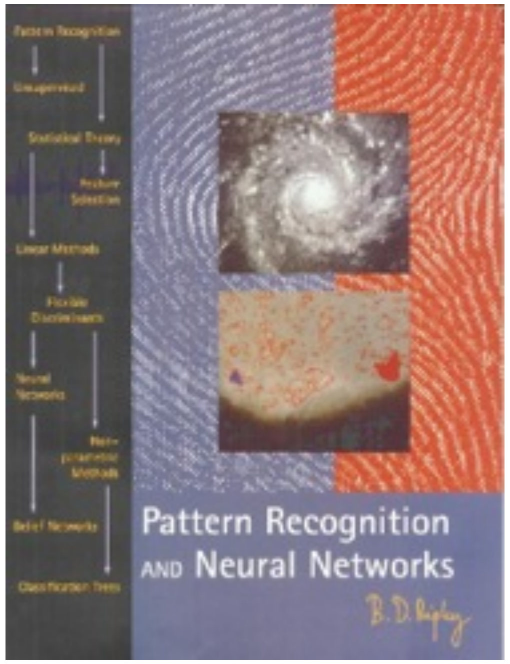
Formalizing reverse inference

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







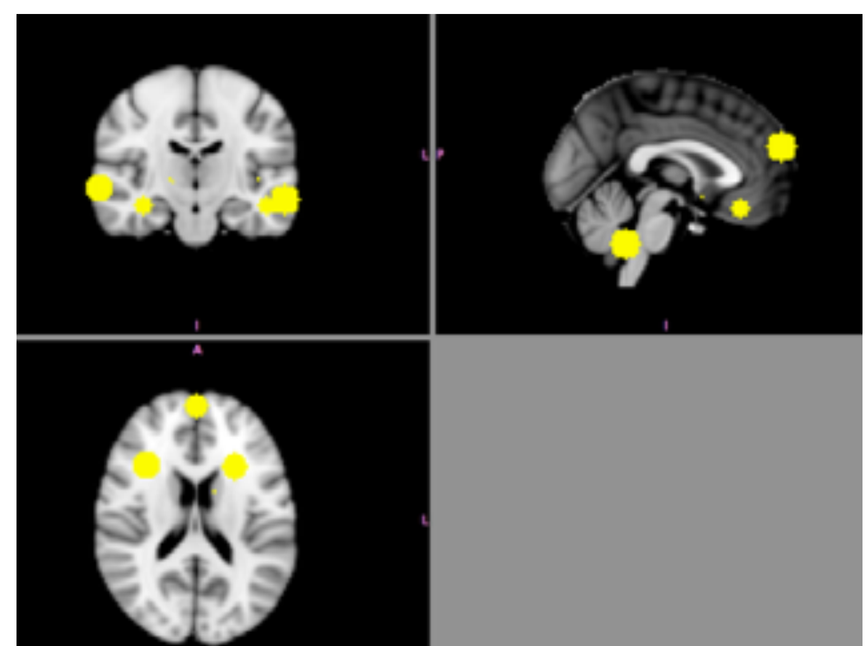


Creating meta-analytic brain maps

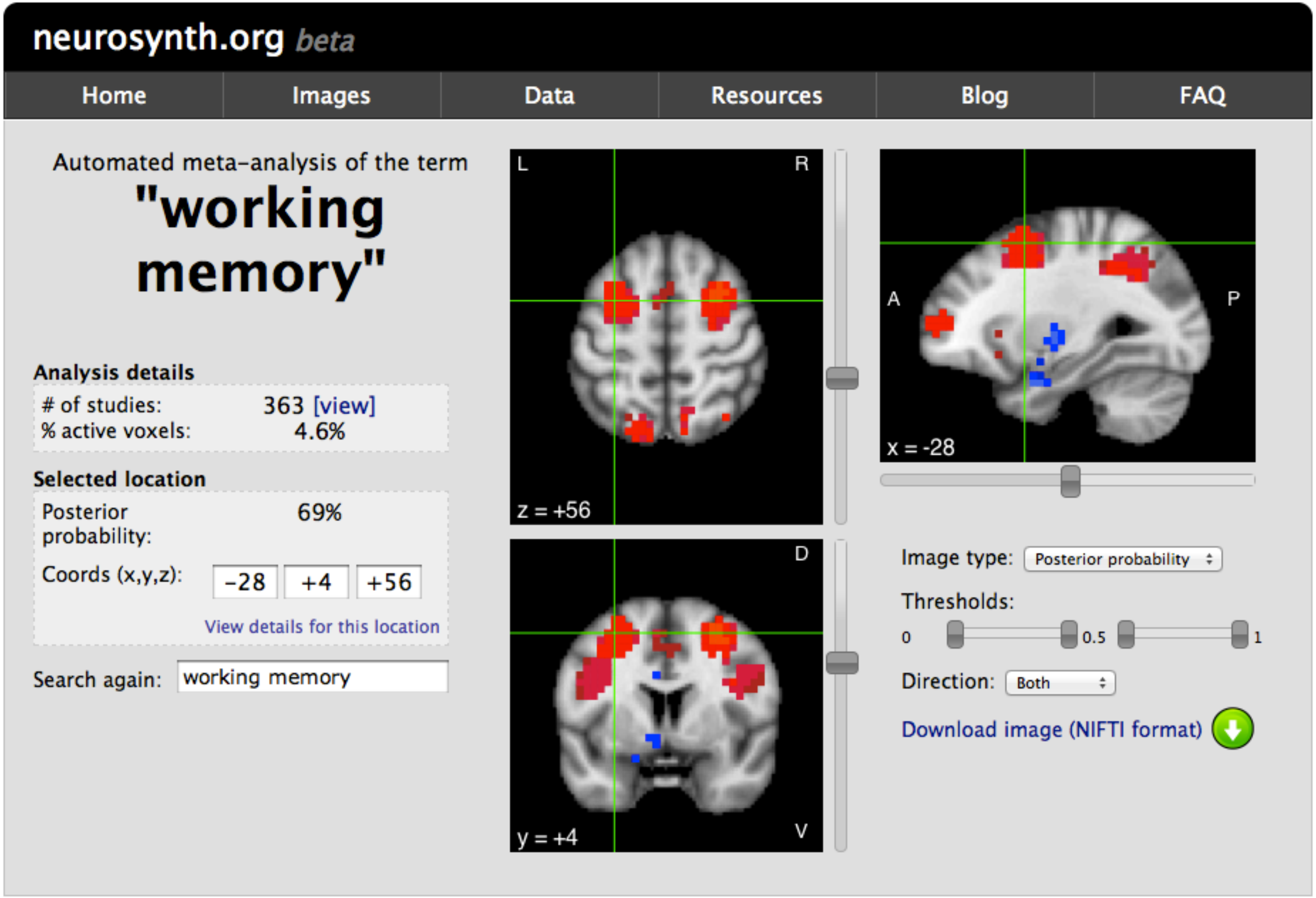
- Automated Coordinate Extraction (Yarkoni et al, 2011, *Nature Methods*)
  - Automatically extracts activation tables from fMRI papers for 17 journals
  - Current database has 5809 papers
  - Good accuracy
    - 84% sensitivity, 97% specificity against SumsDB manual database
- Meta-analytic maps created 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

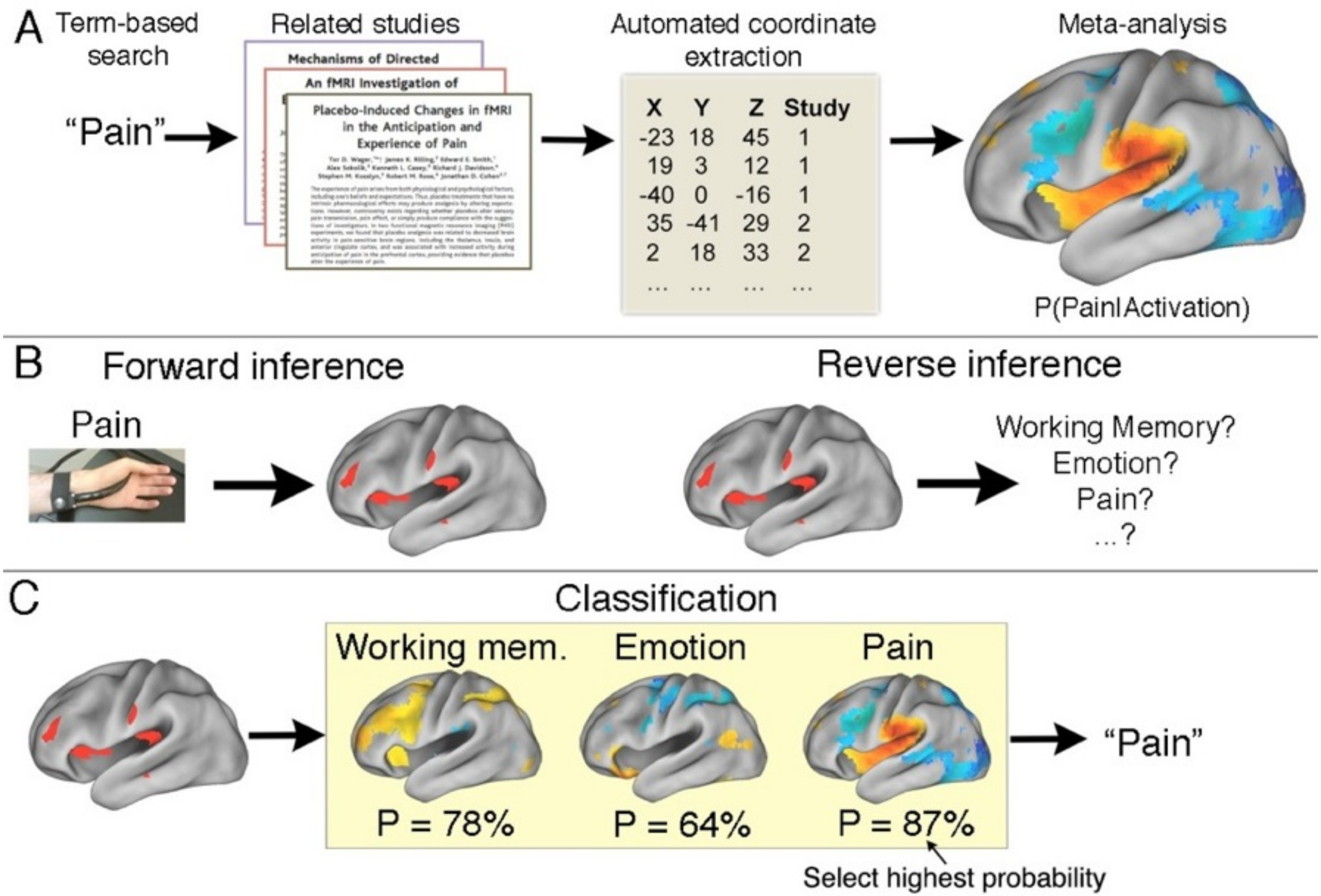
Automated coordinate extraction →



Neurosynth.org

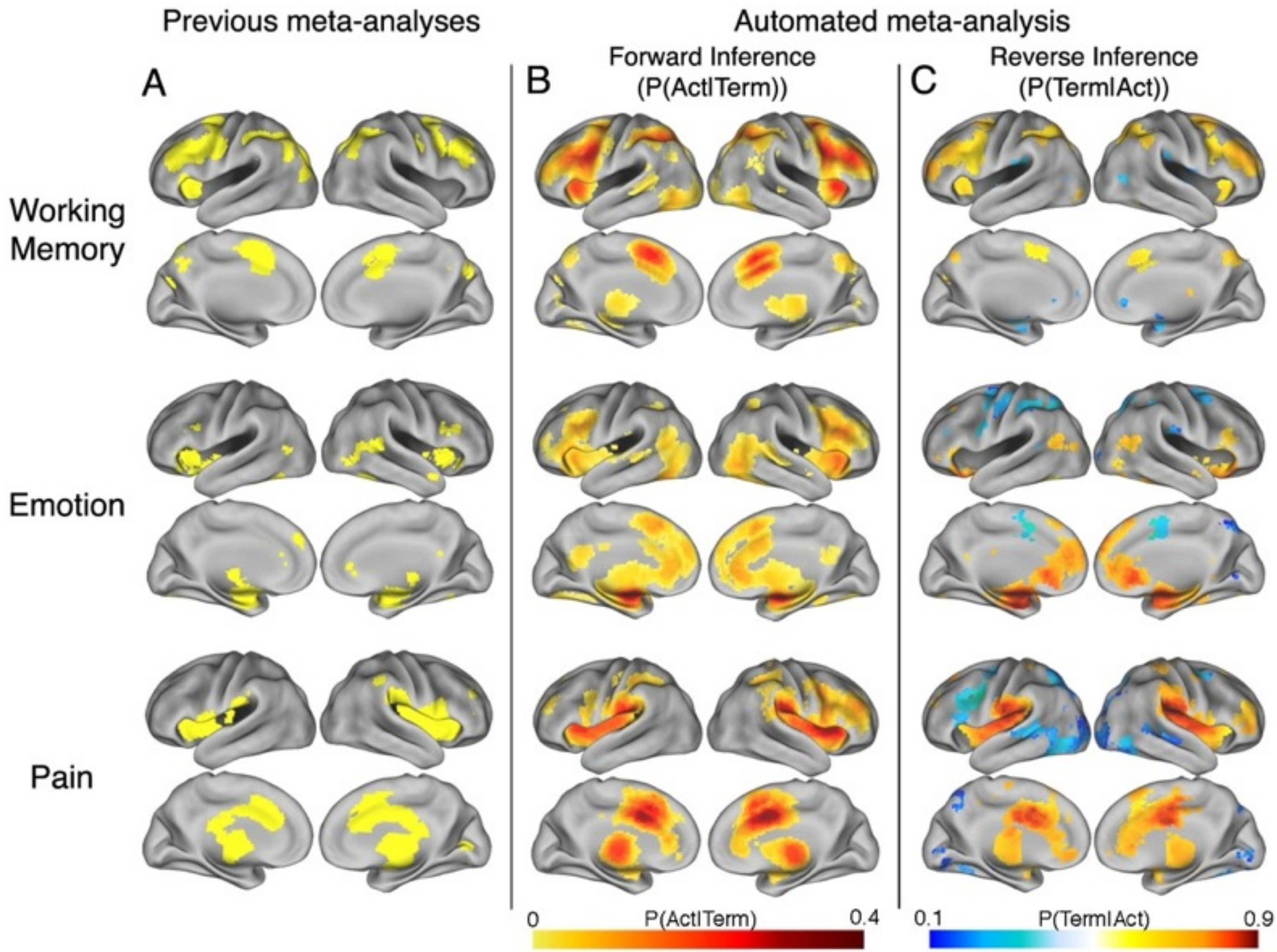


Automated meta-analysis



Yarkoni et al., 2011, *Nature Methods*

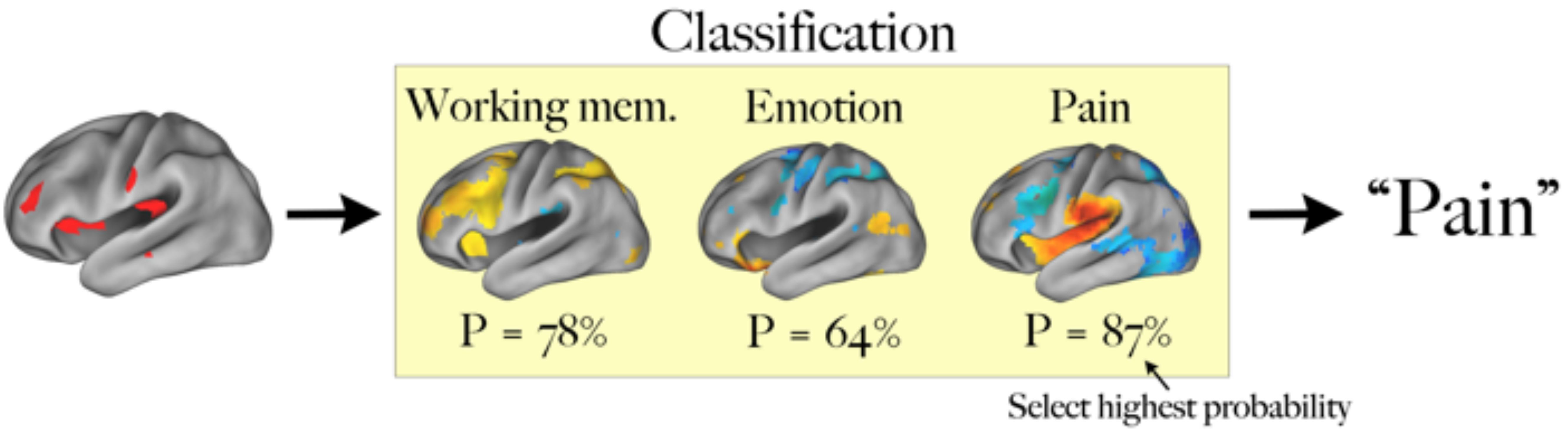
Automated meta-analysis



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.



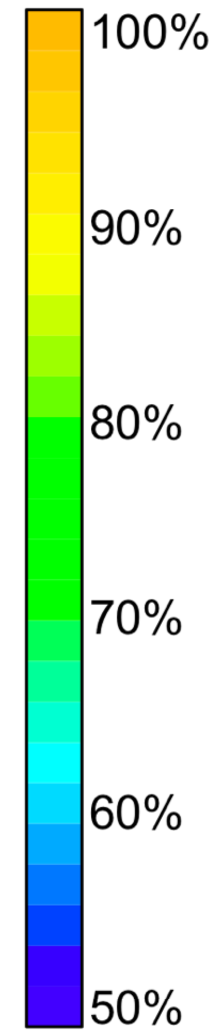
Yarkoni et al, 2011, *Nature Methods*

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?

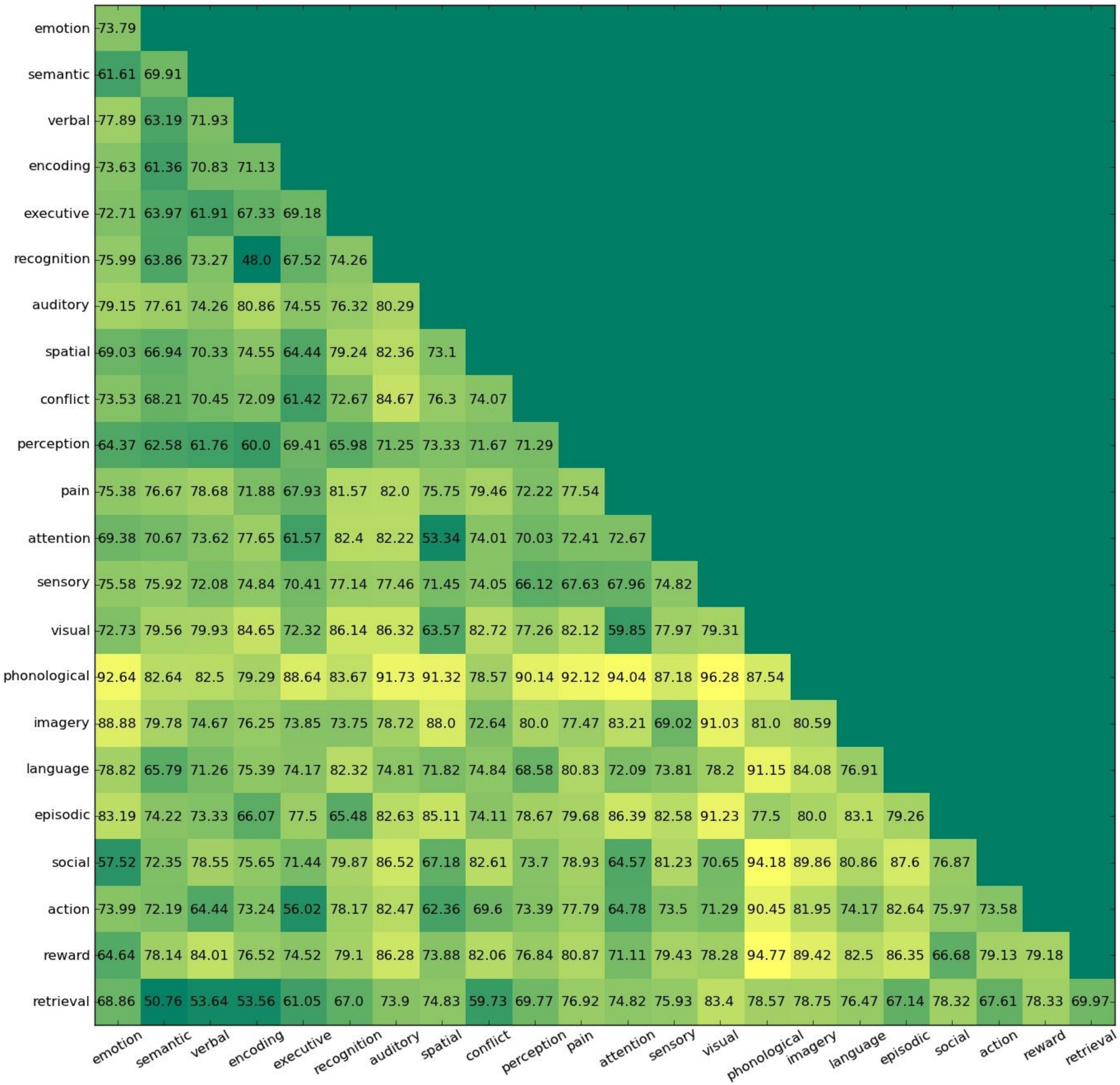
Yarkoni et al, 2011, *Nature Methods*

	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*

Ensemble learning



Use majority vote of naive Bayes, l1-regularized logistic regression, and linear SVM

Madhura Parikh,

Subhashini Venugopalan

Sanmi Koyejo

Automating reverse inference

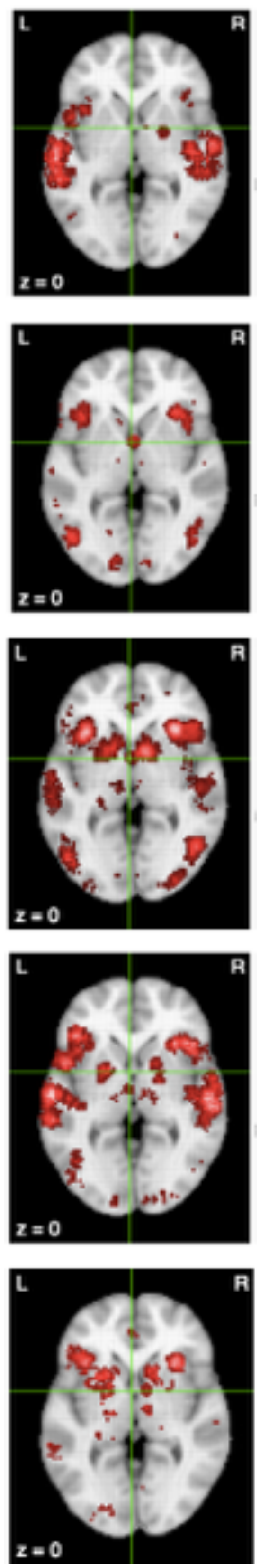


Table 2. Pearson correlations between searchlight classification map and NeuroSynth term-based reverse inference activation maps

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).

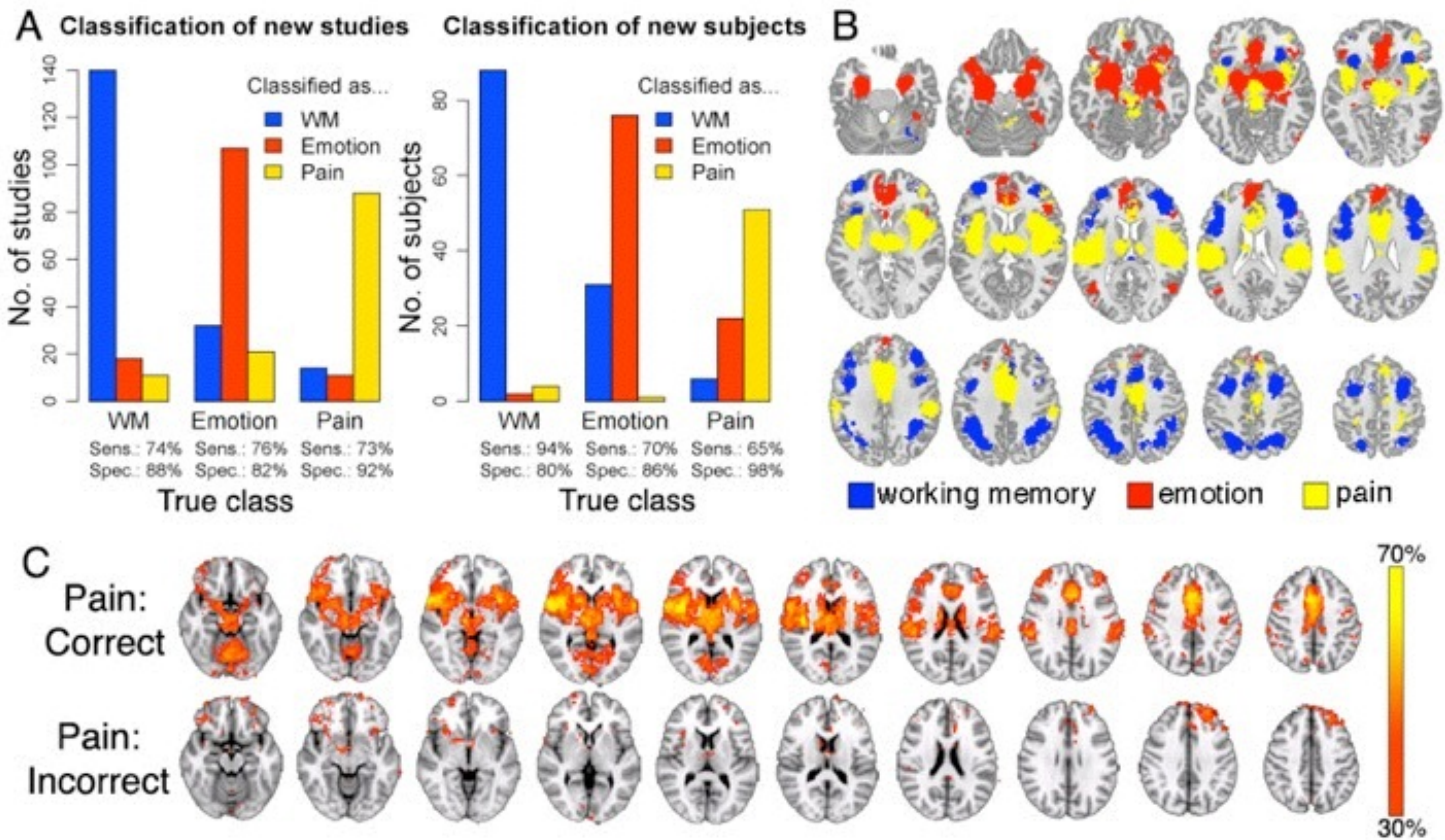
Helfinstein et al, 2014, PNAS

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*

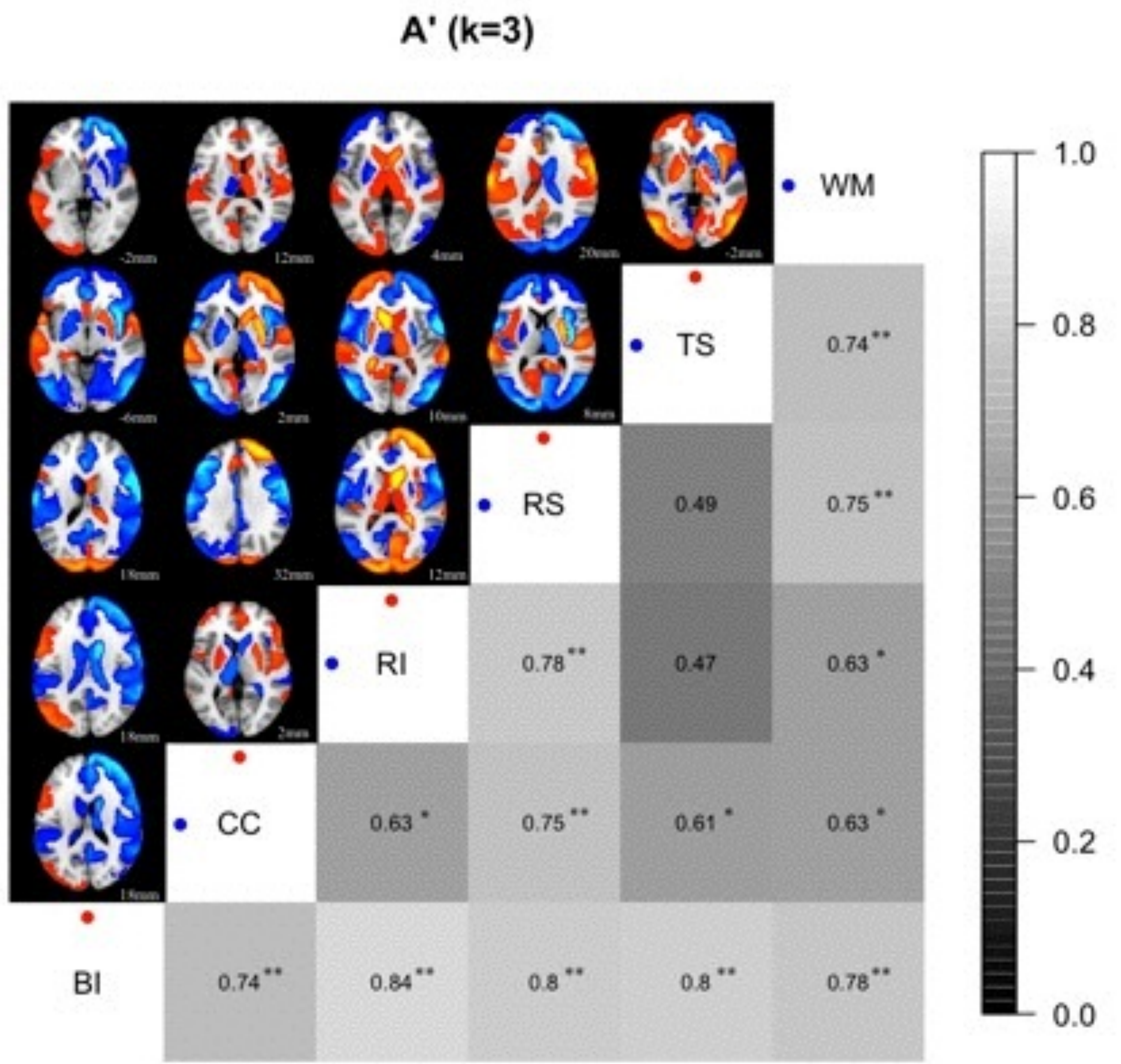
Classifying individual subjects



Yarkoni et al, 2011, *Nature Methods*

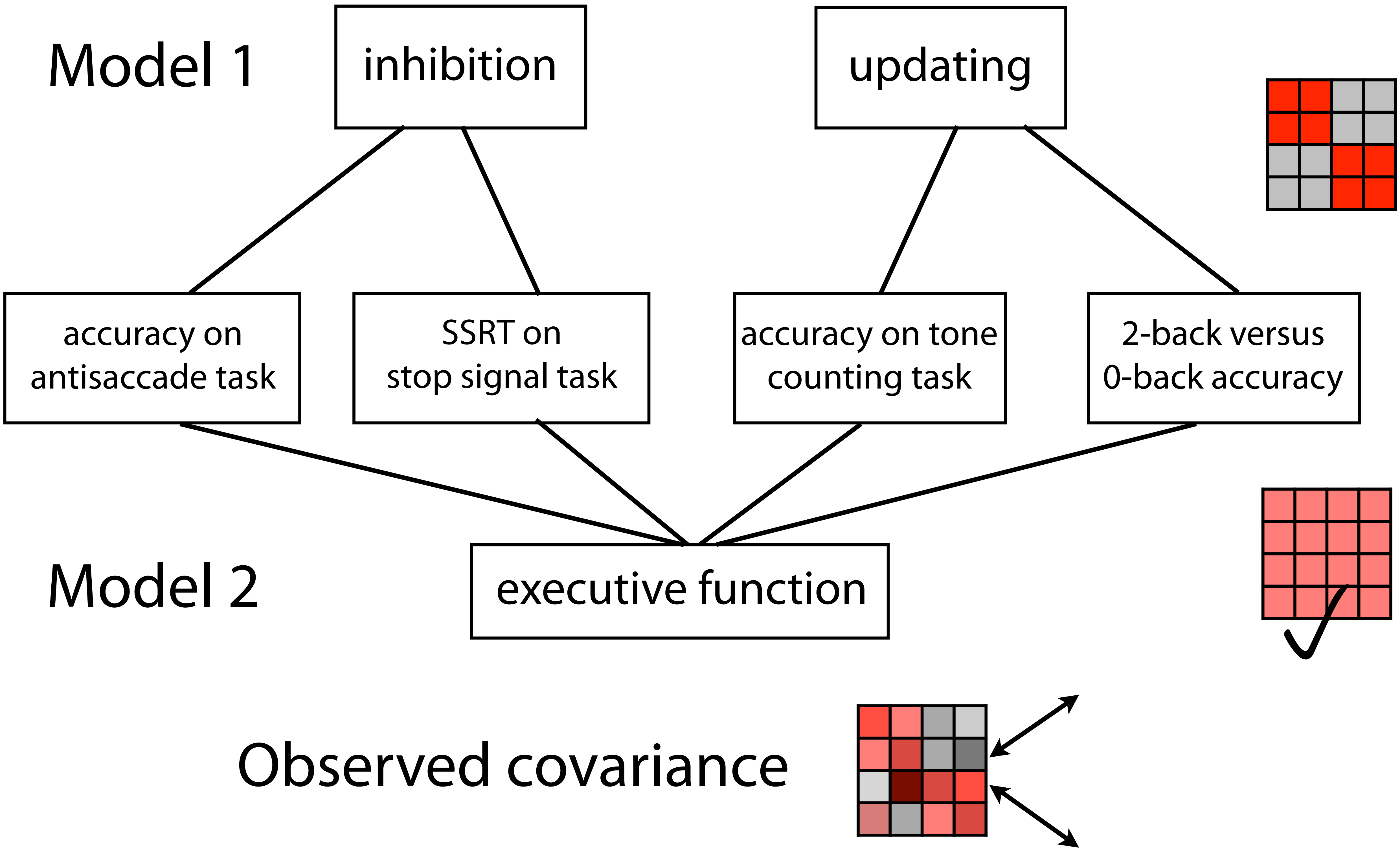
Using classification to understand mental structure

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

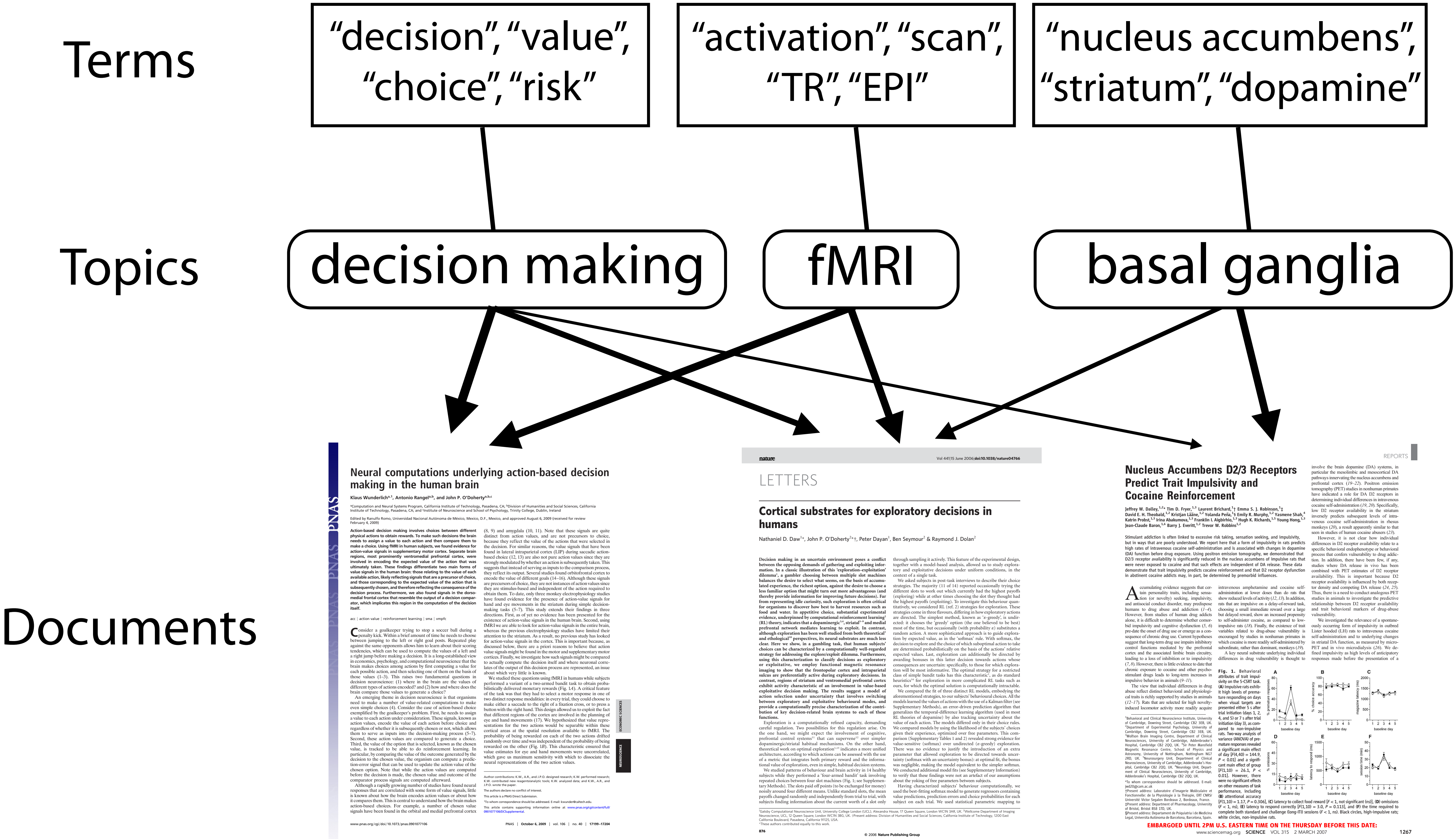


Lenartowicz et al, 2010, *Topics in Cognitive Science*

Towards meta-analytic testing of cognitive theories



Topic modeling



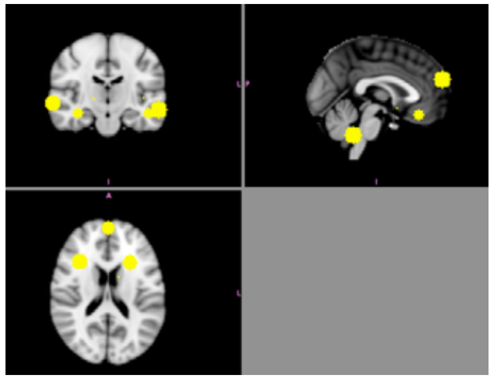
Topic Mapping

- Each document has a loading on each topic
  - On average, each document loads on ~6.5 topics
- Used ACE to extract activation coordinates for all 5,809 papers
- Perform voxelwise chi-square test with FDR correction to examine association between topics and activation

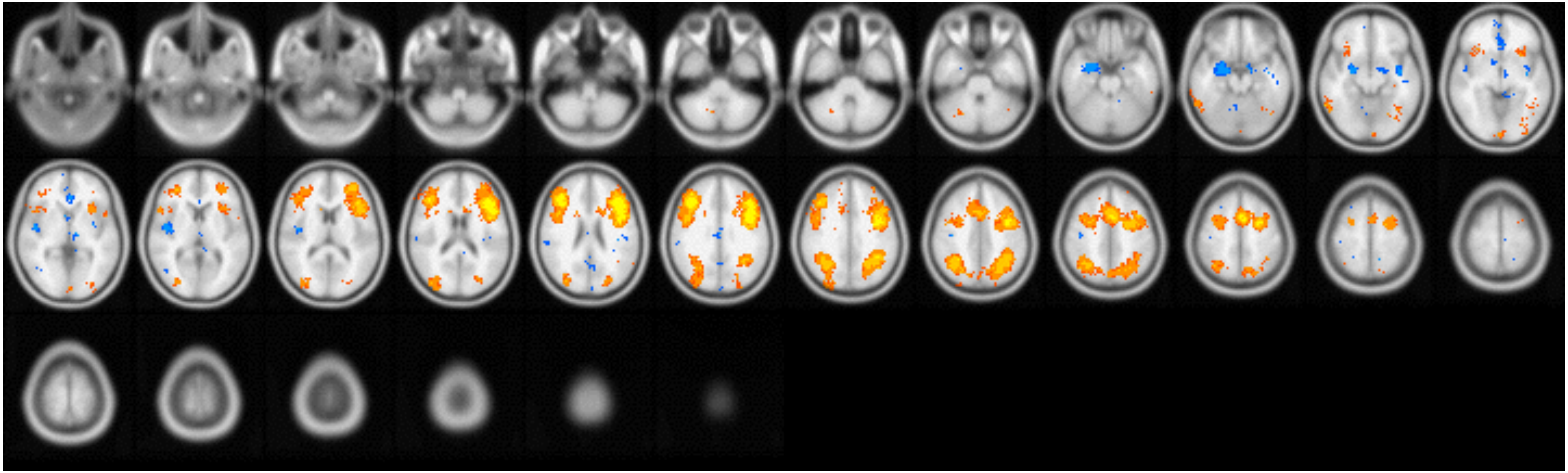
Topic — Documents — Activation Coordinates

emotion
negative
unpleasant

"...amygdala...emotion...negative..."

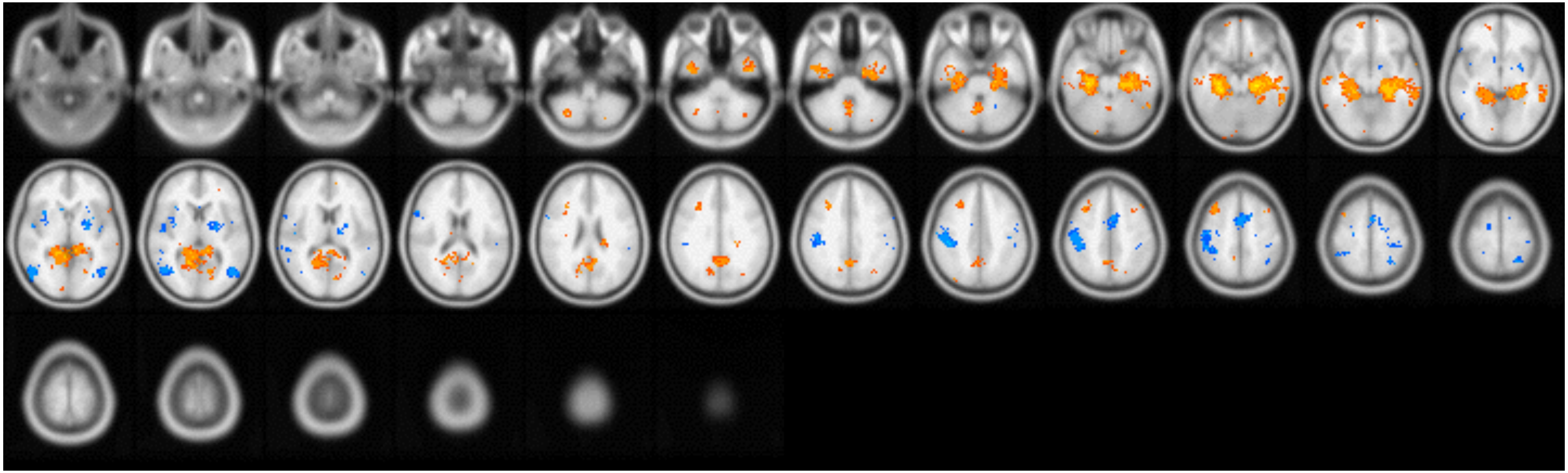


Poldrack et al., 2012, *PLOS Comp Biol*



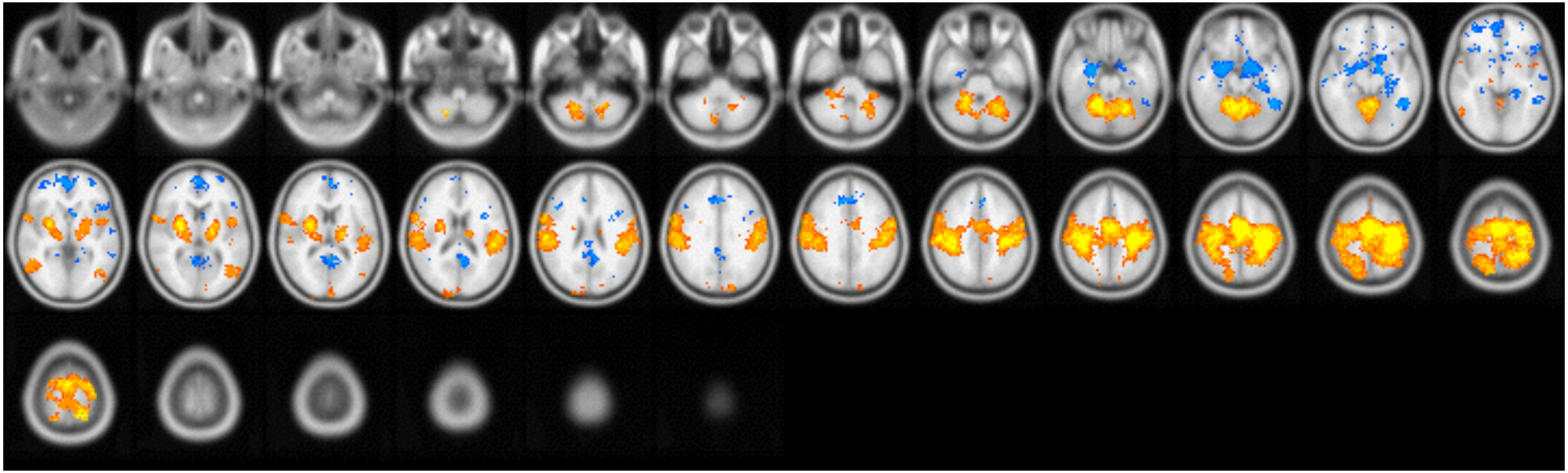
Topic 61 (442 docs): memory working_memory maintenance visual_working_memory spatial_working_memory manipulation episodic_buffer retention rehearsal retrieval

Poldrack et al., 2012, *PLOS Comp Biol*



Topic 3 (389 docs): memory episodic_memory recall learning verbal_memory association encoding risk visual_memory working_memory

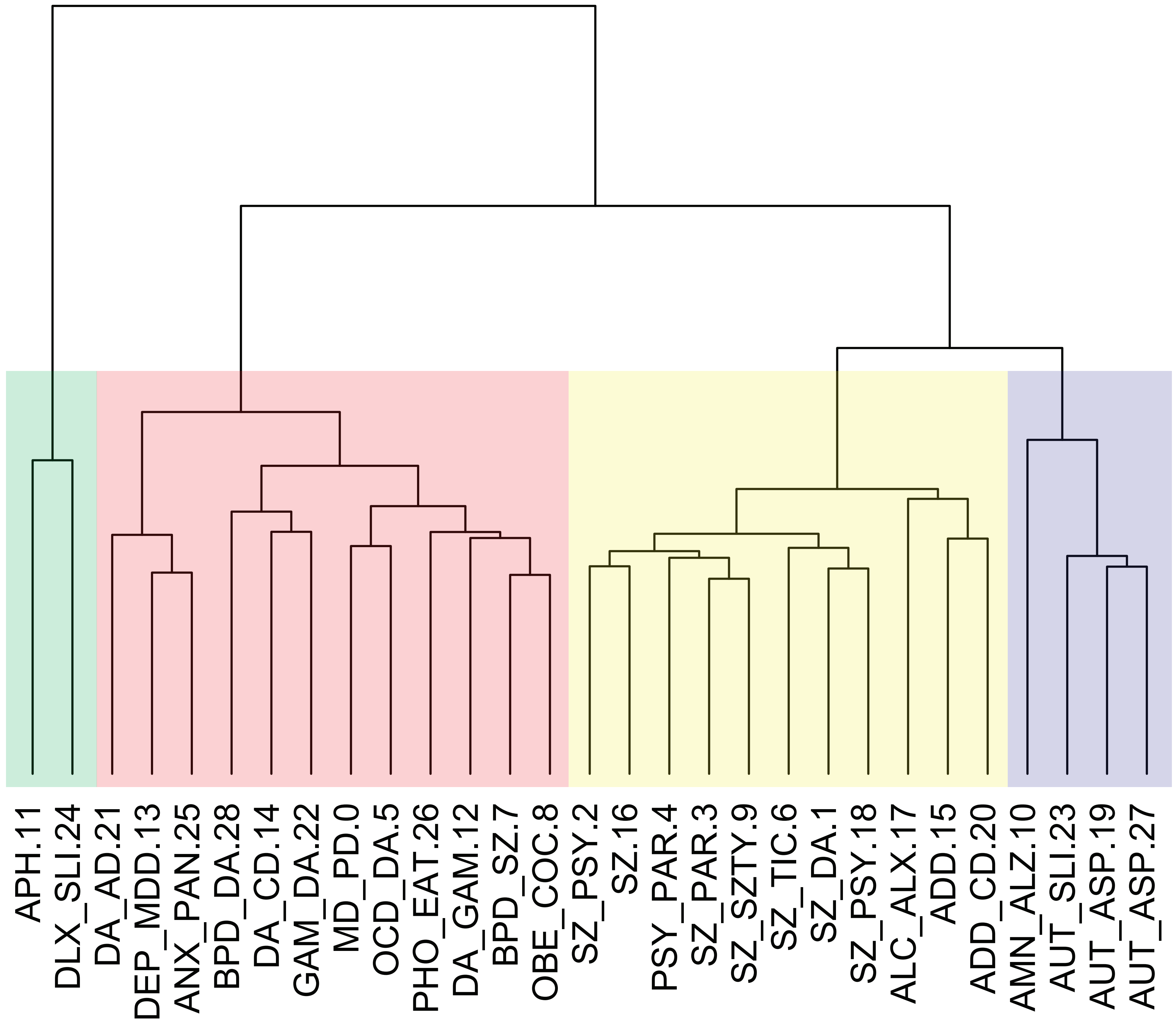
Poldrack et al., 2012, *PLOS Comp Biol*



Topic 106 (391 docs): movement coordination motor_control feedback planning integration goal context knowledge learning

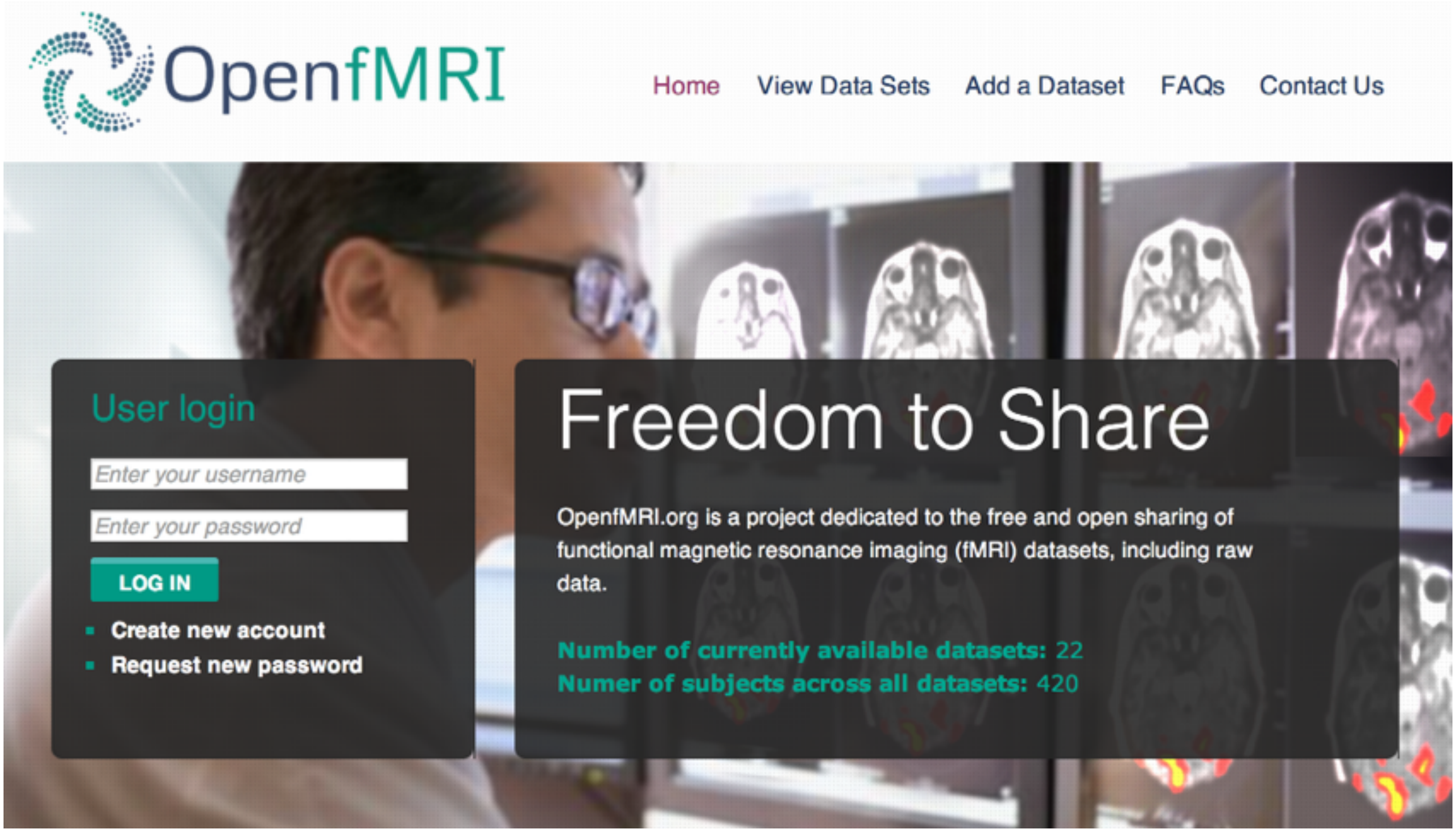
Poldrack et al., 2012, *PLOS Comp Biol*

Clustering disorders by topic maps



Poldrack et al., 2012, *PLOS Comp Biol*

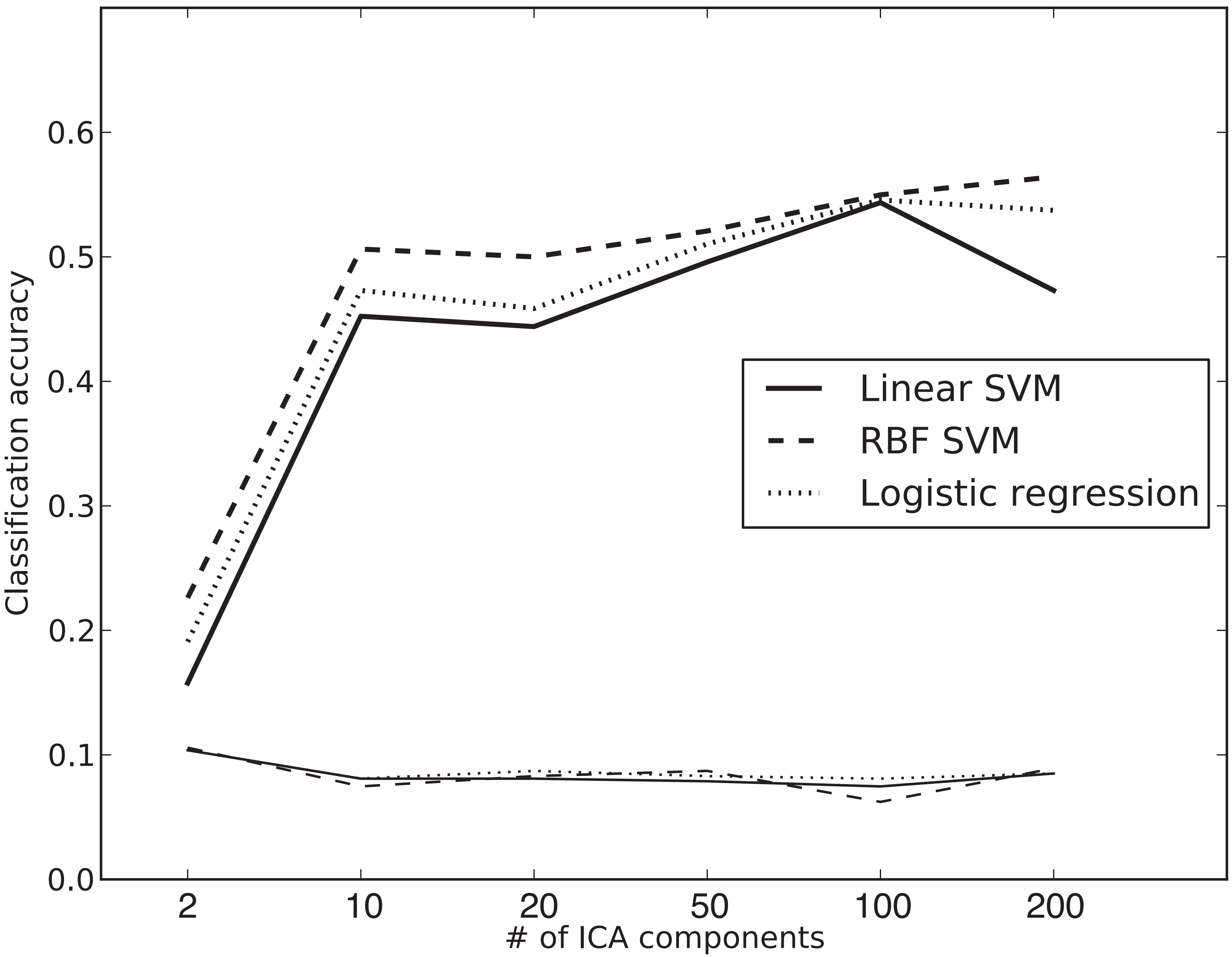
Mega-analysis of fMRI data



26 tasks, 482 images from 338 subjects

Poldrack et al., 2013,
Frontiers in Neuroinformatics

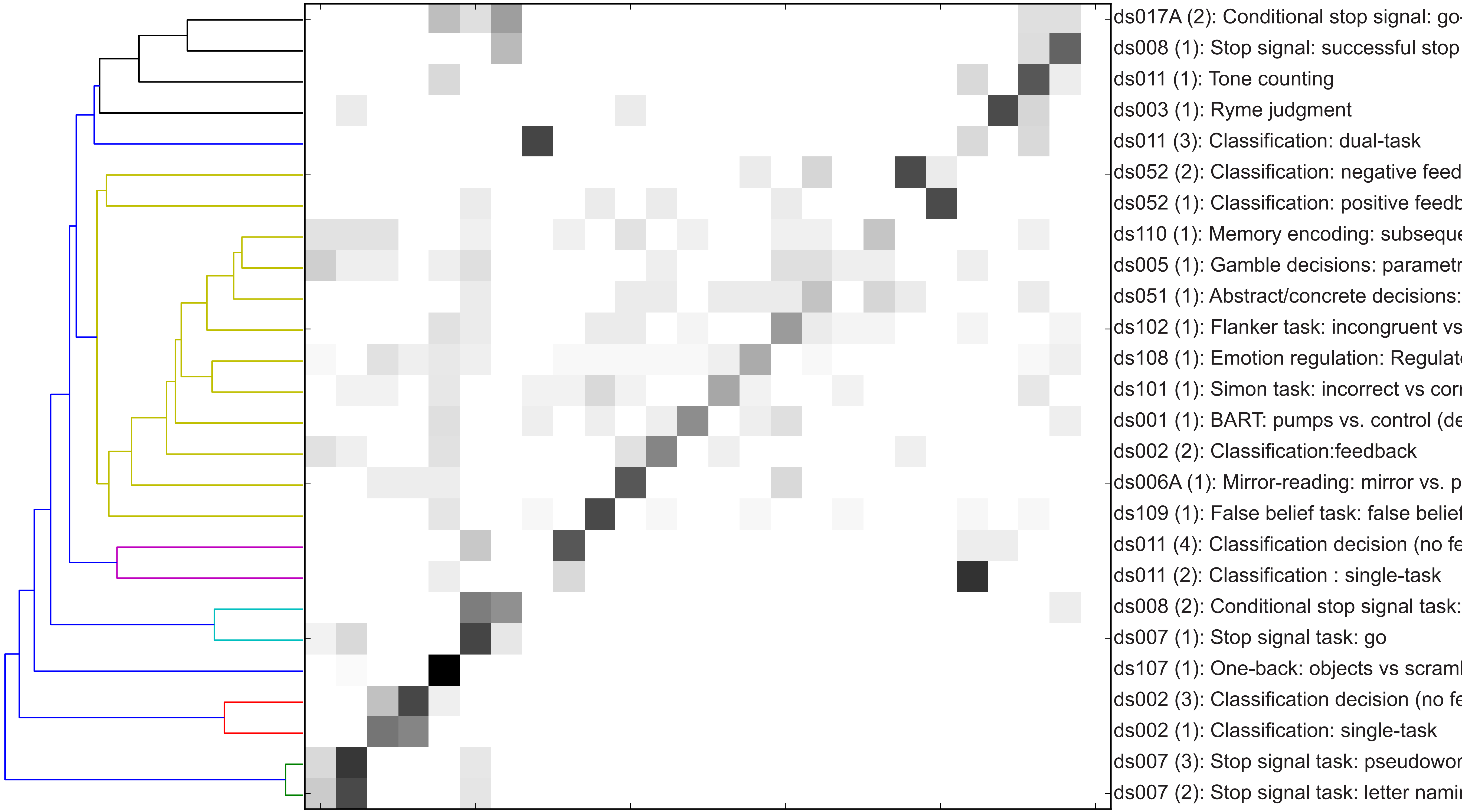
Classification results



Whole-brain with linear SVM: 48% accuracy

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

Larger-scale decoding



Poldrack et al., 2013,
Frontiers in Neuroinformatics

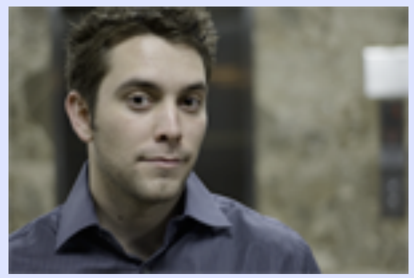
Conclusions

- Cognitive neuroscience needs to get formal about describing the mental processes that are being mapped to brain function
- Much interesting structure can be extracted using text mining, but ultimately progress will require manual annotation by domain experts
- Ontologies plus databases will provide the means to ask whether the claims of psychology regarding mental architecture are respected by the brain

Acknowledgments

UCLA

Robert Bilder



Don Kalar

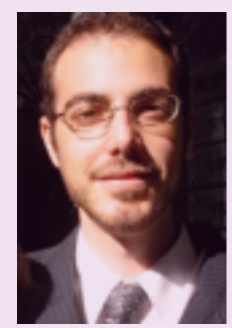
Fred Sabb

D. Stott Parker

UT Austin

Jeanette Mumford

Sanmi Koyejo



Tal Yarkoni

Carnegie-Mellon

Niki Kittur

Get involved!
www.cognitiveatlas.org

NIMH
National Institute of Mental Health

NSF

James S. McDonnell Foundation