Machine learning in neuroimaging: promises and pitfalls

Tal Yarkoni

Department of Psychology, UT-Austin

Machine Learning Summer School 2015

Goals

- Apply some of the concepts/methods you've learned
- Identify some gotchas and caveats
- Compare research objectives in ML vs. many sciences

In the beginning…

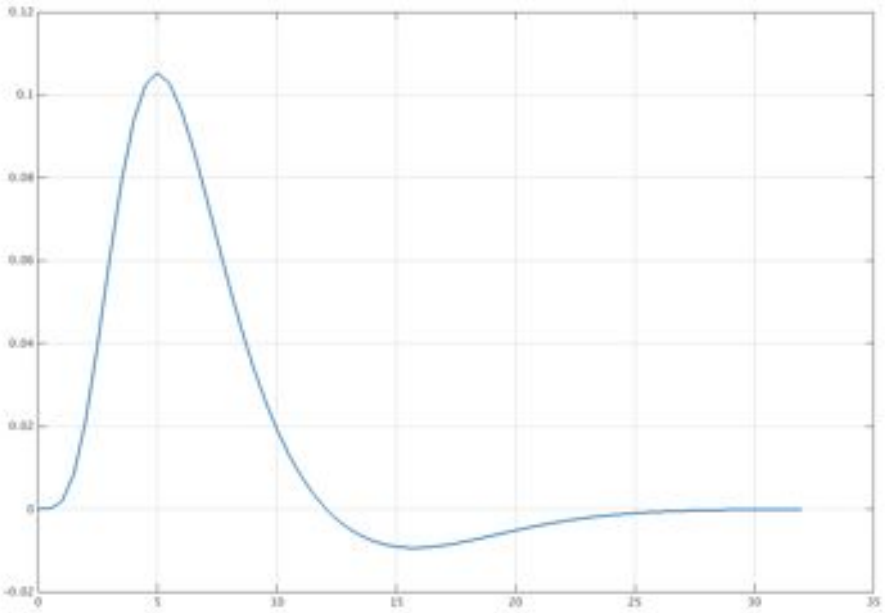
- I.e., in 1990
- There was a scanner

The Blood-Oxygen-Level-Dependent signal (BOLD)

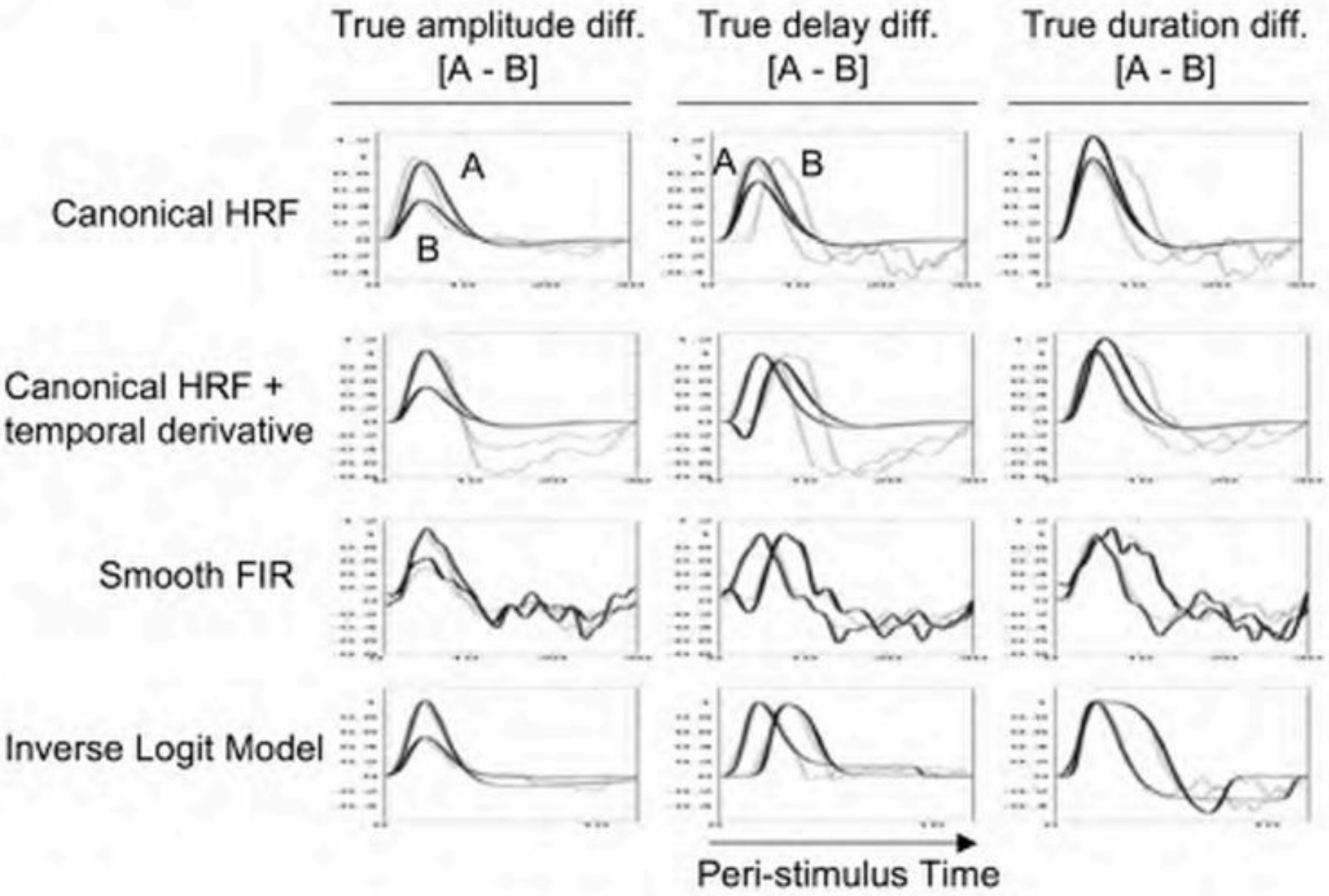
- At rest, hydrogen protons spin freely
- We impose a strong magnetic field to align the protons, then “pulse” with an RF coil
- Neurons temporarily revert to their original states, releasing energy when they relax back to low-energy state
- The MRI machine picks up this signal
- Strength of signal depends on local tissue properties
  - Influenced by ratio of oxygenated to deoxygenated blood
- When neurons are active in an area, local blood flow increases, and so does the amount of oxygenated blood

The hemodynamic response function

- Neurons are fast
- Blood flow is slow (peaks ~6 seconds after activity)
- Need transfer function relating them
- Canonically modeled as a double-gamma function
- Many variants
- HRF varies!

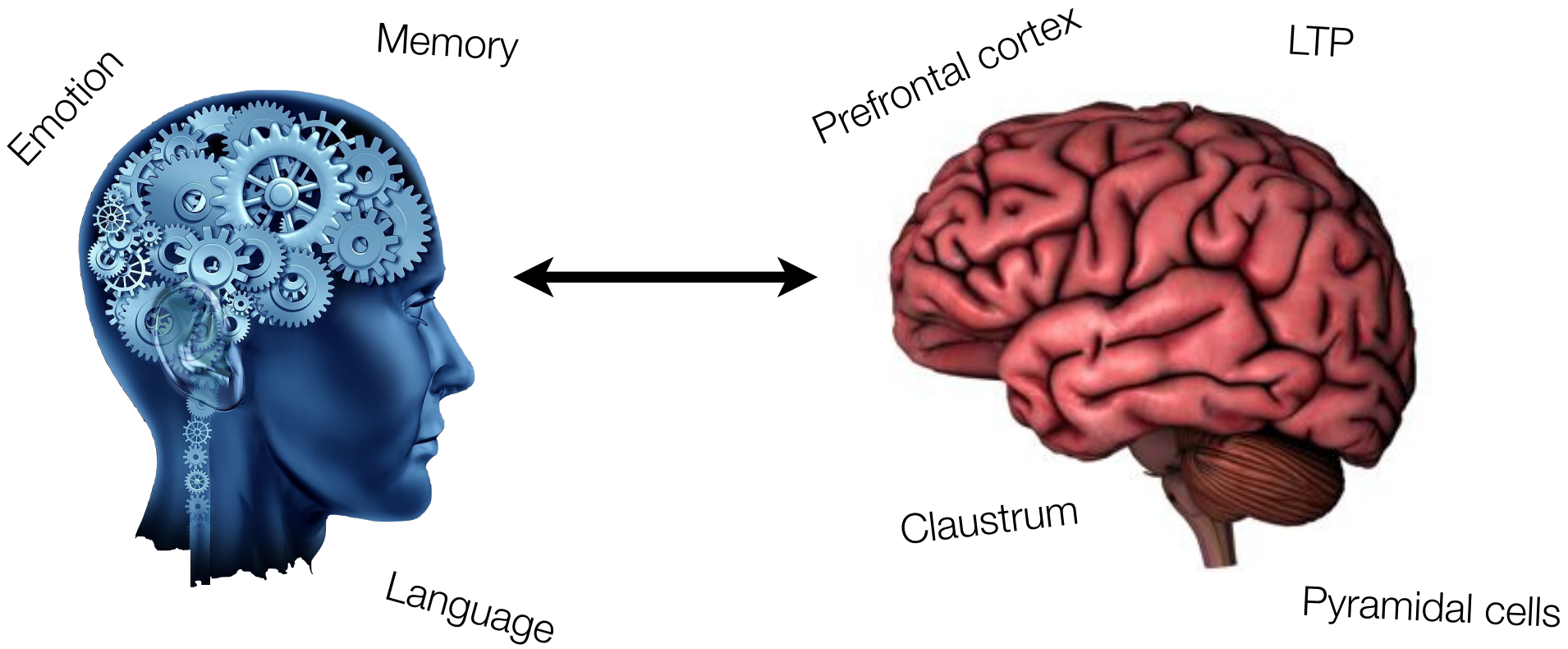


Basis functions



Wager, Hernandez, & Lindquist (2008)

Brain mapping



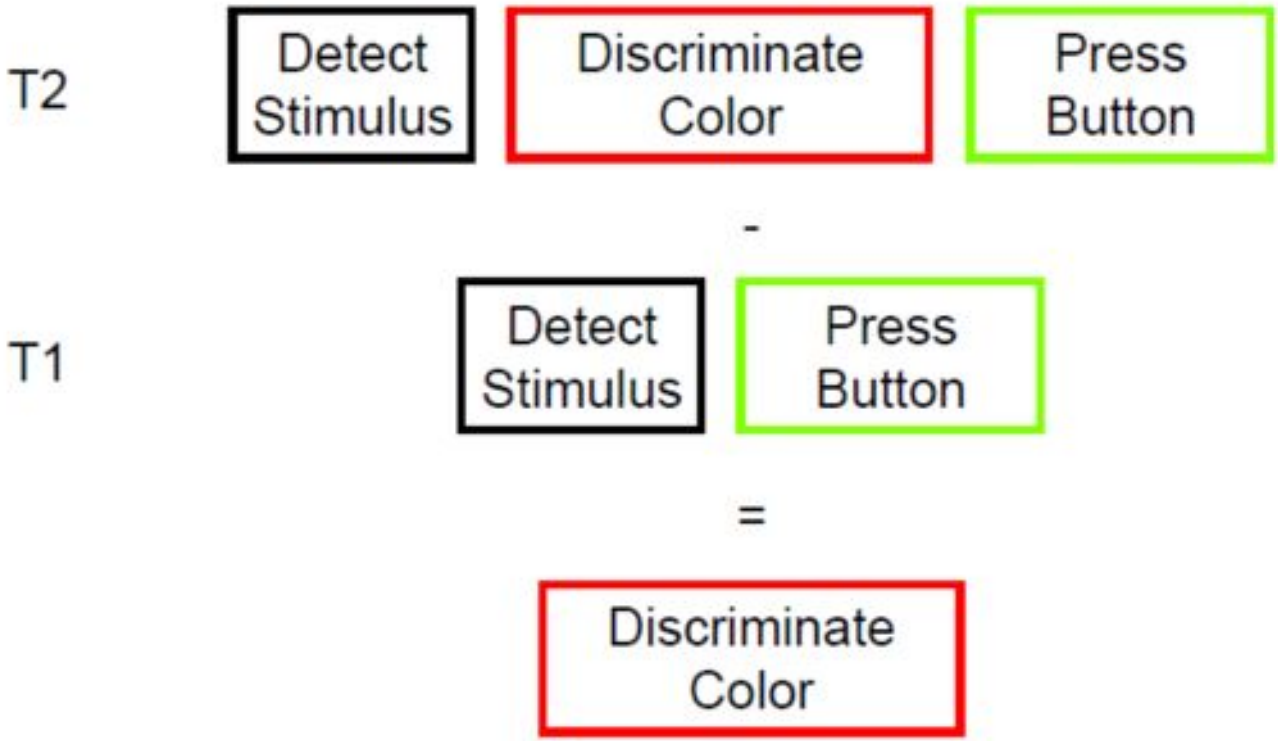
The classical approach

- How can we map specific cognitive processes onto specific brain regions/networks?
- Subtraction logic + mass univariate analysis

Subtraction logic

- Can isolate cognitive processes via experimental manipulation
- Assumption of pure insertion: we can cleanly add a specific processing step to a given task without affecting anything else

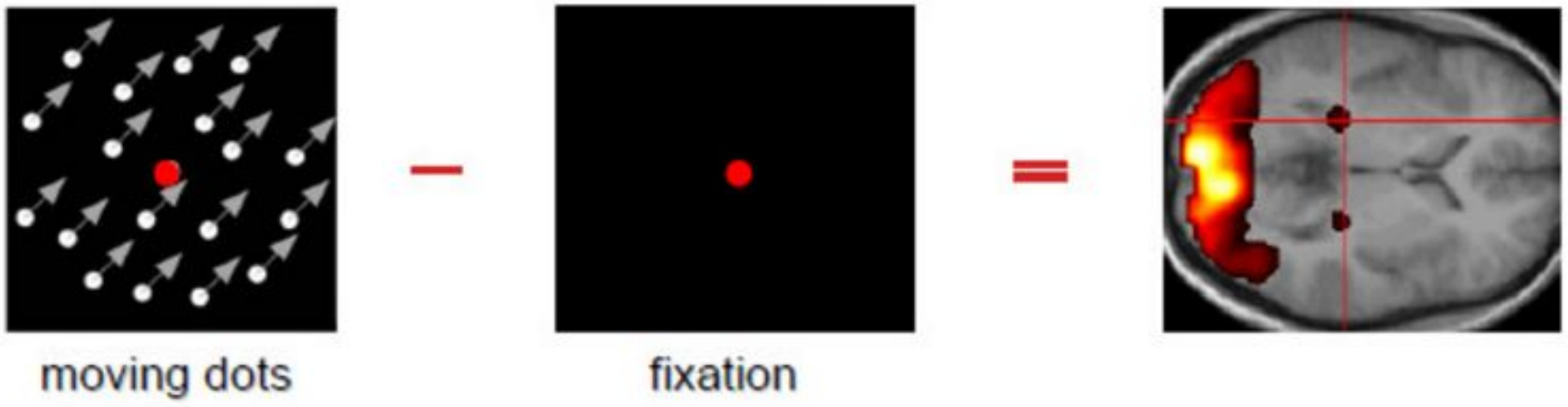
Subtraction logic



http://www.sbirc.ed.ac.uk/cyril/SPM-course/Talks/2013/3-Design_AM.pdf

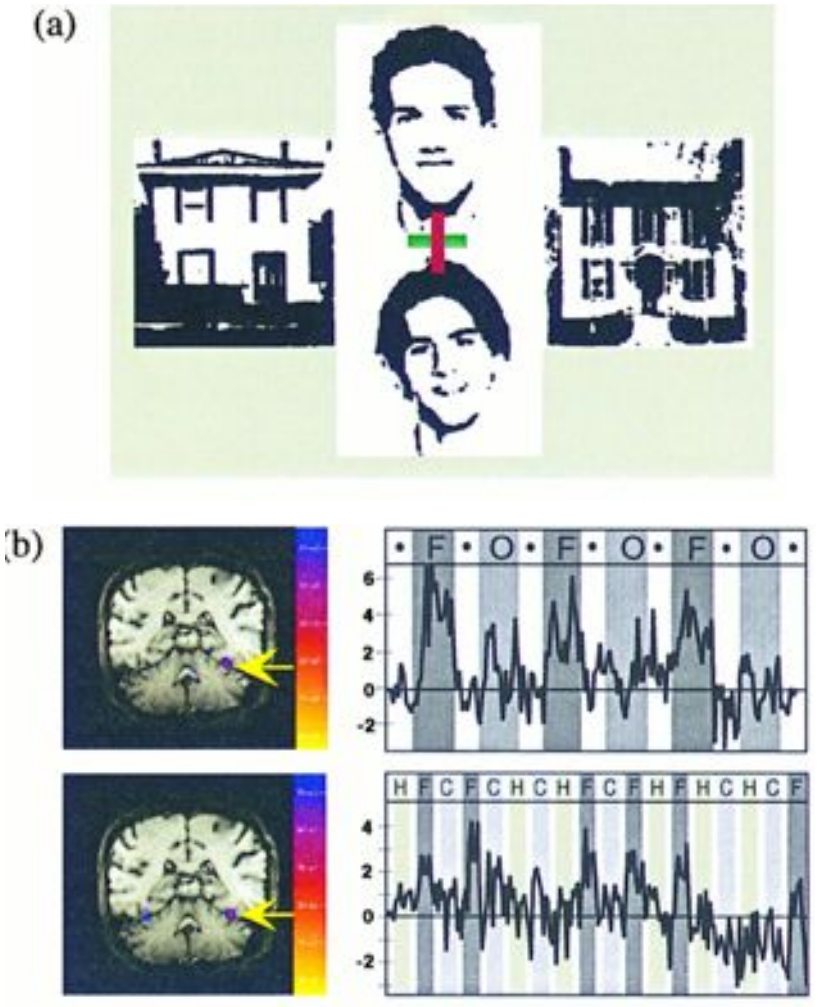
Mass univariate analysis

- Treat every point in the brain (“voxel”) as its own universe
- Run same analysis everywhere (~200k times)
- Create nice colorful whole-brain images

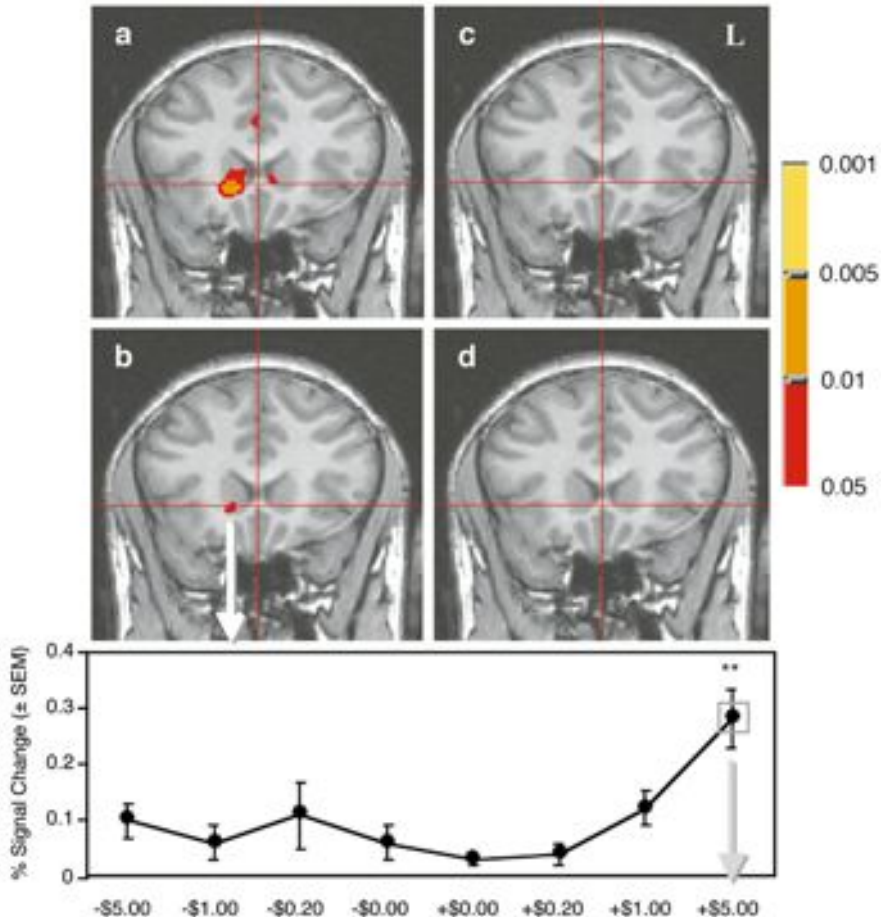


http://www.sbirc.ed.ac.uk/cyril/SPM-course/Talks/2013/3-Design_AM.pdf

Some pretty results



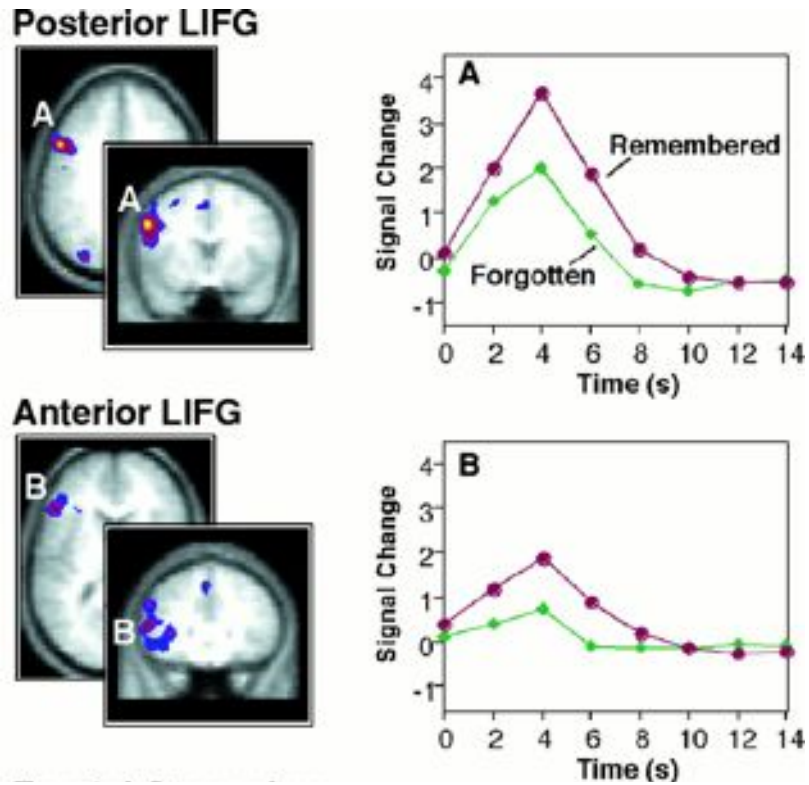
Wojciulik , Kanwisher, & Driver (1998)



Knutson et al (2001)

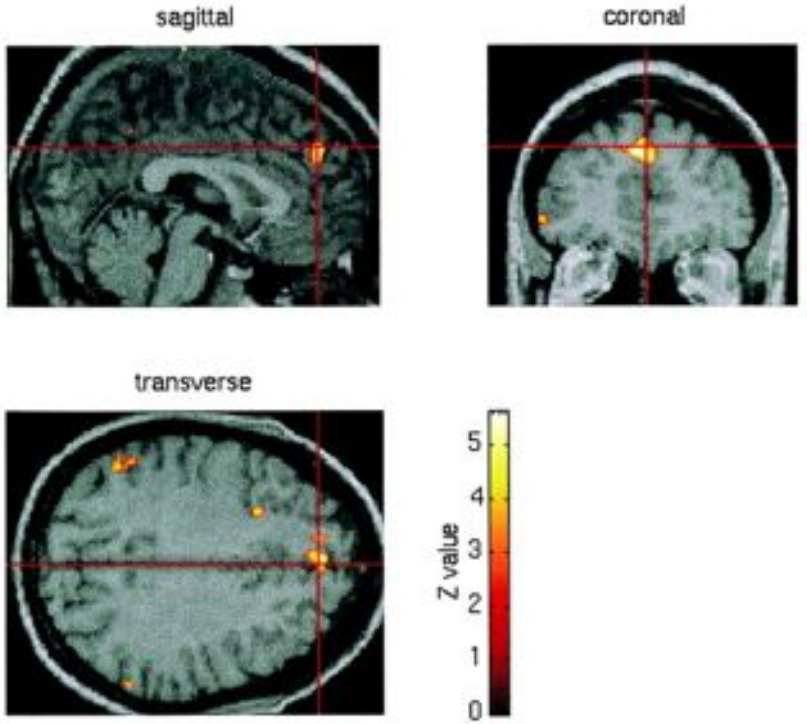
Some pretty results

Subsequent memory

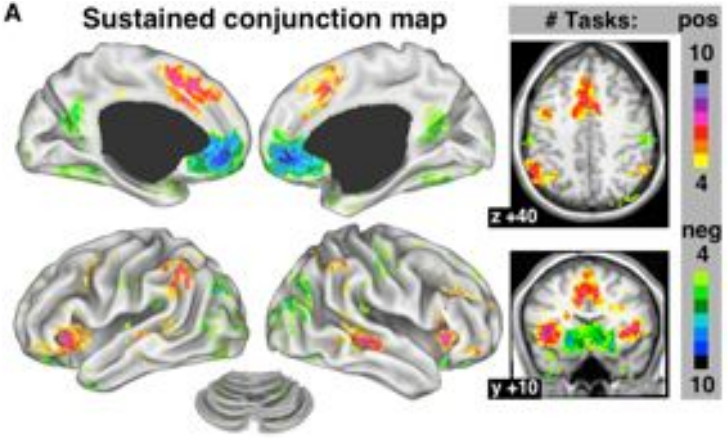


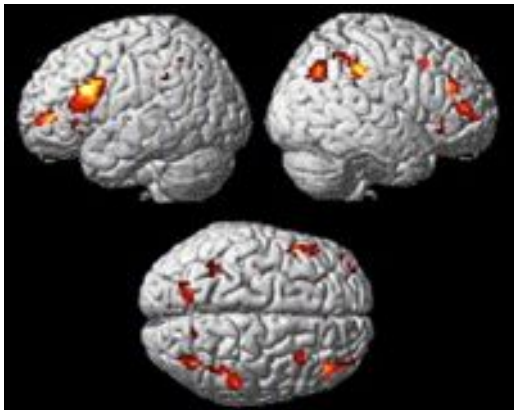
Wagner et al. (1998)

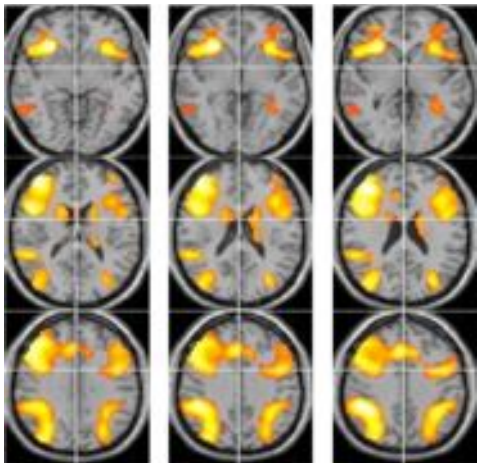
Theory of Mind

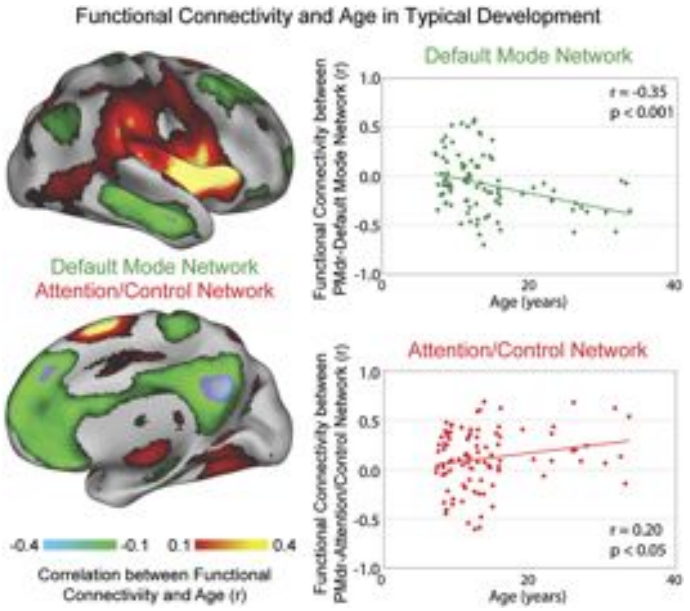


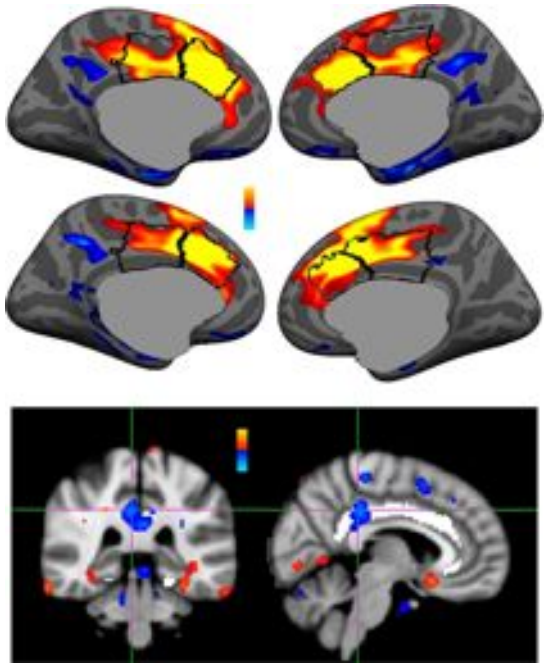
Gallagher et al. (2000)

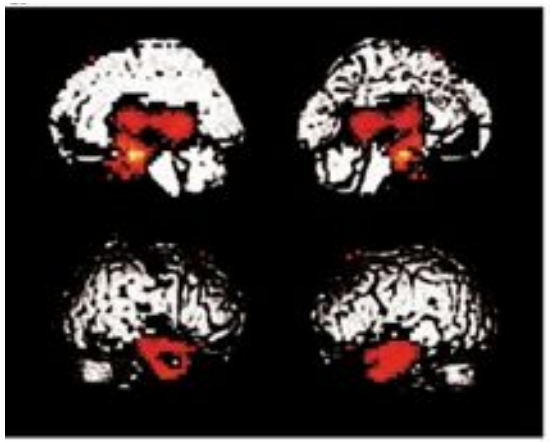




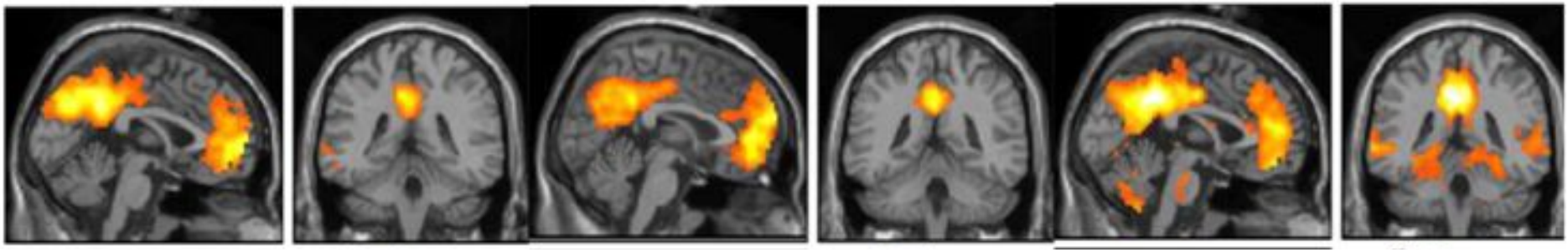






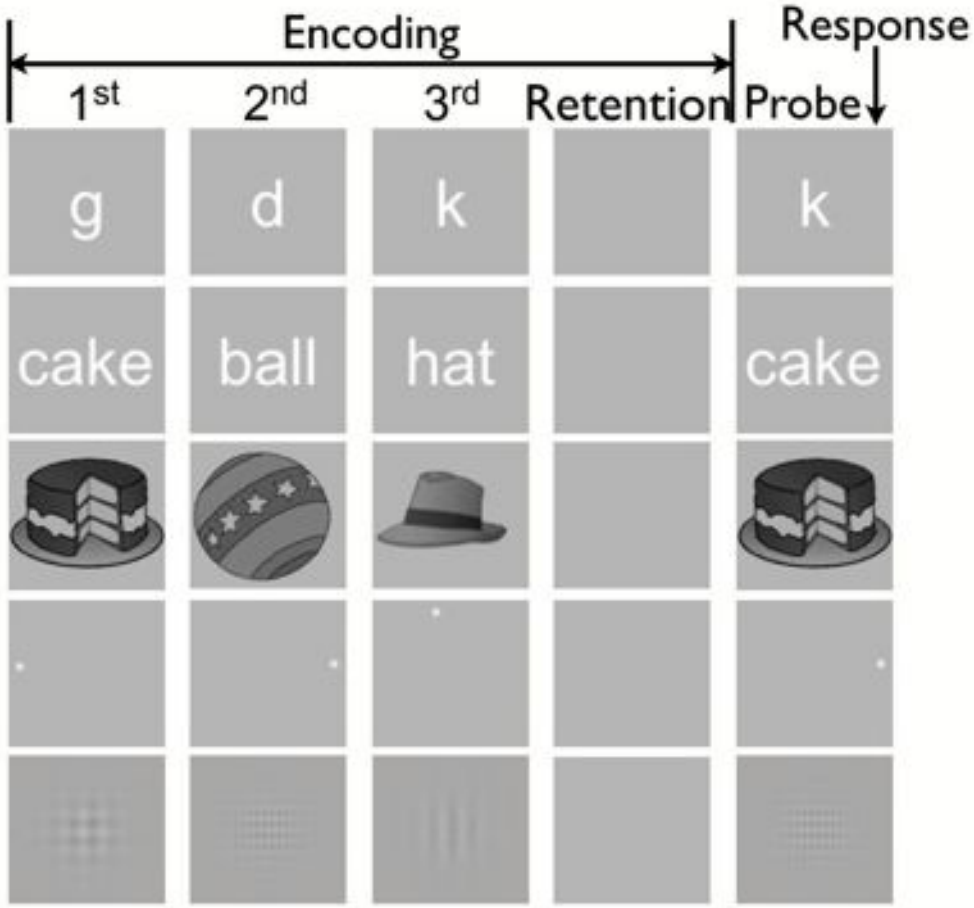


Positive correlation with amygdala



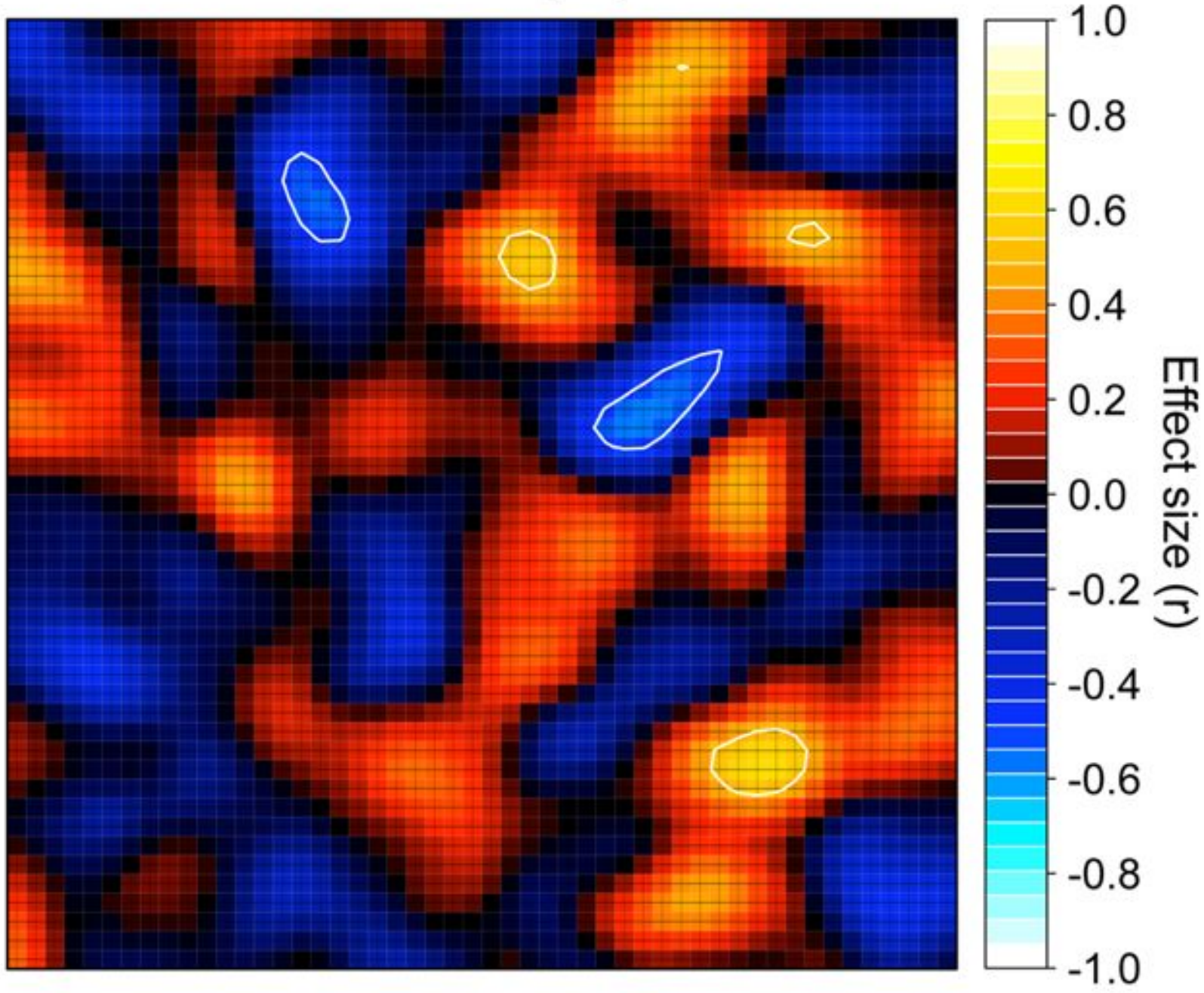
Some problems

- Assumption of pure insertion often fails (Friston et al, 1996)

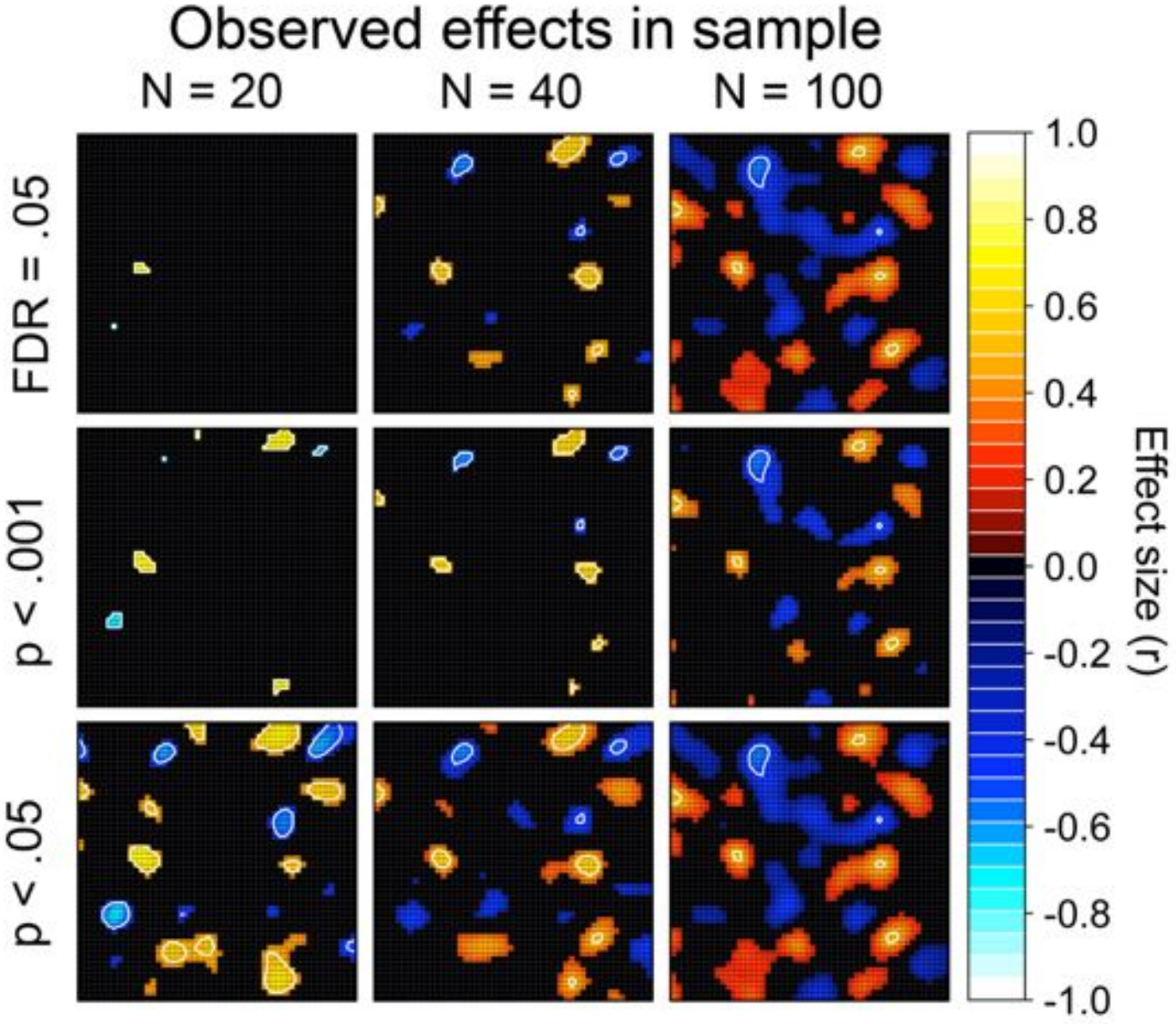


Some problems

- Sensitivity is often very poor
  - Need to correct for thousands of comparisons


Real effects in population
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Effect size (r)

The effects of sampling…

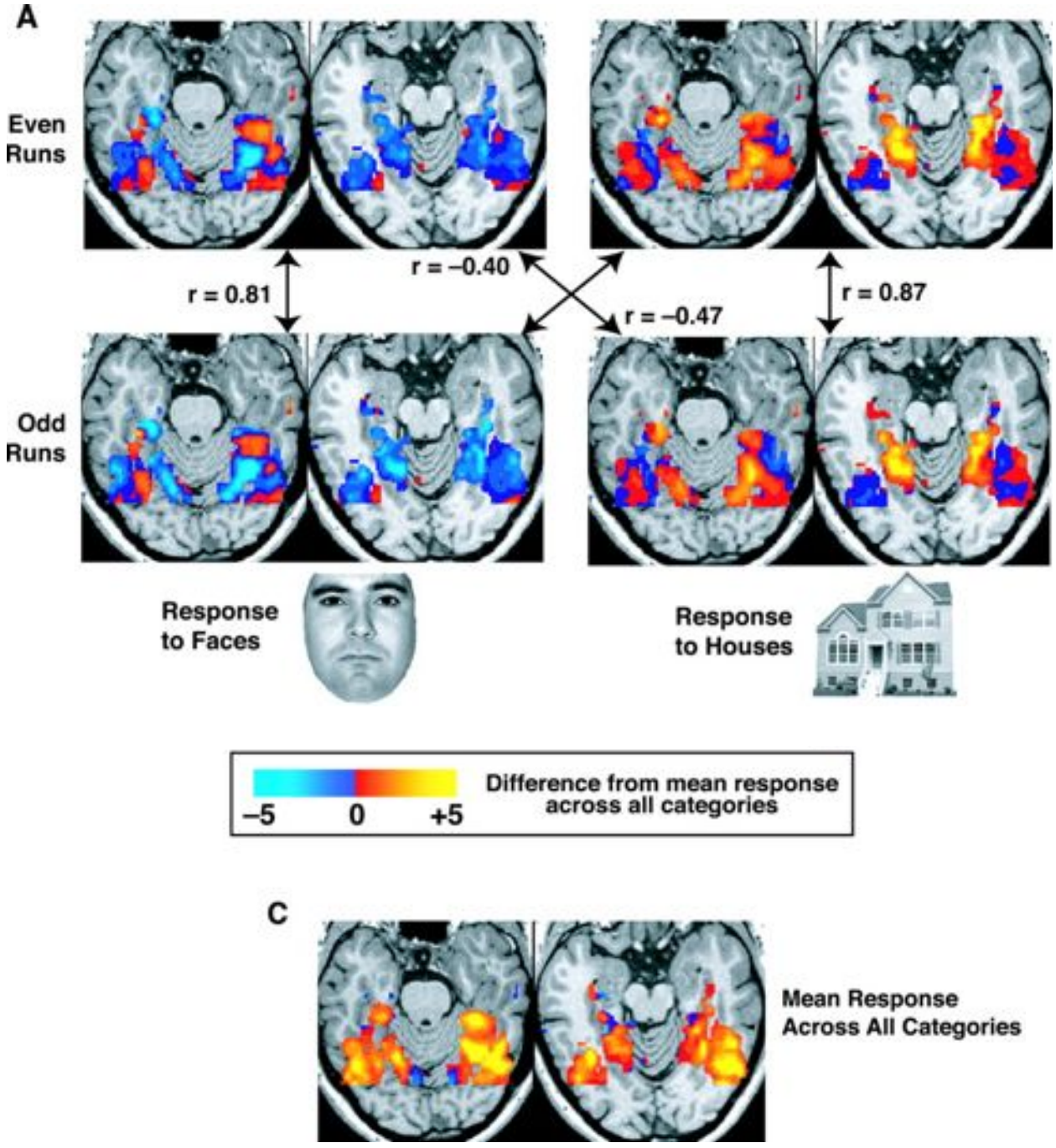


Some problems

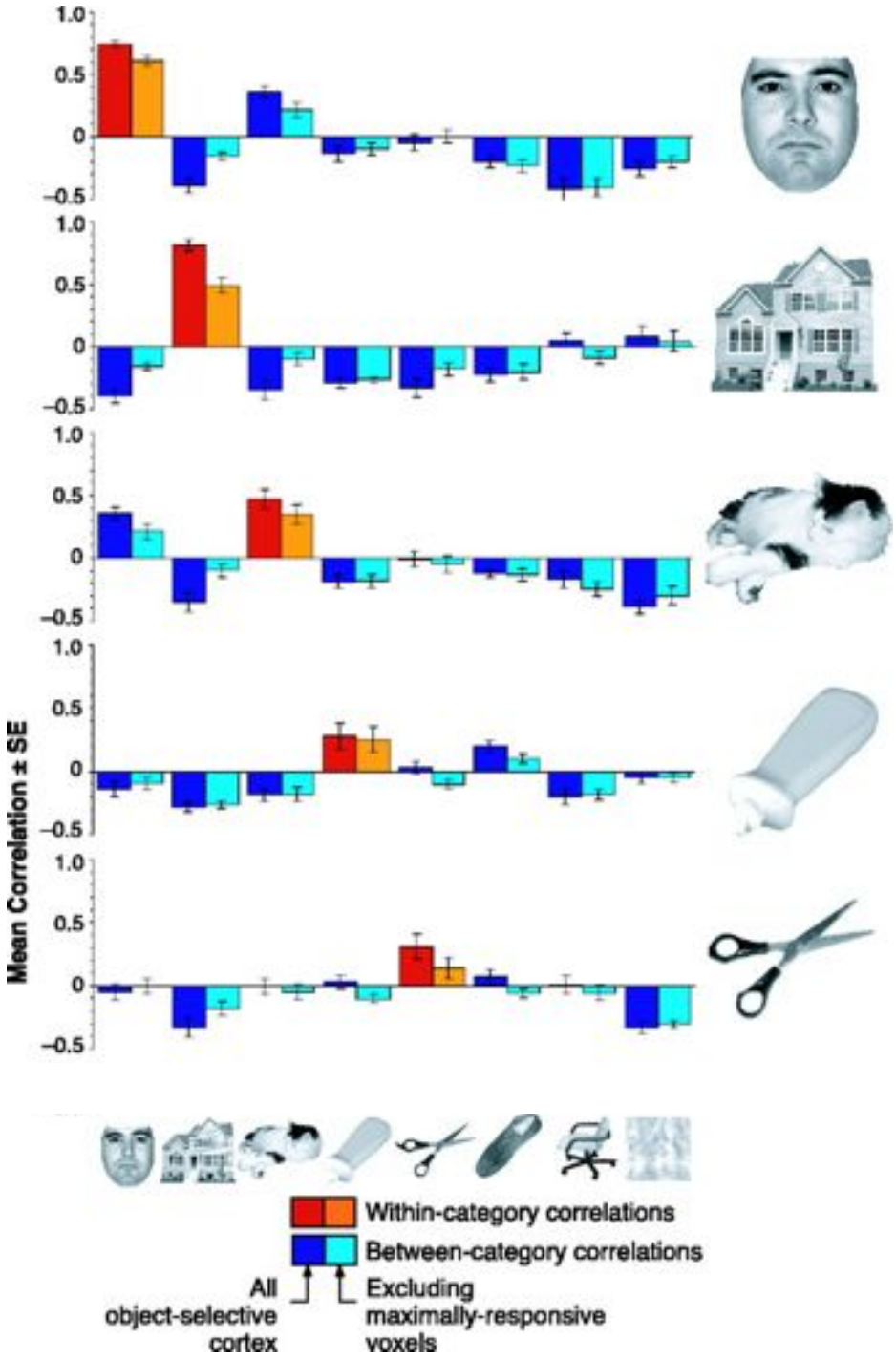
- The brain *isn't* a massively univariate object
- There's all kinds of structure!
  - Local spatial correlations
  - Long-distance networks
  - May remind you of an earlier lecture…
- Shouldn't our analytical methods strive to reflect reality?

The rise of machine learning in fMRI

- Beginning around 2000, people started applying concepts and methods from ML to fMRI data
- First powerful demonstration by Haxby et al (2001)
  - Showed that neural responses to visual categories were widely distributed

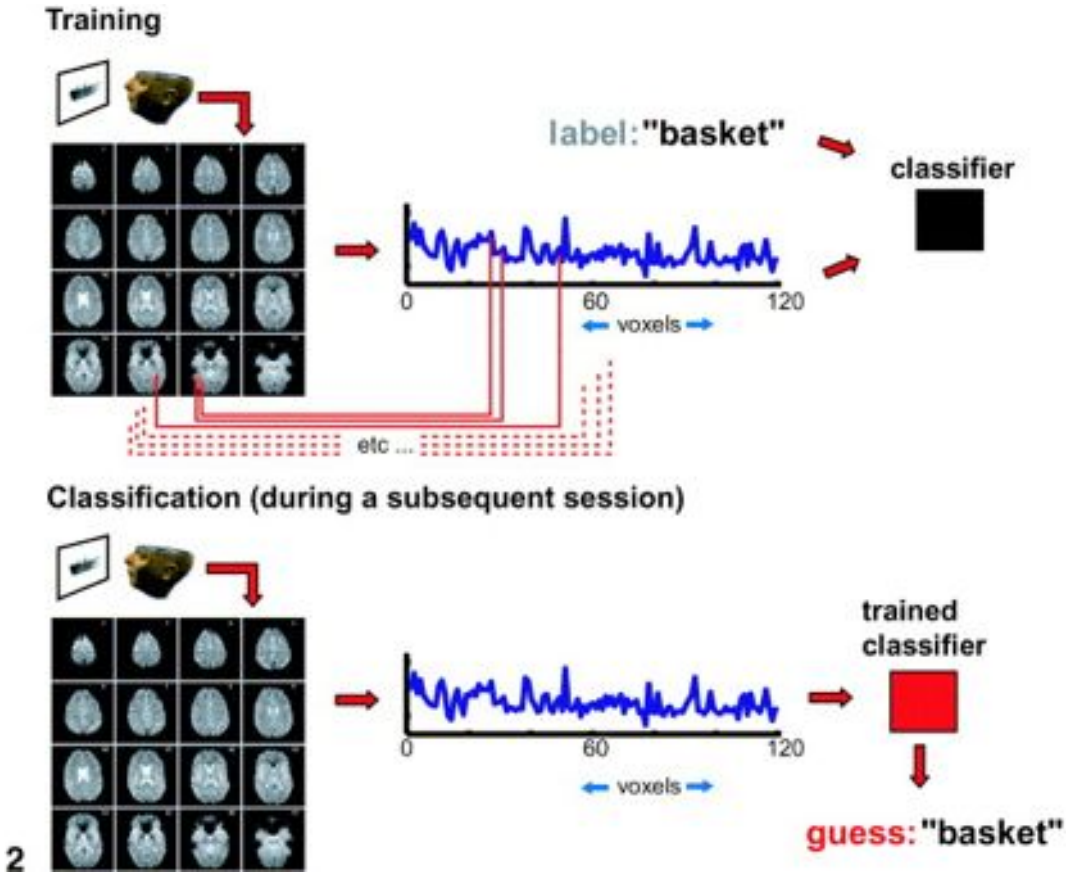


Haxby et al (2001)

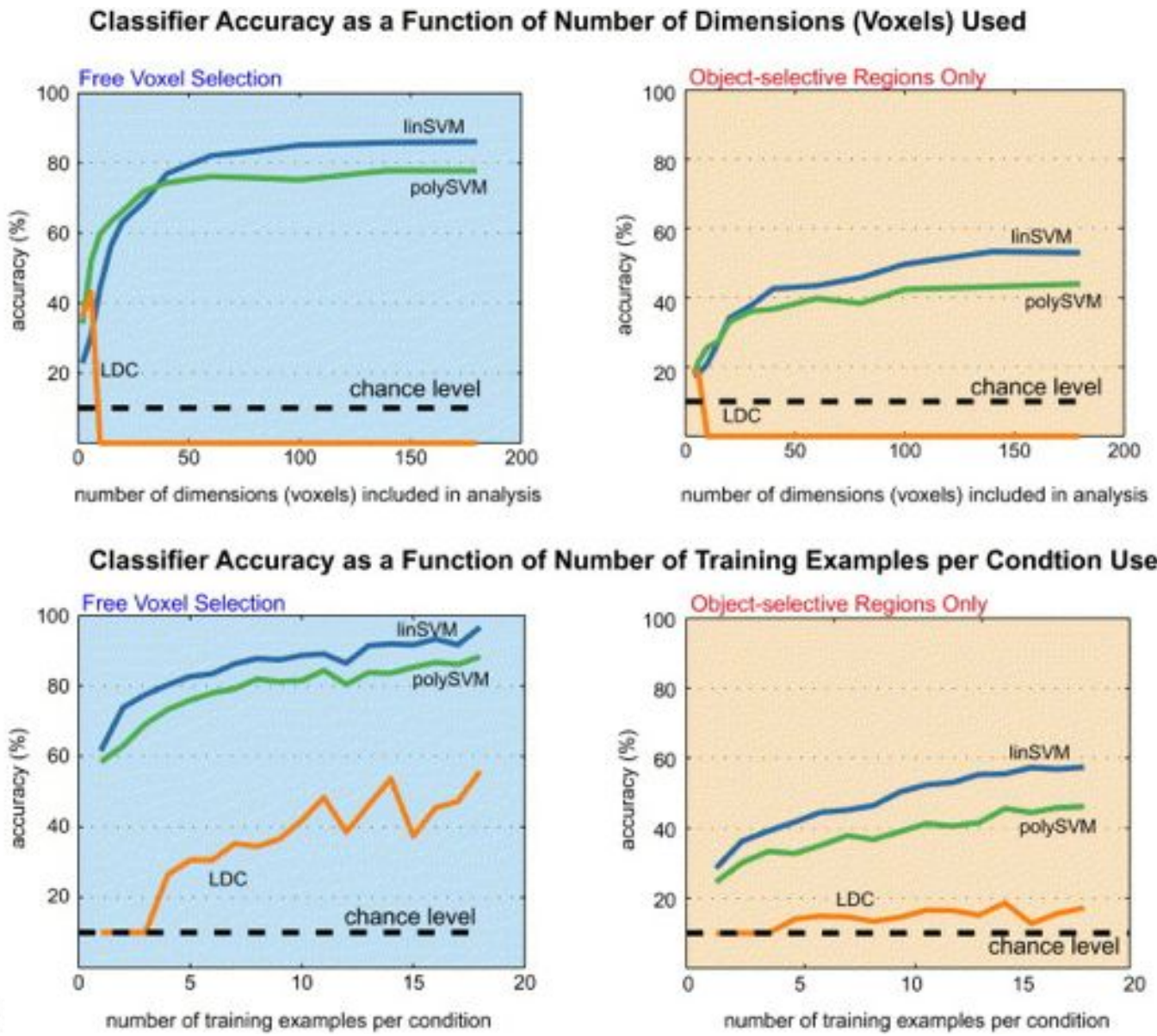

Mean Correlation ± SE
1.0
0.5
0
−0.5
Within-category correlations
Between-category correlations
All object-selective cortex
Excluding maximally-responsive voxels

Early classification studies

- Cox & Savoy (2003)
- Predict which of 10 visual categories subject was looking at
- Feature selection: select category-sensitive voxels
- Several classifiers

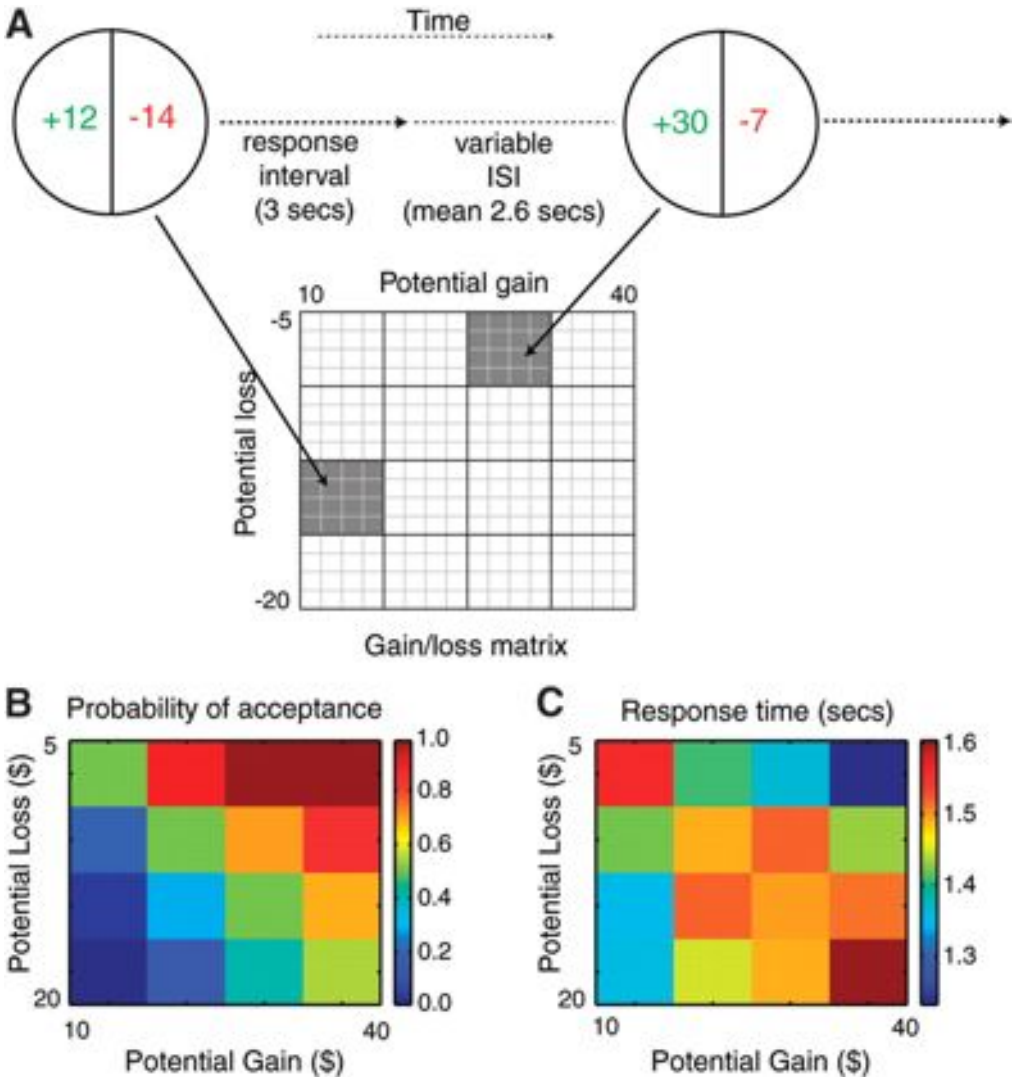


Cox & Savoy (2003)

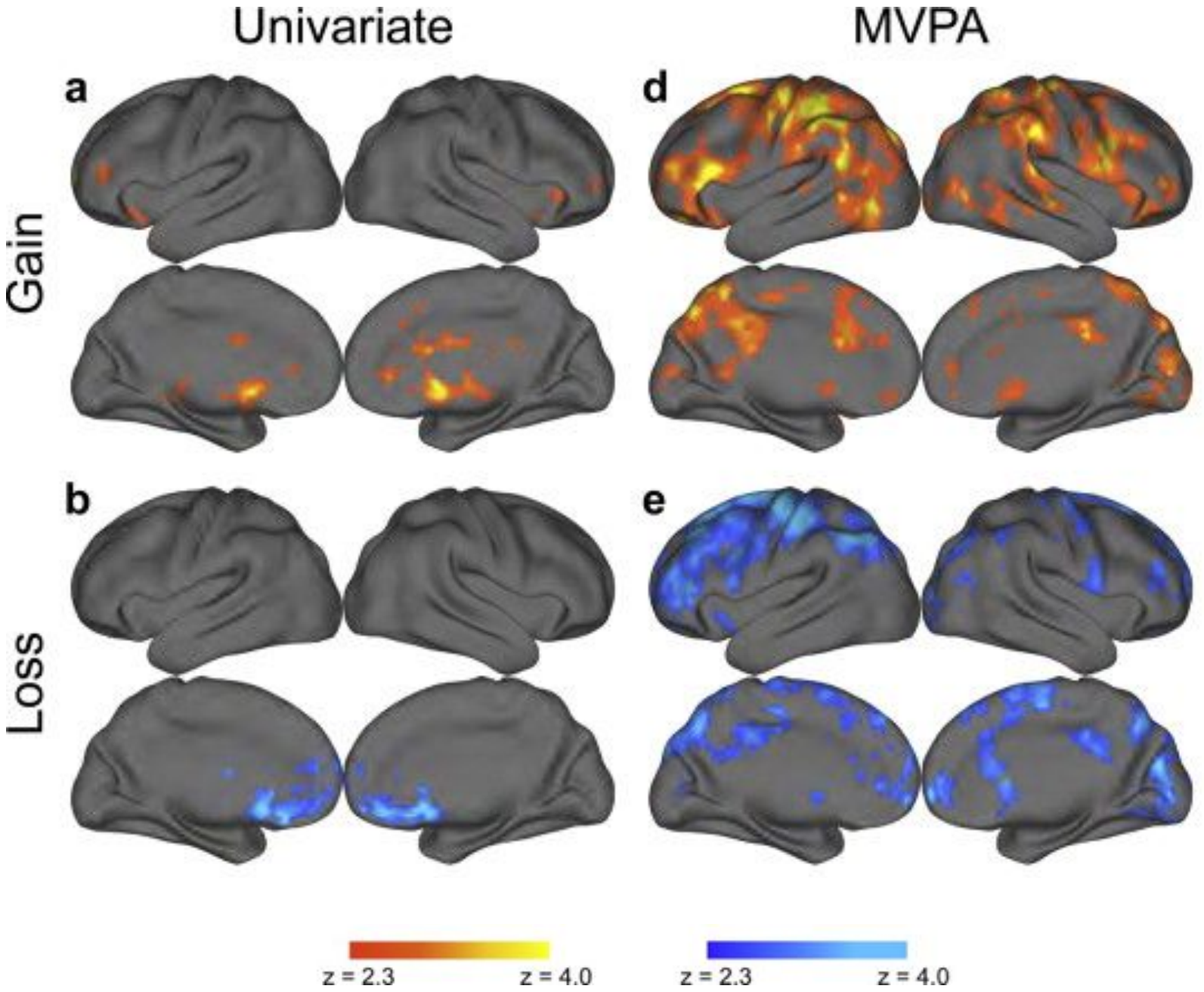


Cox & Savoy (2003)

MVPA vs. univariate approaches

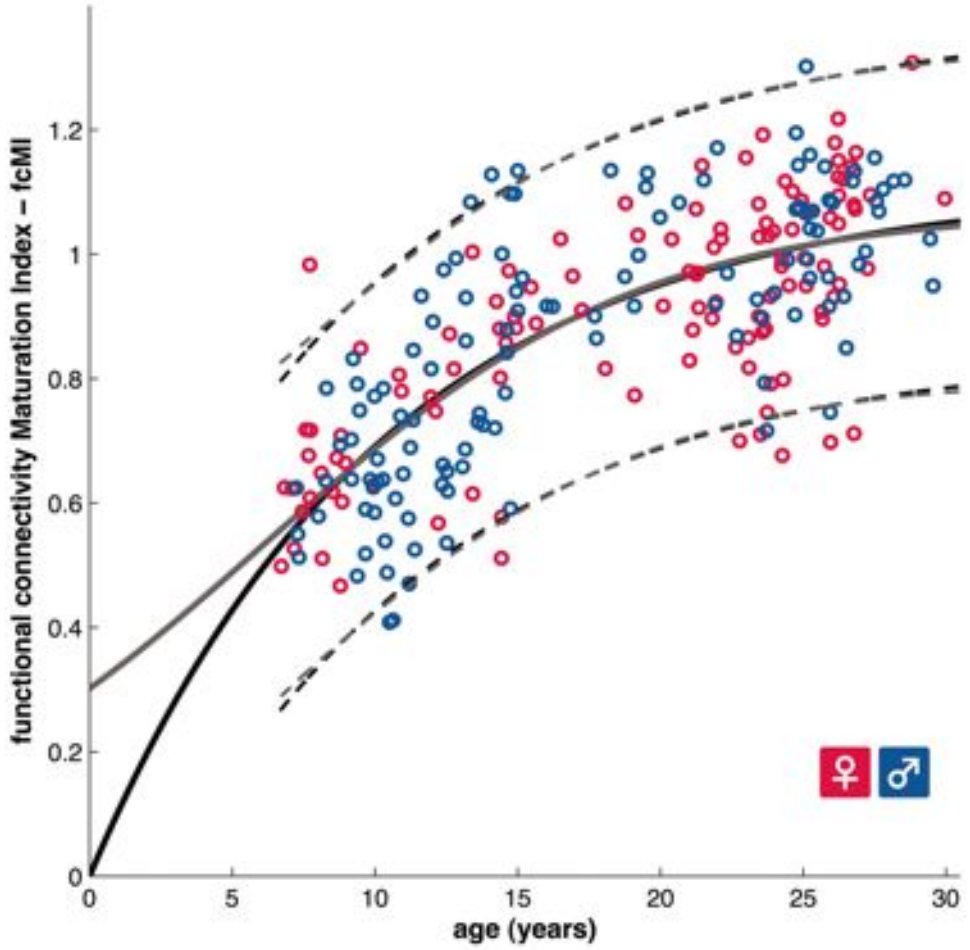


Tom, Fox, Trepel, & Poldrack (2007)



Jimura & Poldrack (2012)

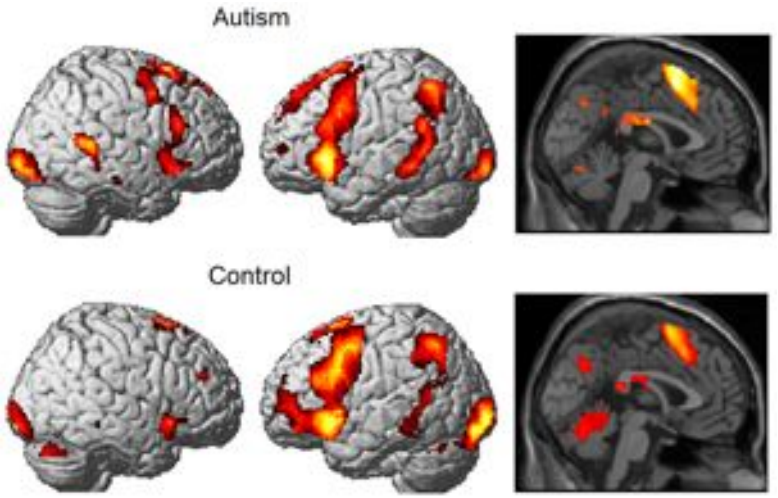
Predicting individual differences…



Dosenbach et al (2010)

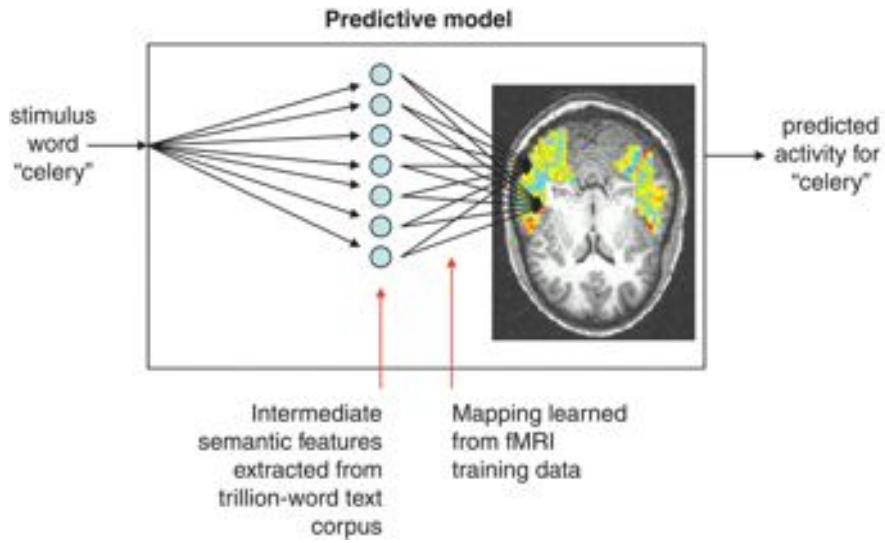
…or diagnoses

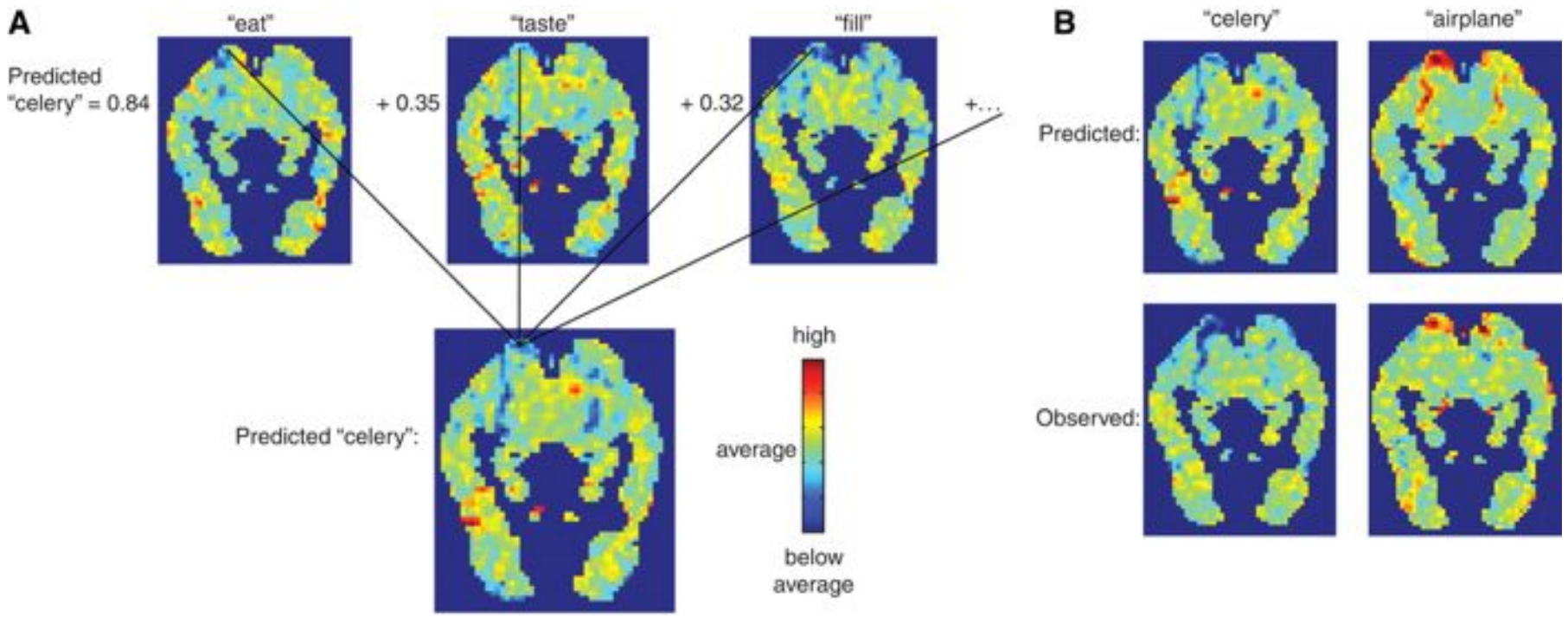
- Hundreds of studies have trained classifiers to discriminate clinical populations from controls
- Often obtain v. high sensitivity & specificity (> 80 - 90%)
- E.g., Just et al (2014) reported 97% accuracy classifying autistic vs. control patients using a GNB classifier



A generative model

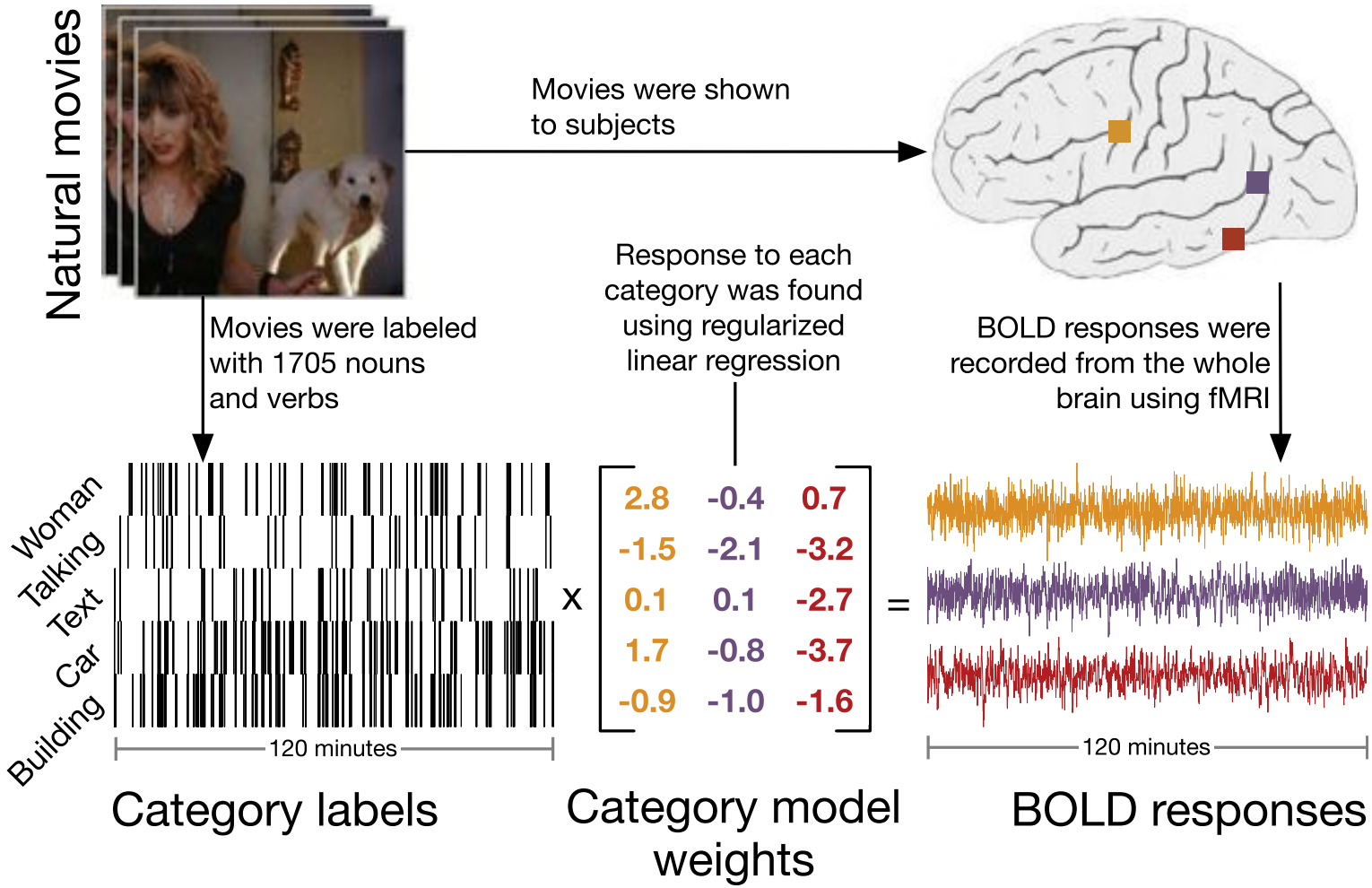
- Mitchell et al (2008) trained a classifier to predict entirely new nouns
- Initially learn (from text corpus) an intermediate set of semantic vectors
- Predict brain activity from learned vectors
- Apply to unseen nouns



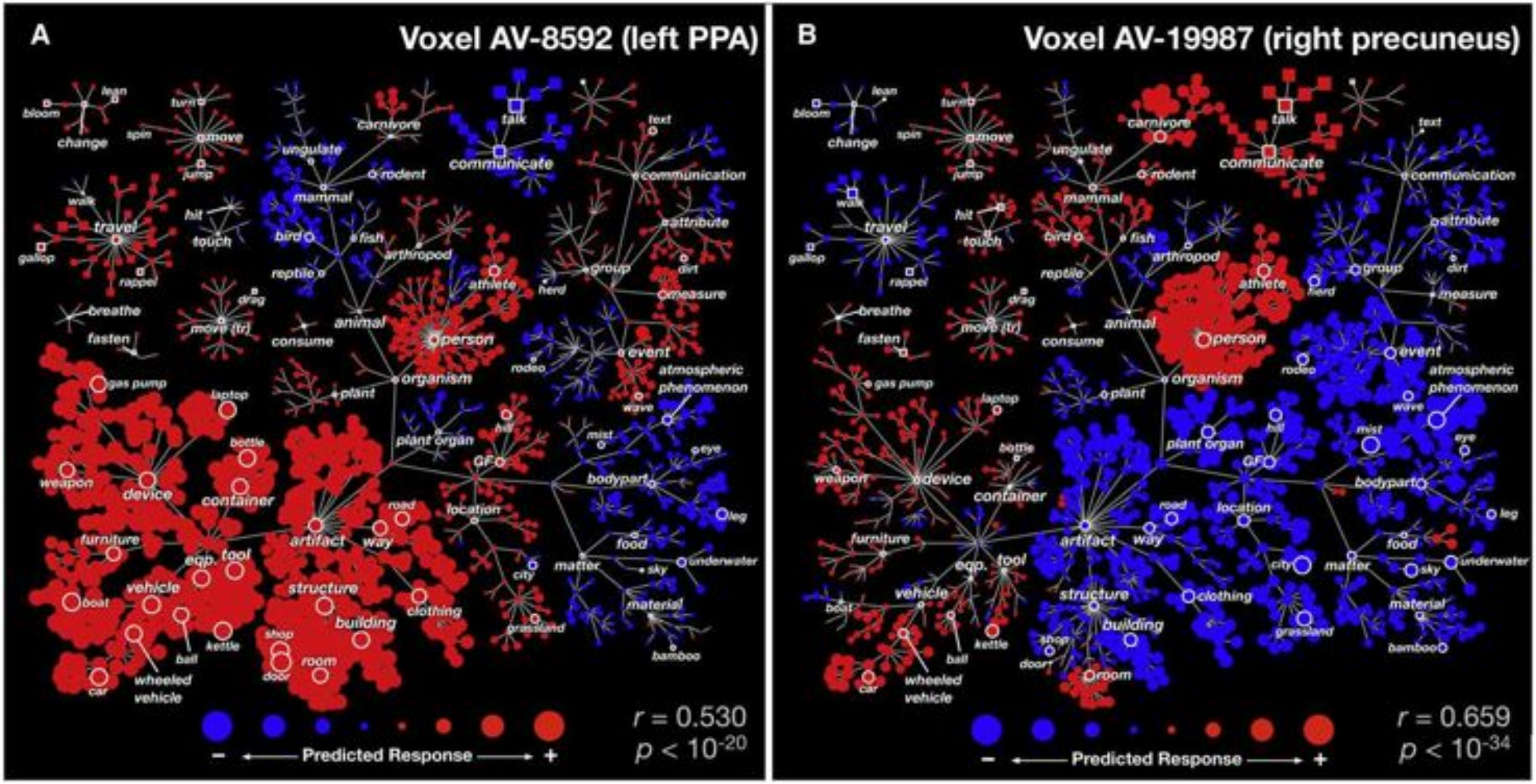


Mitchell et al (2008)

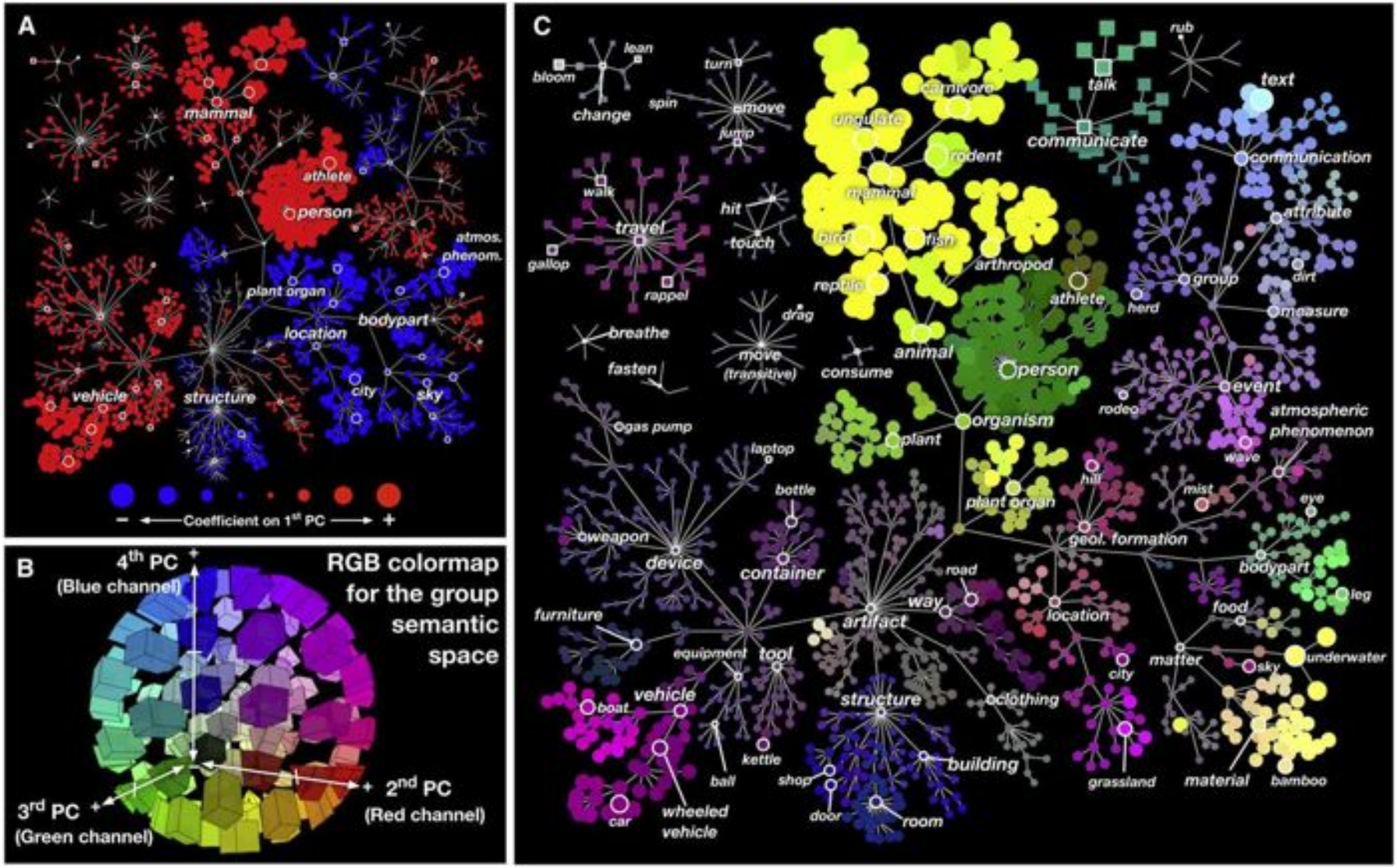
Modeling semantic space



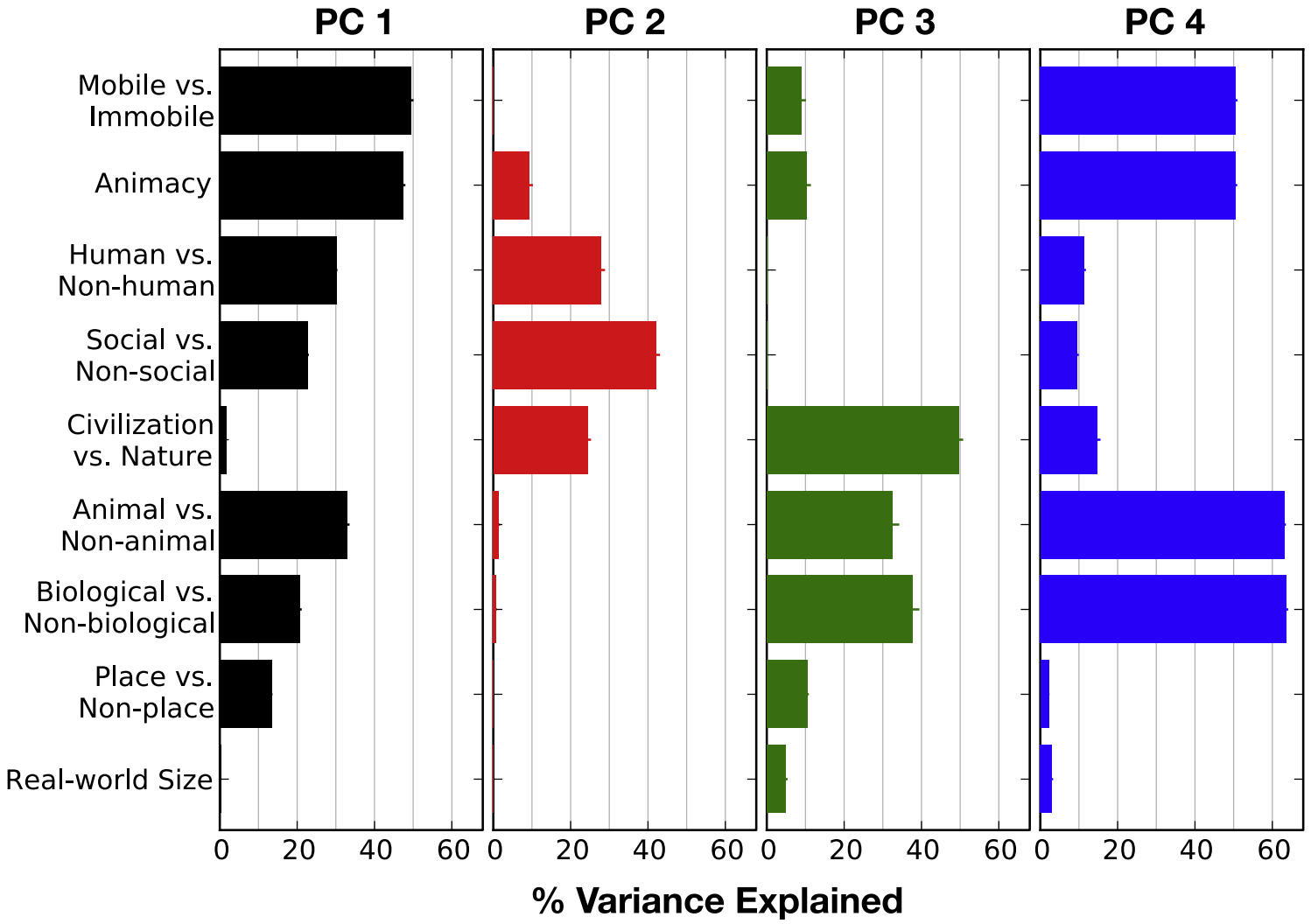
Huth et al (2012)



Huth et al (2012)



Huth et al (2012)



Huth et al (2012)

But keep in mind…

- As always, there are limitations
- We want to avoid that whole “have hammer, must seek nail” thing

What questions are we answering?

- Not exactly the same ones we started out with
- Discussion of neuroanatomy often suspiciously missing from ML papers
  - What does a complex pattern of voxel weights *mean*?
- The famous “black box” in action?

Overfitting, always

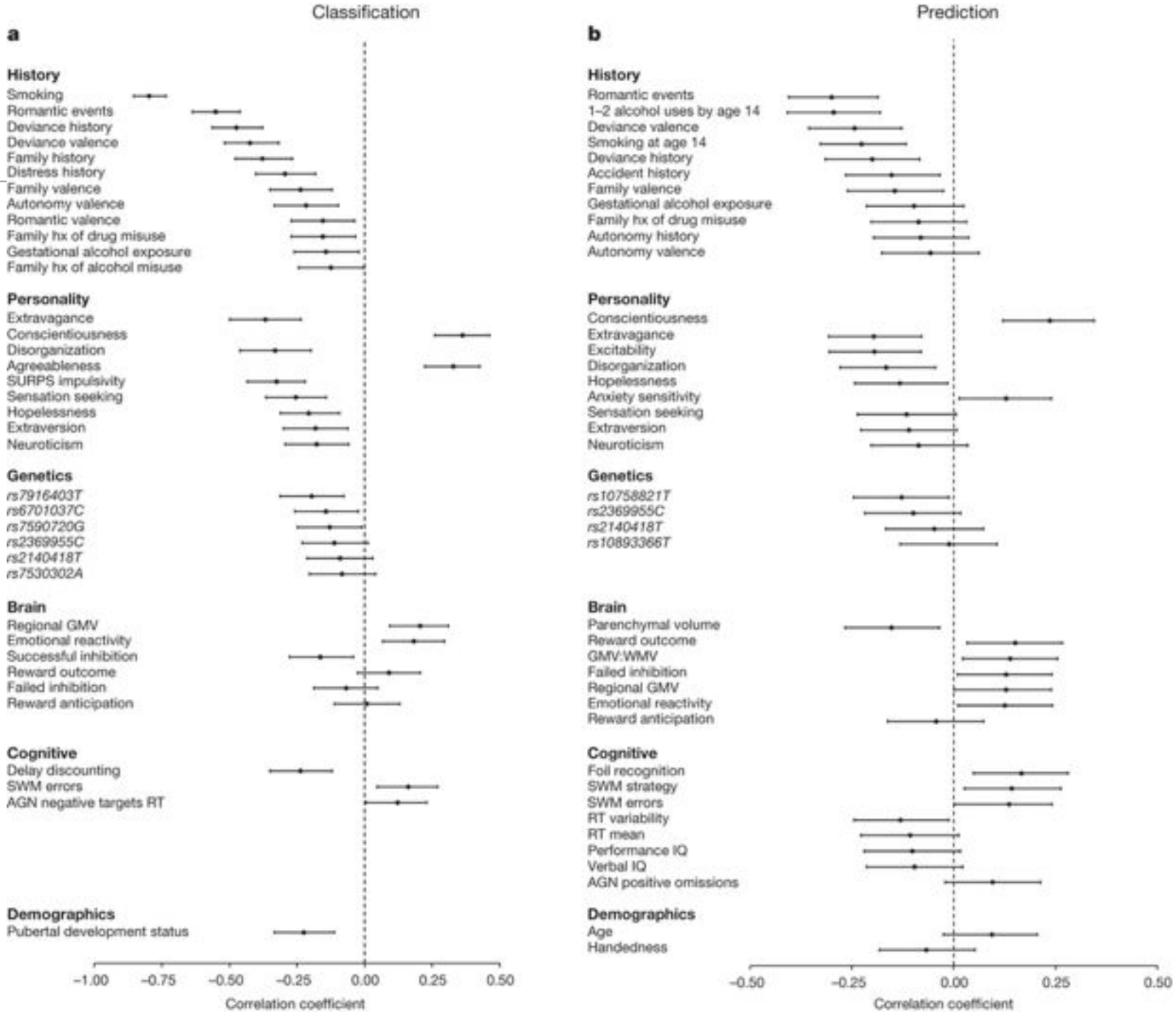
- Cross-validation does not magically protect against overfitting
- Can still overfit during model selection, parameterization, etc.
- Think about reported results in light of prior
  - Performance is bounded by the reliability of the outcome
- Is 95%+ classification of psychiatric phenotypes plausible?
  - Can’t possibly outperform human-generated labels

The ADHD-200 competition

- Challenge: predict ADHD status in ~200 public fMRI datasets
  - Data include resting-state fMRI, anatomical scan, demographics, etc.
- Best performance ~55 - 60% (depending on metric)
- Except… the best team was disqualified
- Why?
  - They didn't use the brain data at all!
  - "For the record, we tried a pile of imaging-based approaches. As a control, we also did classification with age, gender, etc. but no imaging data. It was actually very frustrating for us that none of our imaging-based methods did better than the no imaging results. It does raise some very interesting issues." (Matthew Brown, personal communication)

Prediction is hard

- Whelan et al (2014): predict binge drinking at age 16 from behavioral, genetic, neuroimaging data at age 14 (n = 692)
- Almost everything is useful to some degree
- But overall performance is surprisingly modest (ROC AUCs ~ .75)
- By far the best predictors are drinking/smoking at age 14
  - Incremental value of other features is small



Whelan et al (2014)

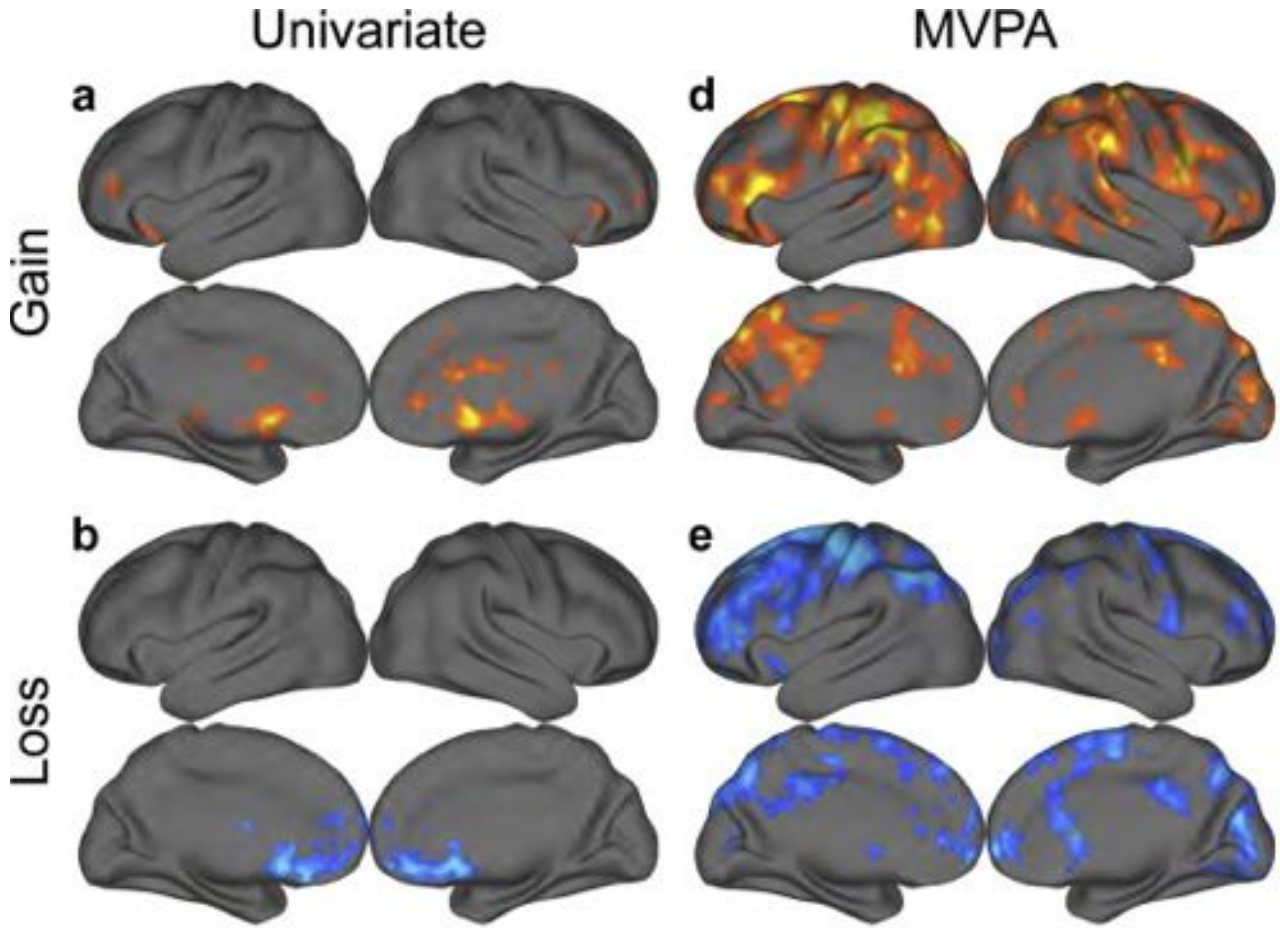
Correlation doesn't imply…

- Do pattern classification approaches bring us closer to causal mechanism?
- A classifier will use *any* information it can to make a prediction!

> While a pattern recognition approach shows great promise for extracting large amounts of information from fMRI data and for guiding multivariate exploration of representation in the human brain, one must always remain cautious about the nature of the information that a classifier is using to distinguish different classes of stimuli. The fact that information can be extracted by our analysis does not necessarily mean that this information is used by the brain or that the information is used in the way that we think it is. One must always remain conscious of this concern for all analysis techniques that are fundamentally correlational (including traditional univariate fMRI data analysis).

Cox & Savoy (2003)

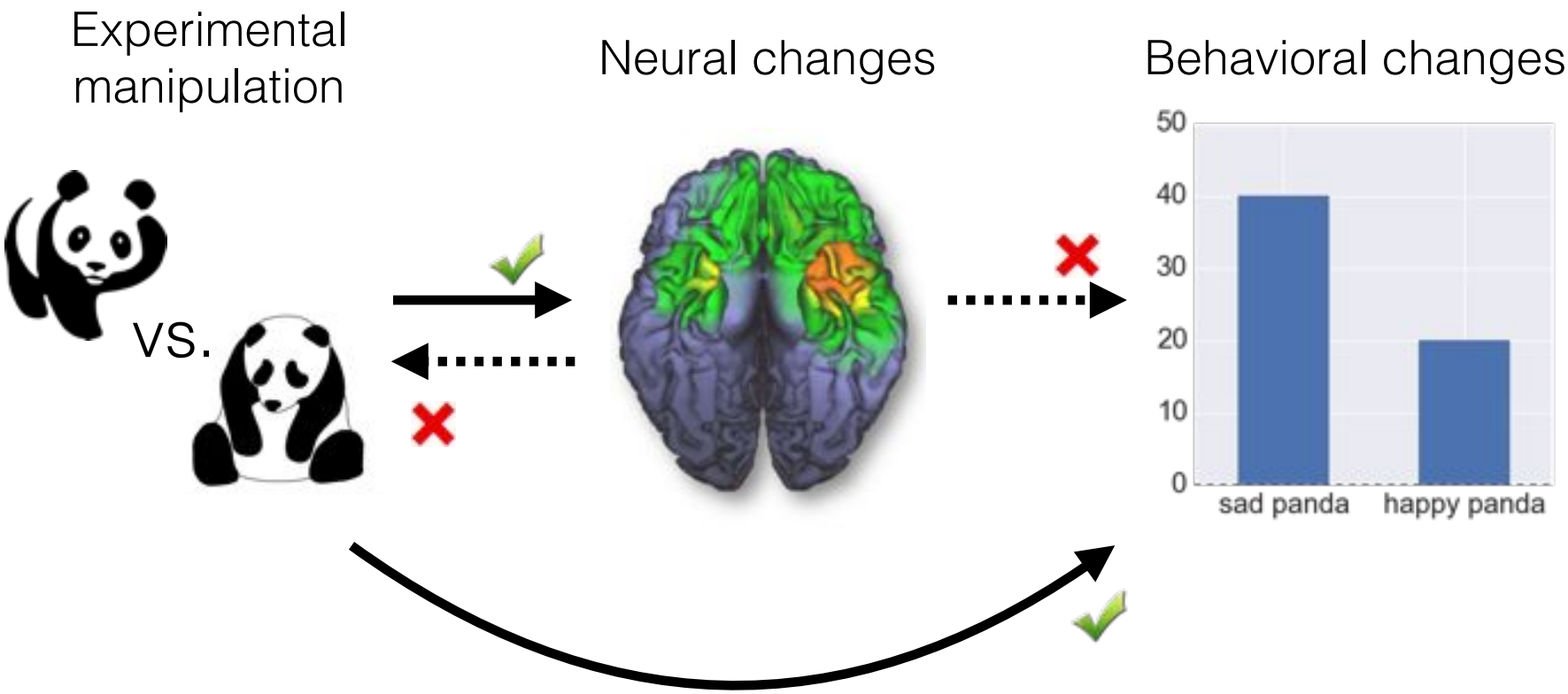
Gains and losses, or…?



Jimura & Poldrack (2012)

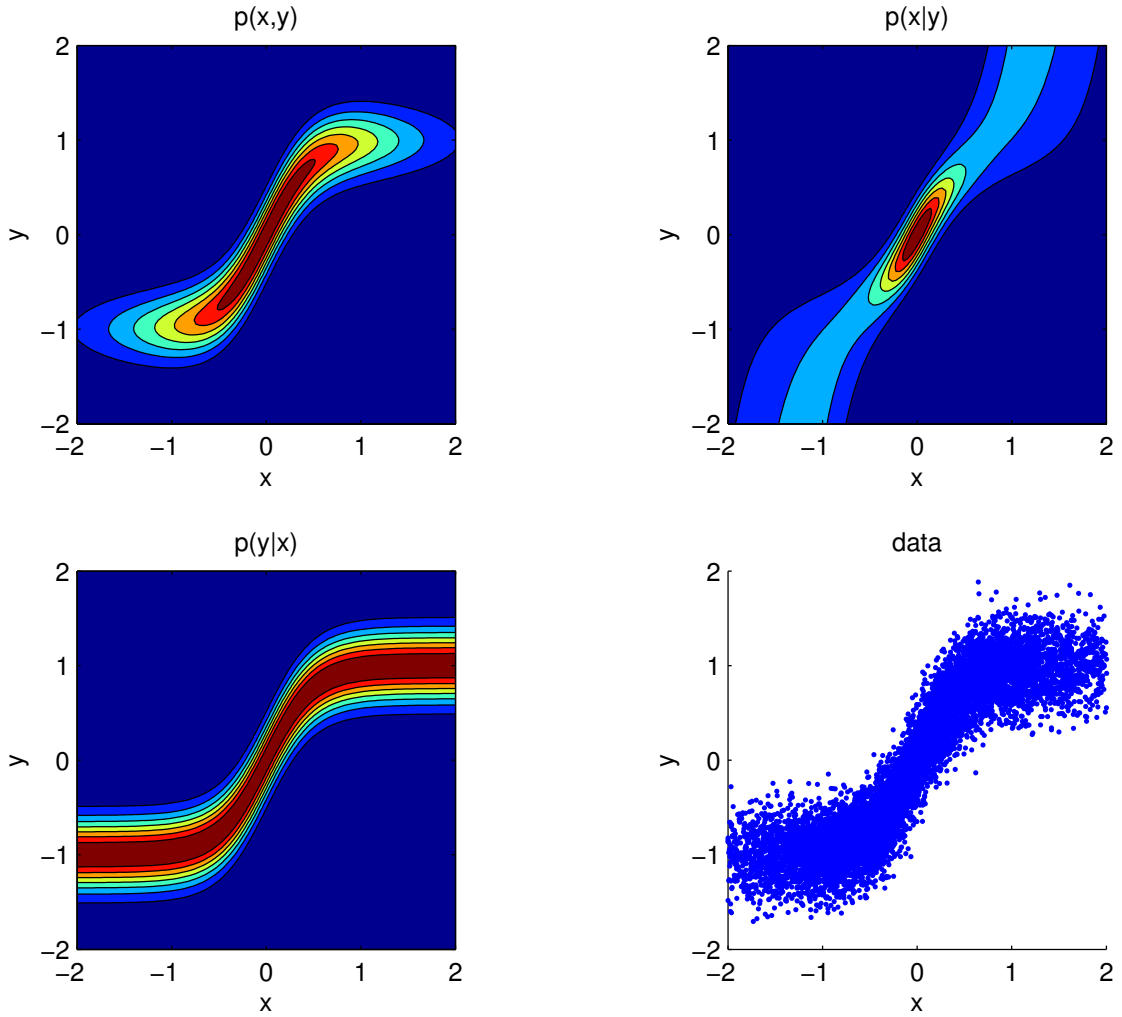
Causal inference in fMRI experiments

- In some cases, predictive approaches actually make it more difficult to draw causal conclusions
- In classical experimental paradigm (manipulation —> brain —> behavior), which links support causal inference?



Can we ever infer causation *just* from correlation?

- Maybe…



Mooij et al. (2014)

Figure 2: Identifiable ANM with $Y = \tanh(X) + E$, where $X \sim \mathcal{N}(0, 1)$ and $E \sim \mathcal{N}(0, 0.5^2)$.

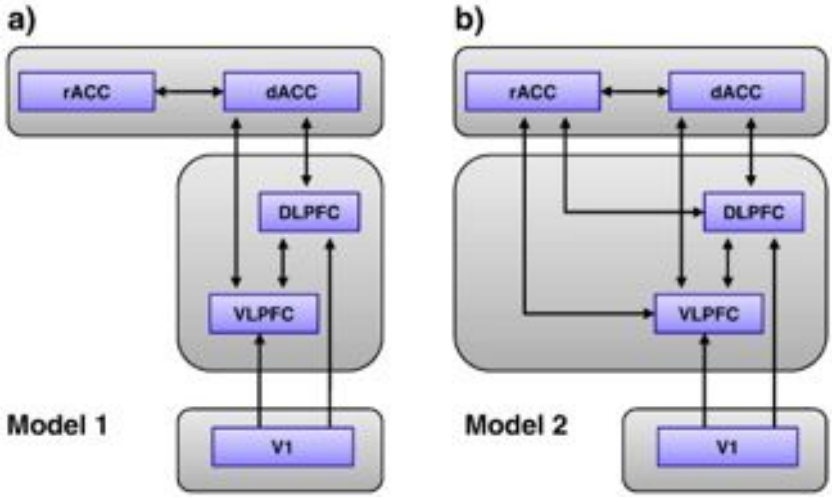
But...

Now suppose that we only have data from the observational distribution $\mathbb{P}_{X,Y}$ (for example, because doing intervention experiments is too costly). Can we then still infer the causal relationship between X and Y? We will simplify matters by considering only (a) and (b) in Figure 1 as possibilities. In other words, we assume that X and Y are dependent (i.e., $\mathbb{P}_{X,Y} \neq \mathbb{P}_X \mathbb{P}_Y$), there is no confounding (common cause of X and Y), no selection bias (common effect of X and Y that is implicitly conditioned on), and no feedback between X and Y (a two-way causal relationship between X and Y). Inferring the causal direction between X and Y, i.e., deciding which of the two cases (a) and (b) holds, using *only the observational distribution* $\mathbb{P}_{X,Y}$ is the challenging task that we consider here. If, under certain assumptions, we can decide upon the causal direction, we say that the causal direction is *identifiable* from the observational distribution.

Mooij et al. (2014)

Meanwhile, in the real world…

- Can we model causal relationships with fMRI?
- Not directly using the BOLD signal!
  - Why?
- So we have to do some deconvolution
  - Causally model deconvolved neuronal responses
  - Dynamic Causal Modeling (Friston, Harrison, & Penny, 2003)



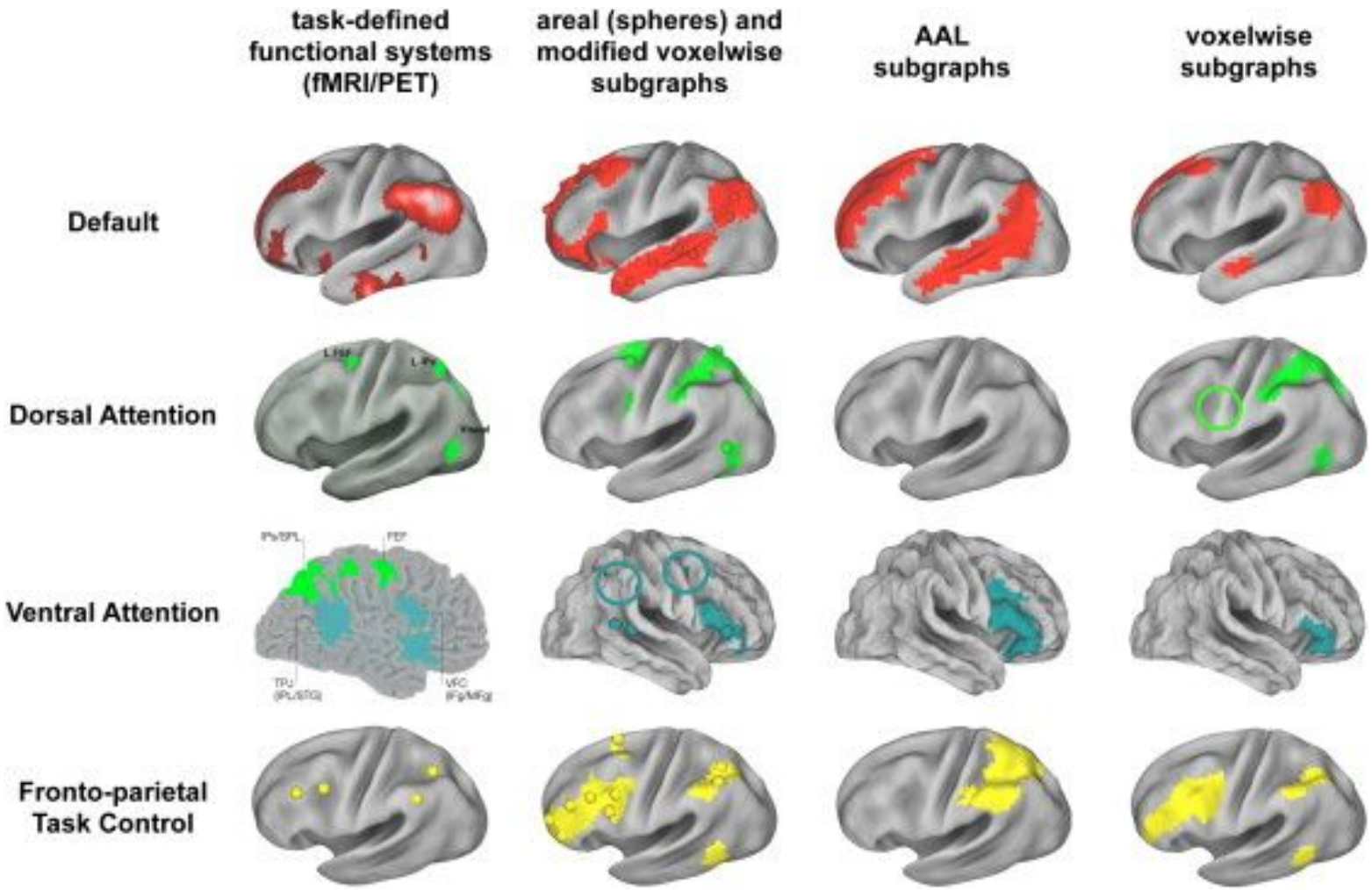
Schlösser et al (2008)

- But… the HRF varies systematically across people, tasks, brain regions, etc.
- Still have standard omitted variables problem
- Is this approach really plausible?

What's the "right" description?

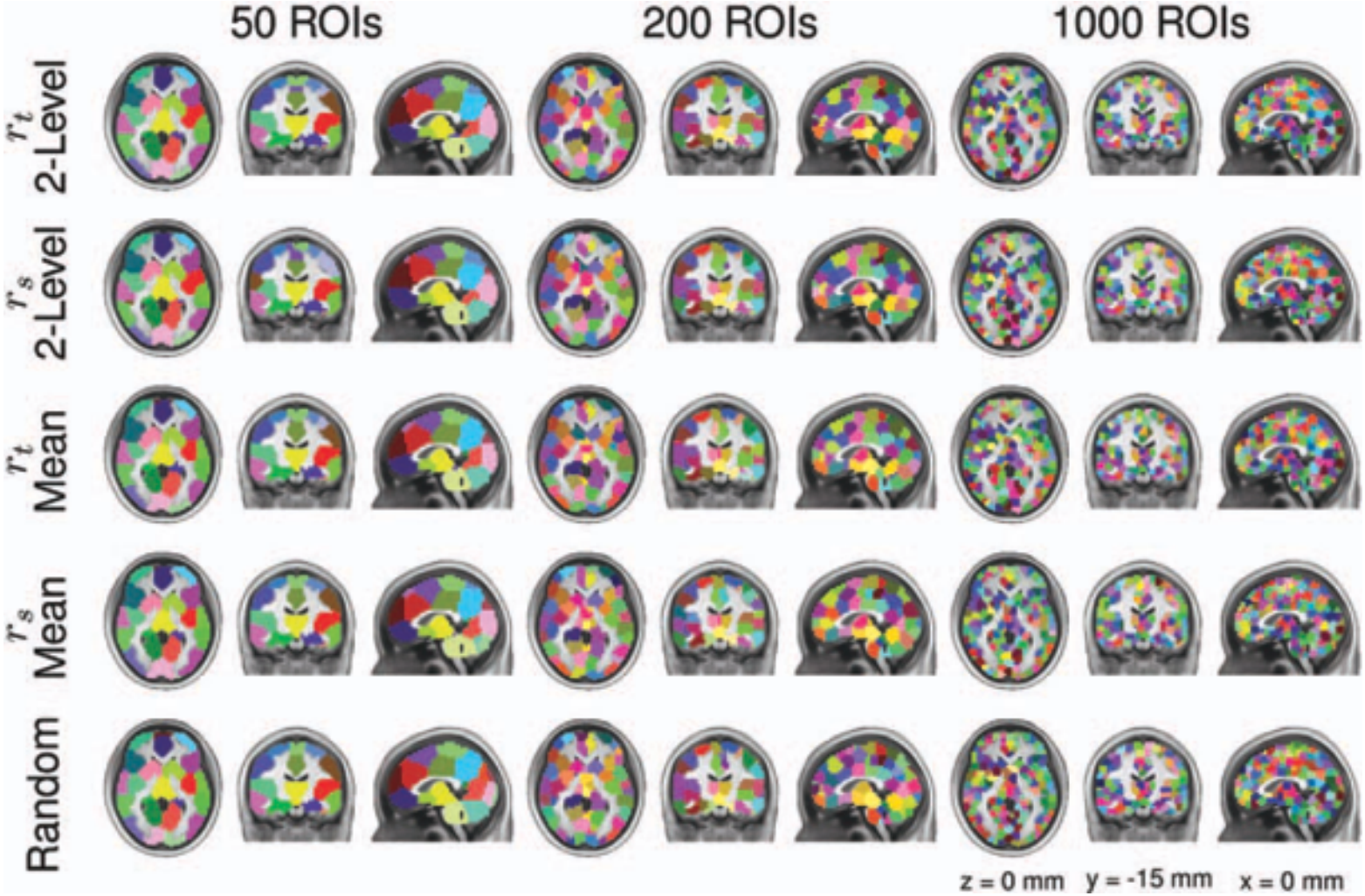
- The brain/mind is a high-dimensional object
- Is there a single optimal low-dimensional description?
  - Absolutely not
- And yet, almost every method gives good results!
- E.g., structure of brain networks

“Functional Network Organization of the Human Brain”



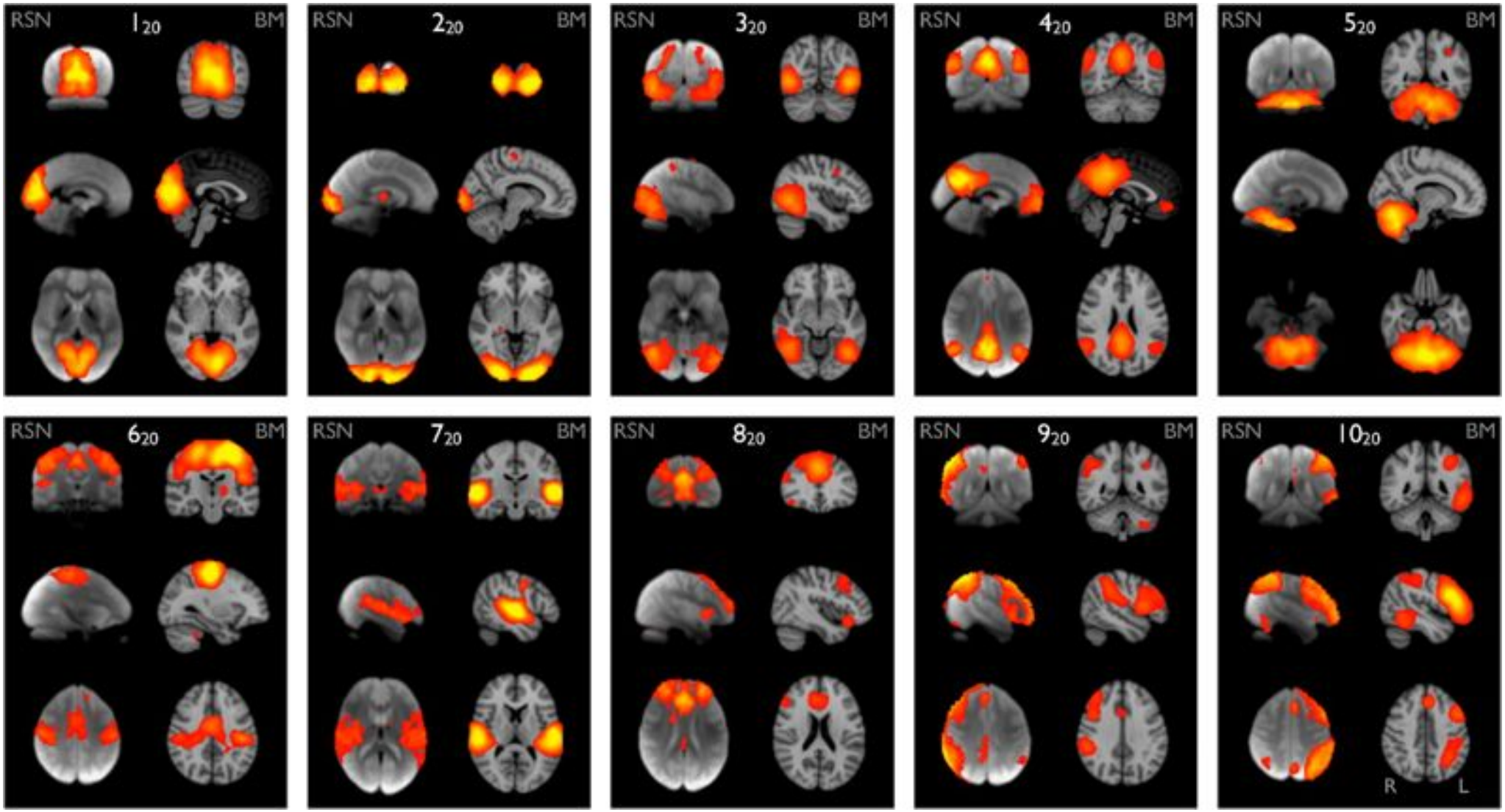
Power et al. (2011)

“A Whole Brain fMRI Atlas Generated via Spatially Constrained Spectral Clustering”



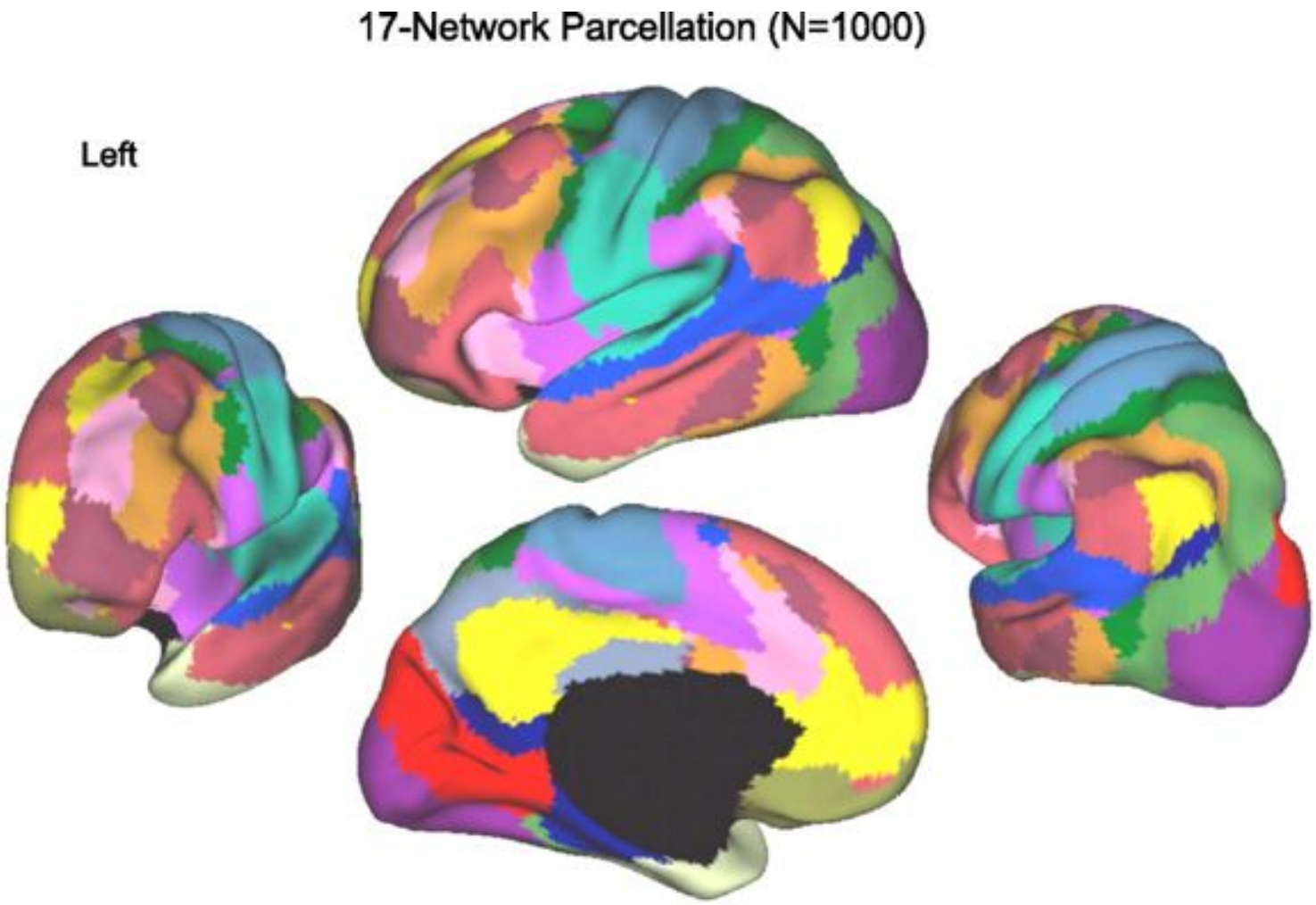
Craddock et al (2011)

“Correspondence of the brain's functional architecture during activation and rest”



Smith et al. (2011)

“The organization of the human cerebral cortex estimated by intrinsic functional connectivity”

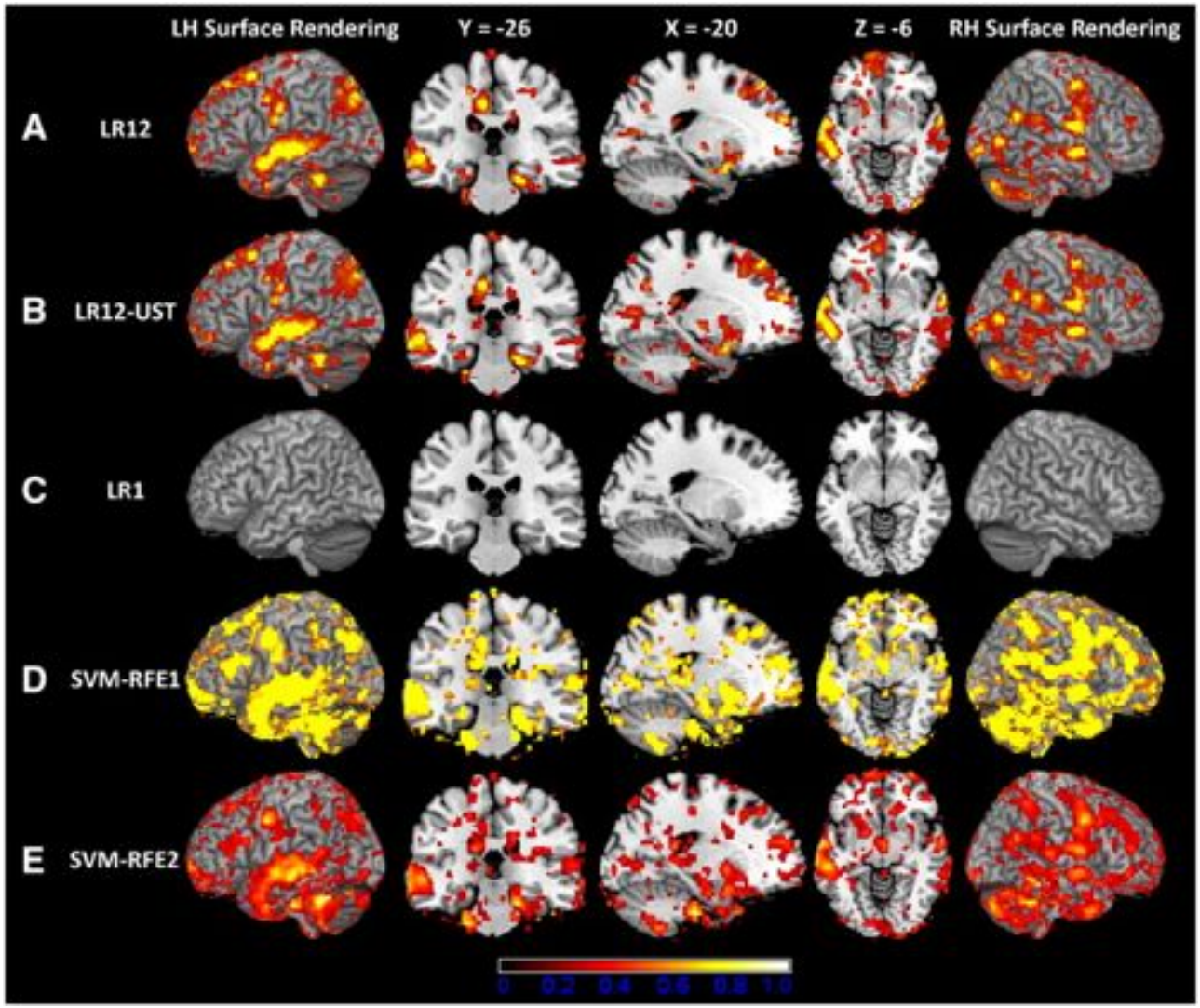


Yeo et al. (2011)

How is this possible?

- Why does almost everything we do look so good?
- Because…
  - A huge amount of structure
  - Massive redundancy
- Remember ‘multiclustering’?
- Most high-D real-world datasets have **many** parsimonious, interpretable low-D descriptions
  - Often reproducible across datasets, contexts
- Doesn’t mean we have the right generative model!

Parsimony is no guarantee



Ryali et al (2010)

Parsimony is no guarantee

- Importance of a feature depends on what else is in the model
- Is it a good idea to select/reduce features prior to estimation?
- Depends…

Interpretation != generation

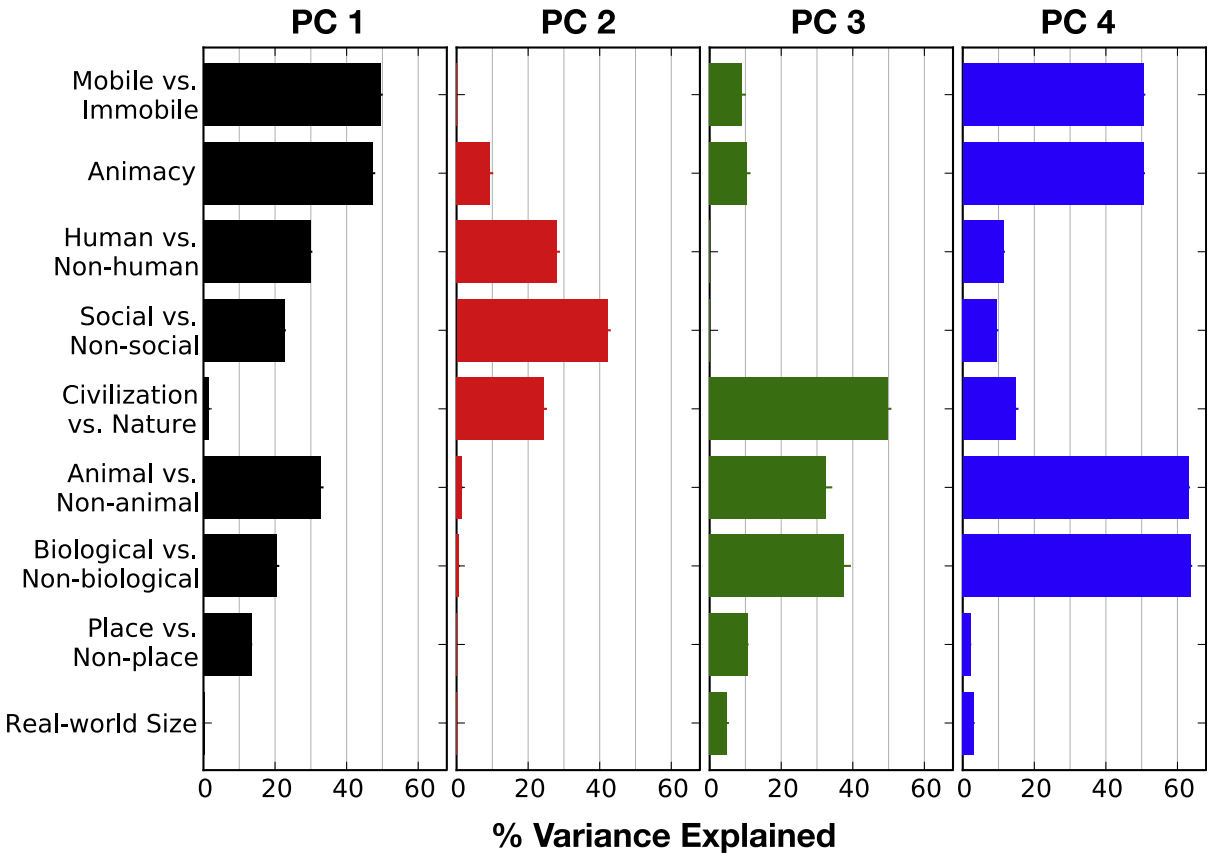


Figure 6. Comparison between the Group Semantic Space and Nine Hypothesized Semantic Dimensions

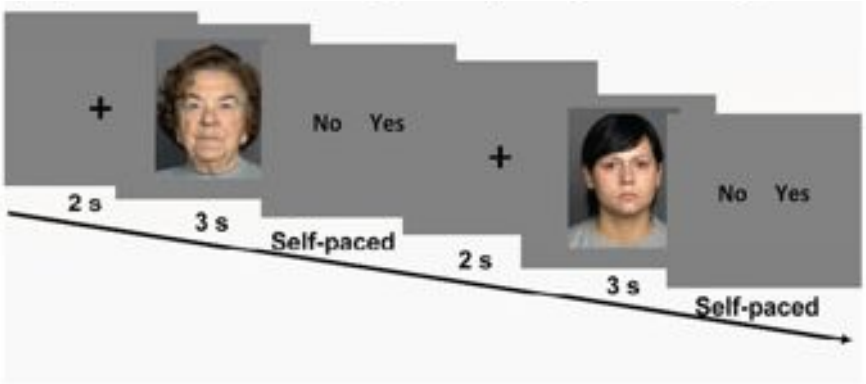
For each hypothesized semantic dimension, we assigned a value to each of the 1,705 categories (see Experimental Procedures for details) and we computed the fraction of variance that each dimension explains in each PC. Each panel shows the variance explained by all hypothesized dimensions in one of the four group PCs. Error bars indicate bootstrap SE. The first PC is best explained by a dimension that contrasts mobile categories (people, nonhuman animals, and vehicles) with nonmobile categories and an "animacy" dimension (Connolly et al., 2012) that assigns high weight to humans, decreasing weights to other mammals, birds, reptiles, fish, and invertebrates, and zero weight to other categories. The second PC is best explained by a dimension that contrasts social categories (people and communication verbs) with all other categories. The third PC is best explained by a dimension that contrasts categories associated with civilization (people, man-made objects, and vehicles) with categories associated with nature (nonhuman animals). The fourth PC is best explained by a dimension that contrasts biological categories (people, animals, plants, body parts, and plant parts) with nonbiological categories and a dimension that contrasts animals (people and nonhuman animals) with nonanimals.

Huth et al. (2012)

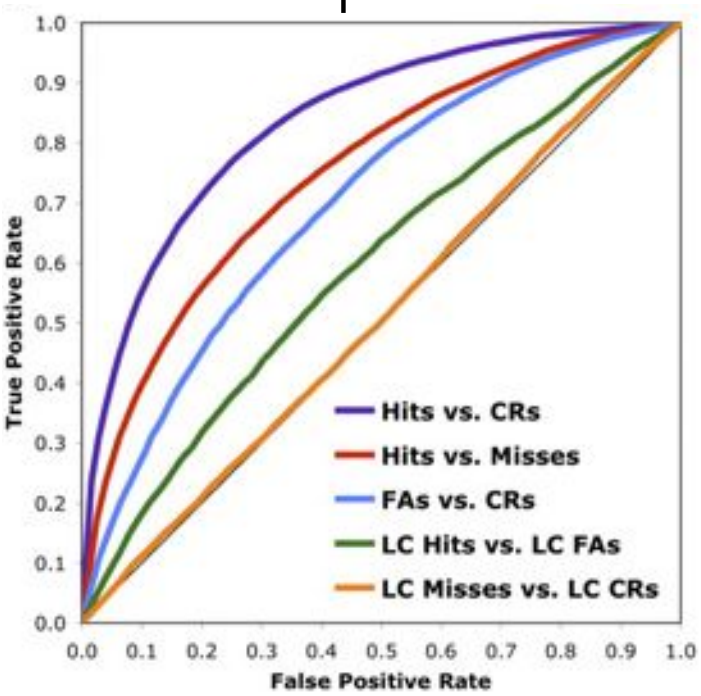
Don’t panic!

- These aren’t ML-specific problems
  - They’re reality problems
- Mostly apply to theoretical science
  - Much less of a concern in applied domains—e.g., clinical prediction
- ML approaches can be used instrumentally to address theoretical questions
  - Often much more powerful than conventional approaches

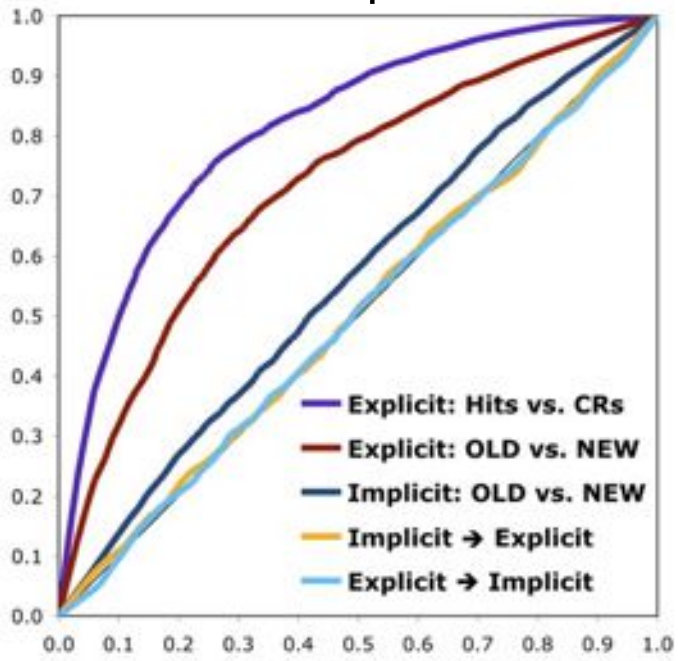
- Is there a detectable neural signature of seen-but-forgotten items?
- I.e., does the brain encode information we can't readily retrieve when prompted?



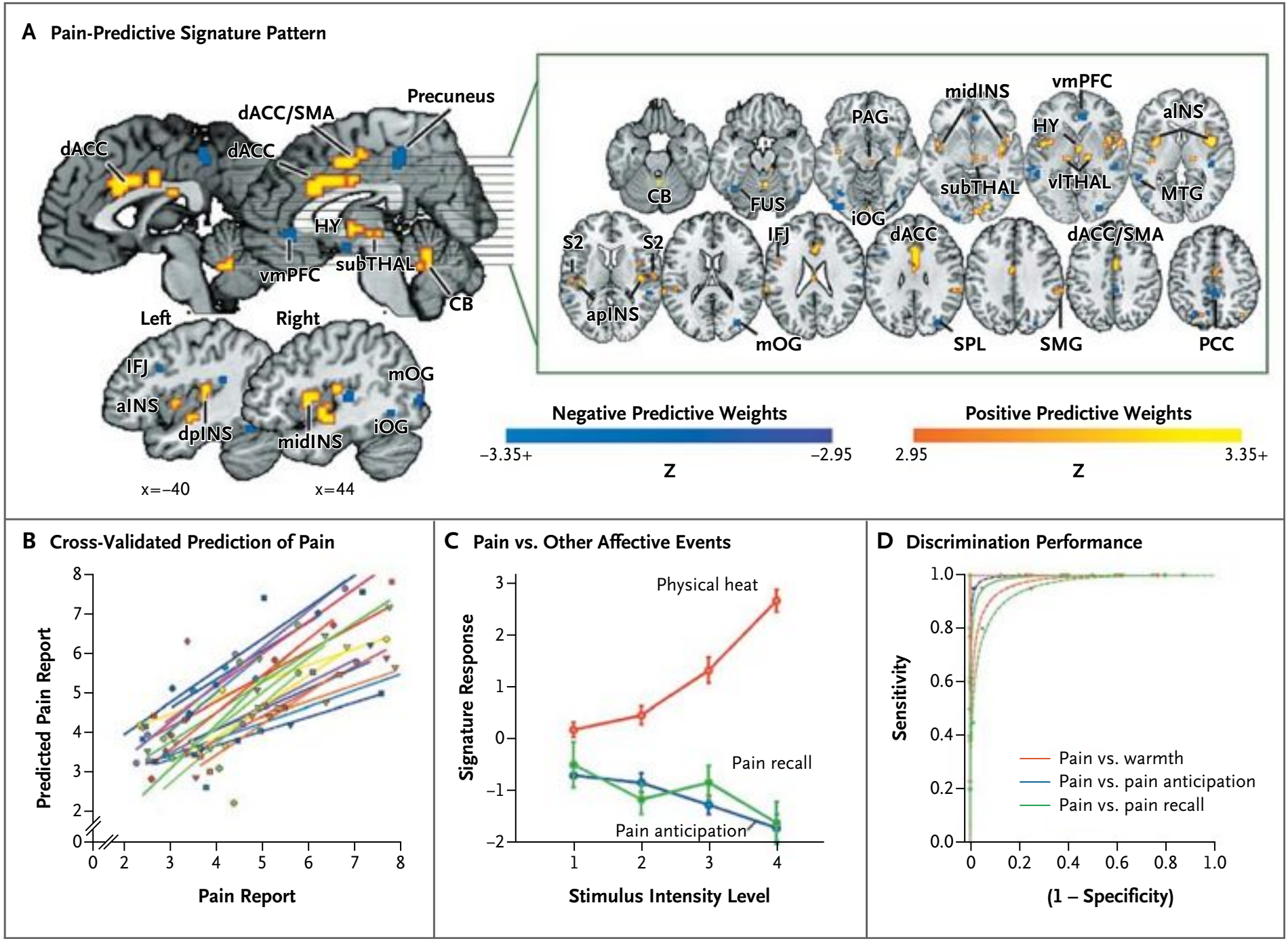
Exp. 1



Exp. 2



Rissman, Greely, & Wagner (2010)



Wager et al (2013)

Conclusions

- ML approaches contribute to neuroimaging in many ways
- In applied contexts (e.g., predicting diagnoses/treatments), prediction is key
- When understanding is more important than prediction, ML can still help
  - Doesn’t have to be a black box
  - The trick is to construct problem in the right way
- But keep pitfalls in mind, and remember that the goal matters!