Computational Models, Machine Learning, and Understanding the Brain

Tom M. Mitchell
Machine Learning Department
Carnegie Mellon University

June 2008

What Will a Theory of the Brain Look Like?

- What is its defining question?
  - "what are the information processing steps that produce perception, decision making, control and learning behaviors, and how are they implemented by the underlying biology?"

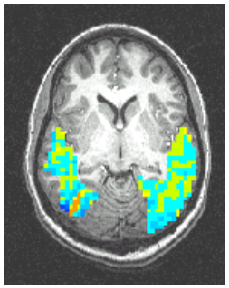
- What will its answer look like?
  - Multi-level from cell biology to circuits to system-wide behavior
  - Complex (**à** online?)
  - Computational (executable)?

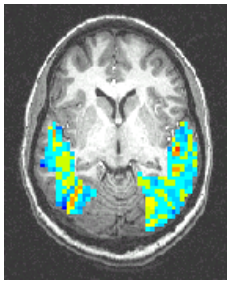
1. Connecting abstract mental states to neural activation

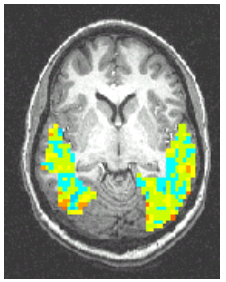
Classifiers of mental state

Training data: pairs of <fMRI images, mental state>

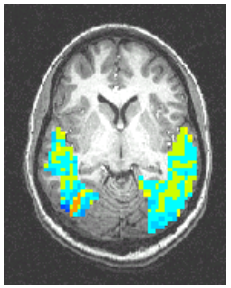

tool


dwelling


dwelling


tool

Resulting classifier:



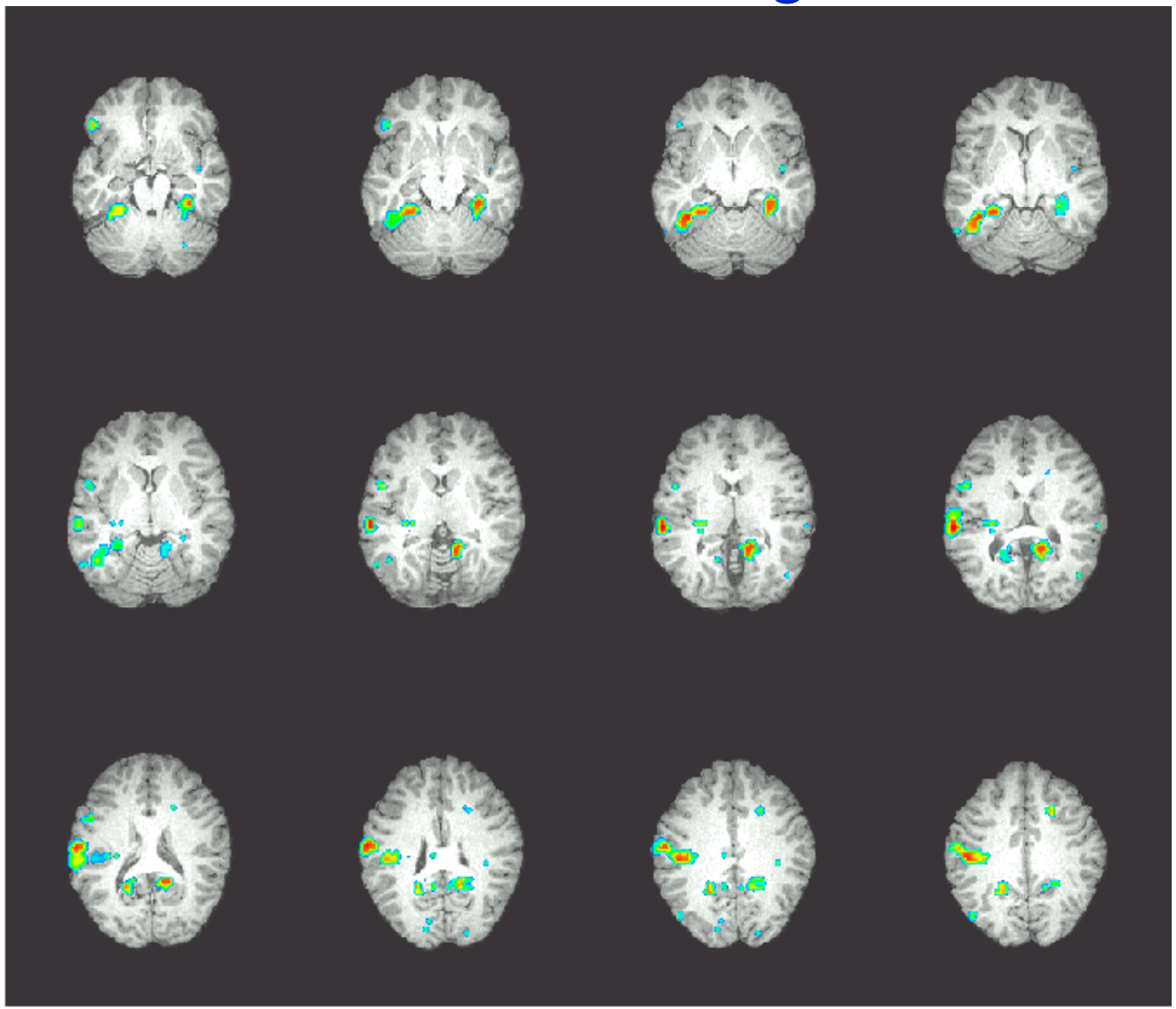
Classifier (e.g., SVM, Bayes, …)

"dwelling"

Where Does the Brain Encode Difference between “tools” and “dwellings”?

“spotlight” classifier uses only the voxel and its immediate neighbors

Accuracy at each significant spotlight [0.7-0.8]



Classifiers: virtual sensors of mental state

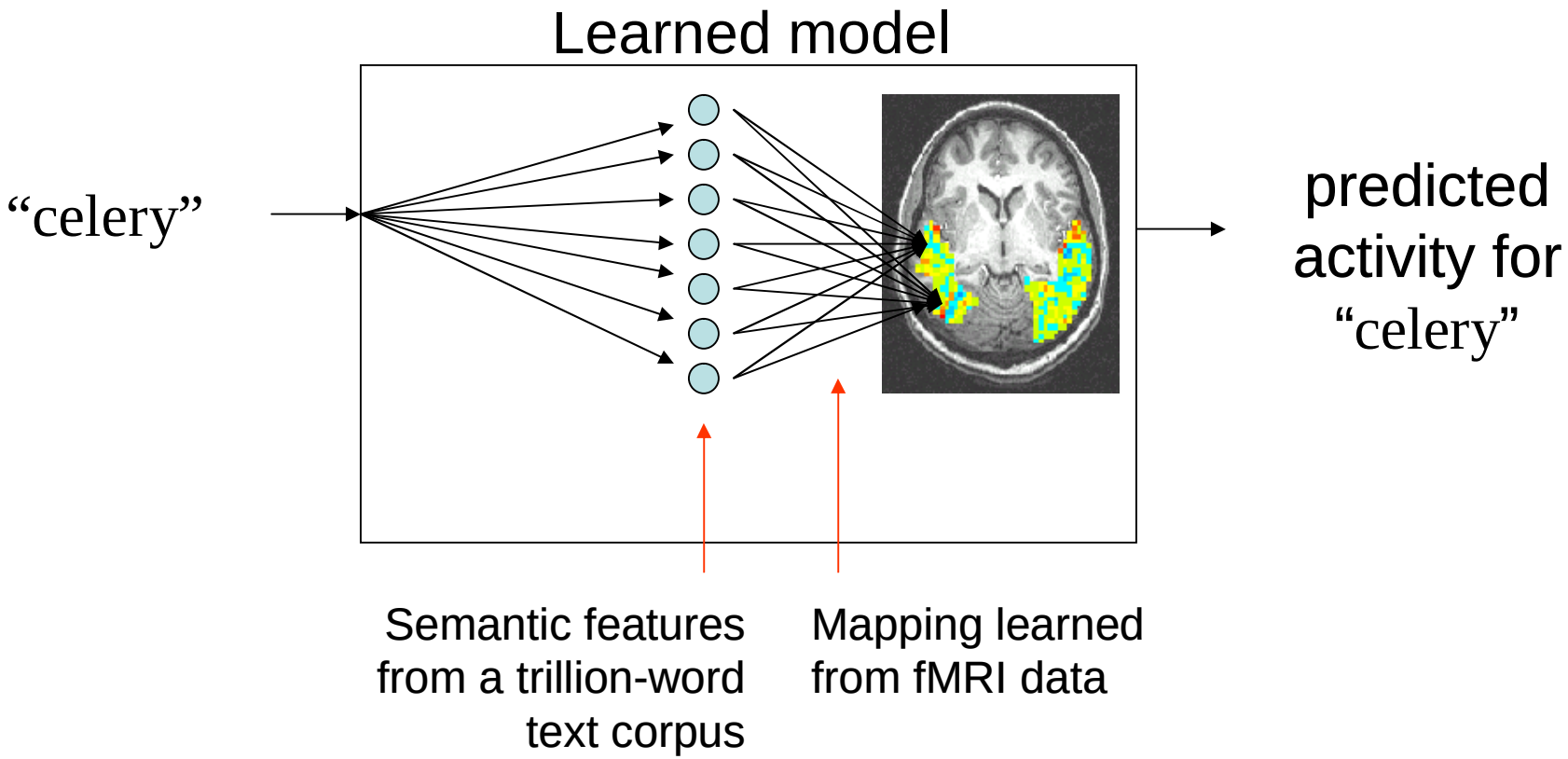
Examples:

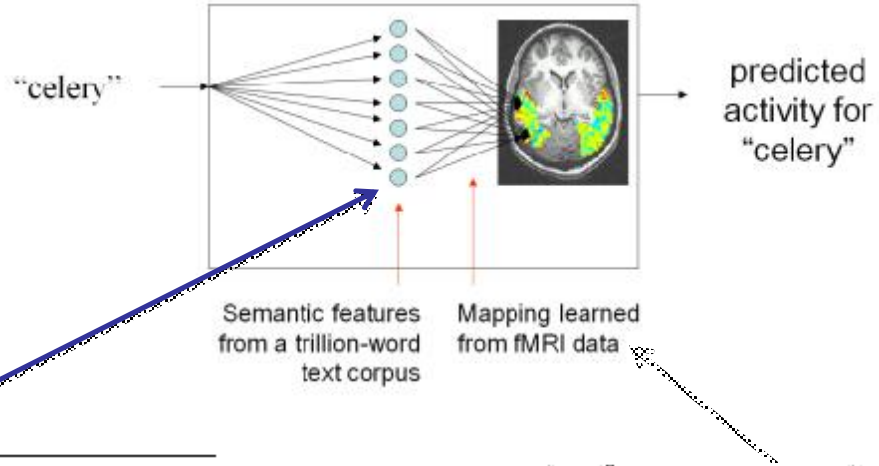
- Distinguish whether person is reading “house” or “hammer”
  - establish that neural encodings are shared across people, by training on one person, decoding another [Shinkareva et al., 2008]
- Distinguish which line orientation person is viewing (“\” vs. “/”)
  - succeeds despite orientation column spacing (~0.5mm) in visual cortex being below fMRI resolution [Kamitani & Tong, 2005]
- Distinguish when and which button a person will press
  - 7 seconds before they are consciously aware [Soon et al., 2008]
  - information encoded in frontal, parietal lobes

2. Machine Learning to help discover predictive theories/models

Predicting Cortical Encodings of Noun Meanings

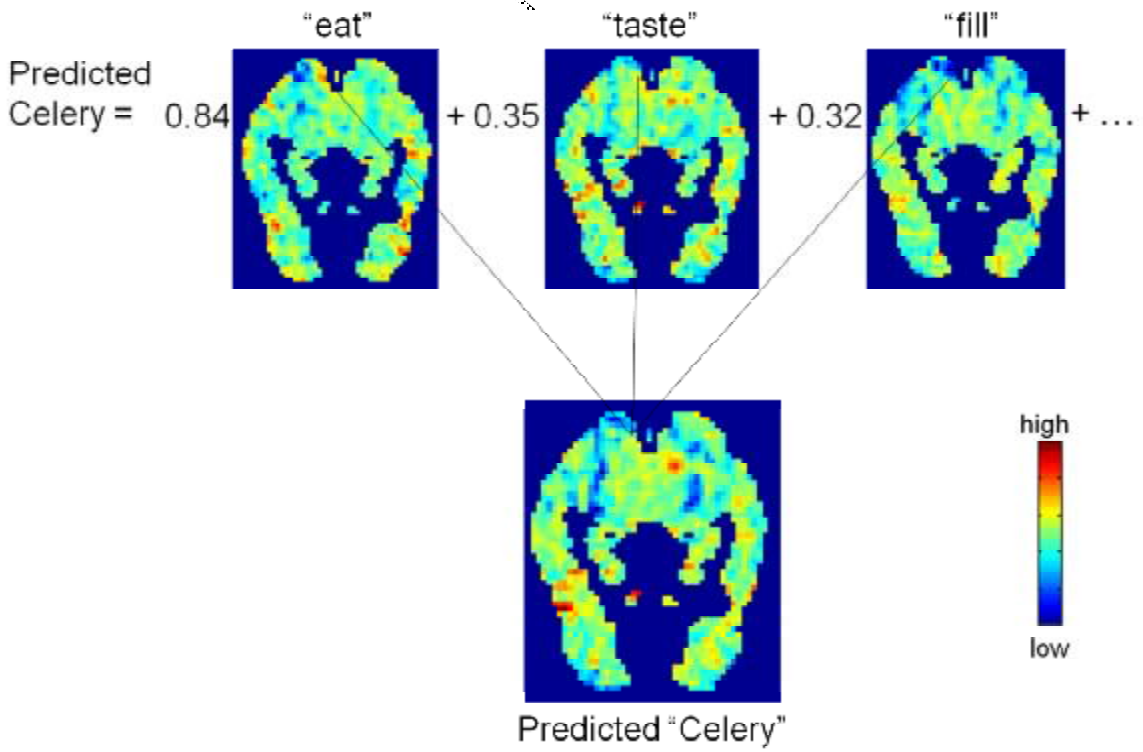
[Mitchell et al., 2008]

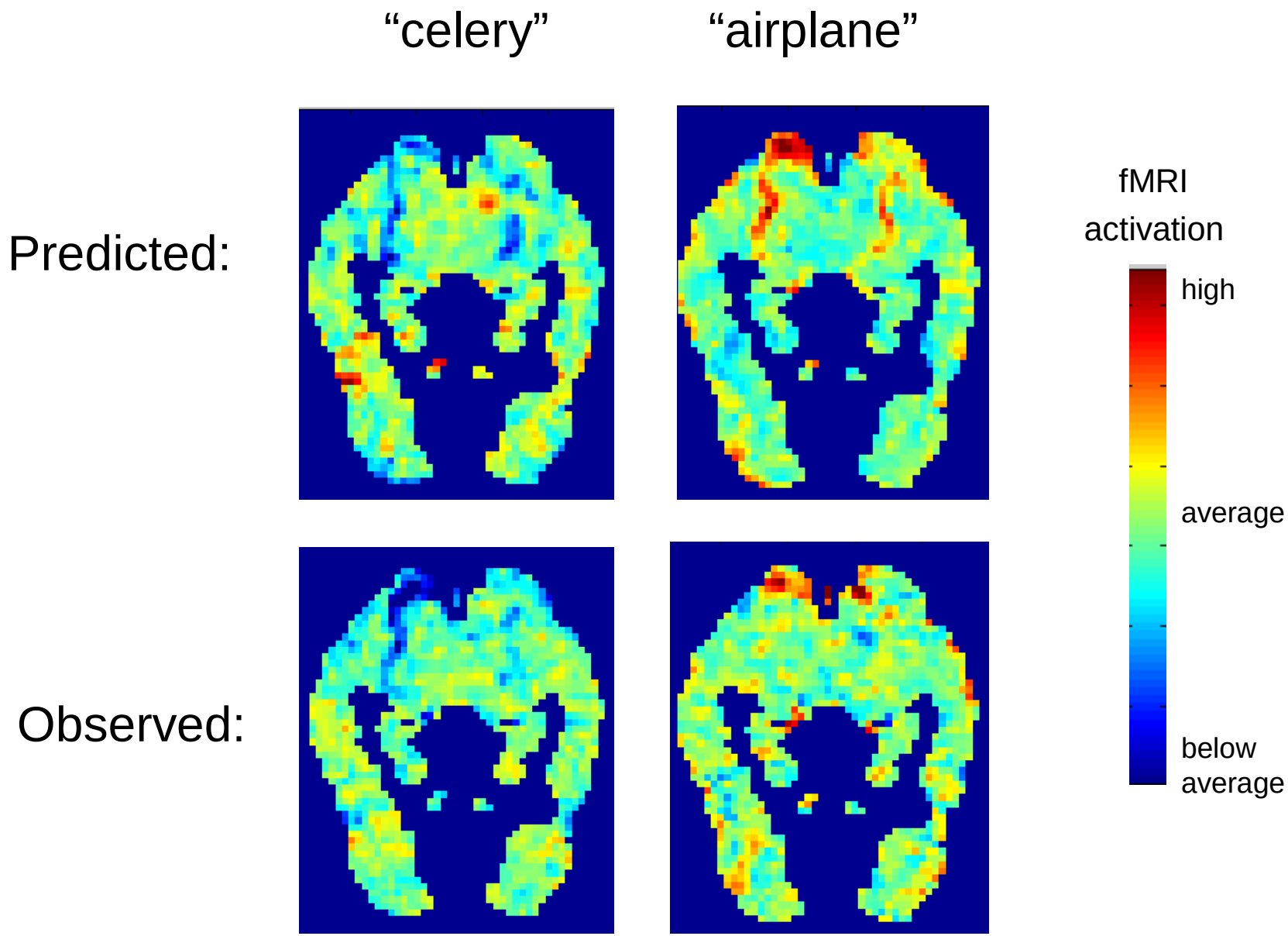




Semantic feature values: **"celery"**

0.8368, eat
0.3461, taste
0.3153, fill
0.2430, see
0.1145, clean
0.0600, open
0.0586, smell
0.0286, touch
…
…
0.0000, drive
0.0000, wear
0.0000, lift
0.0000, break
0.0000, ride

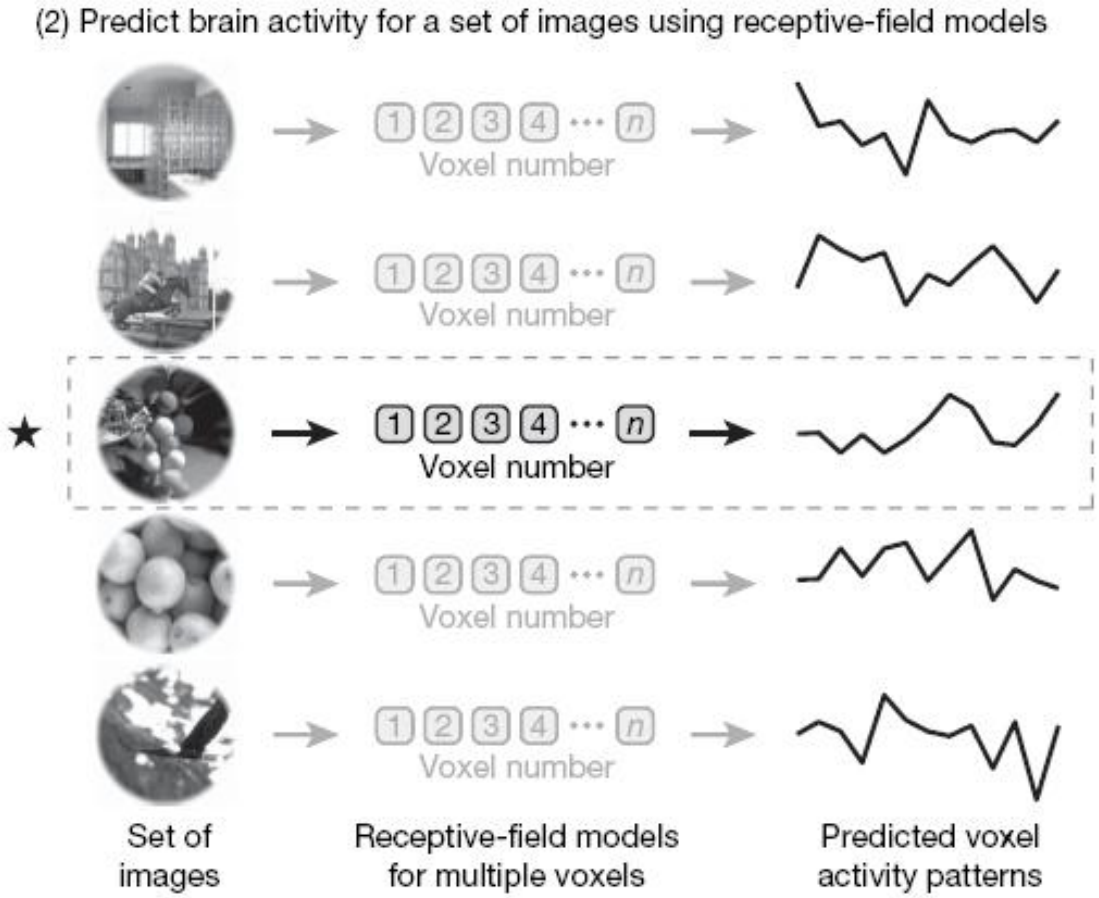




Predicted and observed fMRI images for “celery” and “airplane” after training on 58 other words.

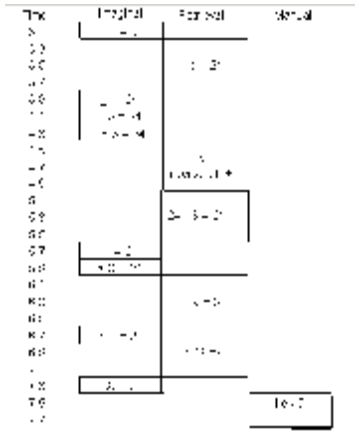
Predicting Visual Cortex Activity from Stimulus Photograph

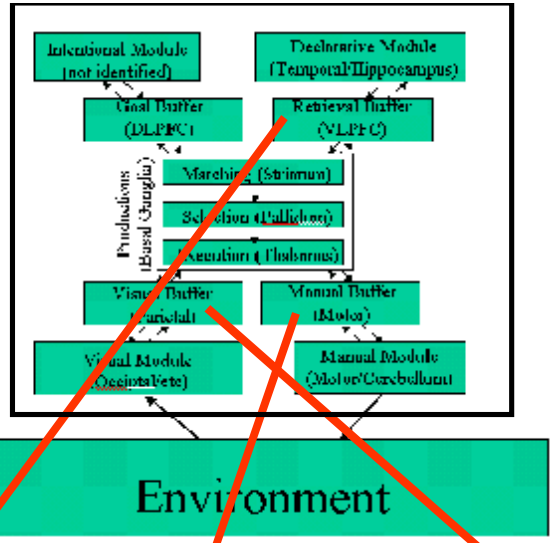
[Kay et al., 2008]



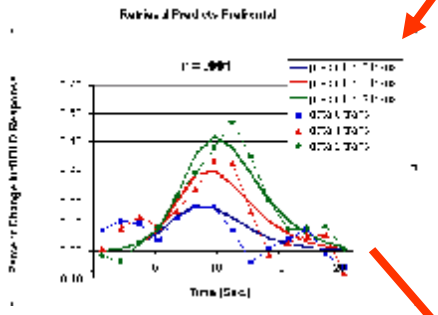
Challenge Problem

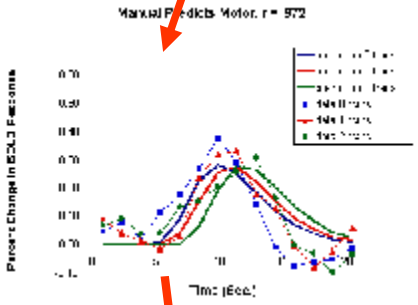
Develop Computational Models of Perception, Control, Decision Making, and Learning Grounded in Observed Brain Activity

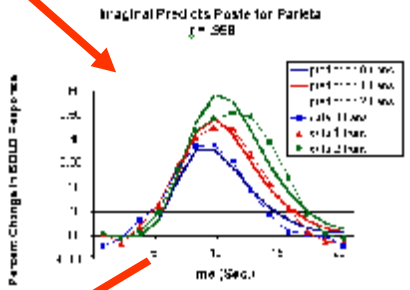


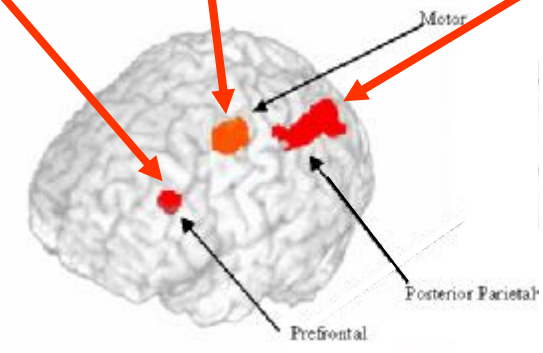


[Anderson, Qin, & Sohn, 2002]









What Will a Theory of the Brain Look Like?

- What is its defining question?
  - "what are the information processing steps that produce perception, decision making, control and learning behaviors, and how are they implemented by the underlying biology?"

- What will its answer look like?
  - Multi-level from cell biology to circuits to system-wide behavior
  - Complex (**à** online?)
  - Computational (executable)?