

# **NEURAL COMPUTATION**

***OR***

# **NEUROSCIENCE FOR COMPUTER SCIENTISTS**

Prof. Alexander Huth  
1/21/2020

# TODAY

- \* Overview of class
- \* Syllabus
- \* A bit about what I do

# BIG QUESTION

- \* Given a system that computes something,
  - \* what does it compute?
  - \* how does it work?
  - \* what are the *internal representations* that support that computation?

# WHAT DOES IT MEAN TO UNDERSTAND?

- \* David Marr (1945 - 1980), *Vision* (1982)
- \* **Levels of analysis:**
  - \* Computational
  - \* Algorithmic/Representational
  - \* Implementation

# COMPUTING SYSTEMS

- \* Biological: brains
- \* Artificial: neural networks

# TOPICS

- \* Biological neural networks (brains)
  - \* What are neurons? How do they compute?
  - \* How are neurons organized into networks?
  - \* How do biological neural networks learn?

# TOPICS

- \* (Computational) Neuroscience
  - \* What methods are used to study biological neural systems?
  - \* How can we design neuroscience experiments?
  - \* How can we deal with noise and uncertainty in data?
  - \* Encoding & decoding models

# TOPICS

- \* System identification (encoding models)
  - \* Given an unknown system  $f(x) \rightarrow y$  and measurements of  $(x, y)$  pairs, can we create a model of  $f$ ?

# TOPICS

- \* Simple artificial neural networks
  - \* Unsupervised learning (Hebb, Oja, Olshausen, etc.)
  - \* Supervised learning (Hopfield, perceptron, etc.)

# TOPICS

- \* Understanding artificial neural networks
  - \* Attribution (gradient methods, input ablation)
  - \* Adversarial methods

# SYLLABUS

- \* Online:  
<https://github.com/alexhuth/neuralcomputation-sp2020>
- \* ~4 problem sets
- \* Final project (alone or in pairs)
  - \* Proposal due Mar. 31
  - \* In-class presentations Apr. 28/30 & May 5/7
  - \* Write-up due Apr. 28

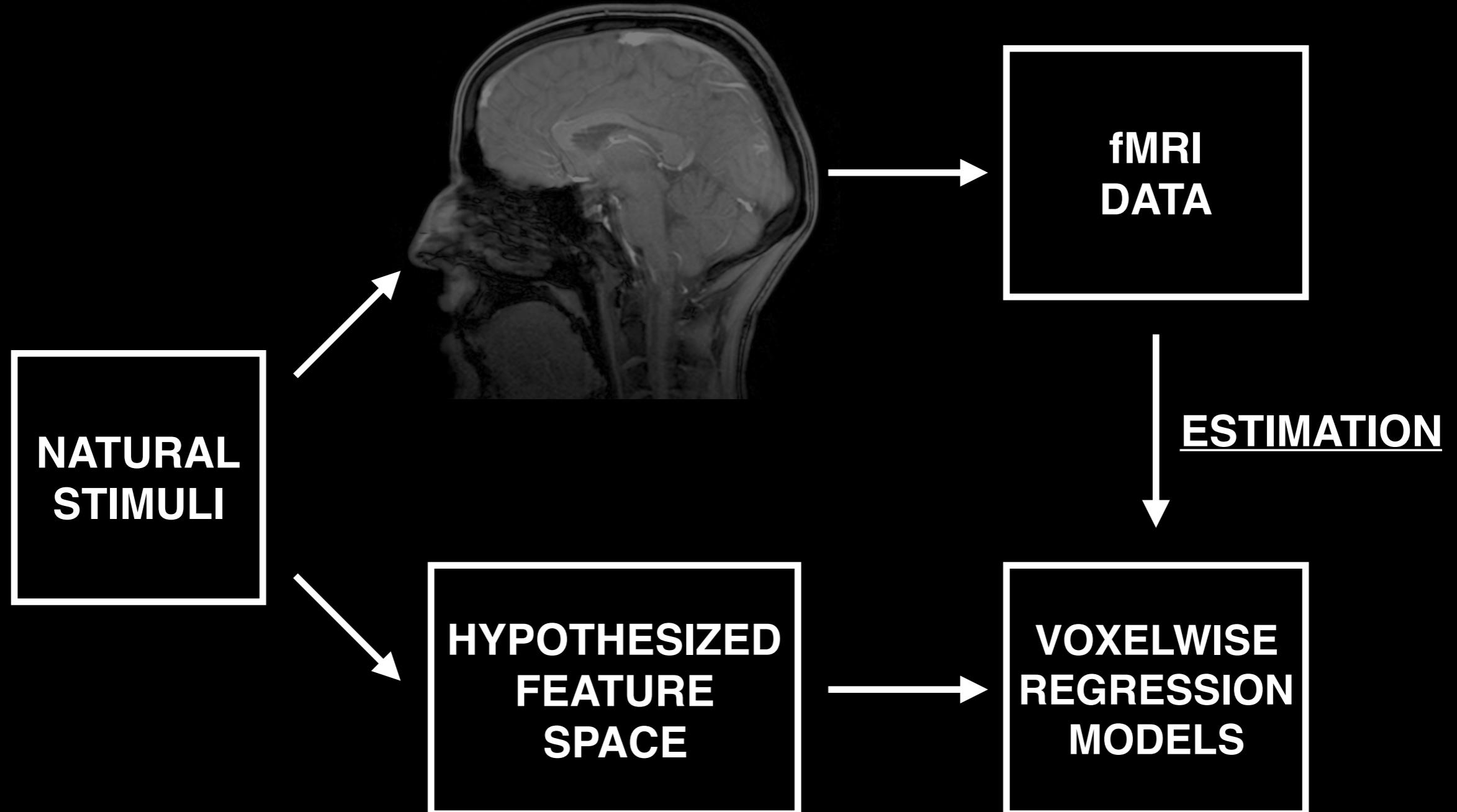
# **MY WORK**

# NATURAL LANGUAGE EXPERIMENT

- Subjects listen to over 2 hours of naturally spoken, narrative stories
- BOLD responses recorded from whole brain using GE-EPI fMRI
  - Voxel size:  $2.24 \times 2.24 \times 4.1\text{mm}$
  - Repetition time (TR): 2.0045 s
  - Custom water-excitation RF pulse

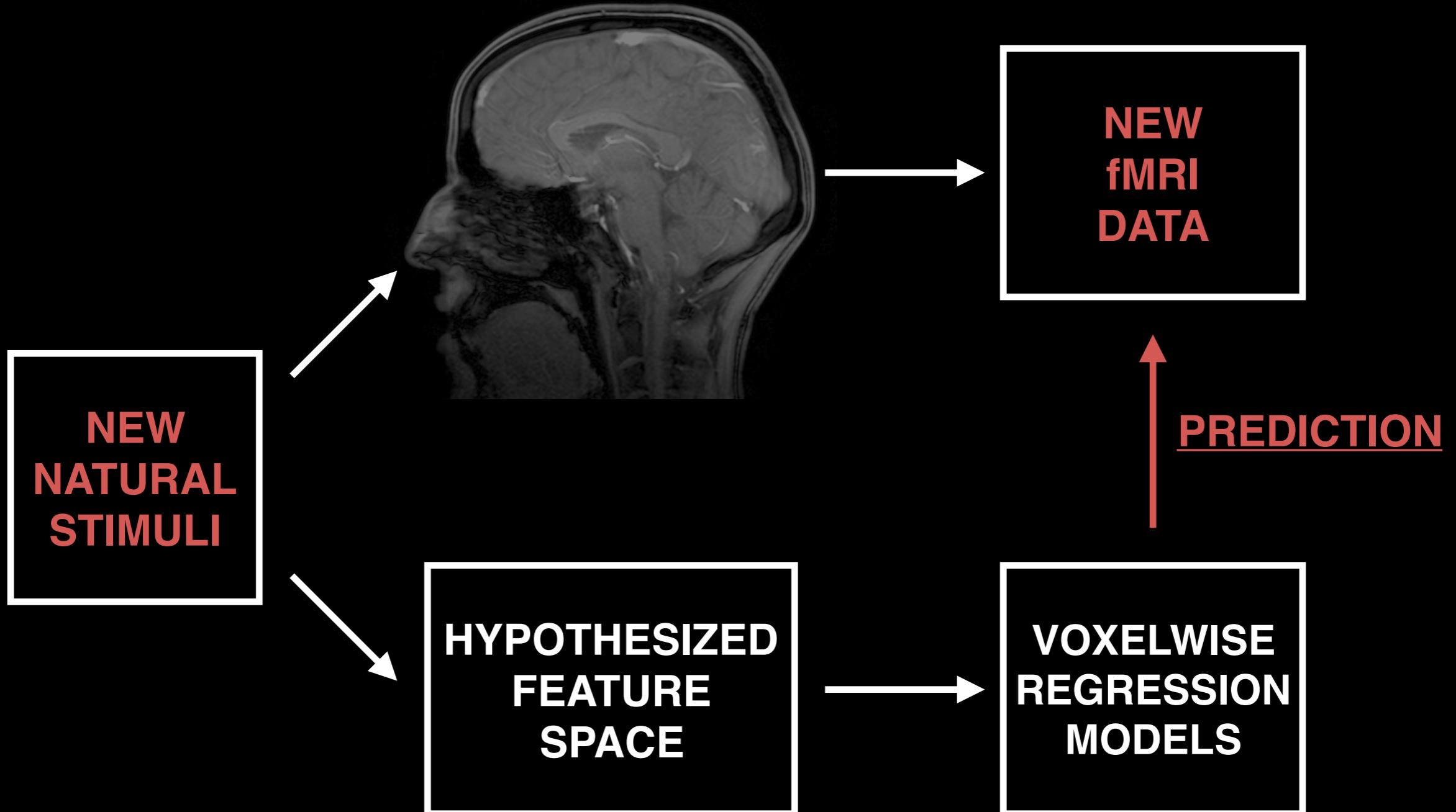


# VOXELWISE MODELING

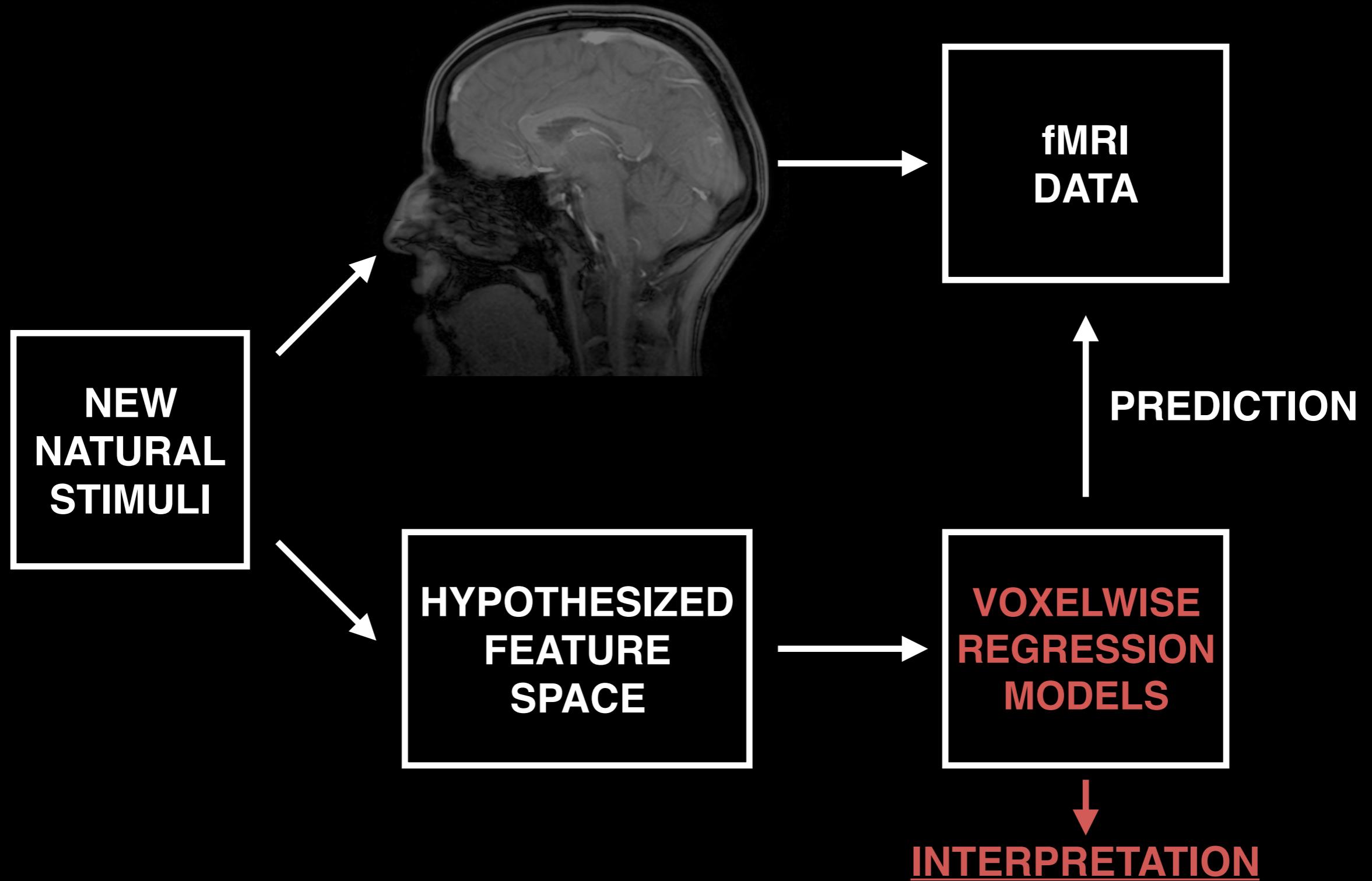


Kay et al. *Nature* (2008), Naselaris et al. *Neuron* (2009),  
Nishimoto et al. *Current Biology* (2011), Huth et al. *Neuron* (2012), etc.

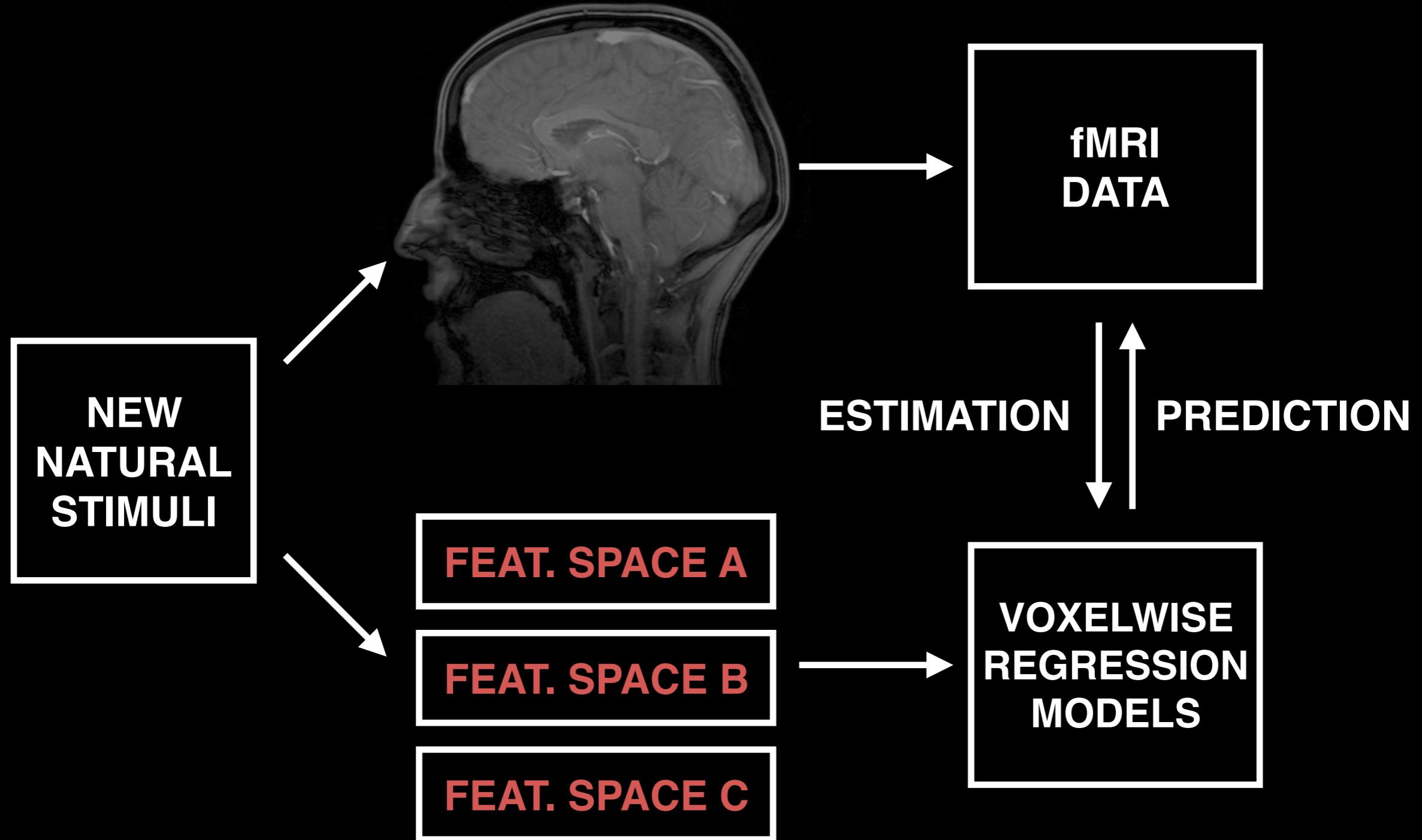
# VOXELWISE MODELING



# VOXELWISE MODELING



# VOXELWISE MODELING



# fMRI DATA VISUALIZATION



# VOXELWISE MODELING

## SIMPLEST MODEL:

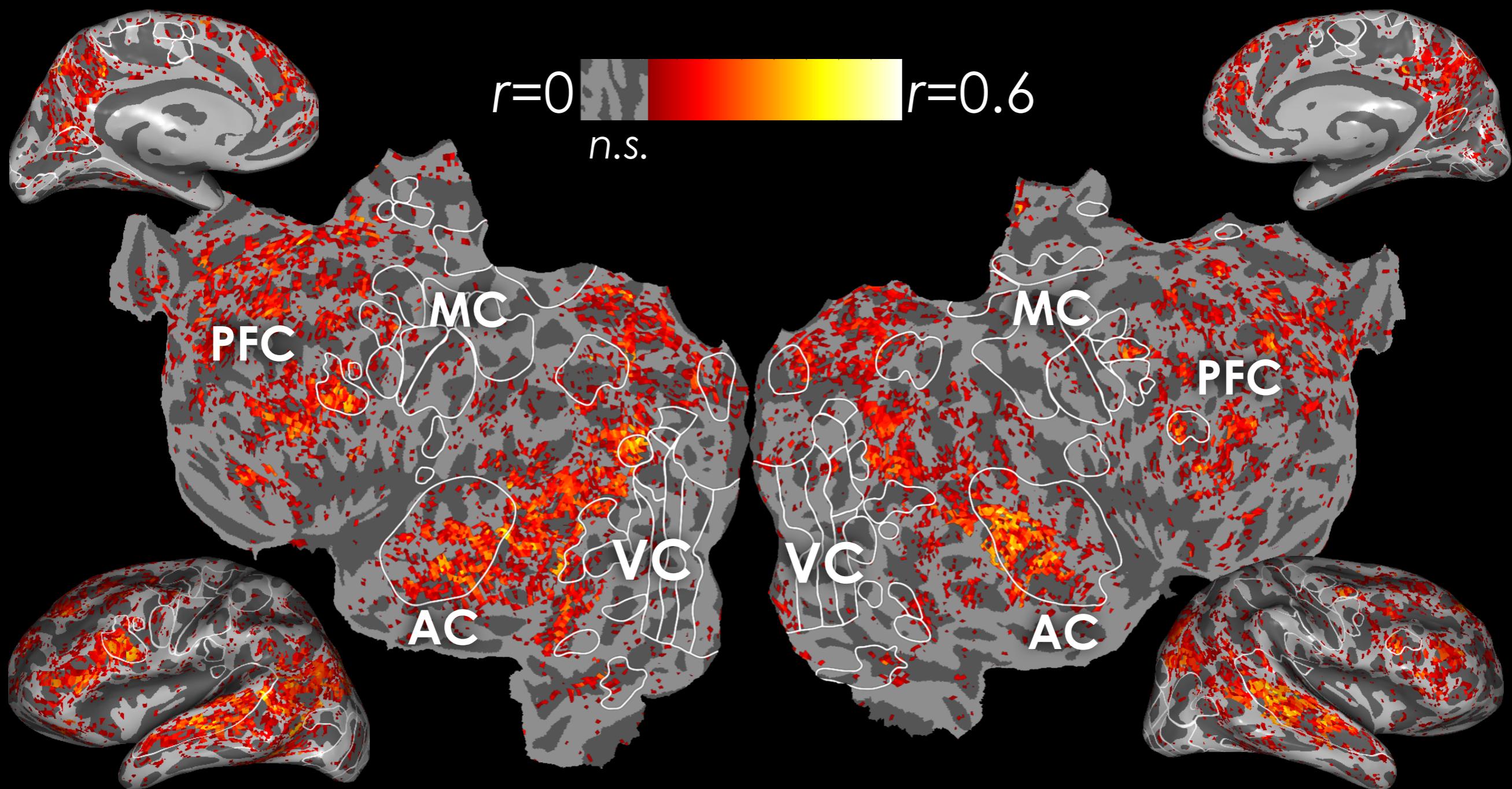
each voxel responds (some amount) to each word

$$R(t) = \sum_{i=0}^N \beta_i W_i(t)$$

$$\hat{\beta} = \operatorname{argmax}_{\beta} P(R|\beta, W)$$

likelihood

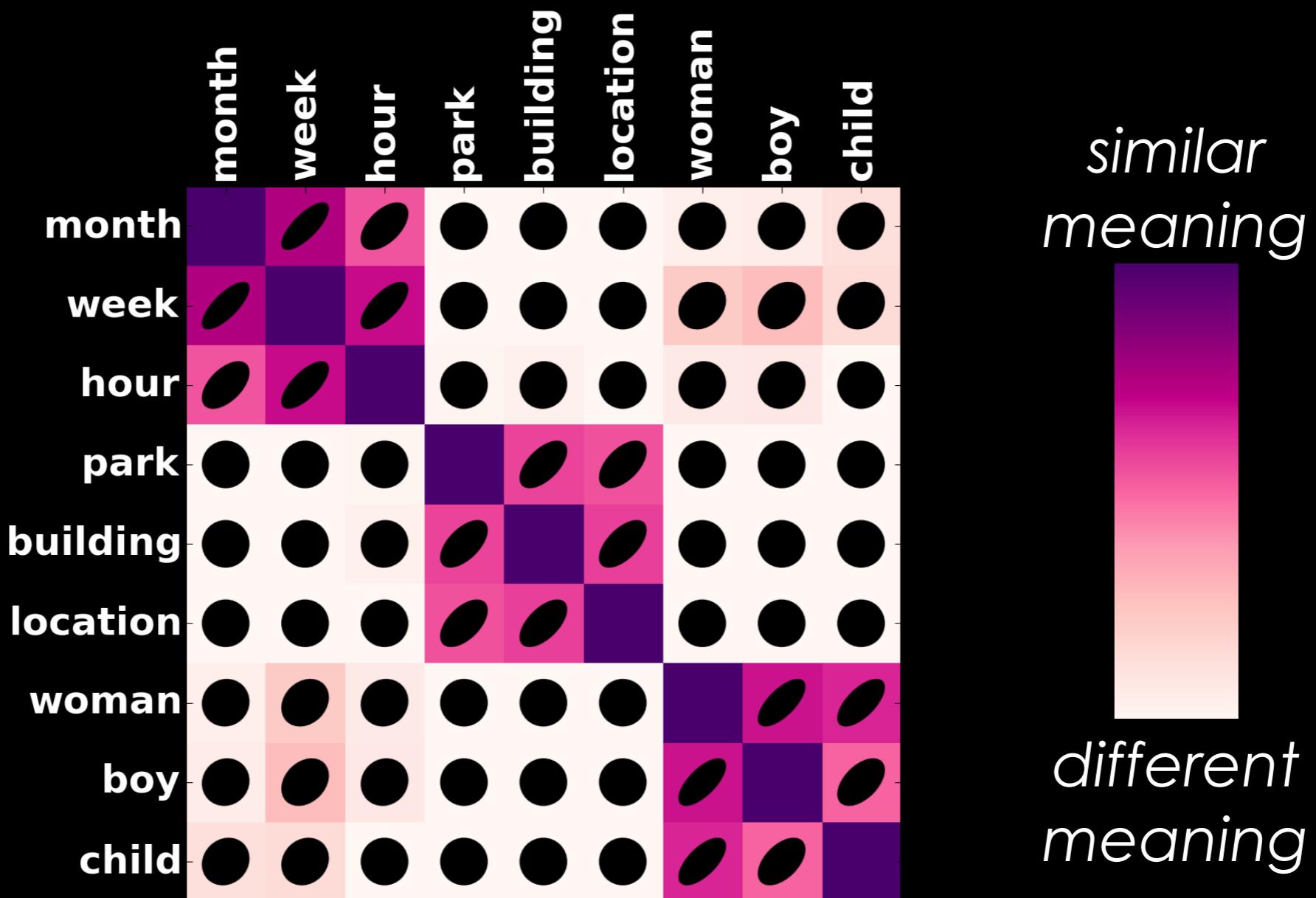
# WORD MODEL PERFORMANCE: MEDIOCRE



# SEMANTIC PRIOR

## IMPROVED MODEL:

similar responses to words with similar meanings



# SEMANTIC PRIOR

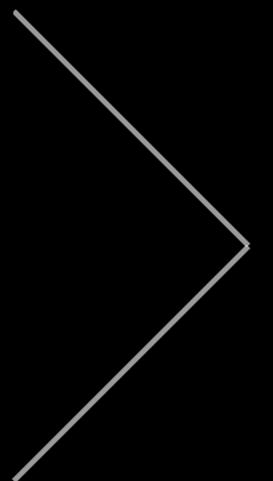
Distributional hypothesis:

“You shall know a word by  
the company it keeps”

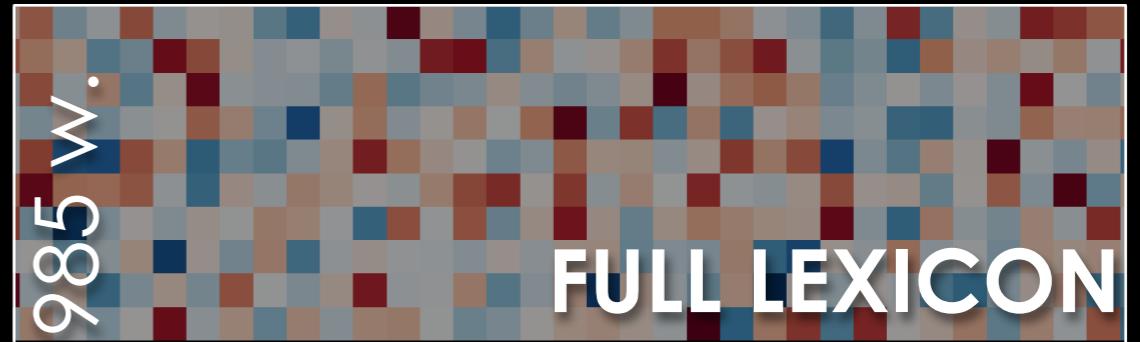
J. R. Firth (1954)

# SEMANTIC PRIOR

...  
difficult  
...  
husband  
...  
potato  
...  
remember  
...



TARGET  
WORDS



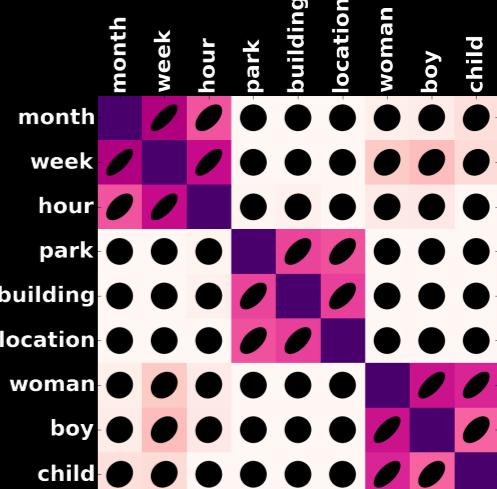
CO-OCCURRENCE  
MATRIX

10,470 words

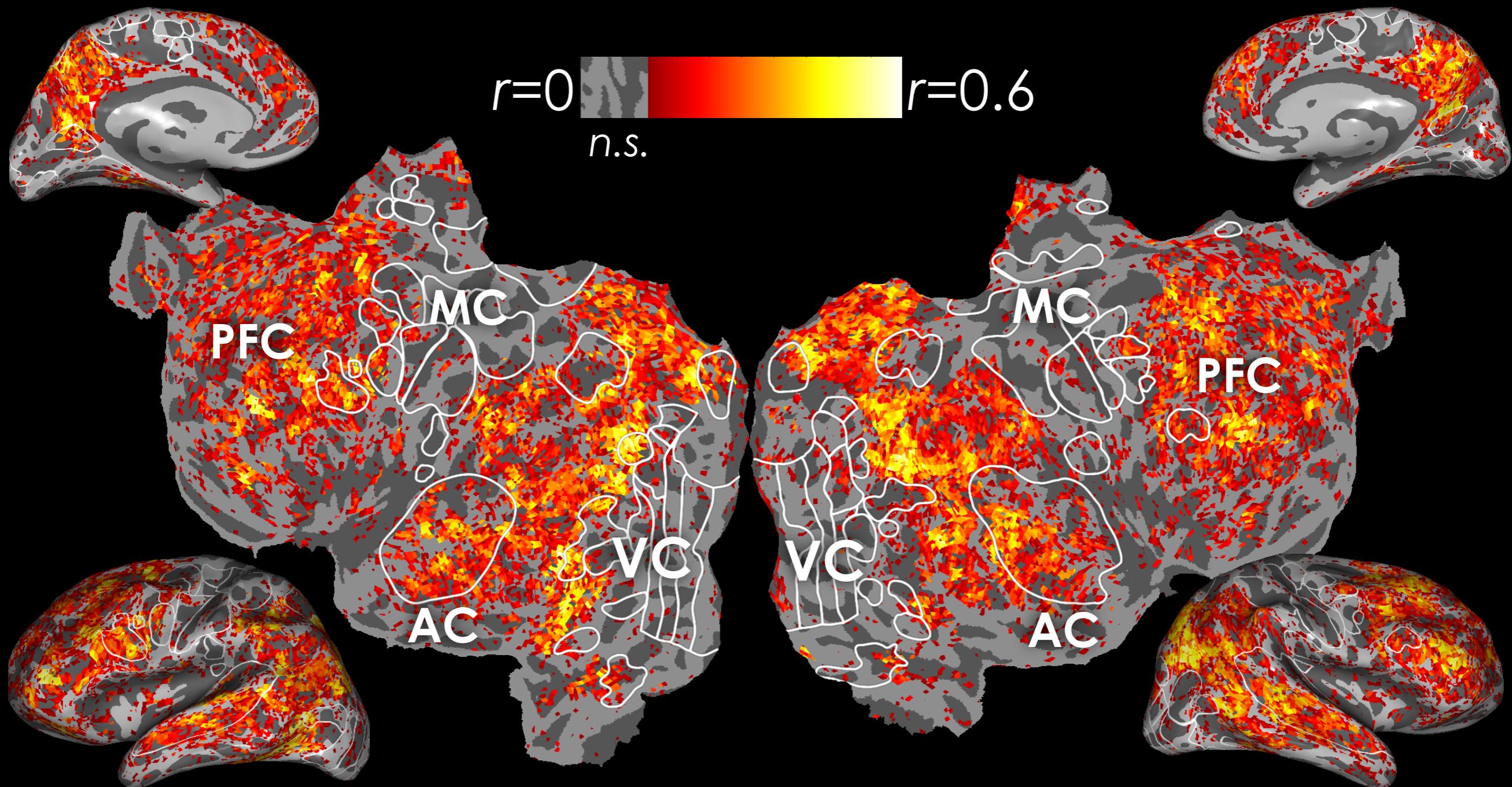
# SEMANTIC PRIOR

# IMPROVED MODEL:

# similar responses to words with similar meanings



# SEMANTIC MODEL PERFORMANCE: EXCELLENT



# MODEL INTERPRETATION

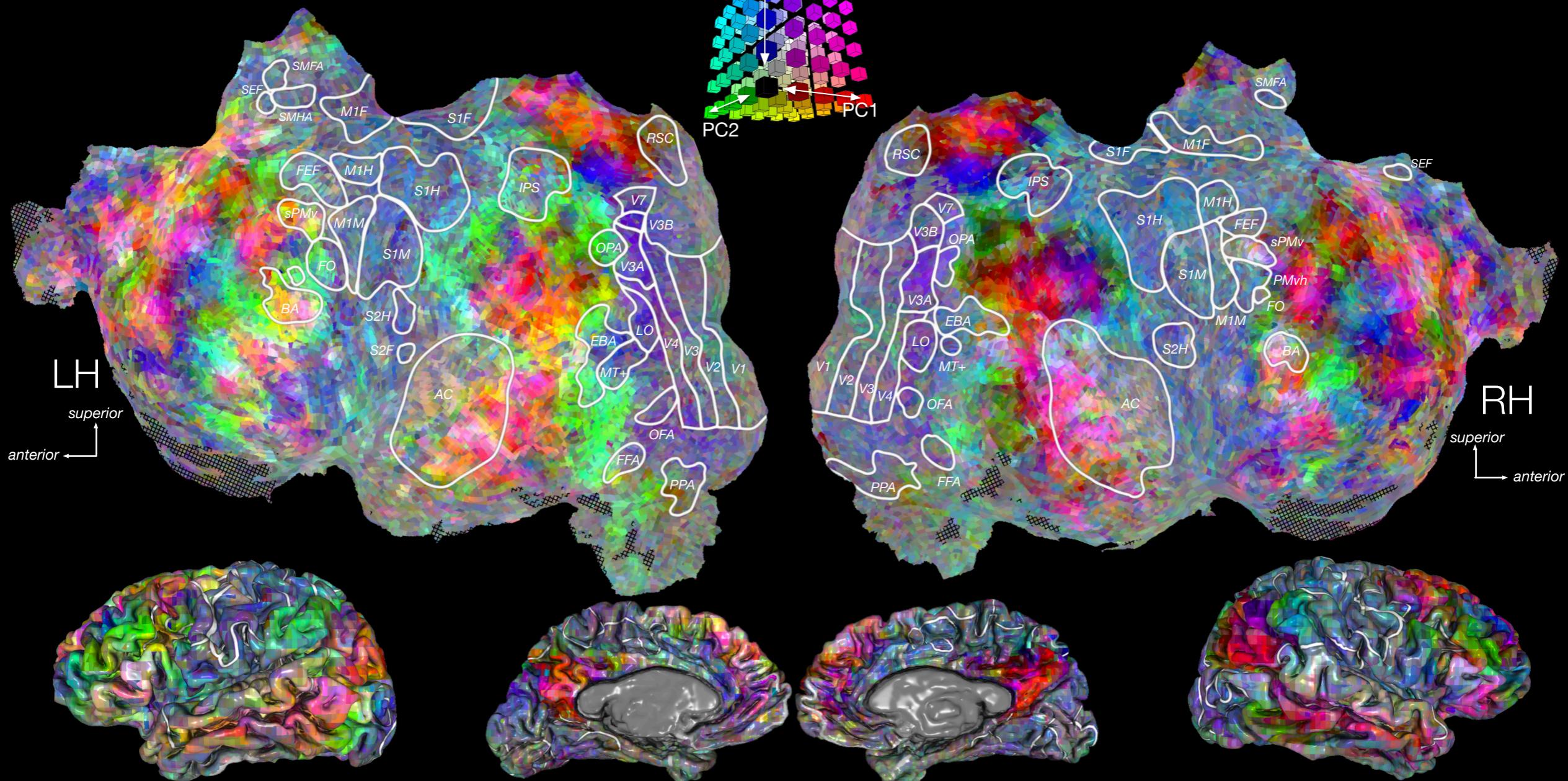
visual  
tactile  
abstract  
numeric

violent  
communal  
emotional  
social

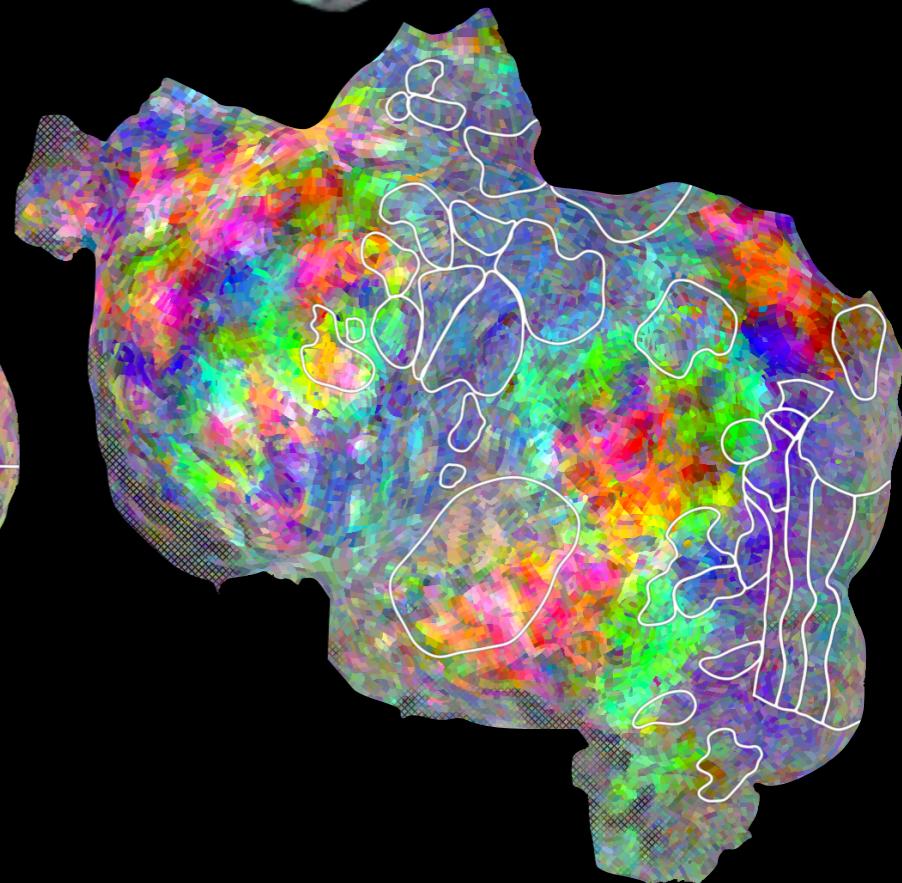
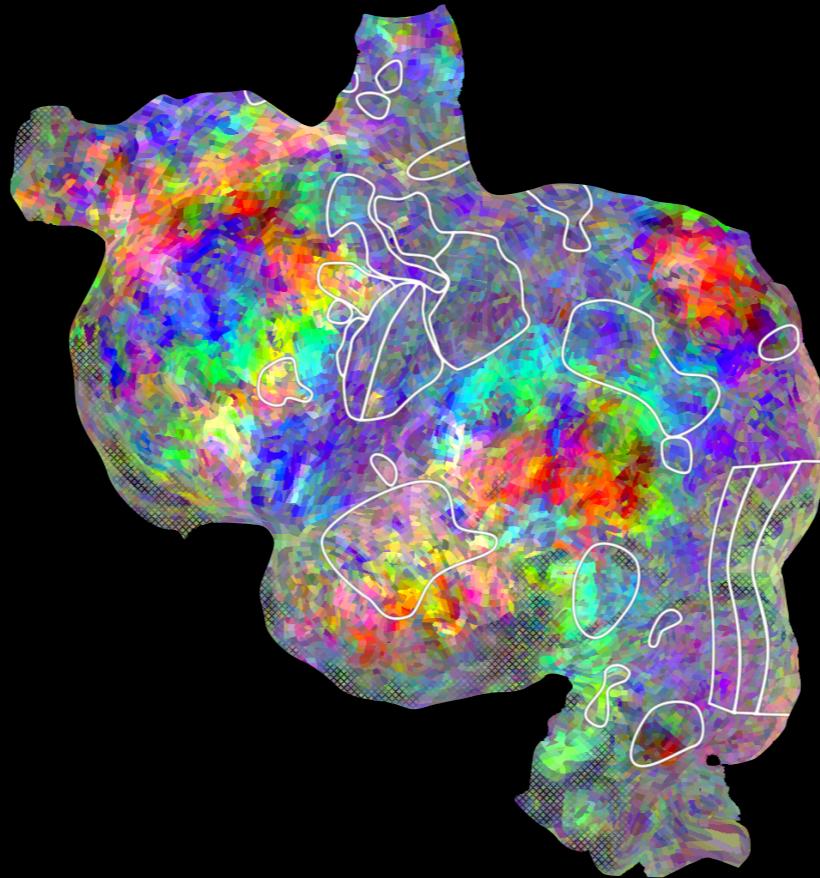
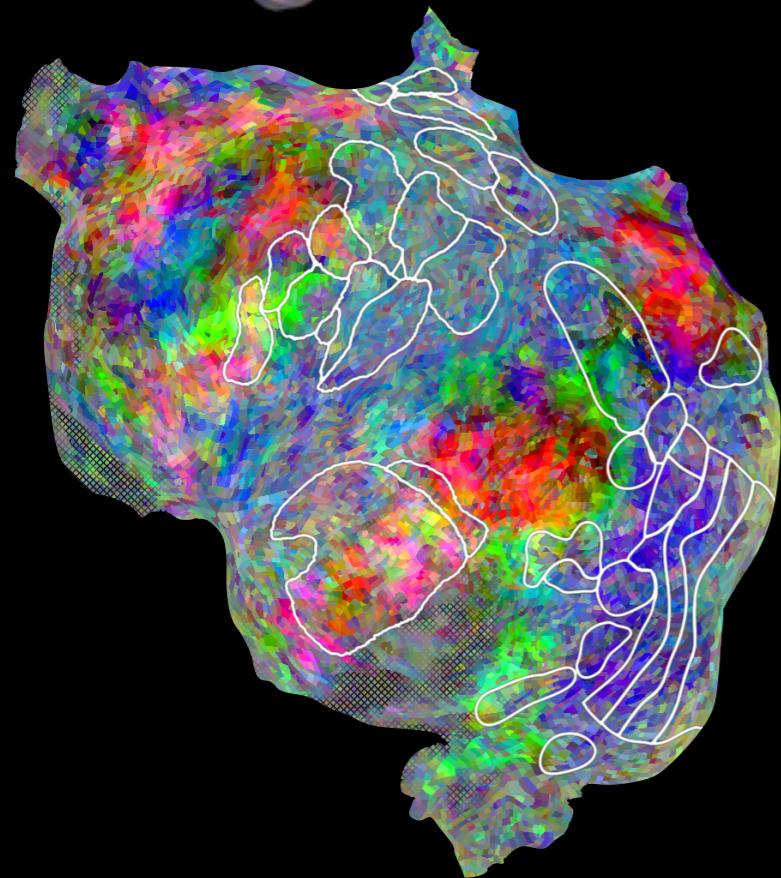
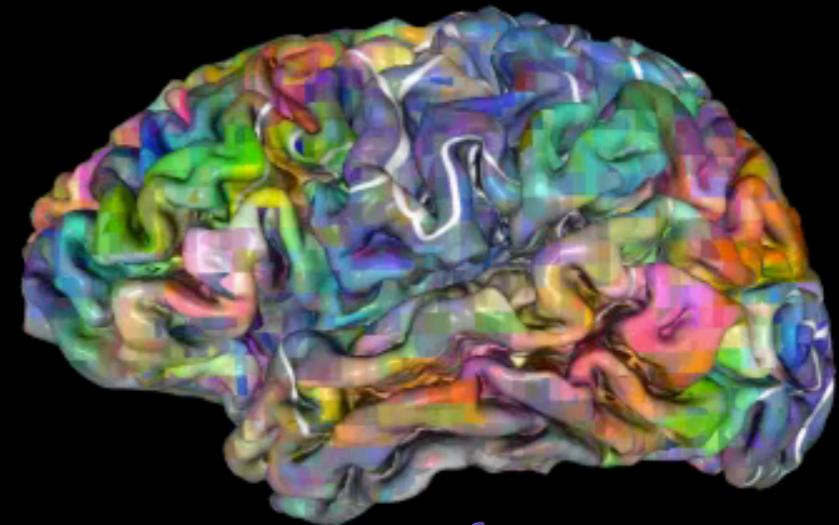
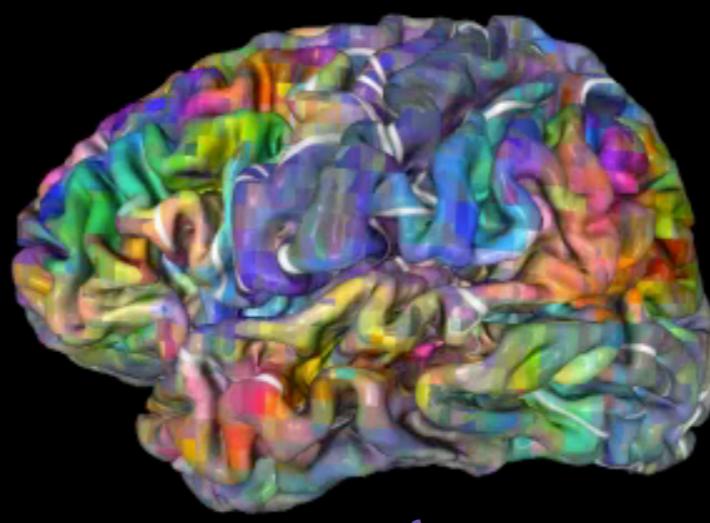
temporal  
professional  
mental

locational

PC2  
PC1



# MAPS ARE CONSISTENT ACROSS SUBJECTS



**THAT'S IT!**