# Validation and Inference

Georg Langs

CIR Lab
Dep. Biomedical Imaging and Image-guided Therapy
Medical University of Vienna

www.cir.meduniwien.ac.at

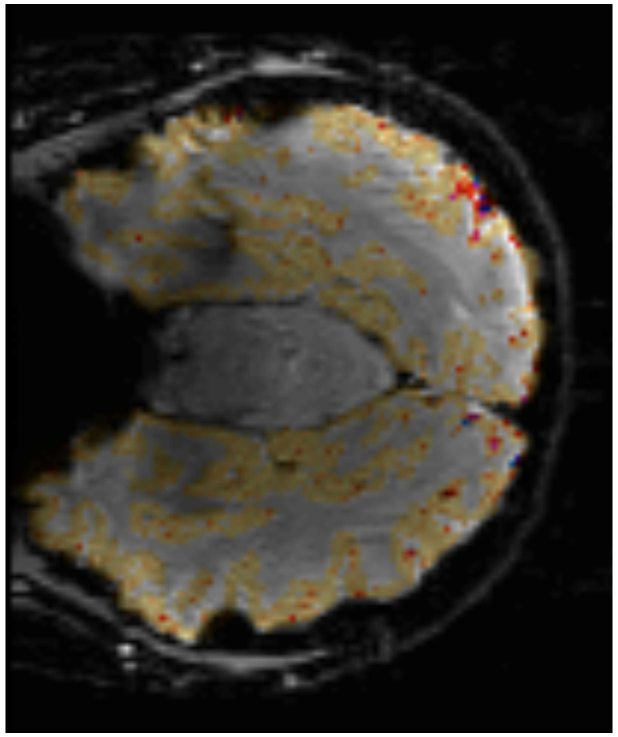
# Outline

***Learning from observations: models and how to know if they capture anything real***

- Validating Relationships:
  - Encoding brain signals from stimuli
  - Decoding stimuli from brain signals
- Validating Structure:
  - Modelling functional architecture
- Significance revisited

# Relationships

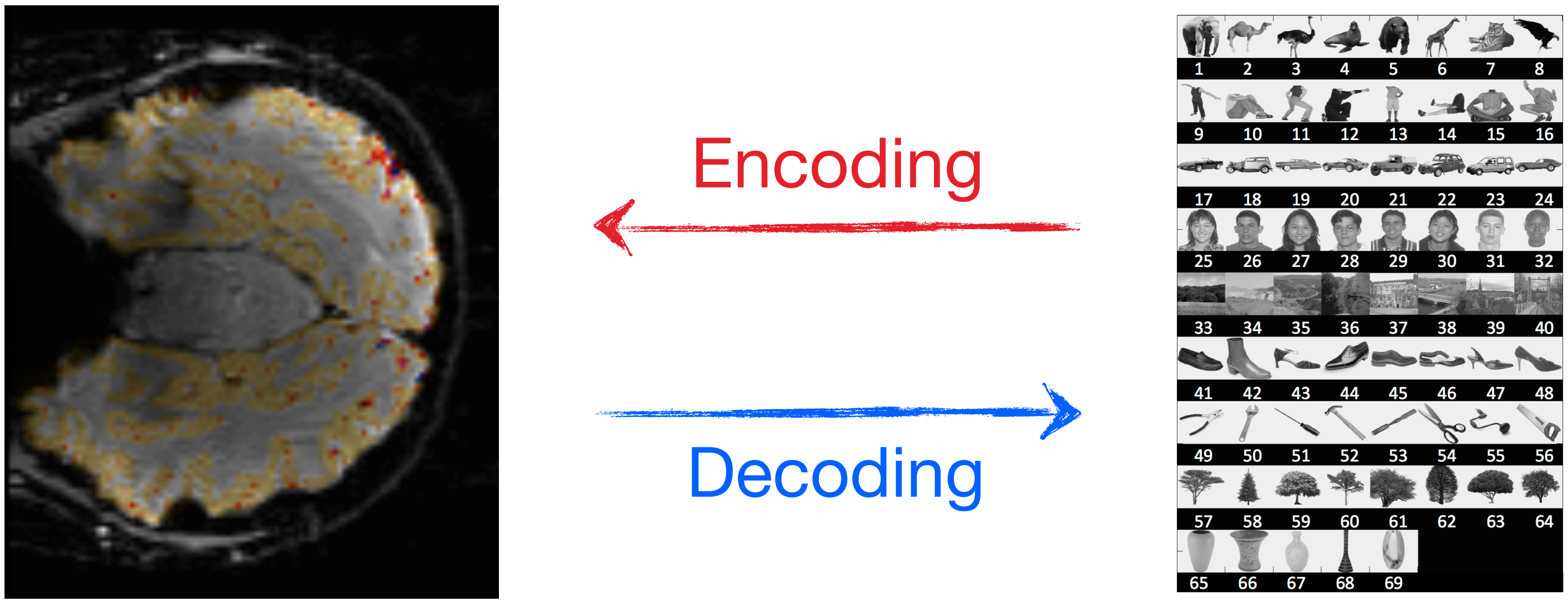
# Stimuli and activity



?

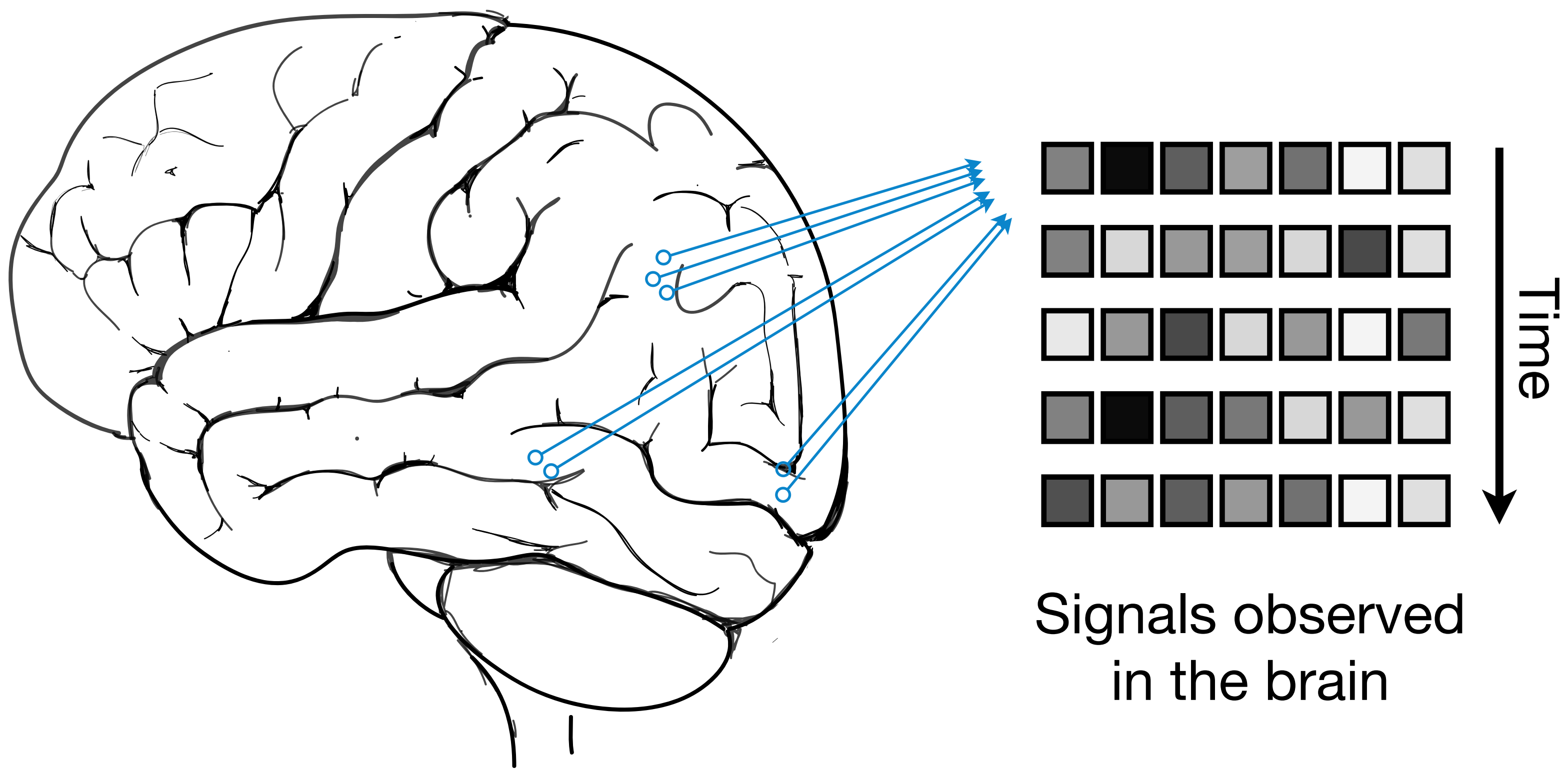

- What is the relationship between the experiment condition and the observed brain activity?
- Does a specific model capture this relationship?

# Stimuli and activity



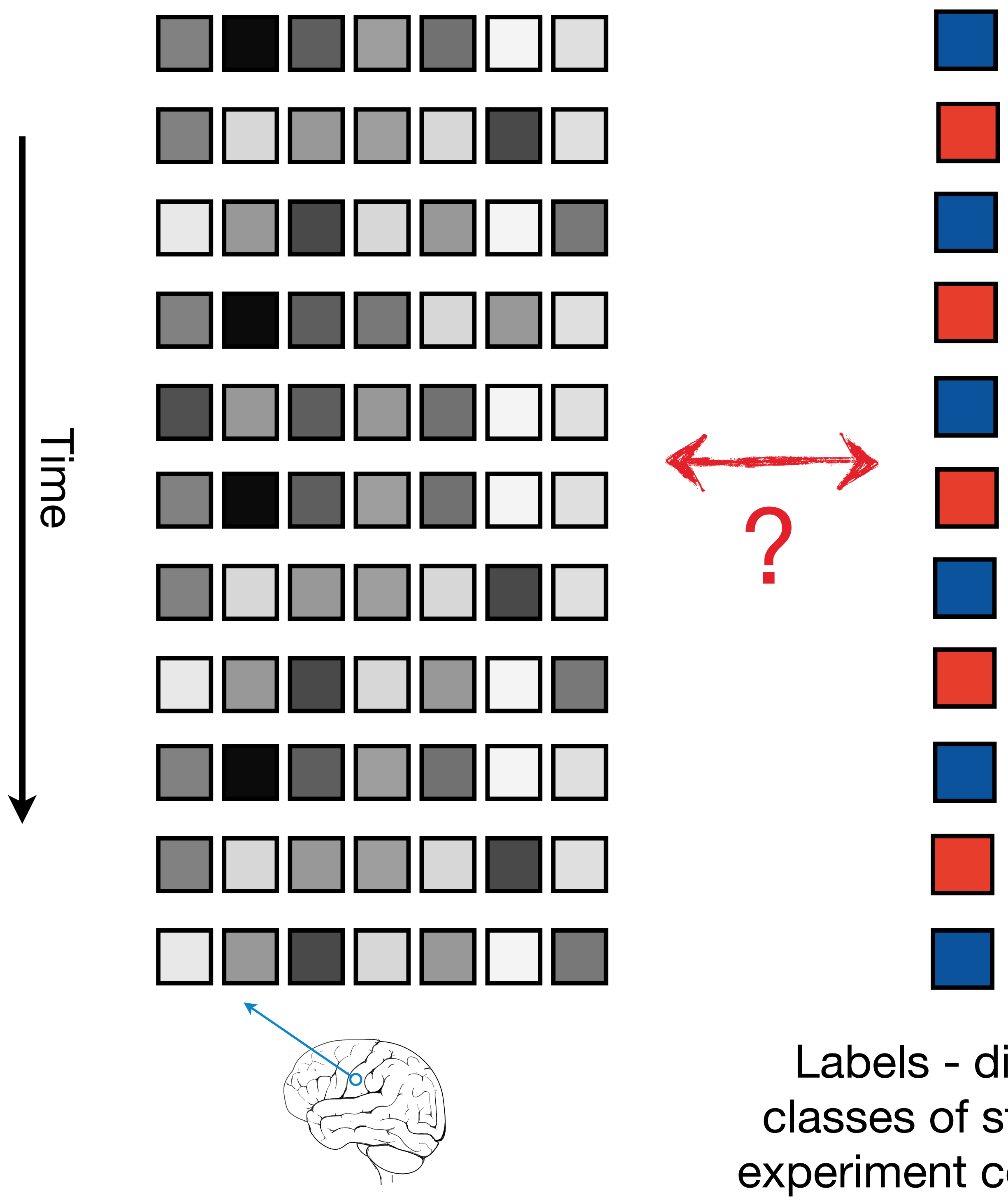

- **Encoding**: Predicting brain activity from experiment condition
- **Decoding**: Predicting experiment condition from brain activity

# Multivariate observations



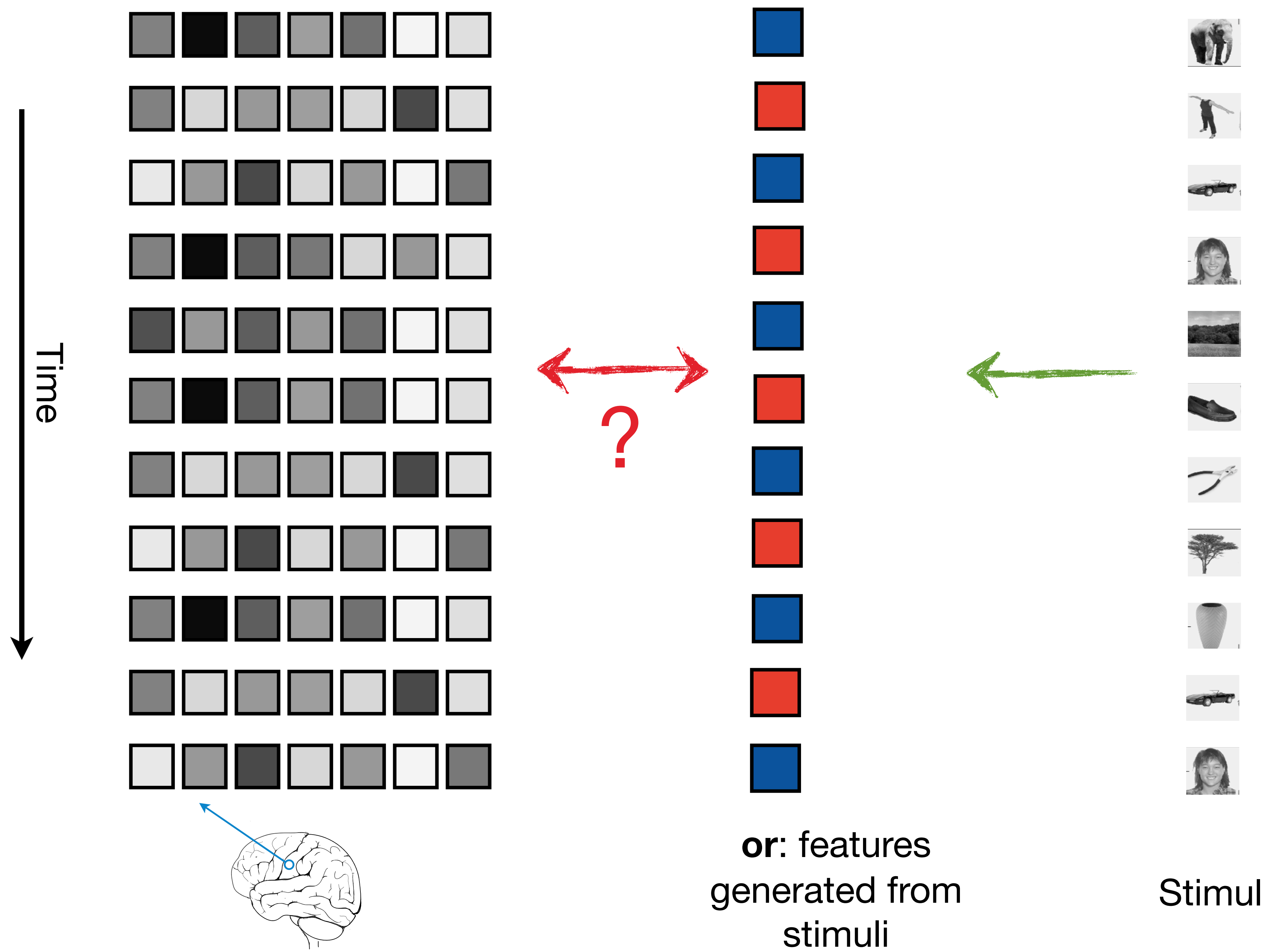

Signals observed in the brain

# Brain and experiment condition

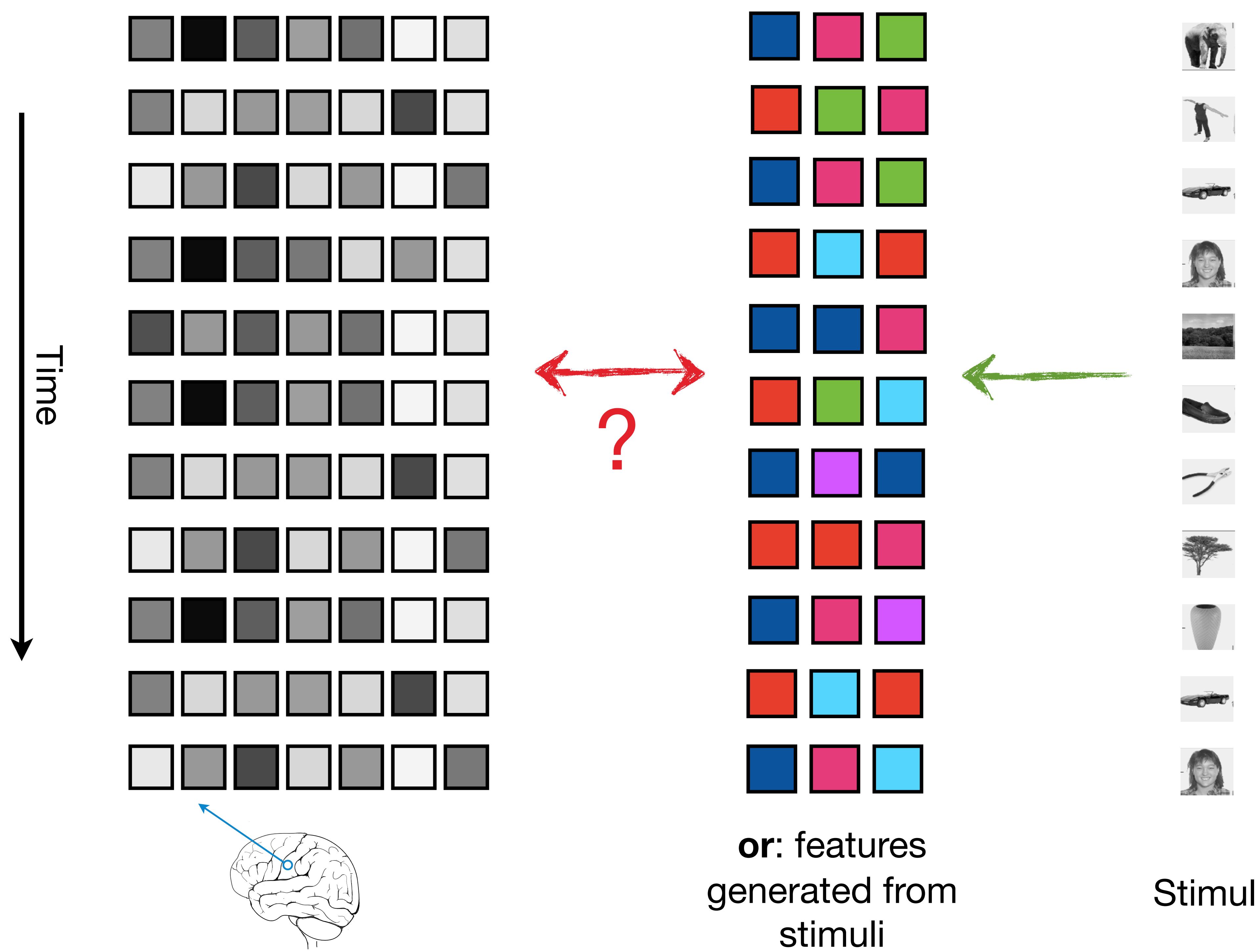



Labels - discrete classes of stimuli or experiment conditions

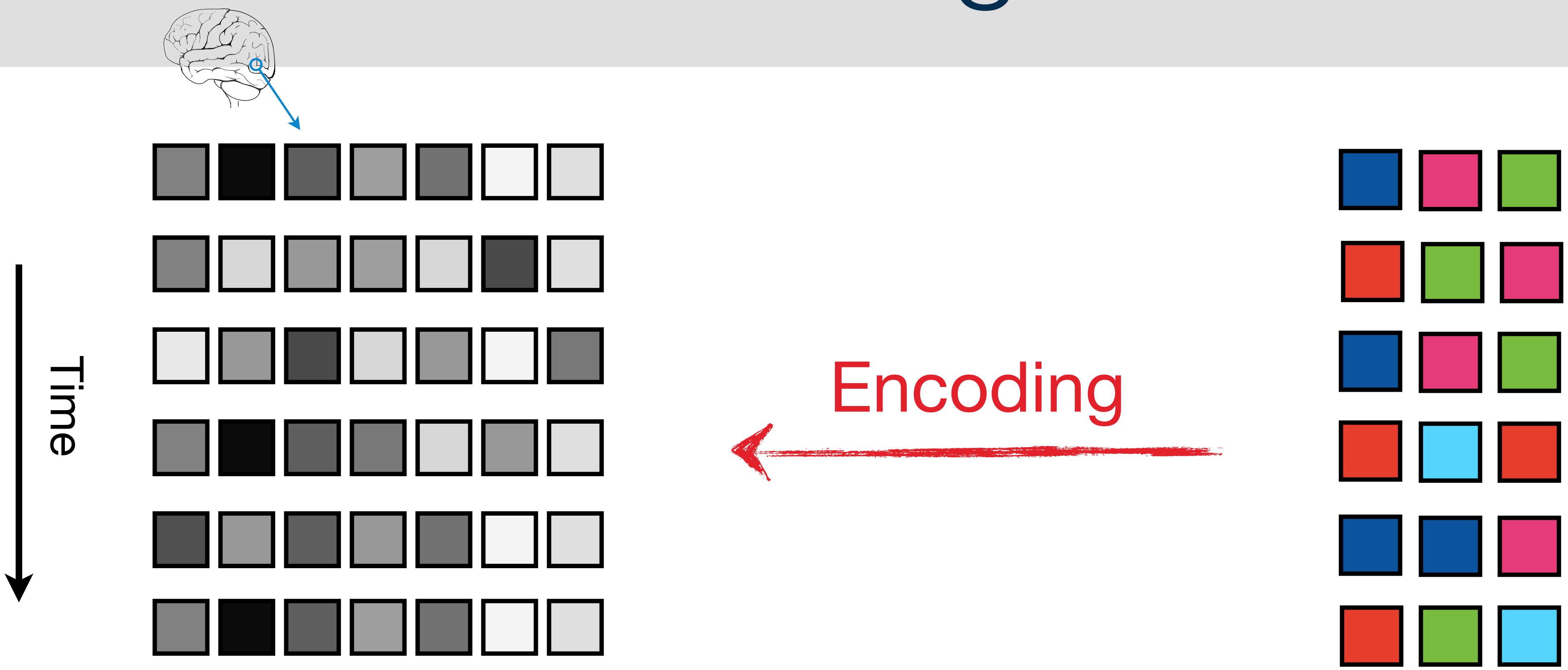
# Rich stimuli



# Rich stimuli



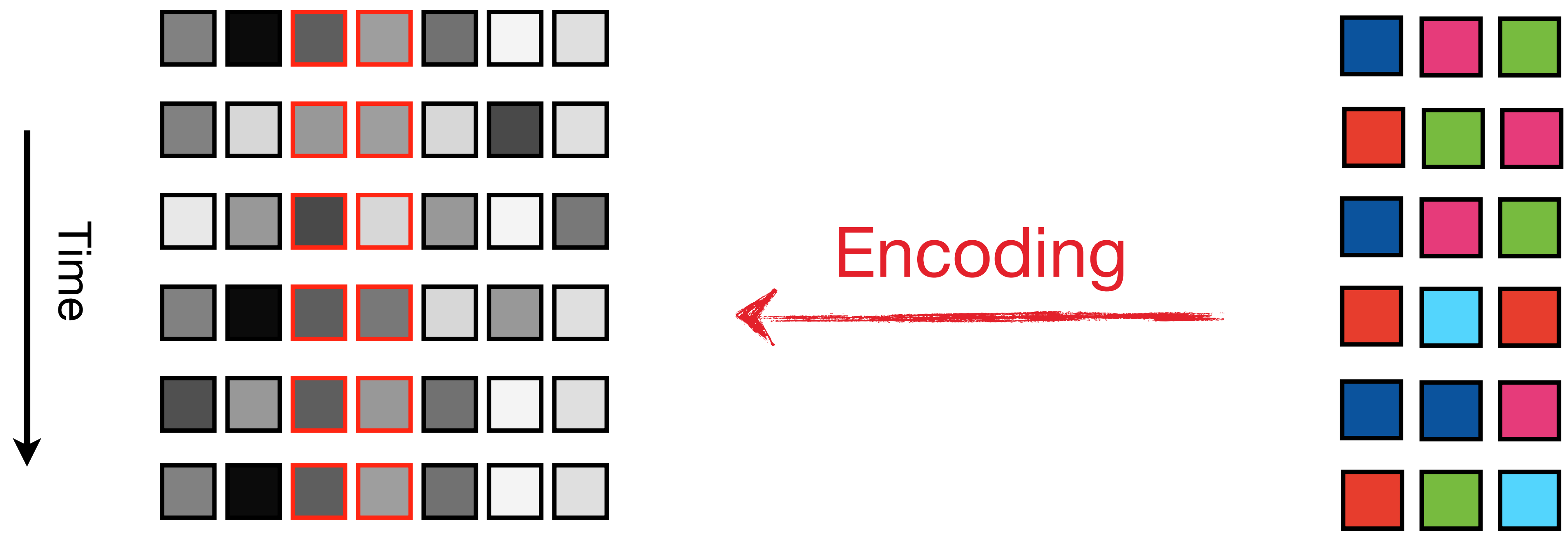
# Encoding




- **Encoding models** predict brain activity from features that correspond to experiment conditions
- Mass-univariate example: GLM
- Multivariate linear encoding model [Kay et al. 2008]

[Example: Kay et al., 2008]

# Encoding 1: represented information




1

- **Does region X represent any information about a stimulus?**
- We choose a model that predicts activity from features
- We test whether we can predict activity significantly better than chance

[Naselaris et al., 2011]

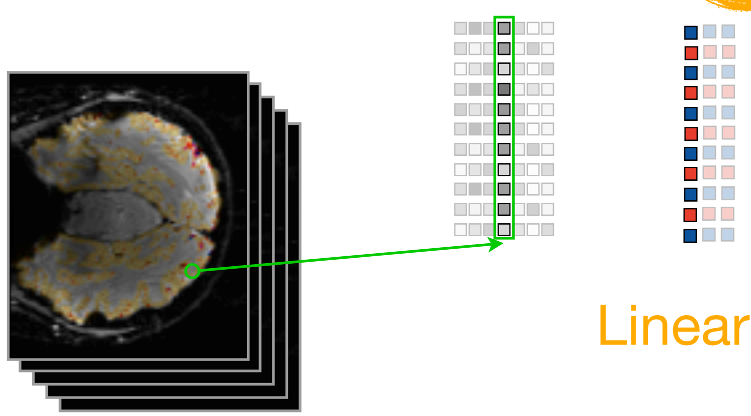
# Example - univariate

- Simple example: mass univariate analysis - GLM

Error

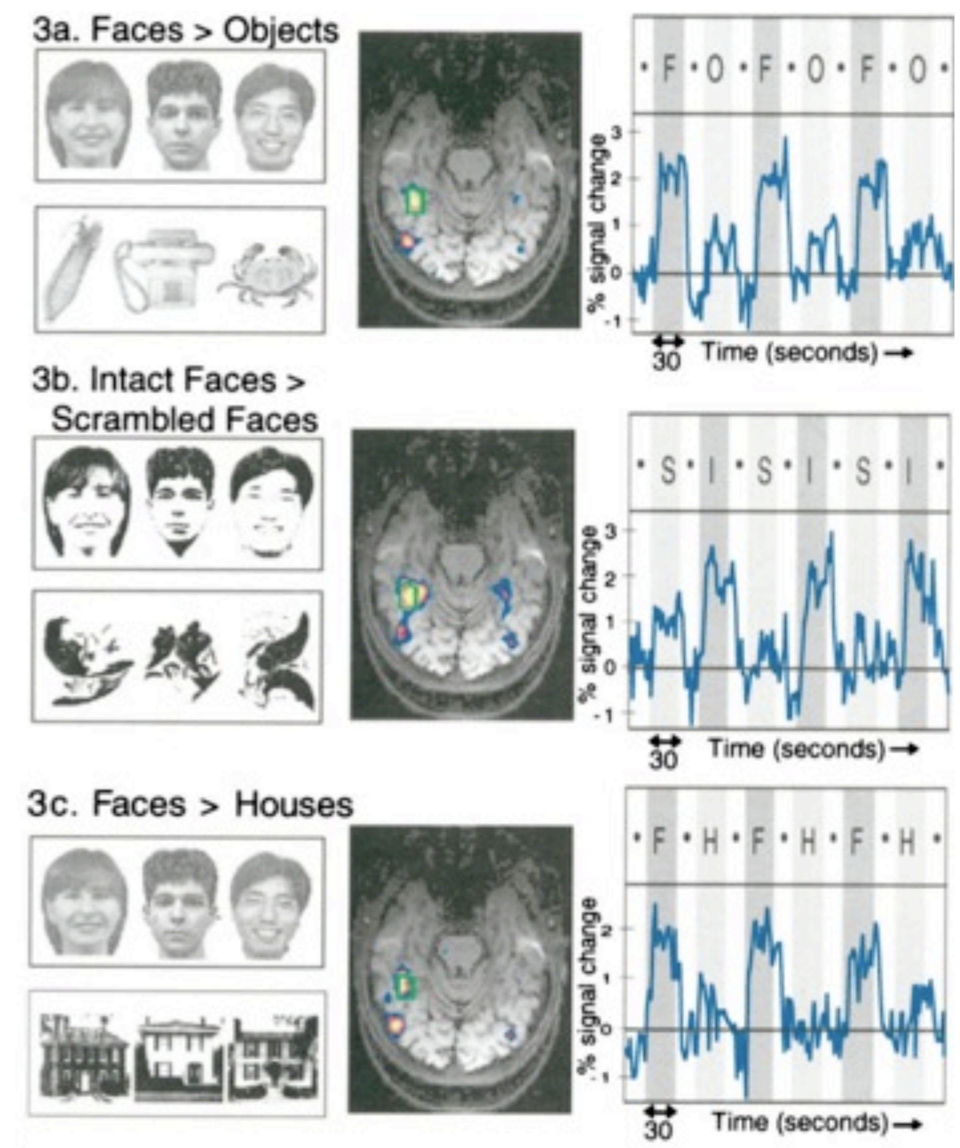
Learned coefficient

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$



Linear predictor

- Estimate **beta** coefficient ~ local effect size for each voxel
- Test if the 0-hypothesis beta = 0 is true or false (t-test)




[image from Kanwisher et al. 1997]

1

# Example - multivariate

- Predict fMRI voxel response from intermediate feature vector of experiment condition (e.g., a noun)
- Results in predictor from feature space to voxel signal space
- Predict response for un-seen stimuli if we can map them into feature space

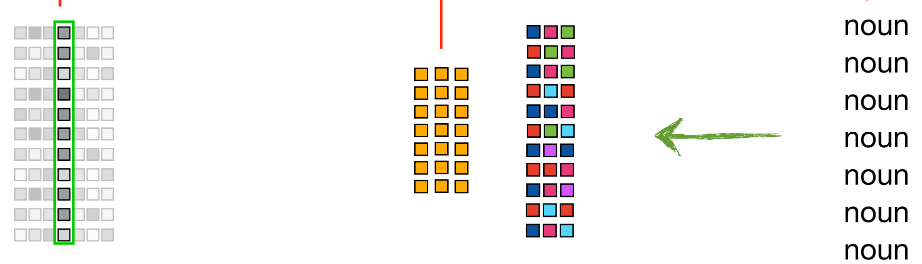
Learned scalar parameter that determines how the feature activates the voxel

Stimulus

$$y_v = \sum_{i=1}^{n} c_{vi} f_i(w)$$

Voxel signal

Feature i of n features in the feature vector




Intermediate semantic features

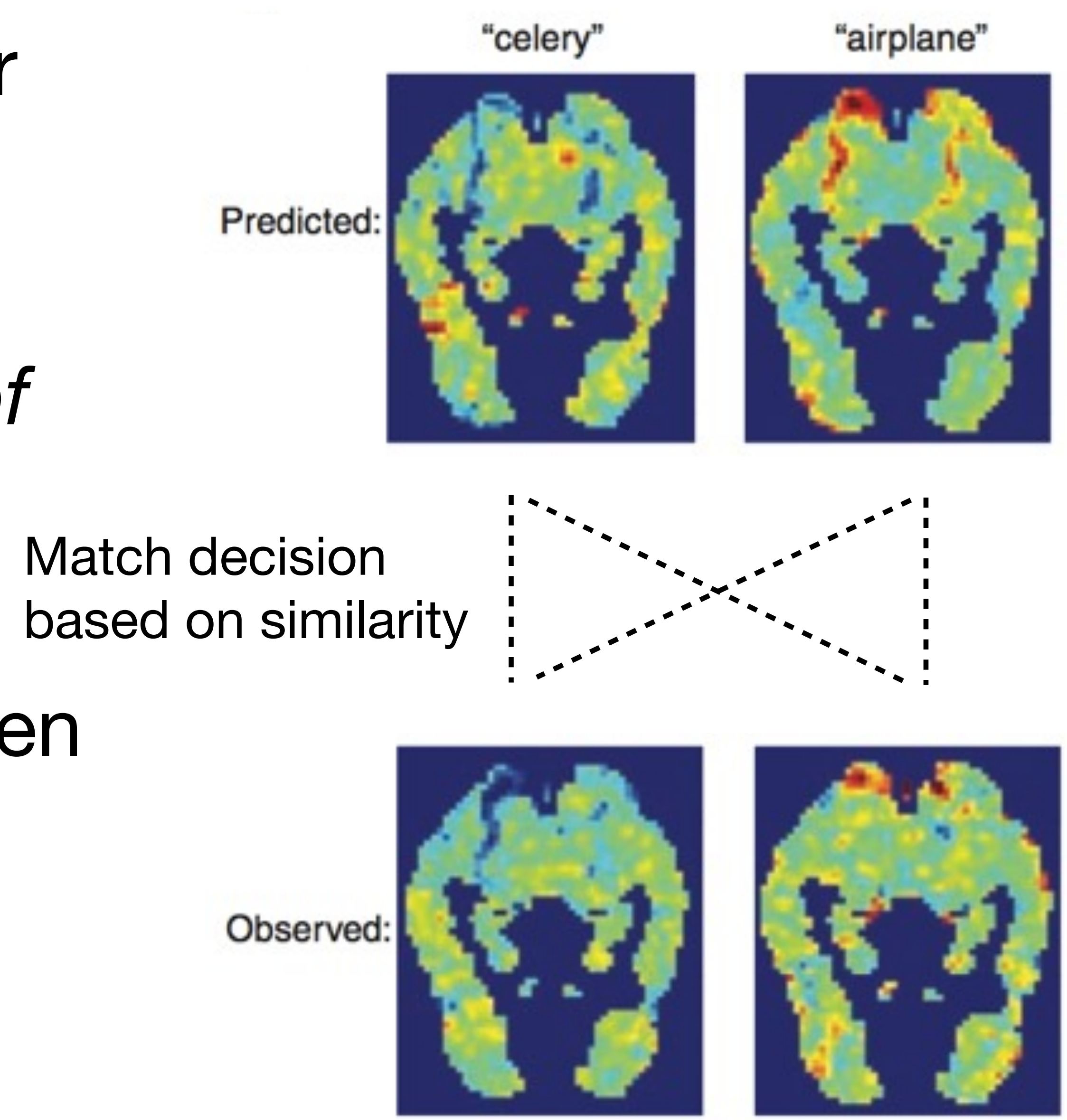
Words

[Mitchell et al. 2008]

1

# Example - multivariate cont’d

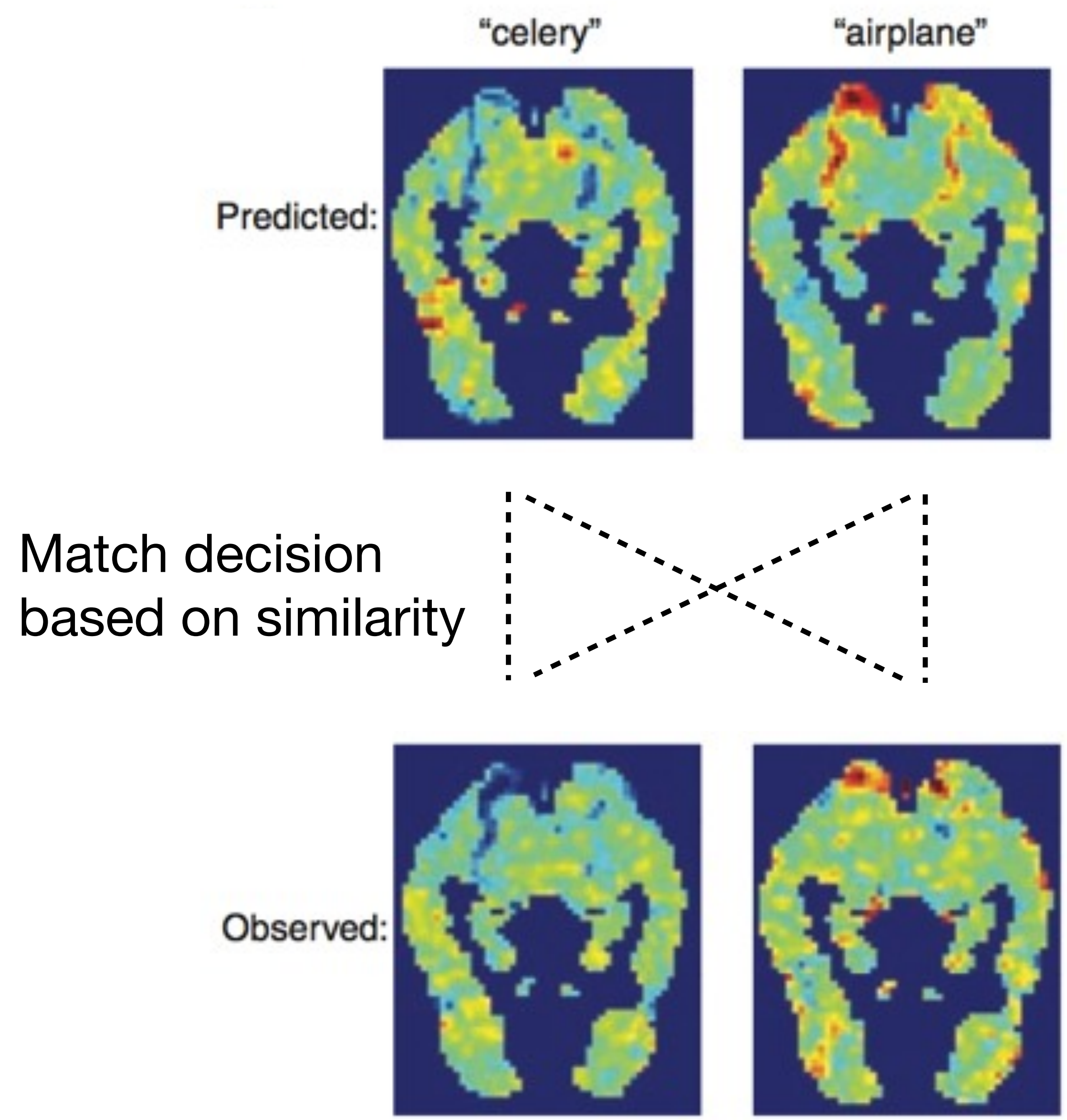
- **Validation**: compare predicted response to observed response for pairs of hold-out words in a cross validation scheme
- Match based on *cosine similarity of the vector of 500 most stable voxels*
- Test: is the match correct more often than chance?
- Chance: 50%
- How far away from 50% is needed to provide for significance of e.g., 0.05




[example from Mitchell et al. 2008]

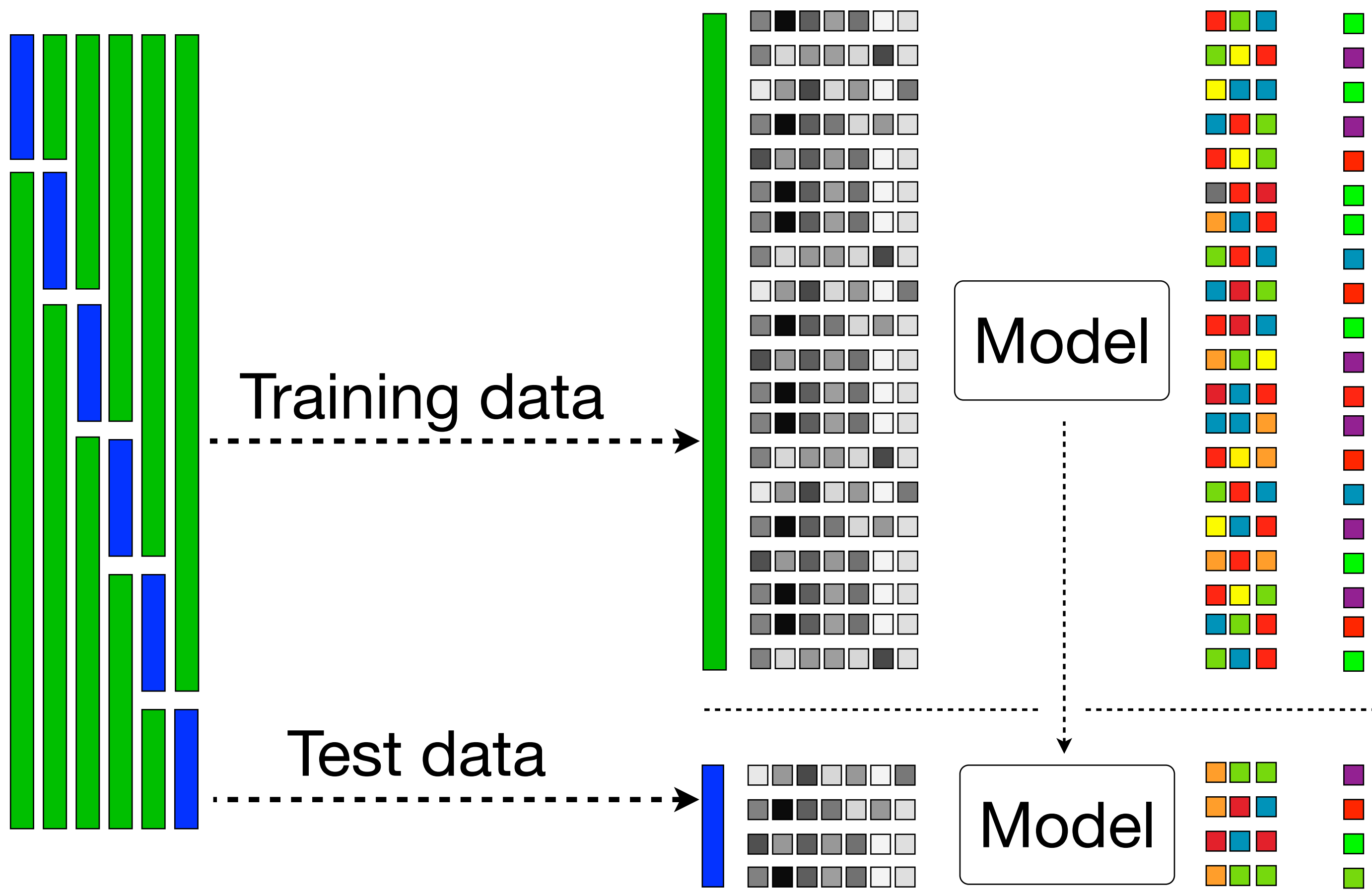
# Example - multivariate cont'd

- Estimate distribution of chance model by
  - permuting words randomly during training of the model
  - **or** choosing other random words, select intermediate features at random during training of the model
- Using the correct labels during testing
- We can determine a level that is significantly above chance based on this distribution




[example from Mitchell et al. 2008]

# Background: cross-validation




1. **Train** the model on the training data

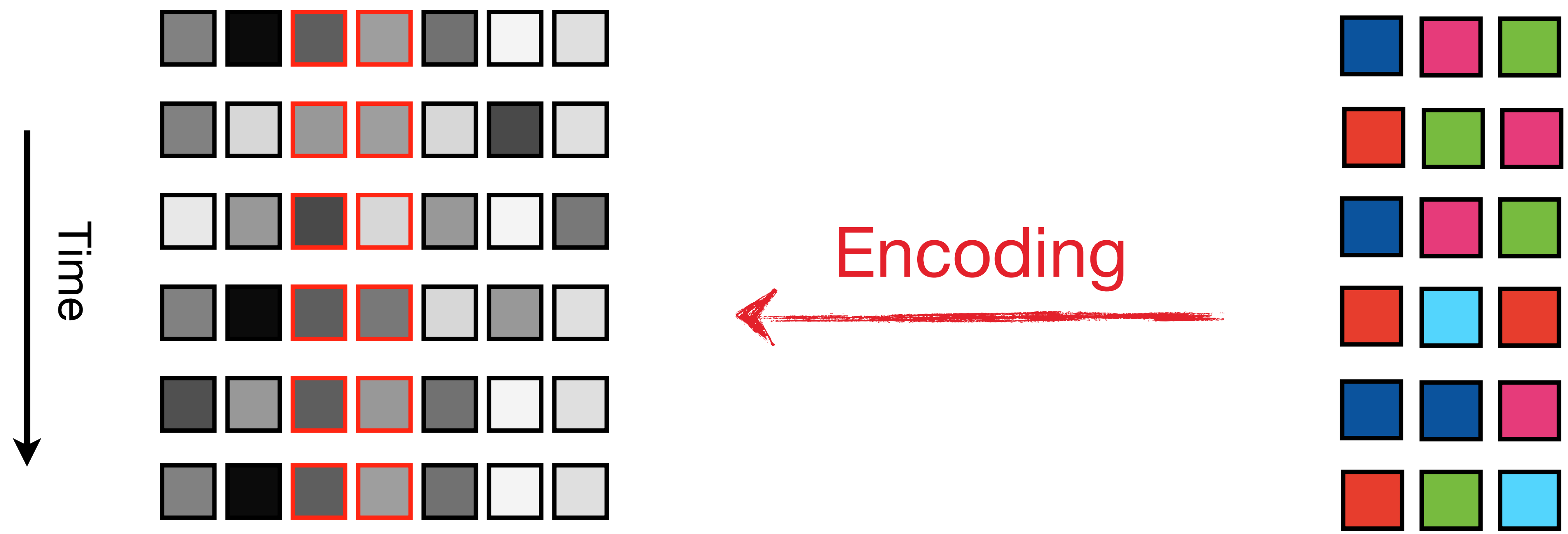
2. **Apply** the model on the test data.
*In the example: predict brain activation from feature vectors*

3. **Validate** result on the test data.
*In the example: check if assignment between simulation and observation is correct.*

Iterate over possible splits. For each split ...

2

# Encoding 2: comparing regions

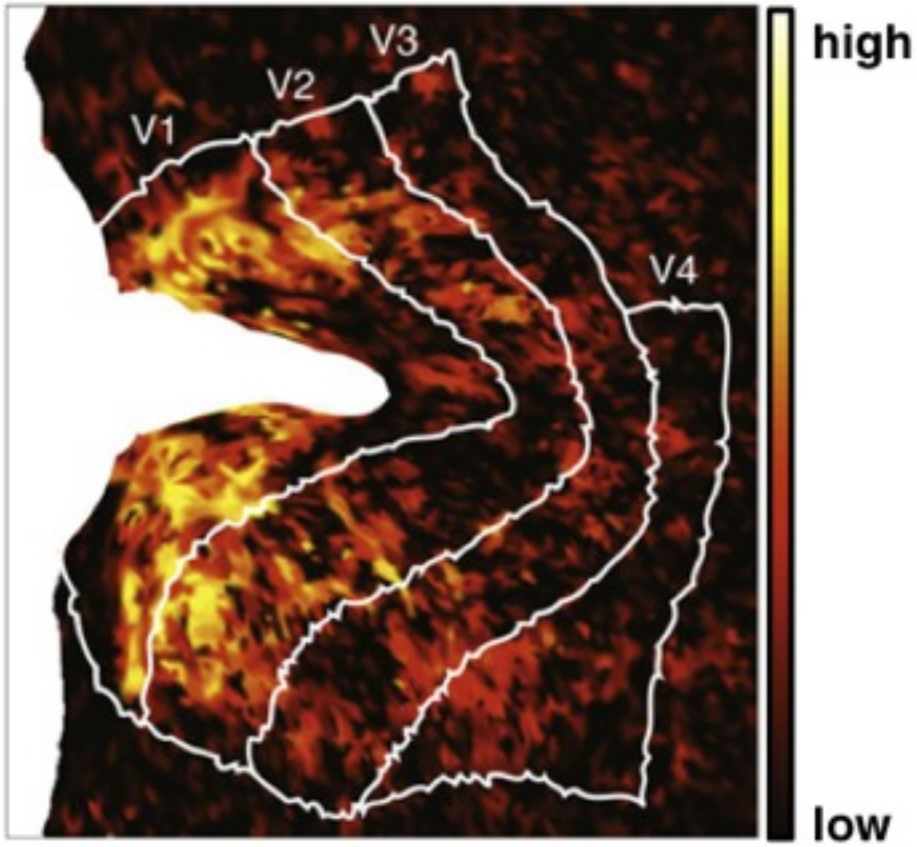



- **Does region X contain more information regarding the stimulus than region Y?**
- Caveat: this is a function of the chosen model, and can analogously be used to compare models. Be aware of your assumptions.

[Naselaris et al., 2011]

# Example

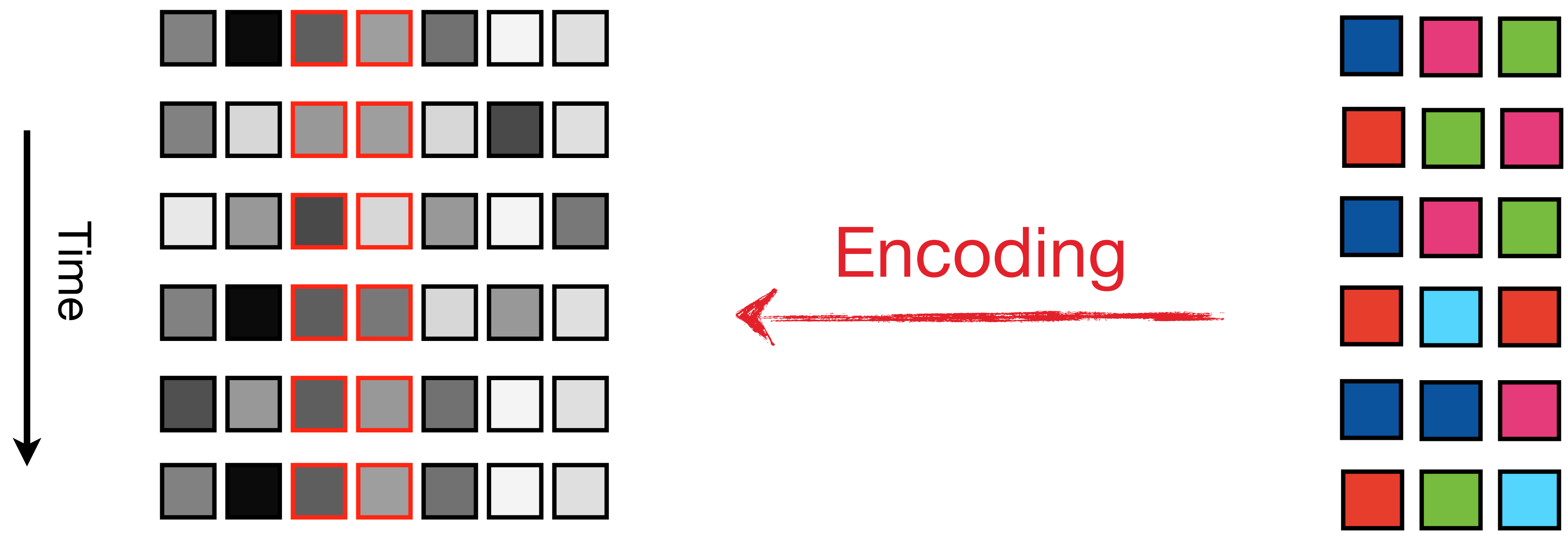
- Determine prediction accuracy at each voxel
- **Validation**: compare prediction accuracy across the brain




Prediction accuracy of fMRI activation with given model. In this case a Gabor wavelet encoding model from visual stimuli

[image from Naselaris et al., 2011]

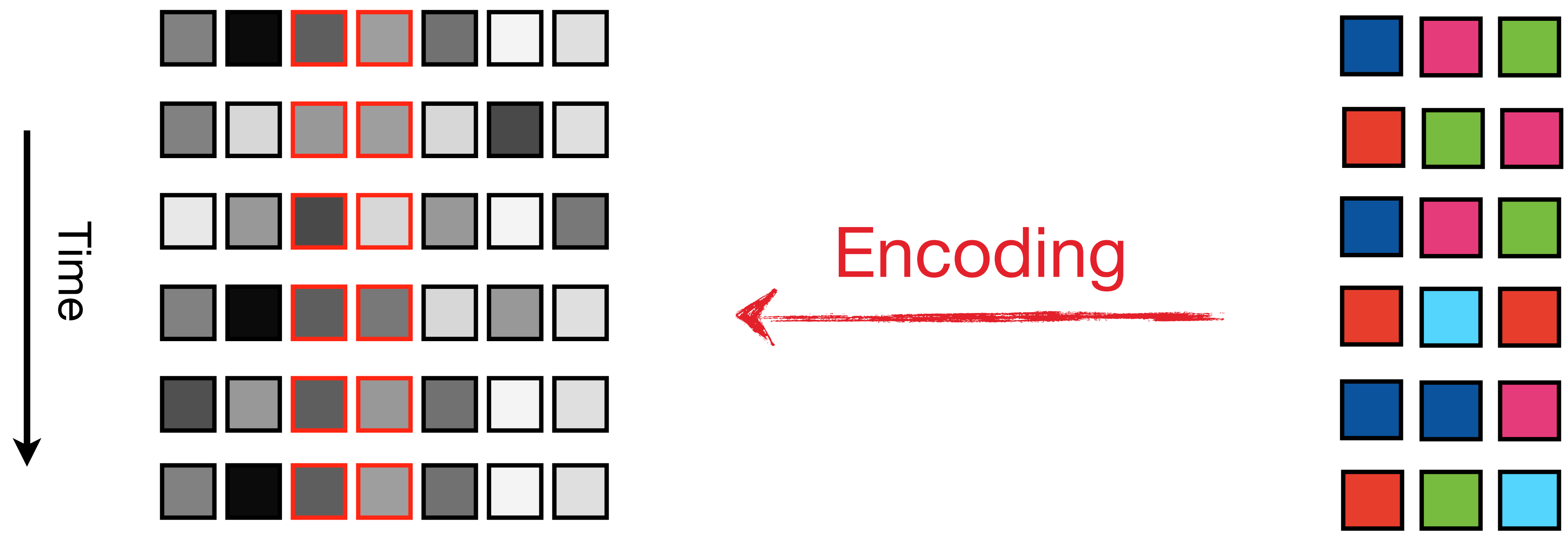
# 3 Encoding 3: comparing models




- **Does region X preferentially represent specific features / information captured by a specific model?**
- Create two encoding models based on different feature extractors, e.g., semantic versus visual, and compare the prediction accuracy

[Naselaris et al., 2011]

# 4 Encoding: complete description?



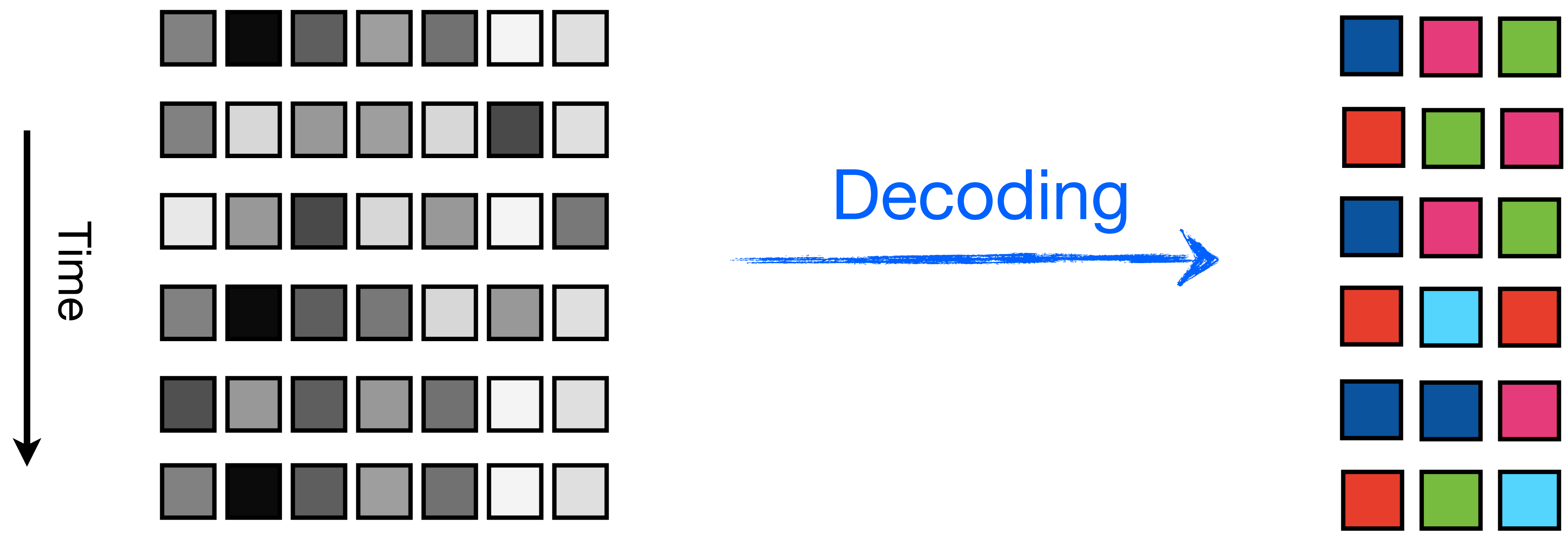

- **Which features describe a region X completely?**
- Can we find a model that explains all variance in region X (except noise)?
- Tricky: this is constrained by the experimental design, since it might well be that our model cannot explain variance that occurs during a different experiment.

[Naselaris et al., 2011]

# Decoding


- **Decoding models** predict stimuli, features or more generally experiment conditions from brain activity
- Very often classifiers are used
- We can determine if a region contains information, and can identify informative regions via feature selection approaches.

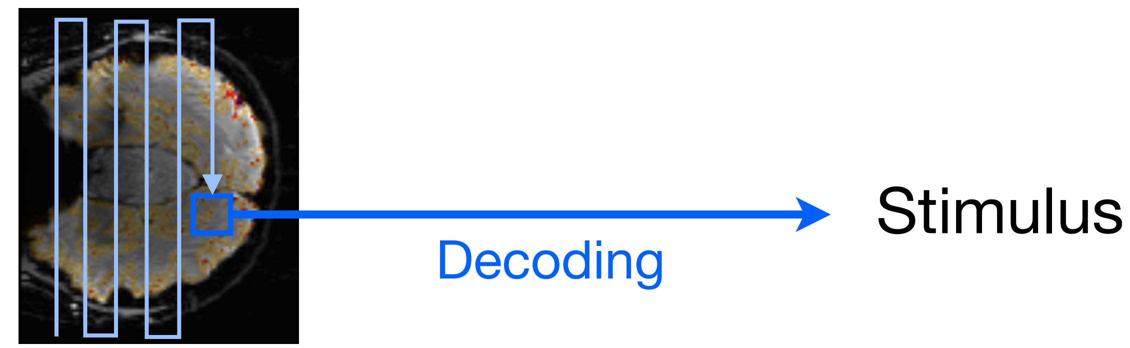
# Decoding

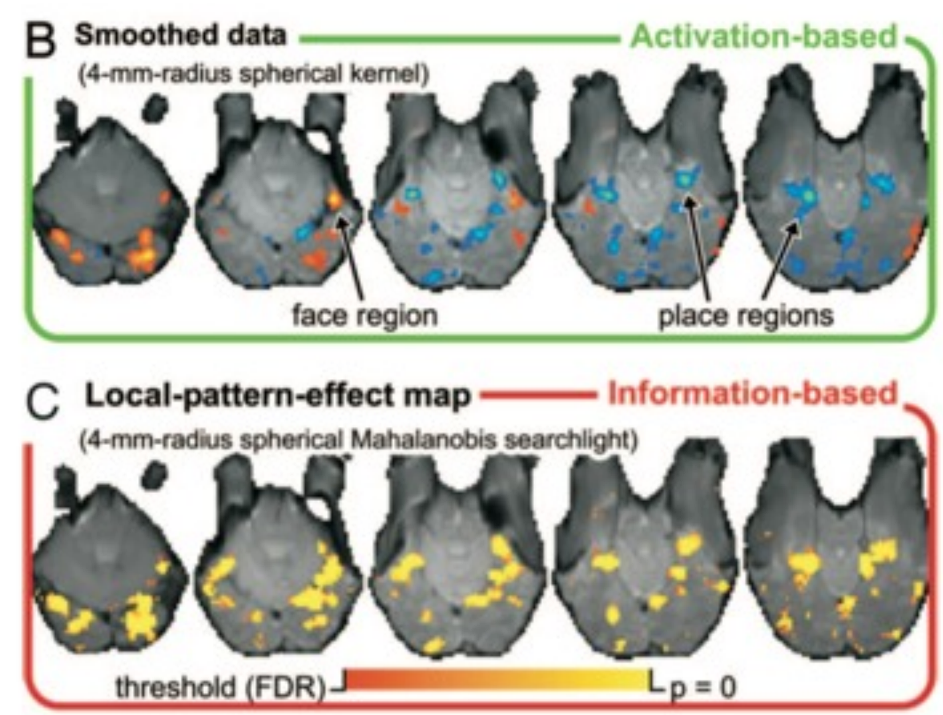



- **Is region X important for processing of a stimulus / behavior?**
- We choose a model that predicts features from activity
- We test whether we can predict stimulus/behavior significantly better than chance

# Example: Search Light

- Perform standard GLM analysis to estimate beta coefficients
- *Scan* the beta values in the brain volume with a sliding window
- Predict stimulus from values in this window by means of multivariate model (e.g., SVM)
- Test if your prediction is above chance

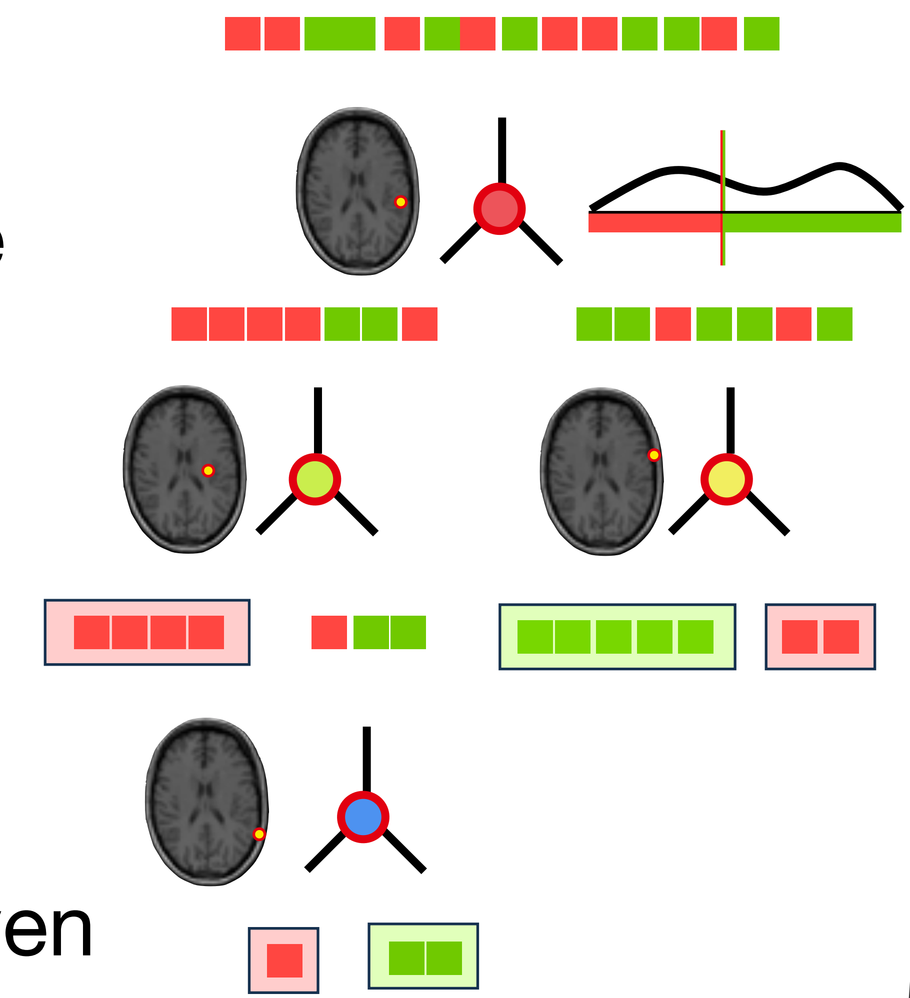



[Figure from Kriegeskorte et al., 2006]

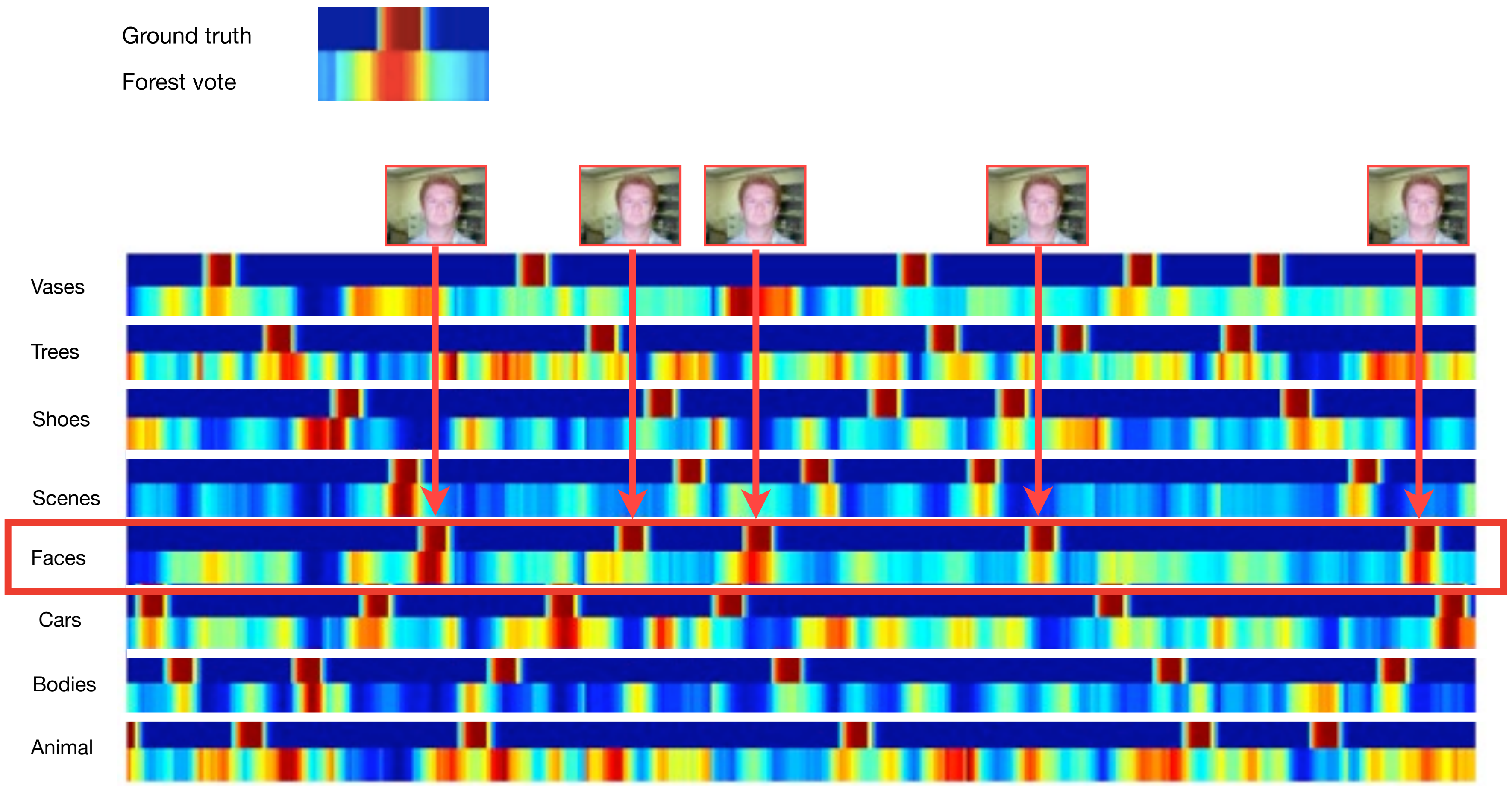
# Example: Gini contrast

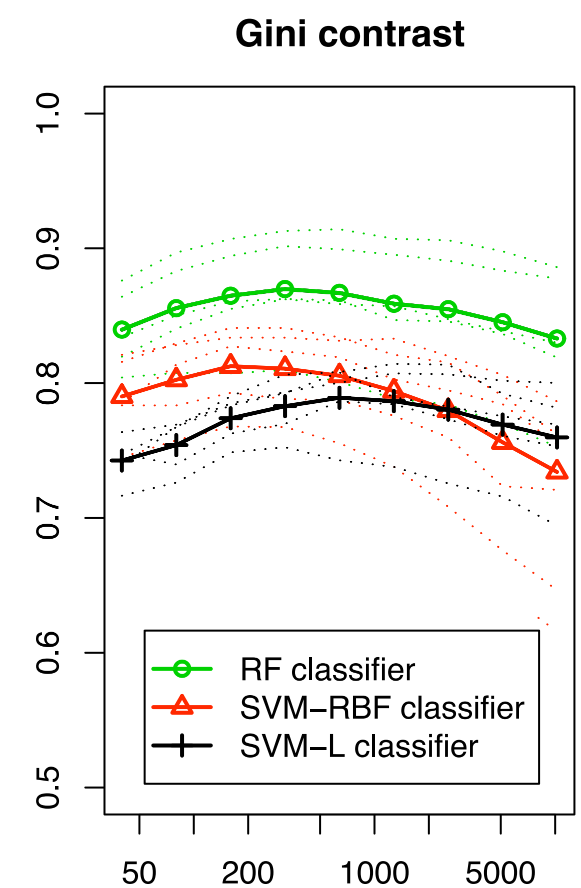
- Detect distributed patterns by learning a classifier on the entire brain / grey matter
- Bagging approaches such as Random forests offer feature scoring
- Grouping effect: informative features are reliably identified even if they are correlated



[Langs et al., 2011]

# Example: Gini contrast



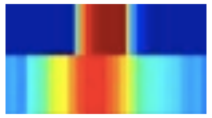



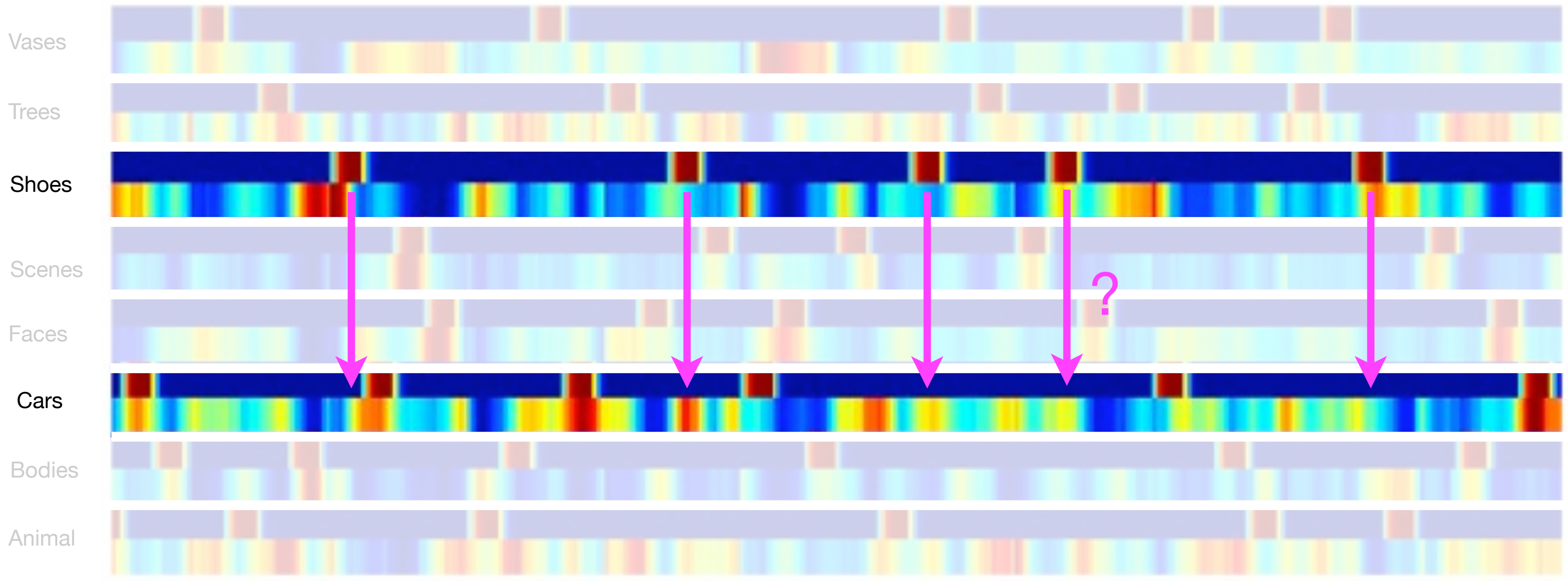

- Validate:
  - Classification accuracy in a cross-validation set-up

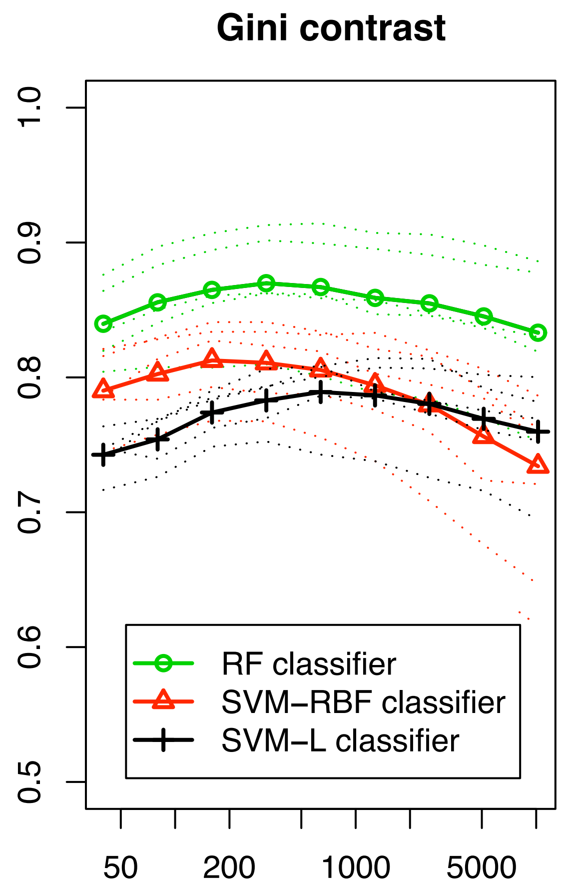
[Langs et al., 2011]

# Example: Gini contrast

Ground truth
Forest vote



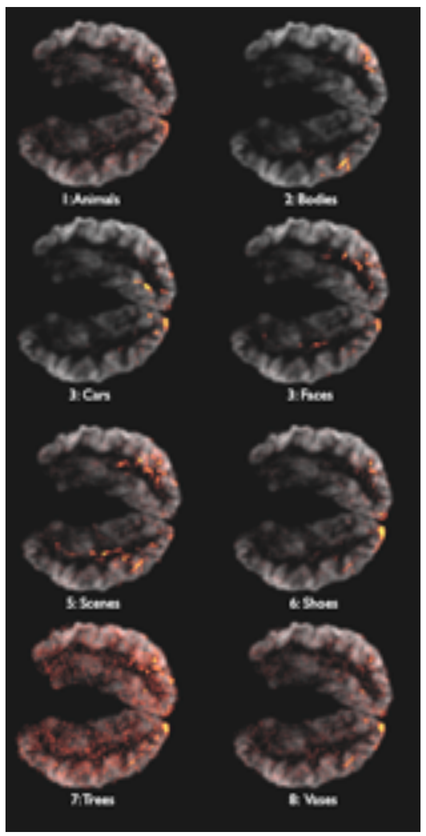


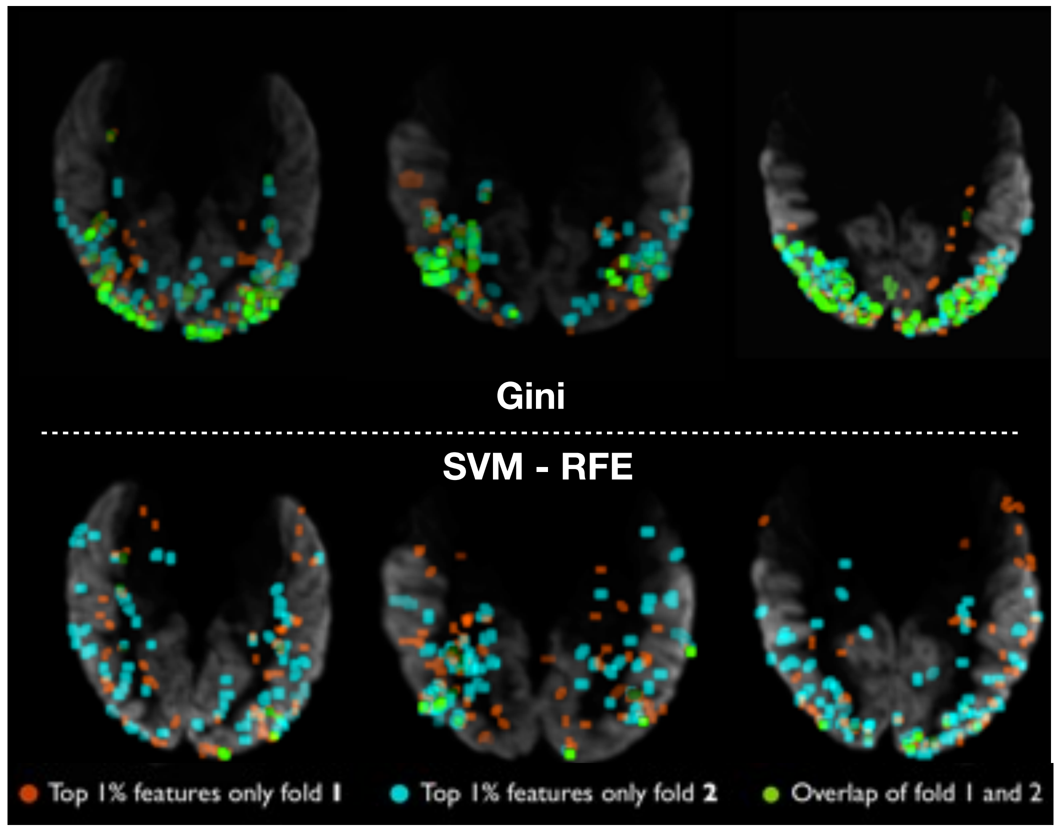



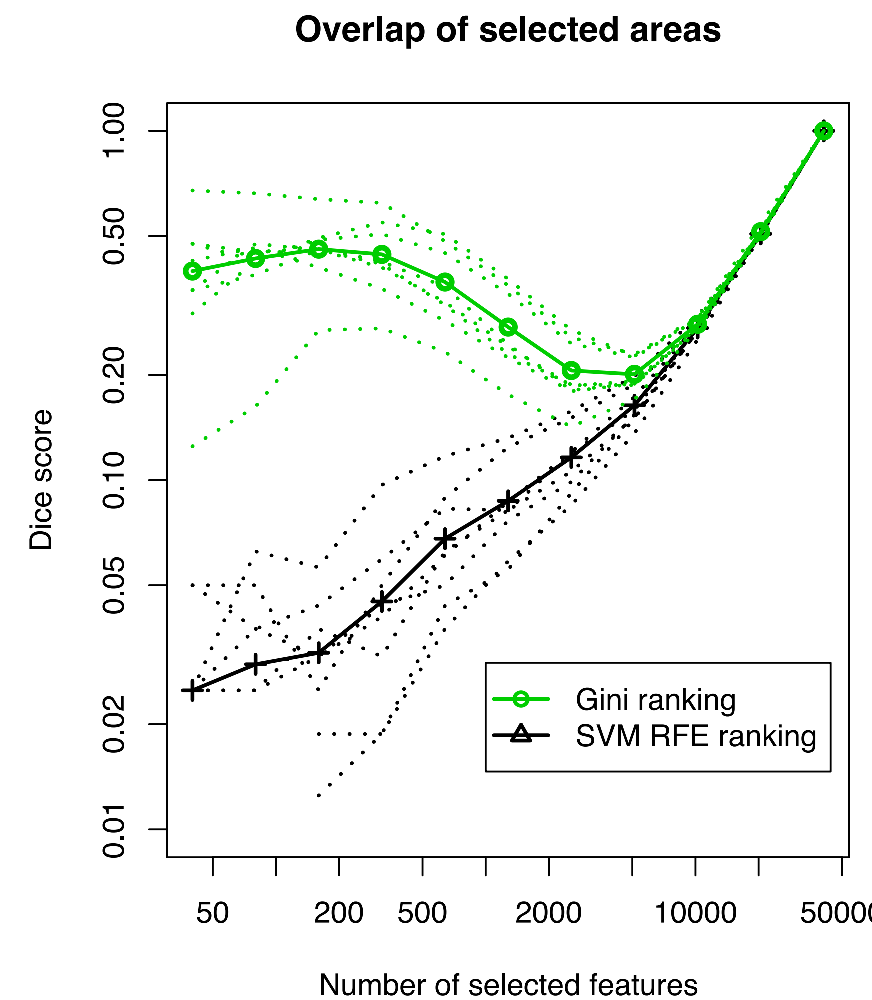

- Validate:
    - Classification accuracy in a cross-validation set-up
    - Confusion among specific pairs of classes

[Langs et al., 2011]

# Example: Gini contrast






- **Validation**:
  - Repeatability of selected features (voxels) if learning is repeated on different data sub-sets

[Langs et al., 2011]

# Structure in the data

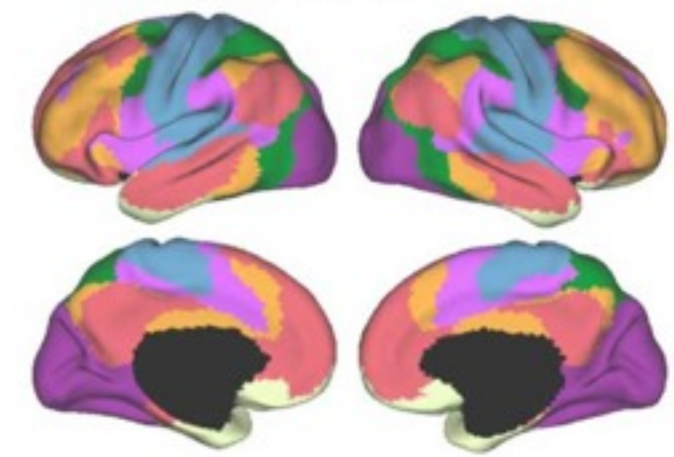
# Finding structure in function

- Is there any structure in the data that is generated by an underlying process of interest?
- We don't have ground-truth, or any out-side relationship to verify if we are detecting anything real
- Reproducibility and Specificity

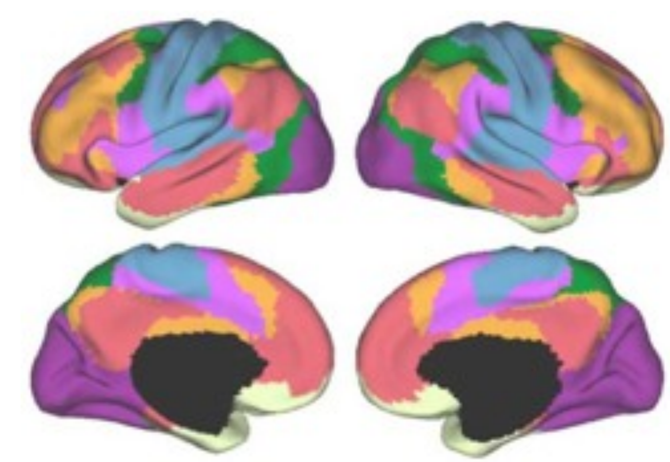
# Example: resting-state networks

- Clustering in the similarity profiles of brain signals to anchor node signals
- Results in parcellation of cortex
- **Validation**:
  - Split into discovery and replication set and comparison of network assignments

Discovery sample (500 subjects)



Replication sample (500 subjects)



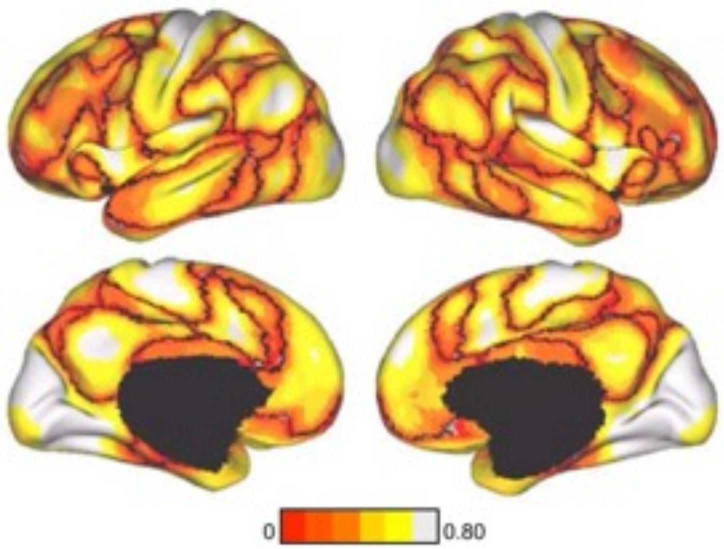
[Images from Yeo et al., 2011]

# Example: resting-state networks

- **Validation**:
- Confidence in the label assignments based on silhouette value for each vertex

$$s(i) = \frac{b(i) - a(i)}{\max_i(a(i), b(i))}$$

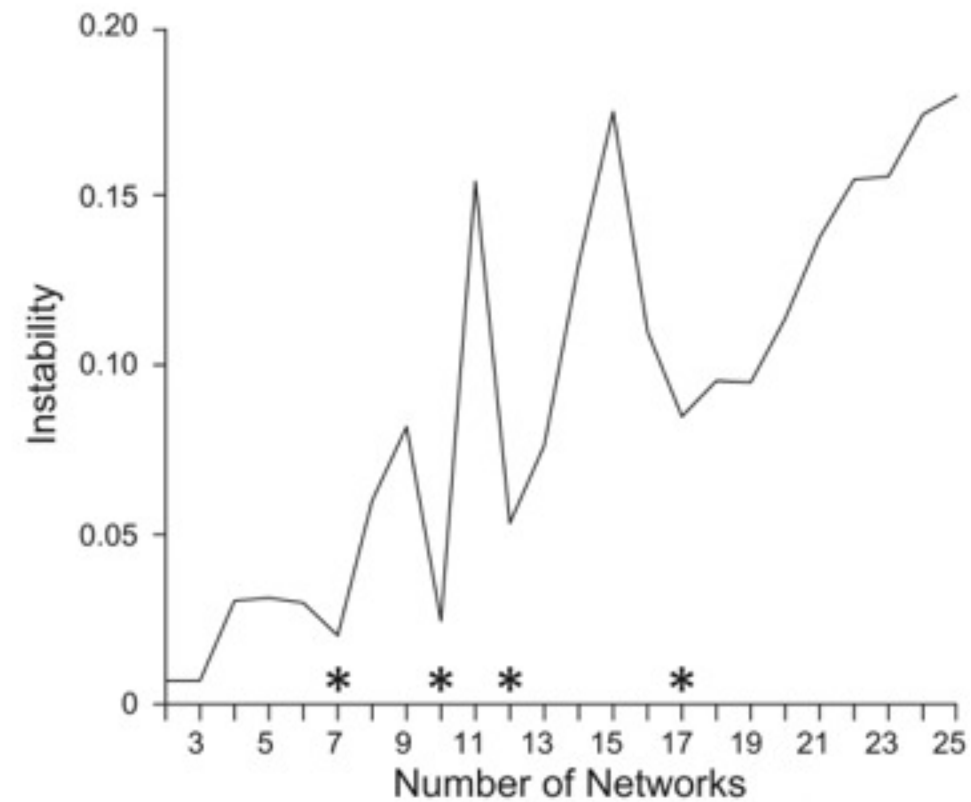
- Accumulate over all vertices: estimate of cluster separation




[Images from Yeo et al., 2011]

# Example: resting-state networks

- **Validation**:
- Evaluate the cluster stability by randomly sub-dividing subjects or vertices in subsets, and estimating clustering on this sub-set
- Transfer labels to complementary set
- Calculate dissimilarity between transferred and native cluster labels

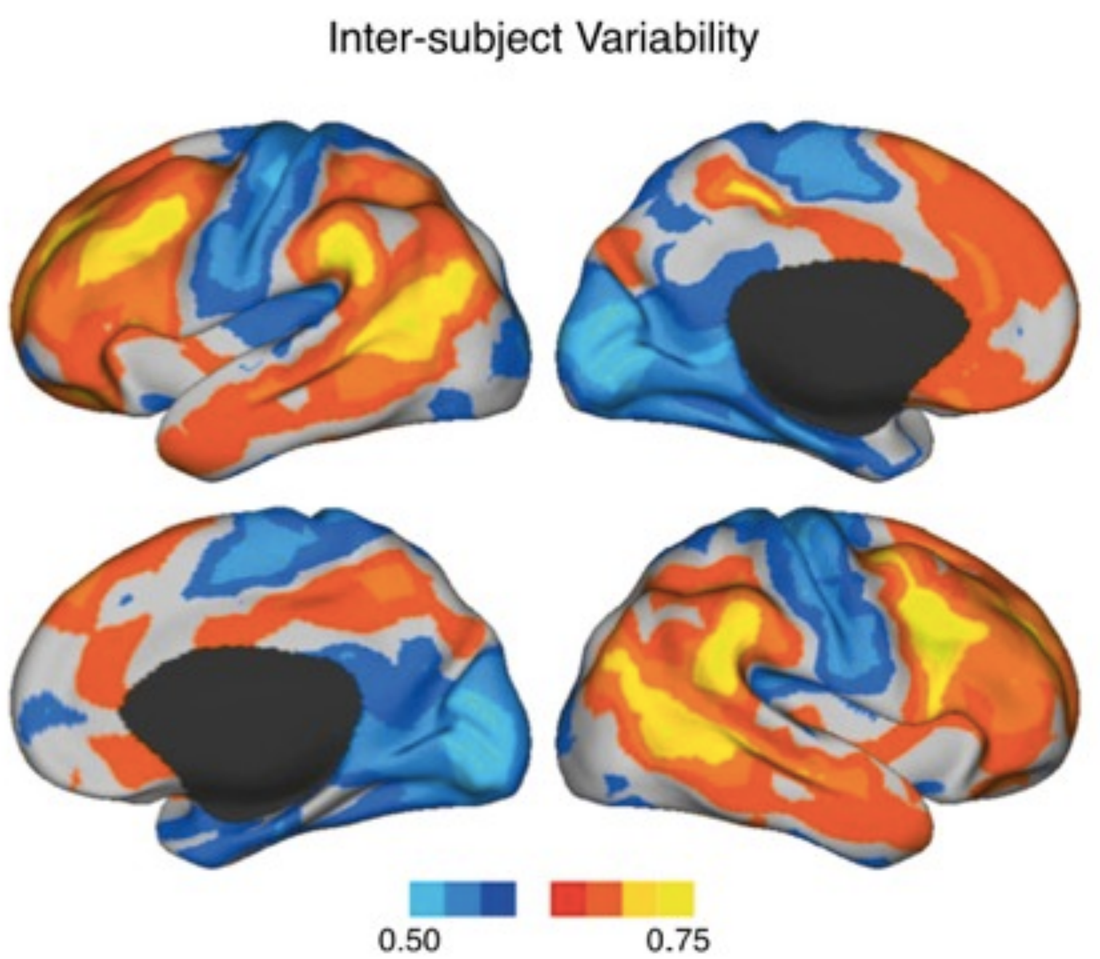



$$d(\Phi(X), Y) = \min_{\pi \in P} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{\pi(\Phi(X_i)) \neq Y_i\}$$

[Lange et al., 2004] [Images from Yeo et al., 2011]

# Specificity

- Are we detecting differences?
- Repeat measurements and control results of differences for variability that is present in repeated measurements (i.e., measurements that should be identical ideally - under your assumption)

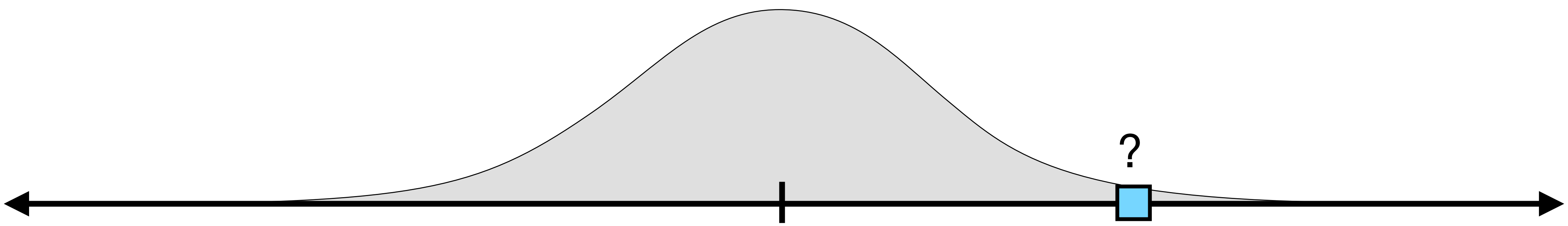



[Images from Müller et al., 2013]

Is there anything to report?

... **significance** revisited

# p-values

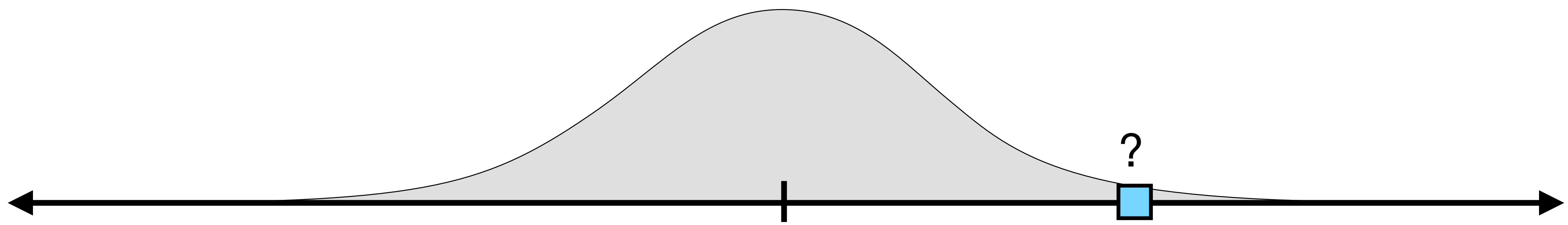



- Significance initially conceived as sanity check: ‘is it worth a second look?’ Fisher 1920
- 0-hypothesis: observations are random
- **Assuming 0-hypothesis is true, what are the odds of getting results at least as extreme as the observation**
- This probability is the p-value

[**read**: Regina Nuzzo, Nature 2014]

[initially suggested by Ronald Fisher 1920]

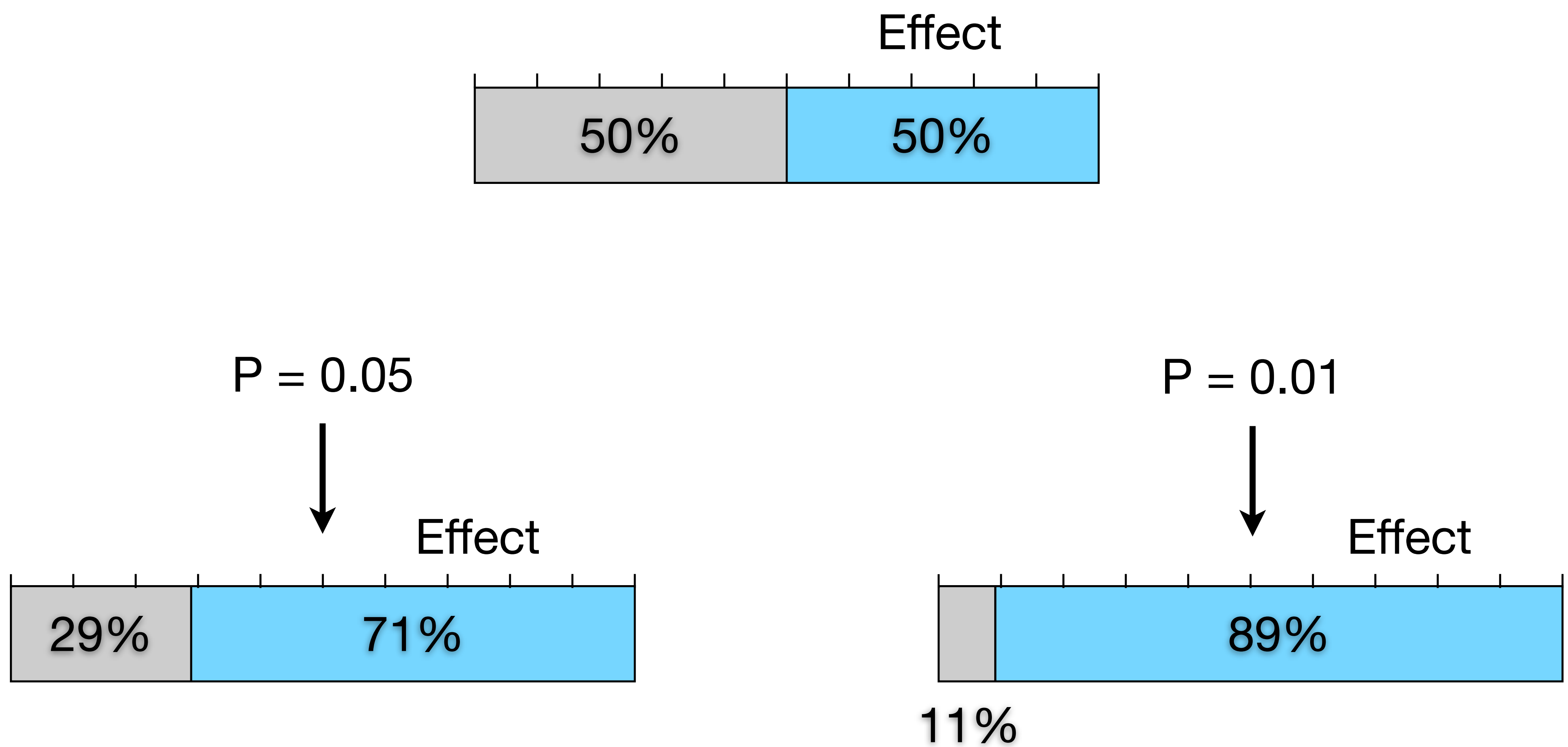
# What p-values cannot do




- p-value: probability that observation could be made despite 0-hypothesis is true
- p-value is **not** ~~probability that 0-hypothesis is true~~
- p-value is **not** the ~~probability that observation was false alarm~~
- We cannot say anything about the underlying mechanism based only on the p-value. We would need additional information for this ...

[**read**: Regina Nuzzo, Nature 2014] [initially suggested by Ronald Fisher 1920]

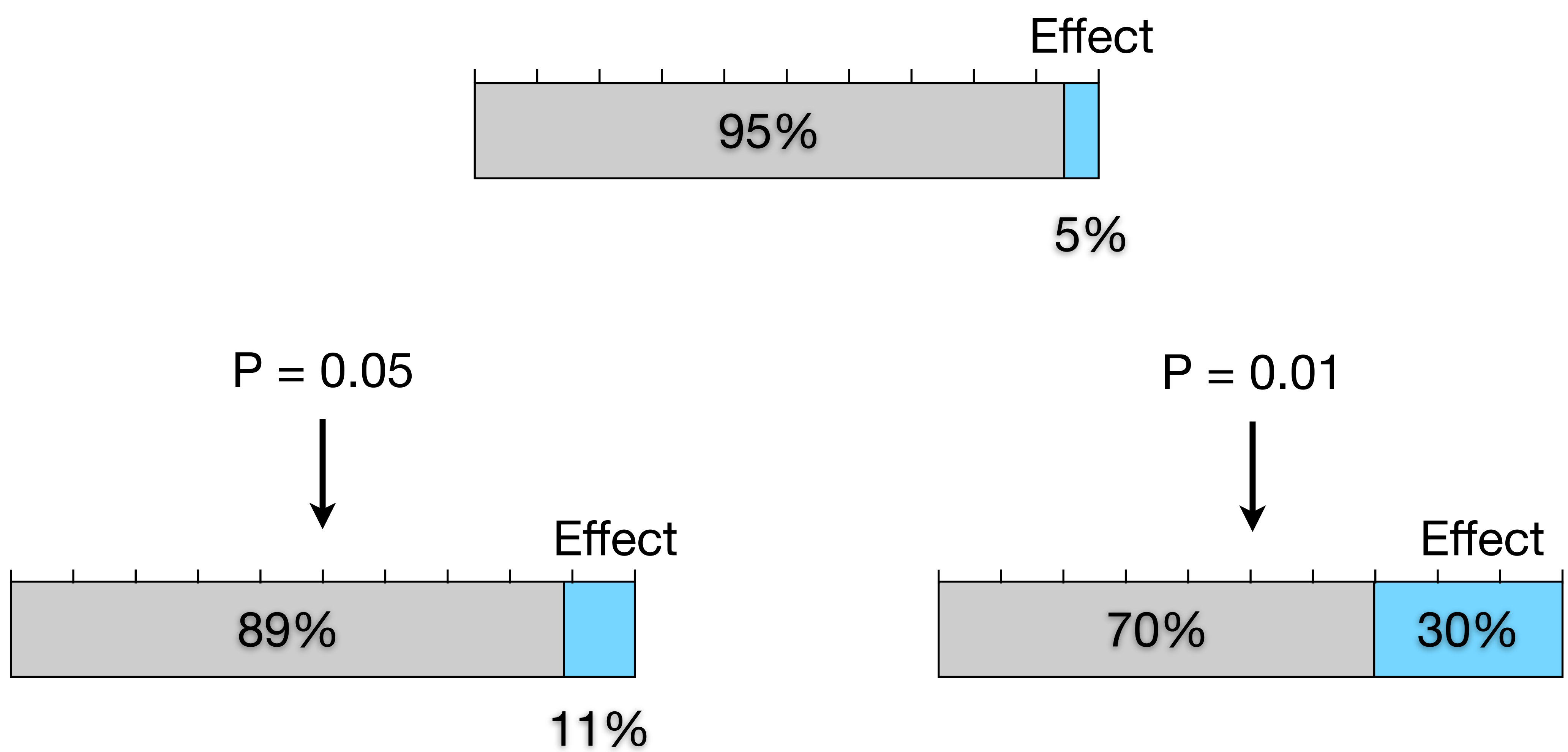
# But what can we get from it?




- Consider the prior probability of the hypothesis.
- Consider the size of the effect

[adapted from Regina Nuzzo, Nature 2014] [Goodman, 2001]

# But what can we get from it?




- Consider the prior probability of the hypothesis.
- Consider the size of the effect

[adapted from Regina Nuzzo, Nature 2014]

[Goodman, 2001]

# Bayes theorem

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} = \frac{p(B|A)}{p(B)} p(A)$$

What we learn from our observation

What we knew before

- p(A|B) Posterior probability of **hypothesis/model** A given **observation** B
- p(A) prior probability of **hypothesis/model** A
- p(B) prior probability of **observation** B
- p(B|A) Probability of **observation** B assuming **hypothesis/ model** A is true

# Summary

- Validation and Inference: what do we learn from the observations?
- Multivariate relationships between functional neuroimaging data and experiment conditions
- Structure in functional neuroimaging data
- Significance - do we have anything to report?

## Thank you!

# Literature


A. G. Huth, S. Nishimoto, A. T. Vu, and J. L. Gallant, “A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain,” Neuron, vol. 76, no. 6, pp. 1210–1224, Dec. 2012.

T. Naselaris, K. N. Kay, S. Nishimoto, and J. L. Gallant, “Encoding and decoding in fMRI,” NeuroImage, vol. 56, no. 2, pp. 400–410, May 2011.

K. J. Friston, “Modalities, Modes, and Models in Functional Neuroimaging,” Science, vol. 326, no. 5951, pp. 399–403, Oct. 2009.

T. M. Mitchell, S. V. Shinkareva, A. Carlson, K.-M. Chang, V. L. Malave, R. A. Mason, and M. A. Just, “Predicting human brain activity associated with the meanings of nouns.,” Science, vol. 320, no. 5880, pp. 1191–1195, May 2008.

K. N. Kay, T. Naselaris, R. J. Prenger, and J. L. Gallant, “Identifying natural images from human brain activity,” Nature, vol. 452, no. 7185, pp. 352–355, Mar. 2008.

B. Thirion, E. Duchesnay, E. Hubbard, J. Dubois, J.-B. Poline, D. Lebihan, and S. Dehaene, “Inverse retinotopy: inferring the visual content of images from brain activation patterns.,” NeuroImage, vol. 33, no. 4, pp. 1104–1116, Dec. 2006.

K. Norman, S. Polyn, G. Detre, and J. V. Haxby, “Beyond mind-reading: multi-voxel pattern analysis of fMRI data,” Trends in Cognitive Sciences, vol. 10, no. 9, pp. 424–430, 2006.

G. Langs, B. H. Menze, D. Lashkari, and P. Golland, “Detecting stable distributed patterns of brain activation using Gini contrast.,” *NeuroImage*, vol. 56, no. 2, pp. 497–507, May 2011.

N. Kriegeskorte, R. Goebel, and P. Bandettini, “Information-based functional brain mapping,” *Proc Natl Acad Sci USA*, vol. 103, no. 10, pp. 3863–3868, Mar. 2006.

B. T. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, B. Fischl, H. Liu, and R. L. Buckner, “The organization of the human cerebral cortex estimated by intrinsic functional connectivity.,” *J Neurophysiol*, vol. 106, no. 3, pp. 1125–1165, Sep. 2011.

T. Lange, V. Roth, M. Braun, and J. Buhmann, “Stability-based validation of clustering solutions,” *Neural Comput*, vol. 16, no. 6, pp. 1299–1323, 2004.