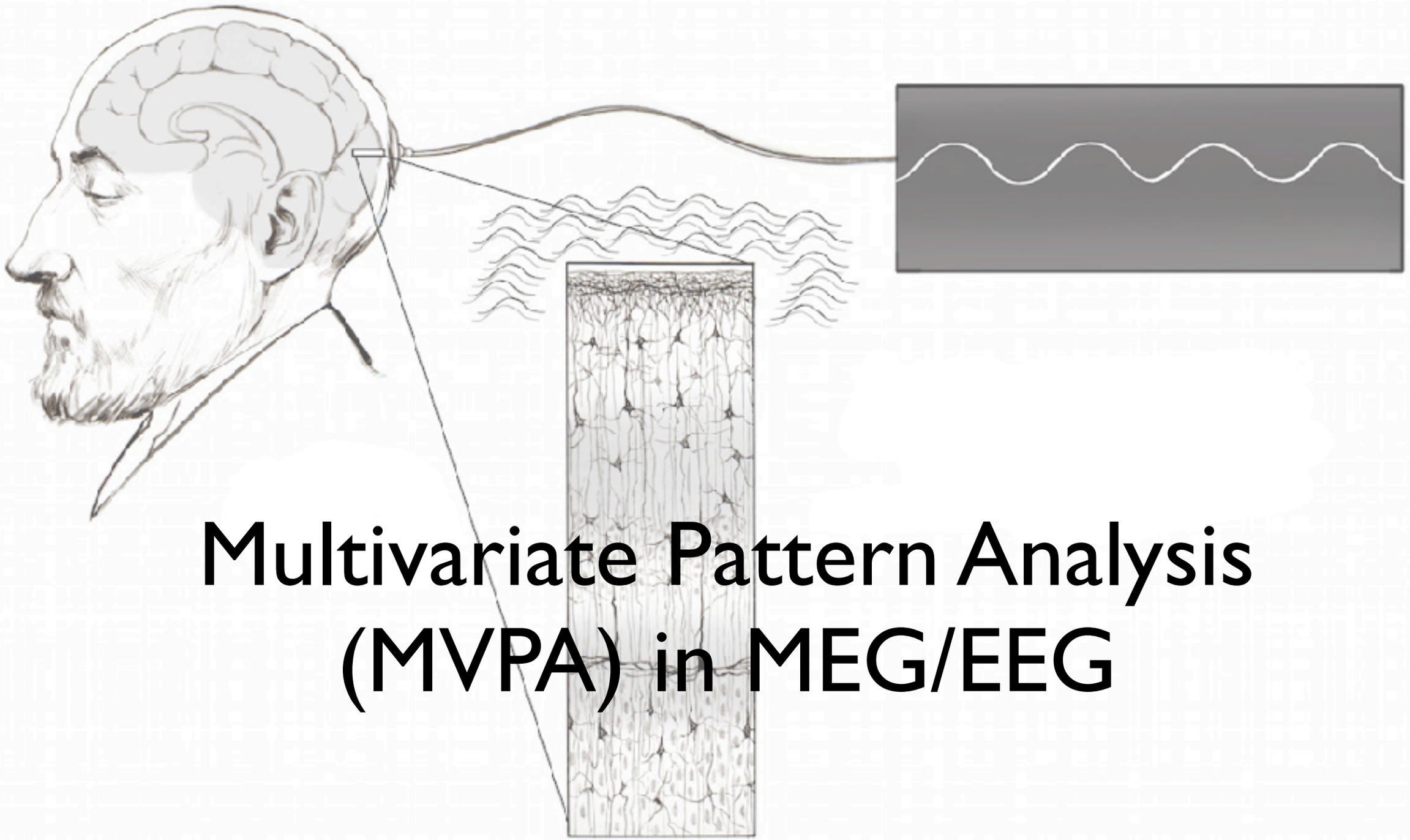


Multivariate Pattern Analysis (MVPA) in MEG/EEG

Johannes Fahrenfort

epdos

Graduate education network

Course overview

- **Day 1 (introduction)**
  - lecture 1: History and electrophysiological basis of EEG
  - lecture 2: Backward decoding models in MVPA: concepts and analytical approach
  - lecture 3: Advantages of MVPA, the temporal generalization method
  - lecture 4: Architecture of the ADAM toolbox and explaining the experiment of the practical
  - *Afternoon: practical*
- **Day 2 (advanced)**
  - lecture 1: Multiple comparisons, MVPA experimental design, mapping brain to brain/behavior
  - lecture 2: Forward encoding models in MVPA: how do they work, concepts and analytical approach
  - *Afternoon: practical, analyze your own data and/or a supplied dataset*

Lecture 1

- Multiple comparison correction
- MVPA experimental design: decoding-specific confounds
- Ways to map brain to brain, or brain to behavior

Type I and Type II errors

- Type I error is the incorrect rejection of a true null hypothesis (also known as a "false positive" finding)
- Type II error is incorrectly retaining a false null hypothesis (also known as a "false negative" finding)

The multiple comparisons problem (MCP)

The multiple comparisons problem (MCP)

- Statistical tests return the **probability** of the observed data under the **null-hypothesis** (no effect).
- Often people use a threshold of 0.05, meaning that 1 in 20 tests will be significant even when there is no actual effect
- For 1 second of EEG (one ERP), recorded in 64 electrodes at 512 Hz, we compute 64 * 512 = **32768 statistical tests**
- This means that even without an effect, 32768 * 0.05 = **1638 tests** are going to be significant **because of normal random variation (Type I false positives)**
- How to tell which tests are significant by chance, and which tests are significant because of a real experimental effect?
- This is called **the multiple comparisons problem**

Solutions (1)

- **Bonferroni correction**: divide your statistical threshold by the number of statistical tests you want to perform, e.g. $P = 0.05 / 32768 = 0.0000015$

- The chance that even **a single test among *all* these tests** is spuriously significant under this threshold is $32768 * 0.0000015 = 0.05$

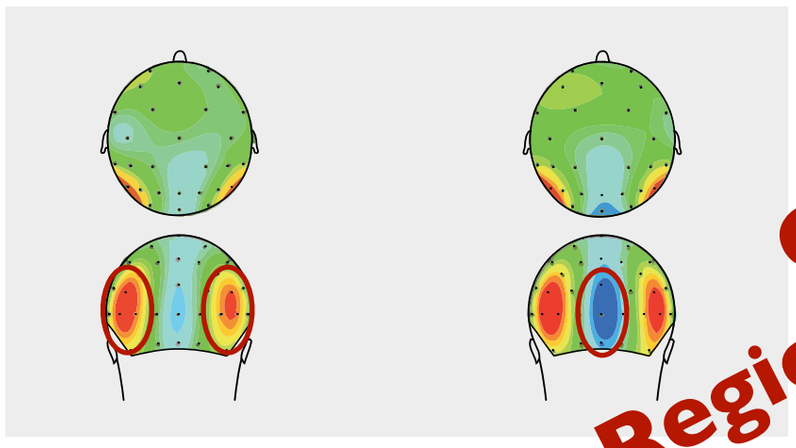
- Bonferroni correction in EEG research is usually considered **overly conservative**

Solutions (2)

- Restricting the number of comparisons by pre-selecting time windows and/or electrodes of interest
- To prevent double-dipping, selection has to be based on *independent data* (e.g. on the literature or on a *split-half procedure*)

Conceptually similar to Region of Interest analysis in fMRI

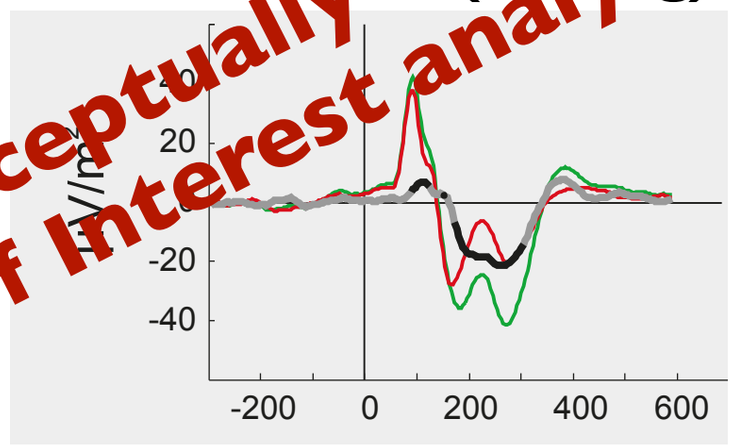
***EVEN* TRIALS (selection)**



78 - 109 ms

109 - 141 ms

***ODD* TRIALS (testing)**

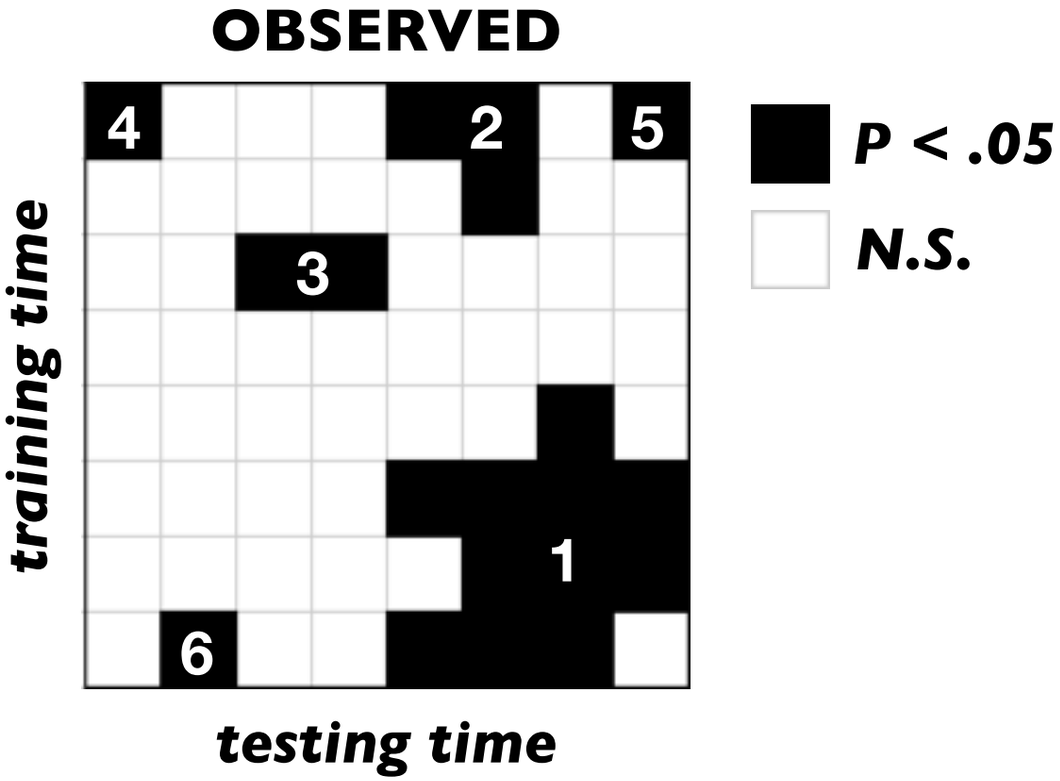


testing only in occipitotemporal and occipital pooling

Solutions (3)

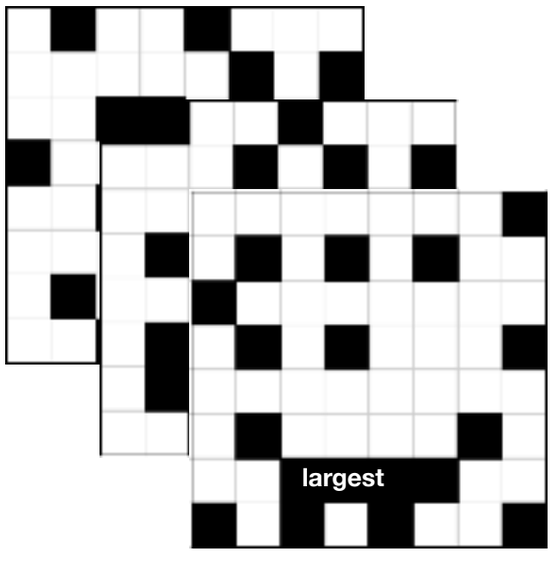
- in **Cluster based permutation-testing** the test statistic is based on *clusters* rather than on individual *samples*
- It computes the probability that a *cluster* of the observed size occurs by chance
- *P*-values refer to significance of clusters, *not* to significance of individual samples

Cluster-based permutation



RECIPE

1. Perform statistical tests for all samples
2. Determine the clusters (temporally or spatially contiguous significant samples) using some threshold (e.g. *P*<.05), and give each a number
3. Count the **size** of each of the **observed clusters** (C1 = **11**, C2 = **3**, C3 = **2**, etc.)
4. Create a counter for each of the clusters
5. Repeat the following for **n** iterations:
   - Permute the labels of the conditions
   - Re-compute all tests
   - Determine the size of the largest cluster
   - Increase the counters of all observed clusters for which this permuted cluster is larger
6. Divide all counters by **n**. These are the *P*-values for your clusters.

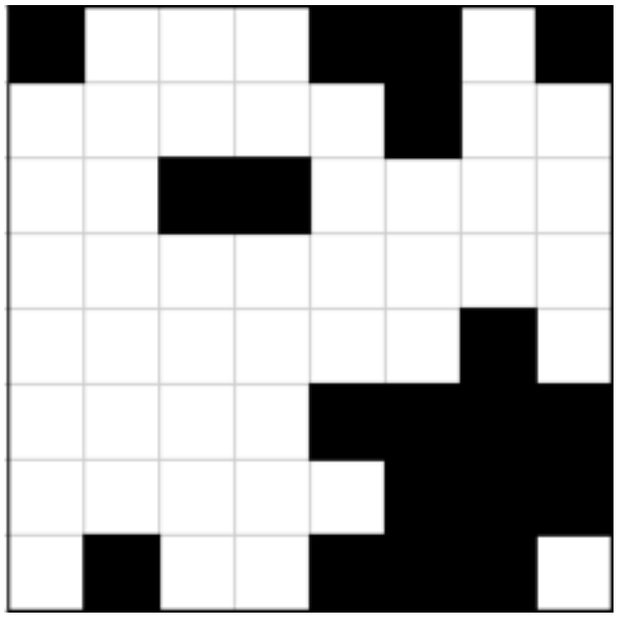


COUNTER	*P*-vals
1: 0	0.009
2: 455	0.455
3: 763	0.763
4: 897	0.897
5: 897	0.897
6: 897	0.897

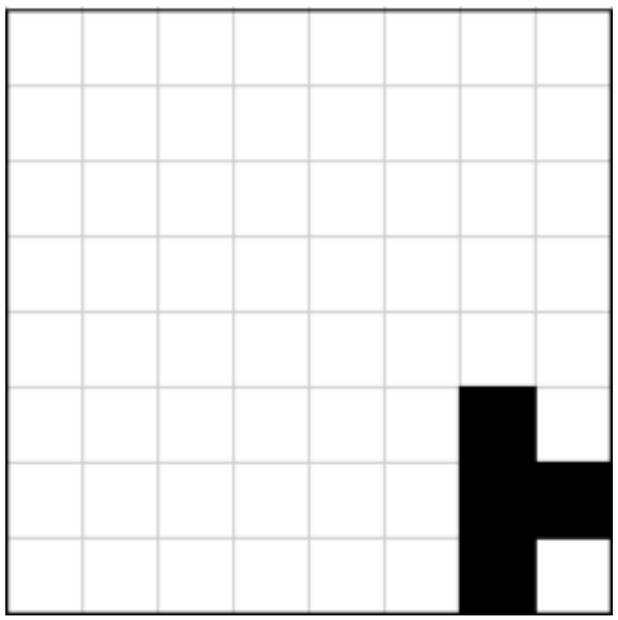
.... n = 1000

Hypothetical outcome

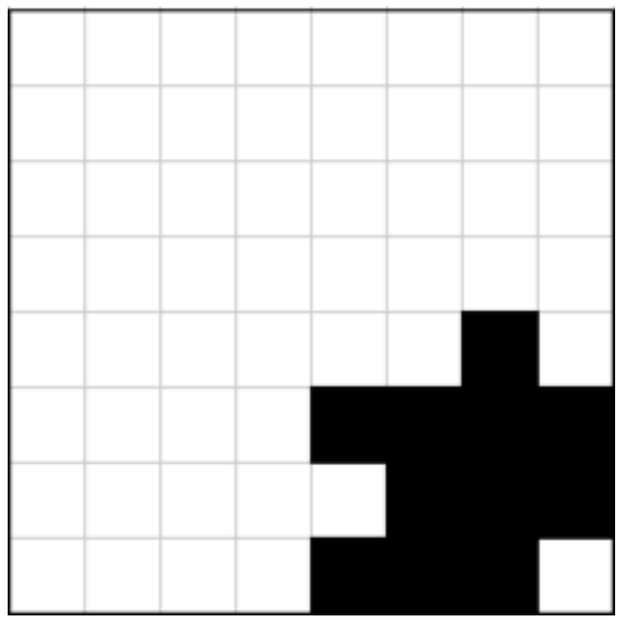
UNCORRECTED



BONFERRONI



CLUSTER-BASED



Solutions (4)

- The **FDR (False Discovery Rate)** controls the expected proportion of incorrectly rejected null hypotheses ("false positives")
- The FDR determines a cutoff *P*-value under which no more than a set percentage of tests q (q is usually 5%) is likely to reflect spurious false positives
- Procedure: the *P*-values are ordered from large to small, a cut-off is determined under which q = 5%

False Discovery Rate (FDR)

list of *P*-values (10 tests)

0.6892	0.0794	0.1656	0.0311	0.7482	0.0162	0.2630	0.6020	0.5285	0.6541

sorted *P*-values

0.0162	0.0311	0.0794	0.1656	0.2630	0.5285	0.6020	0.6541	0.6892	0.7482

threshold *P*-values = false discovery rate q (usually .05) * (index_of_test/nr_of_tests) =

(1/10)*.05	(2/10)*.05	(3/10)*.05	(4/10)*.05	(5/10)*.05	(6/10)*.05	(7/10)*.05	(8/10)*.05	(9/10)*.05	(10/10)*.05
0.0050	0.0100	0.0150	0.0200	0.0250	0.0300	0.0350	0.0400	0.0450	0.0500

Bonferroni correction

Uncorrected

Find the first sorted *P*-value that is smaller than or equal to the threshold *P*-value =
sorted *P*-values <= threshold *P*-values

sorted *P*-values	0.0162	0.0311	0.0794	0.1656	0.2630	0.5285	0.6020	0.6541	0.6892	0.7482
threshold *P*-vals	0.0050	0.0100	0.0150	0.0200	0.0250	0.0300	0.0350	0.0400	0.0450	0.0500
	X	X	X	X	X	X	X	X	X	X

False Discovery Rate (FDR)

list of *P*-values (10 tests)

0.0794	0.6020	0.0002	0.6892	0.6541	0.1656	0.7482	0.2630	0.5285	0.0311

threshold *P*-values = false discovery rate q (usually .05) * (index_of_test/nr_of_tests) =

(1/10)*.05	(2/10)*.05	(3/10)*.05	(4/10)*.05	(5/10)*.05	(6/10)*.05	(7/10)*.05	(8/10)*.05	(9/10)*.05	(10/10)*.05
0.0050	0.0100	0.0150	0.0200	0.0250	0.0300	0.0350	0.0400	0.0450	0.0500

Bonferroni correction

Uncorrected

Find the first sorted *P*-value that is smaller than or equal to the threshold *P*-value =
sorted *P*-values <= threshold *P*-values

sorted *P*-values	0.0002	0.0311	0.0794	0.1656	0.2630	0.5285	0.6020	0.6541	0.6892	0.7482
threshold *P*-vals	0.0050	0.0100	0.0150	0.0200	0.0250	0.0300	0.0350	0.0400	0.0450	0.0500
	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗

False Discovery Rate (FDR)

list of *P*-values (10 tests)

0.0285	0.0482	0.0156	0.0094	0.0130	0.0003	0.0392	0.0311	0.0320	0.0011

threshold *P*-values = false discovery rate q (usually .05) * (index_of_test/nr_of_tests) =

(1/10)*.05	(2/10)*.05	(3/10)*.05	(4/10)*.05	(5/10)*.05	(6/10)*.05	(7/10)*.05	(8/10)*.05	(9/10)*.05	(10/10)*.05
0.0050	0.0100	0.0150	0.0200	0.0250	0.0300	0.0350	0.0400	0.0450	0.0500

Bonferroni correction

Uncorrected

Find the first sorted *P*-value that is smaller than or equal to the threshold *P*-value =
sorted *P*-values <= threshold *P*-values

sorted *P*-values	0.0003	0.0011	0.0094	0.0130	0.0156	0.0285	0.0320	0.0311	0.0392	0.0482
threshold *P*-vals	0.0050	0.0100	0.0150	0.0200	0.0250	0.0300	0.0350	0.0400	0.0450	0.0500
	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

MCP corrections summary

- Bonferroni correction: reduces alpha so that the results reflect the odds that there are ***no false positives among any of the corrected significant tests*** (zero false positives)

- Split-half (or other selective procedures): reduce the MCP problem by ***restricting the number of tests based on independent data***

- Cluster-based permutation testing: restricts the number of tests by using ***clusters rather than samples as the relevant unit*** for which *P*-values are computed

- FDR correction: reduces alpha to ***restrict the expected number of false positives among all the corrected significant tests to a fixed proportion*** (q, usually .05)

MVPA experimental design & confounds

- In principle, very similar principles apply to experimental design for regular ERPs and for decoding analyses, BUT:

  - Decoding is very sensitive, so many confounds can drive above chance classification performance

  - Subject responses can drive classification performance

  - If you compare different contrasts (decoding analyses), each contrast has to have the same number of trials, same power etc.

  - Having from two different classes that are collected in different blocks or even sessions can drive classification performance

Lecture 1

- Multiple comparison correction
- MVPA experimental design: decoding-specific confounds
- Ways to map brain to brain, or brain to behavior

MVPA experimental design

- Pretty much the same as for regular EEG/MEG experiments
- BUT… the design *can* impact how you can/should analyze your data, which in turn should make you think about your design

EEG experimental design ⟷ **MVPA analytical approach**

Confounds

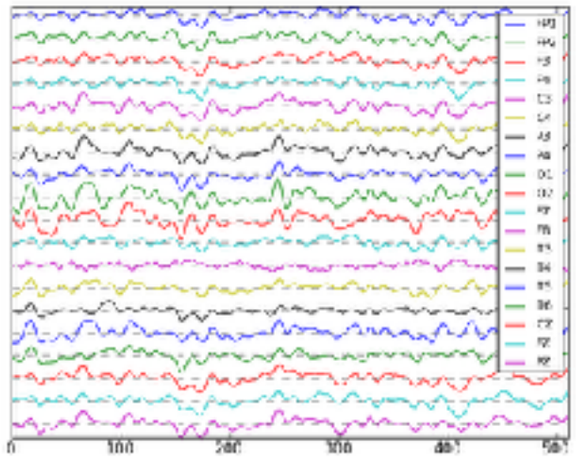
- Overfitting and related confounds can cause above chance decoding. This is why we have **k-fold cross-validation**.

But does this solve all problems!

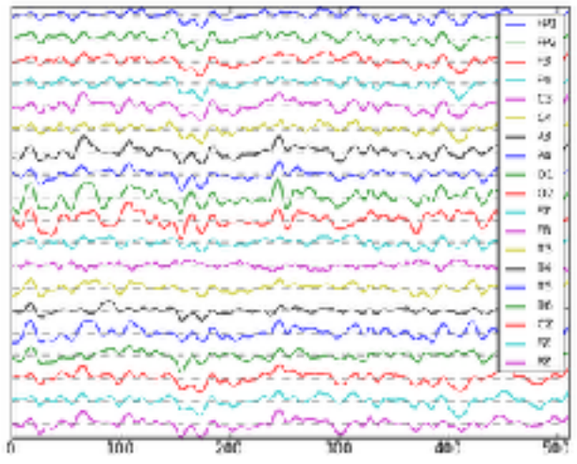
Confounds

- **k-fold cross-validation:** Works well for *event-related designs*, but does not protect from confounds intrinsic to your dataset, e.g. when you have a strongly *blocked* design, or if you have one condition/class in one session and the other condition/class in another session. Can also be problematic if you select trials contingent on *subject responses* and/or if you compare different decoding analyses based on unequal trial counts.

Session 1, task 1

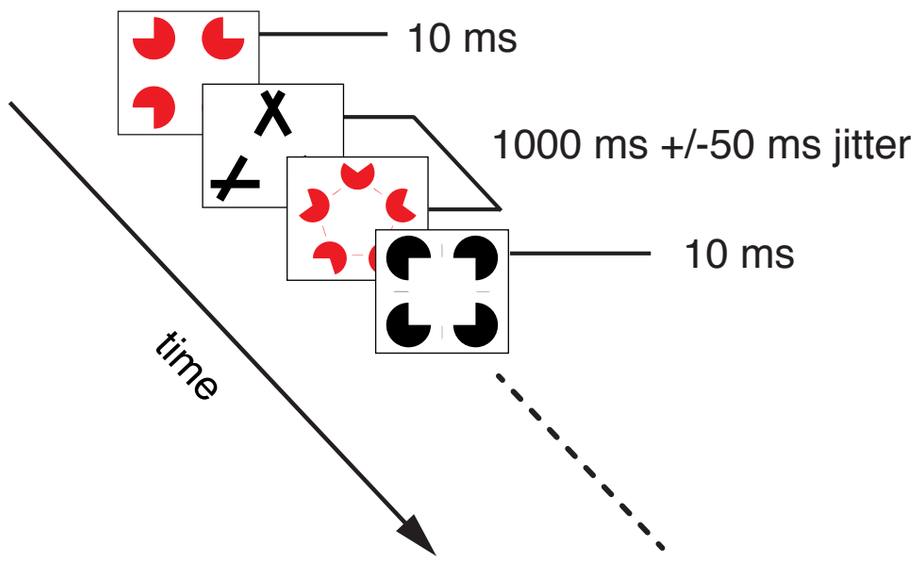


Session 2, task 2

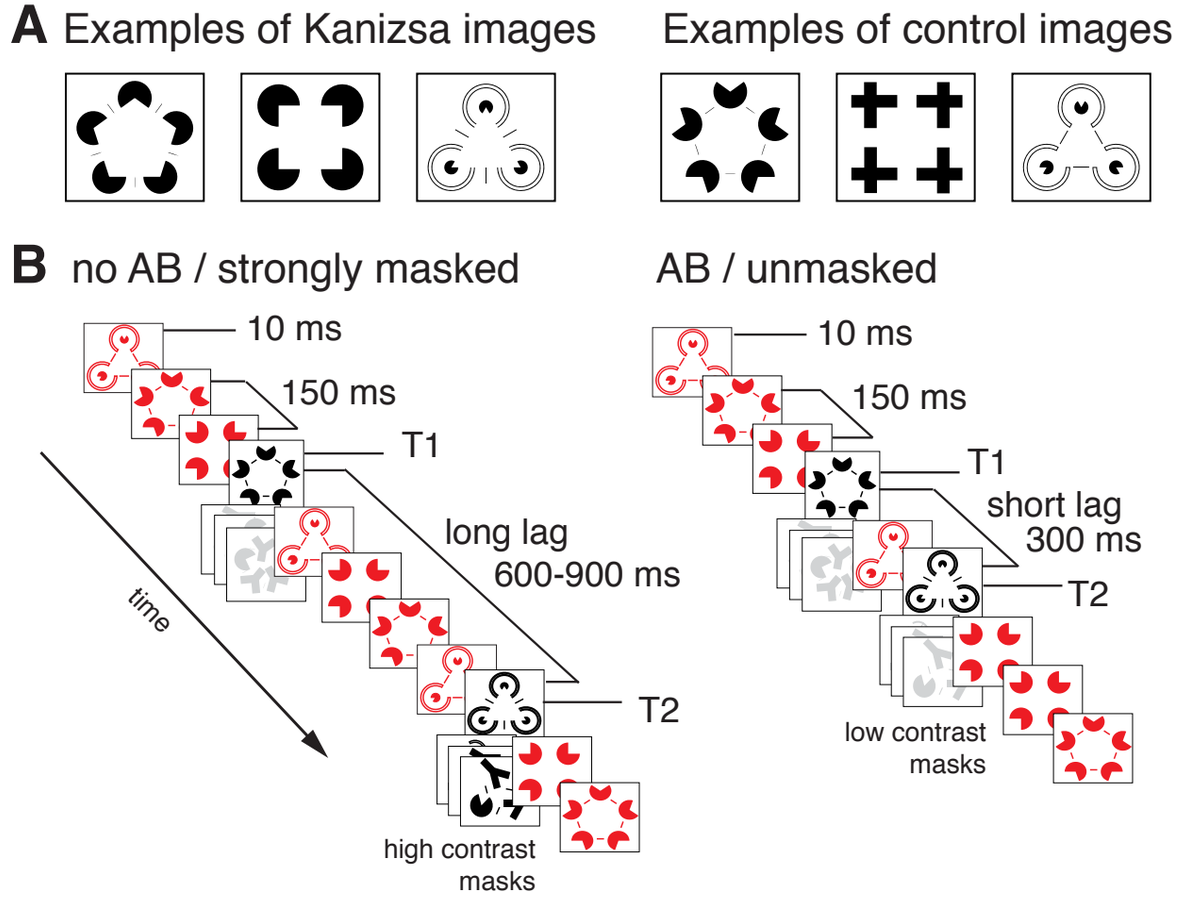


Confounds

- A **separate training** set to train the classifier solves many of these problems. Such a training set should be balanced, event related, and not contain responses that are relevant to your experimental conditions (often a 1-back task works well).



Training data



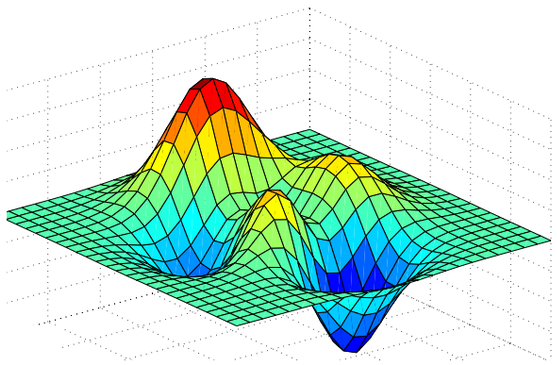
Testing data (experimental task)

Lecture 1

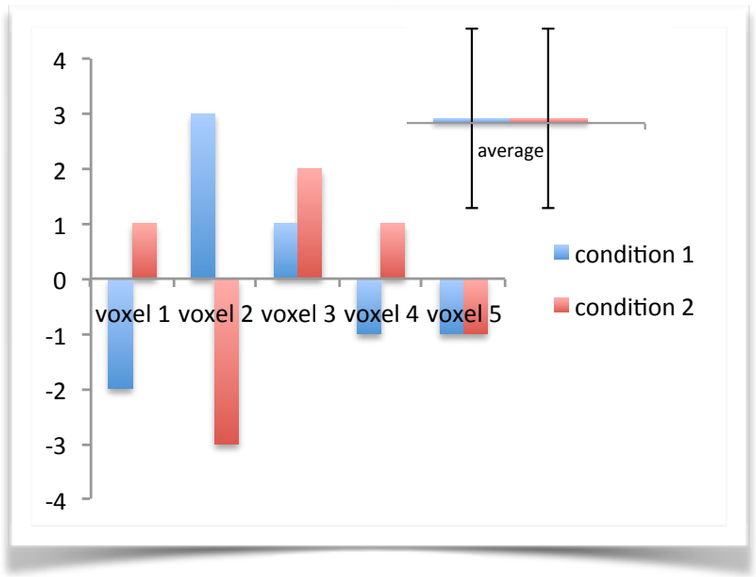
- Multiple comparison correction
- MVPA experimental design: decoding-specific confounds
- Ways to map brain to brain, or brain to behavior

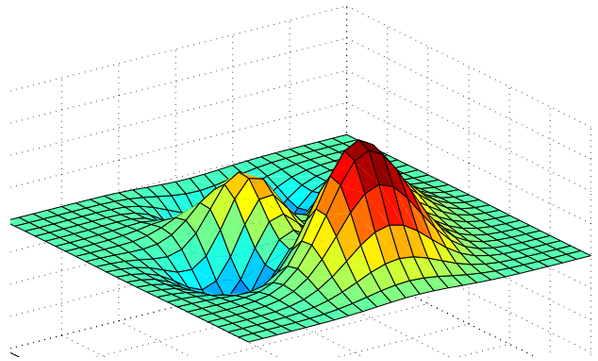
Representational Dissimilarity

Similarity: correlation



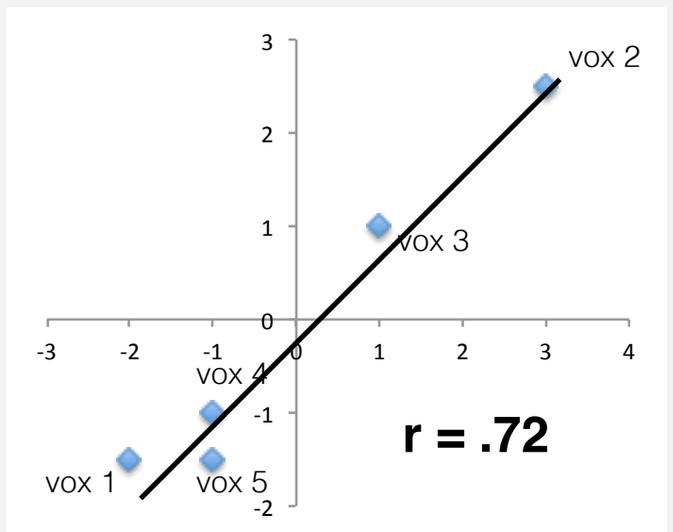
condition 1





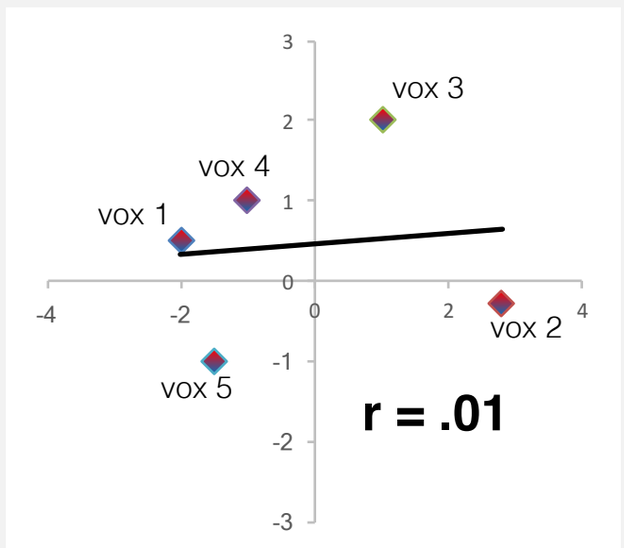
condition 2

condition 1 (trial 1)



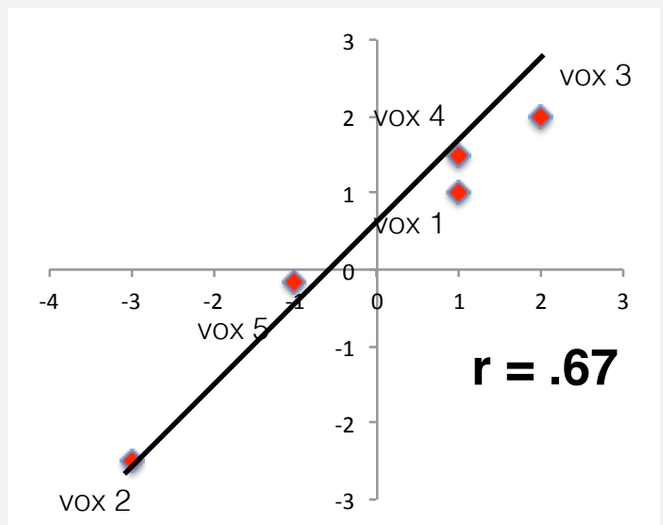
condition 1 (trial 2)

condition 1



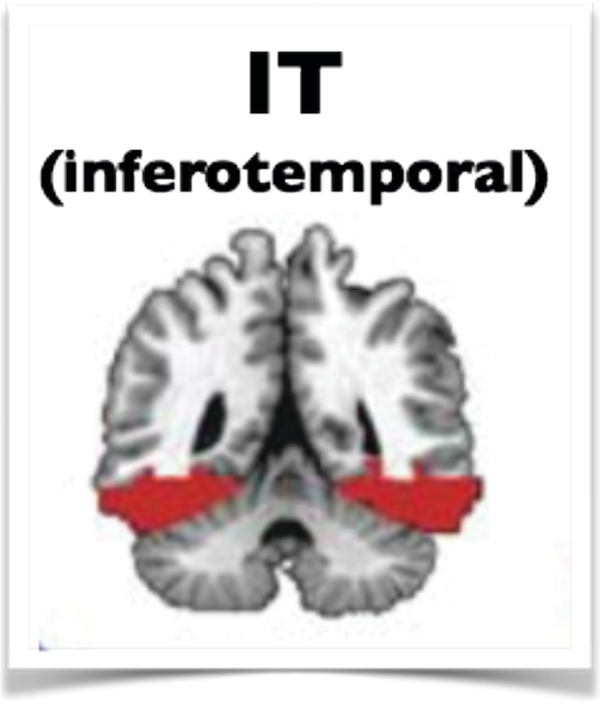
condition 2

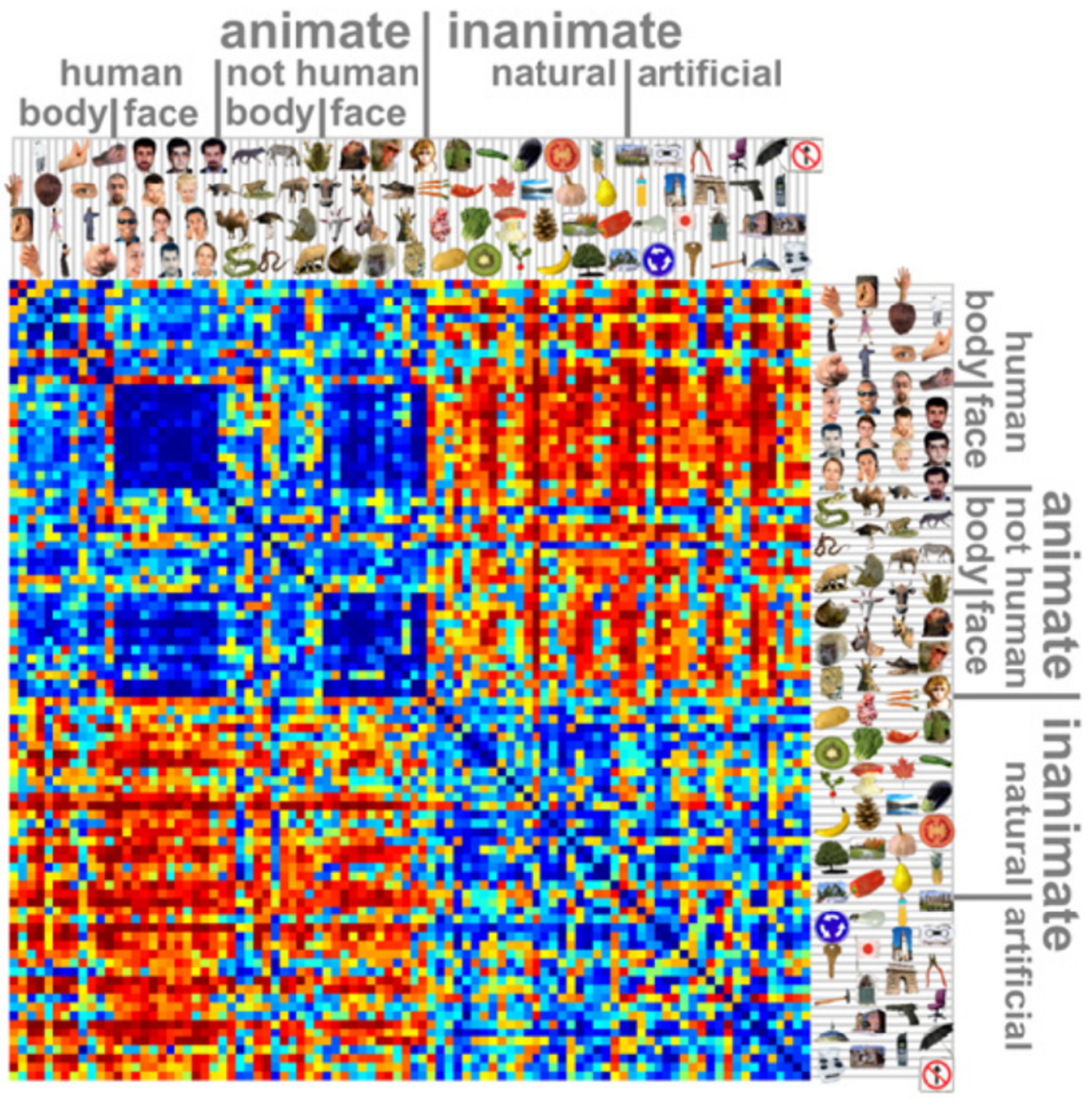
condition 2 (trial 1)

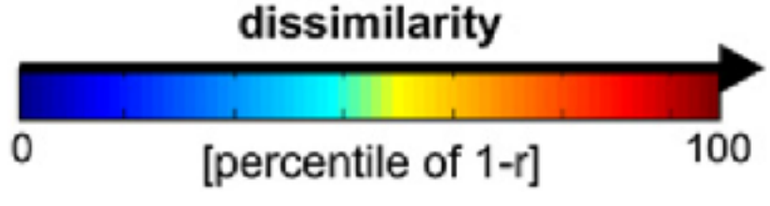


condition 2 (trial 2)

the Representational Dissimilarity Matrix
RDM = 1 - correlation

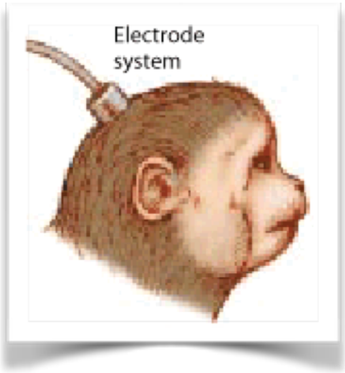






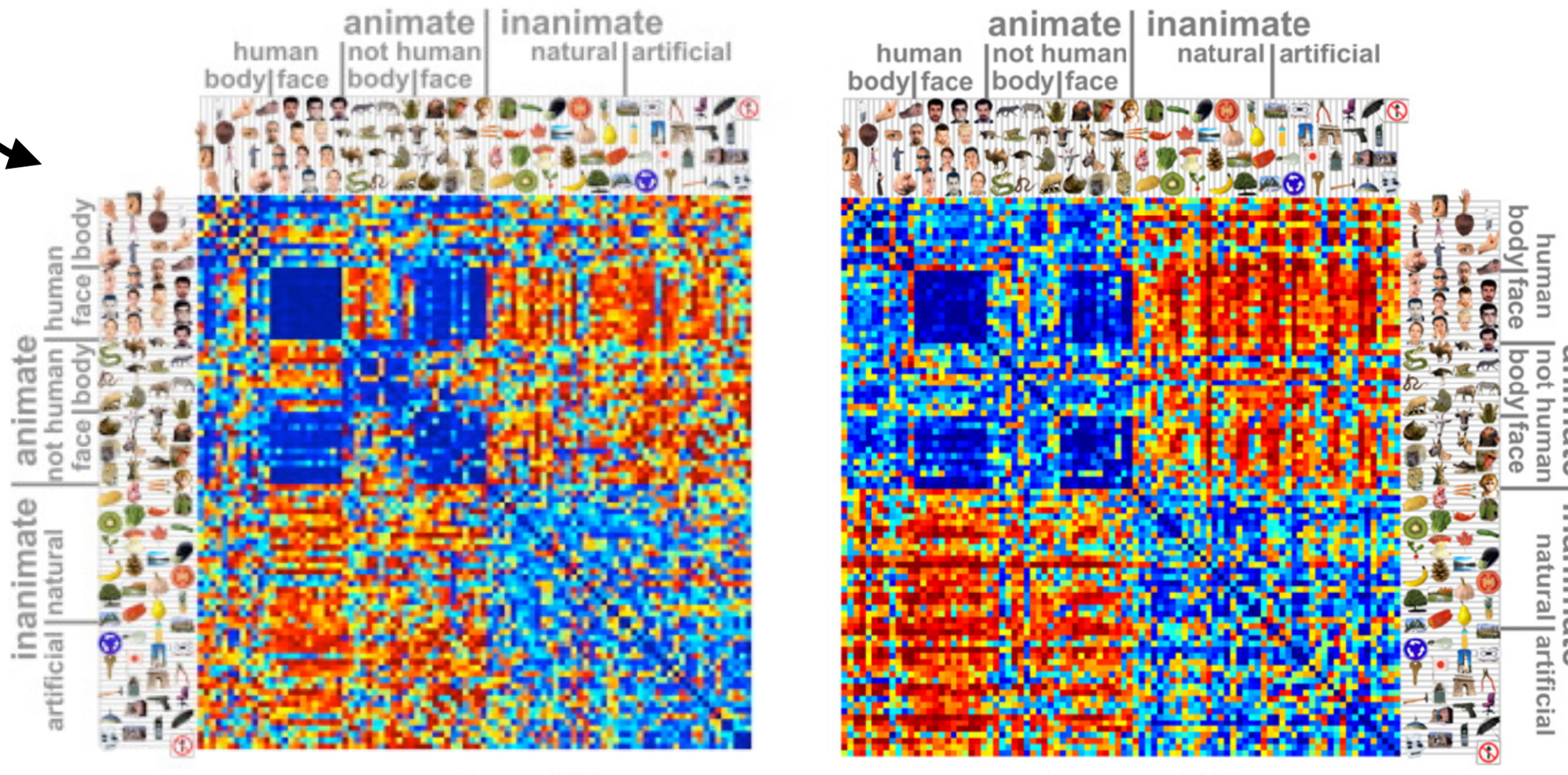
fMRI

the RDM reveals the "information structure"

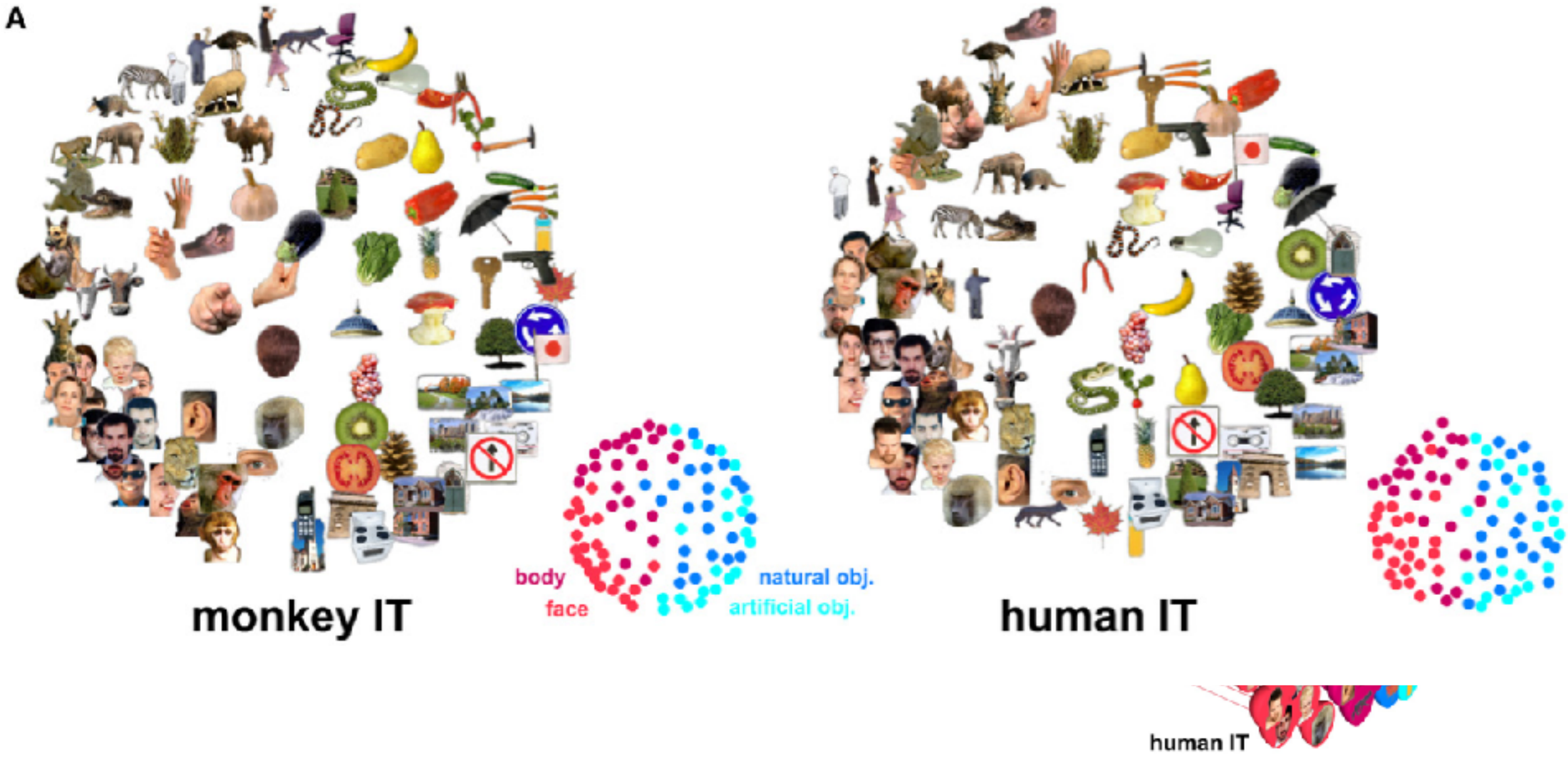


neurophysiology

fMRI



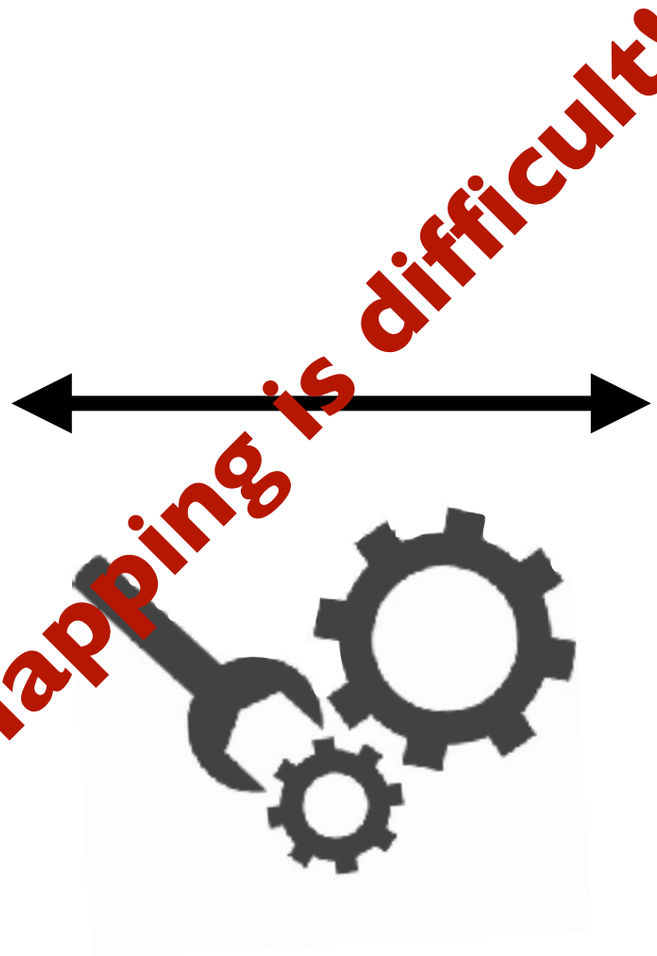
Mapping information structure between different measures and physical substrates

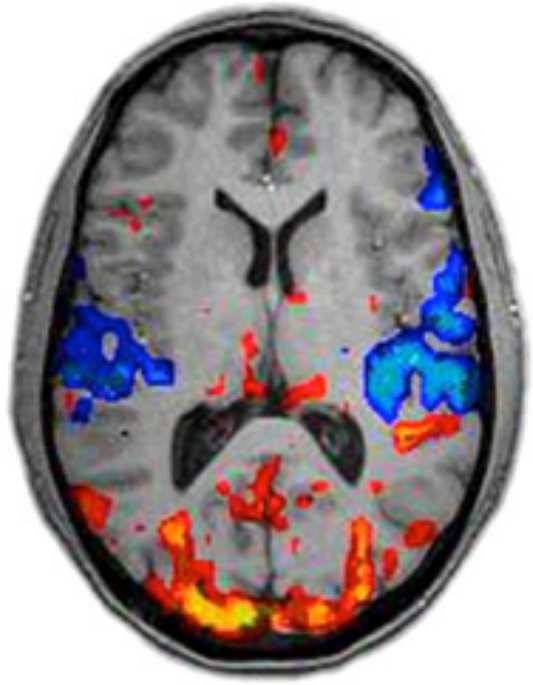


Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*, 60(6), 1126–1141.

Mapping fMRI to EEG

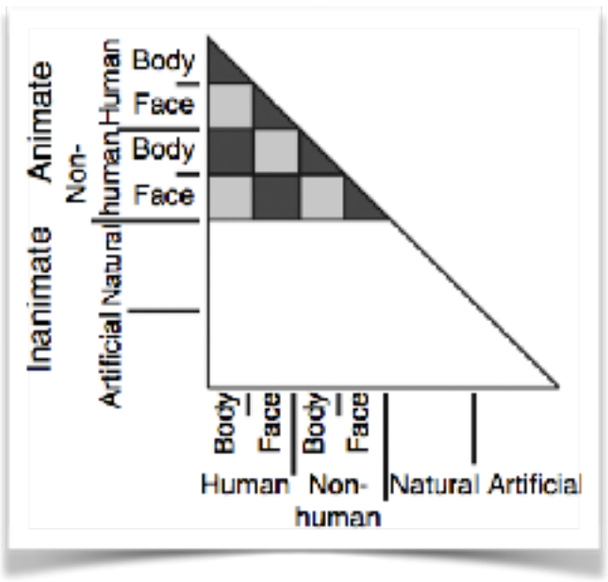
- Good **temporal** resolution
- Bad spatial resolution
- Polarity unrelated to excitation or inhibition (ambiguous), influenced by cortical folding
- Prominent inverse problem due to neuroanatomy

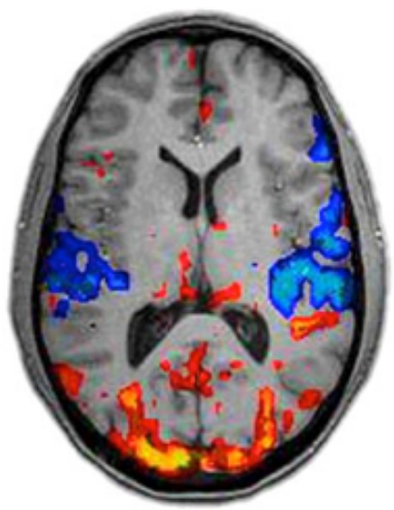


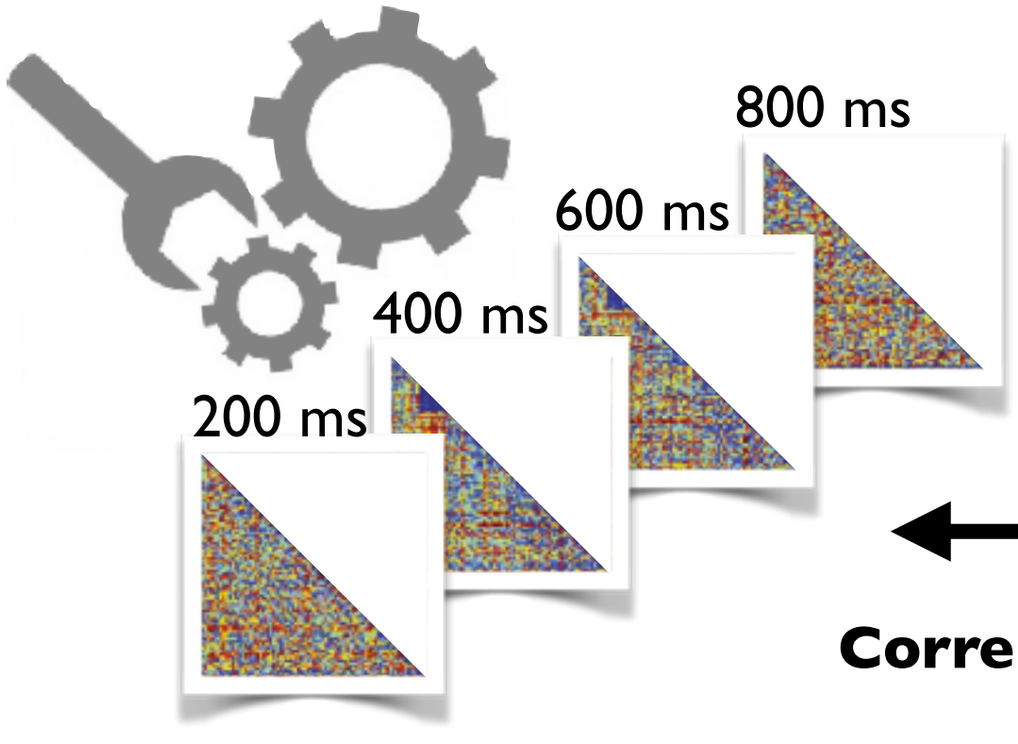


- Good **spatial** resolution
- Bad temporal resolution
- Sign of parameter estimate unrelated to excitation or inhibition (ambiguous)
- Minor inverse problem due to relationship BOLD-LFP

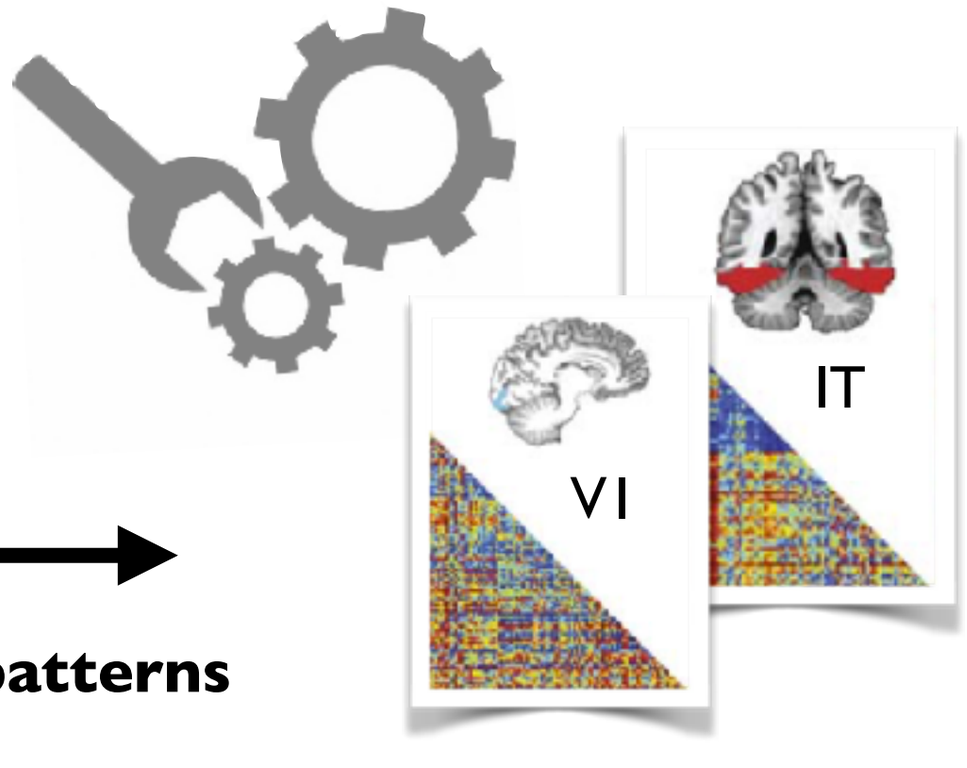
Solution: compute RDMs



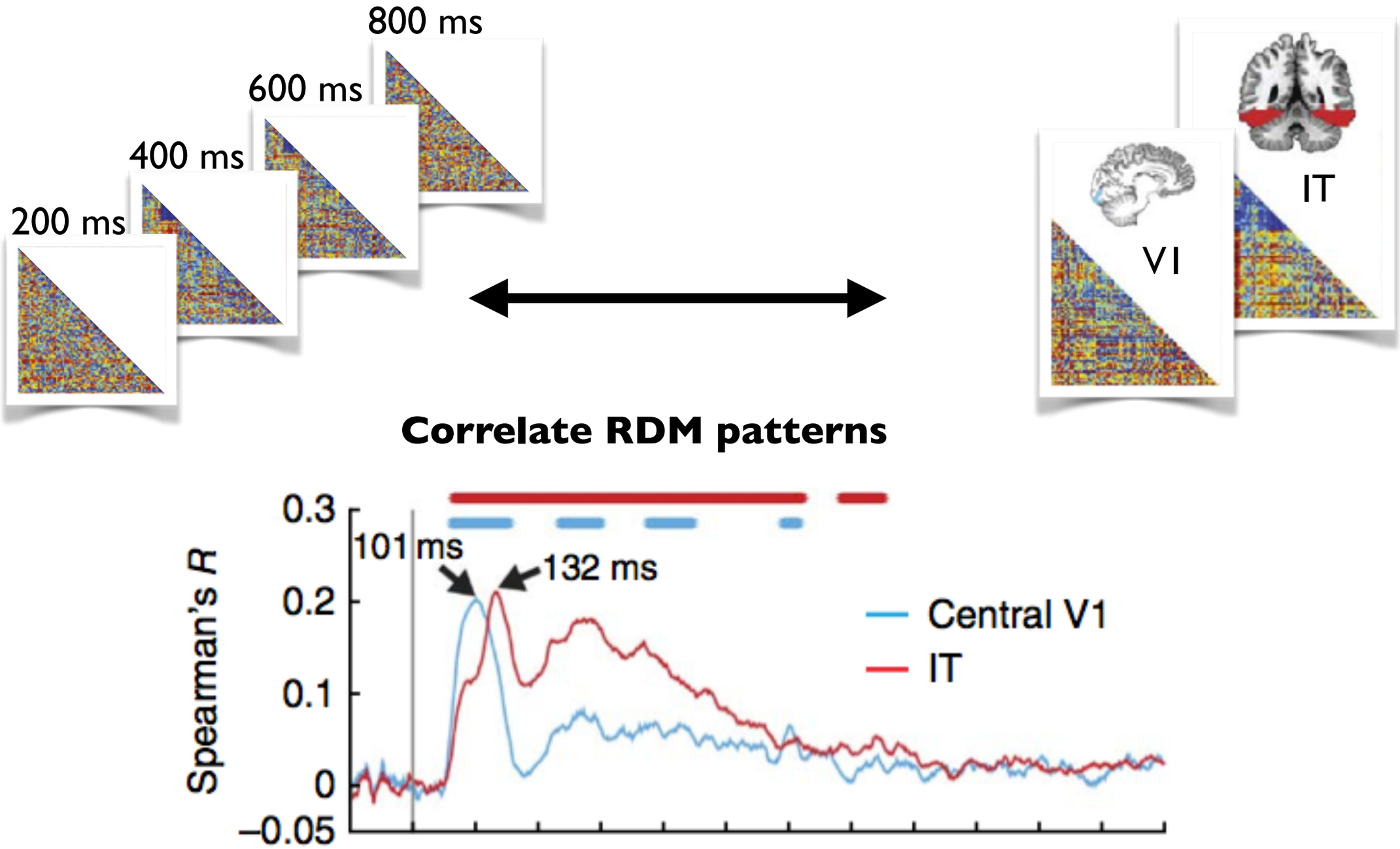




Correlate RDM patterns

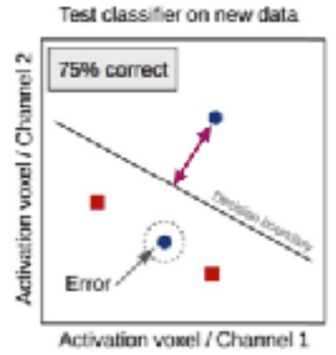


Matching up the spatial and the temporal domain between fMRI and EEG

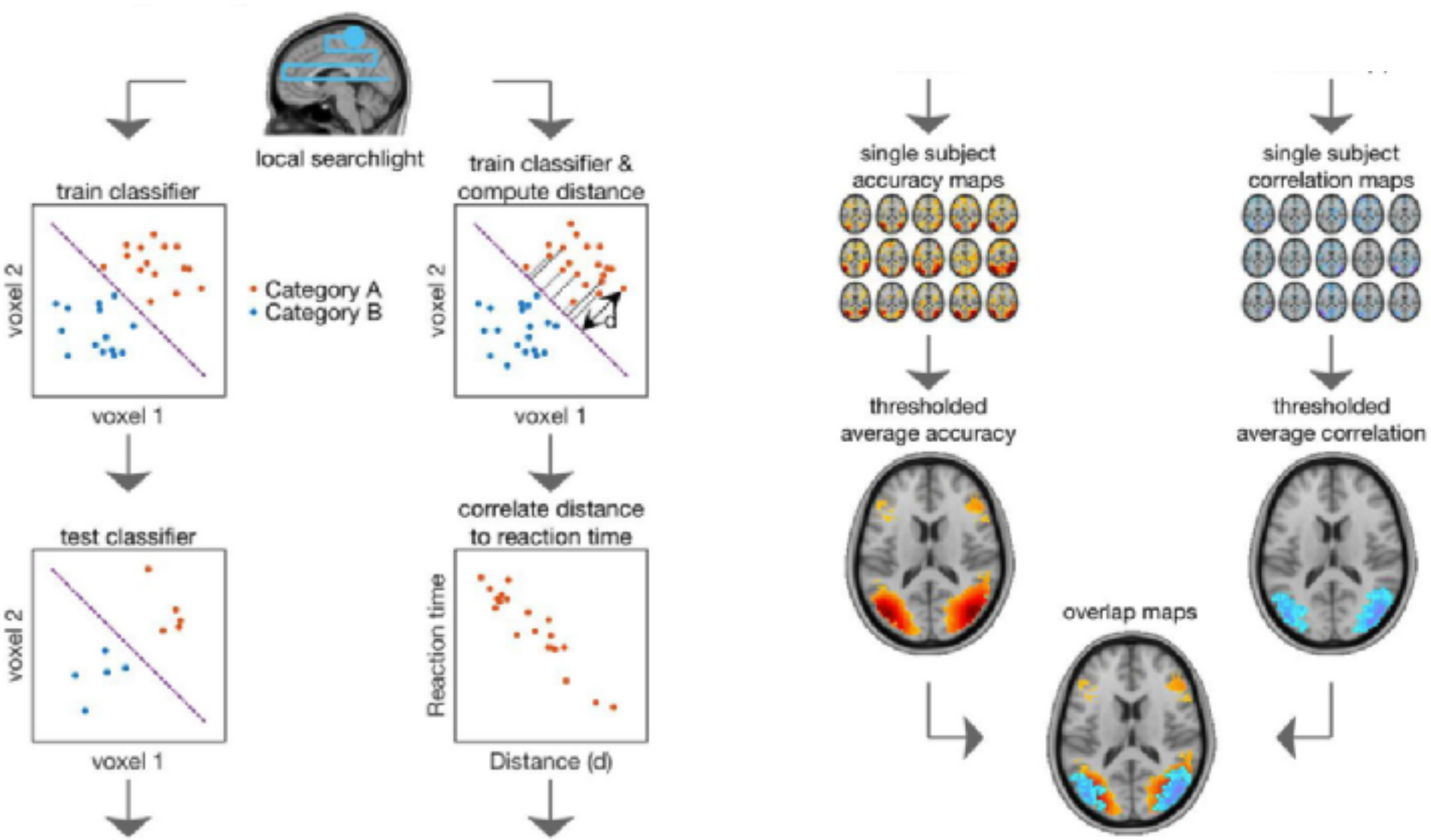


Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, 17(3), 455–462.

Using the distance from the decision boundary as a metric



distance from decision boundary



Grootswagers, T., Cichy, R. M., & Carlson, T. A. (2018). Finding decodable information that can be read out in behaviour. *Neuroimage*, 179, 252–262.

Questions?

Course overview

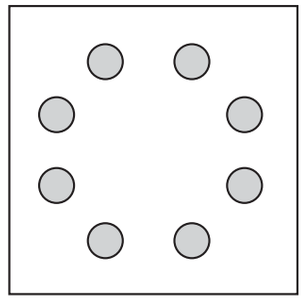
- **Day 1 (introduction)**
  - lecture 1: History and electrophysiological basis of EEG
  - lecture 2: Backward decoding models in MVPA: concepts and analytical approach
  - lecture 3: Advantages of MVPA, the temporal generalization method
  - lecture 4: Architecture of the ADAM toolbox and explaining the experiment of the practical
  - *Afternoon: practical*
- **Day 2 (advanced)**
  - lecture 1: MVPA experimental design and various ways to map brain to brain or to behavior
  - lecture 2: Forward encoding models in MVPA: how do they work, concepts and analytical approach
  - *Afternoon: practical, analyze your own data and/or a supplied dataset*

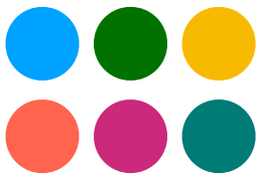
Lecture 2

- Forward encoding models (versus decoding)
- Brain Computer Interfaces (BCIs)

Backward decoding versus Forward encoding models

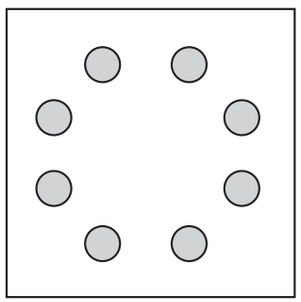
- Backward decoding models work with *discrete* stimulus classes (e.g. object categories, but classes that are continuous can also be treated as discrete).

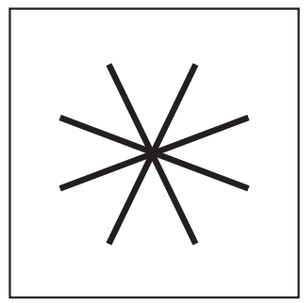




- The model *cannot* make new predictions for stimuli that were never used to train the model

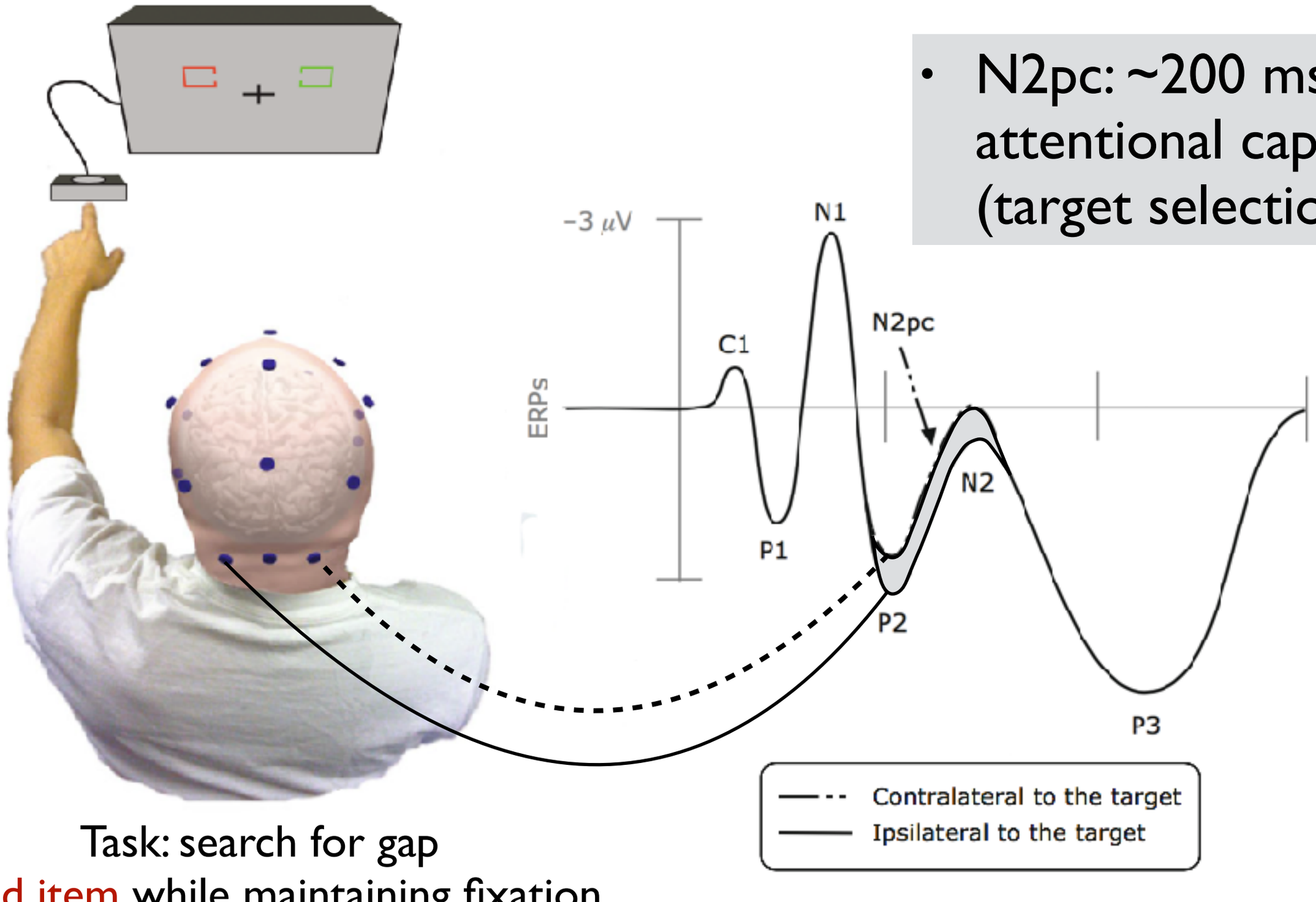
- Forward encoding models only make sense when using *continuous* stimulus classes (e.g. position on a circle, orientation of a bar, color etc)





- Allow you to make *new predictions* about cortical responses for stimuli that were never used to create the model

Attentional selection: N2pc

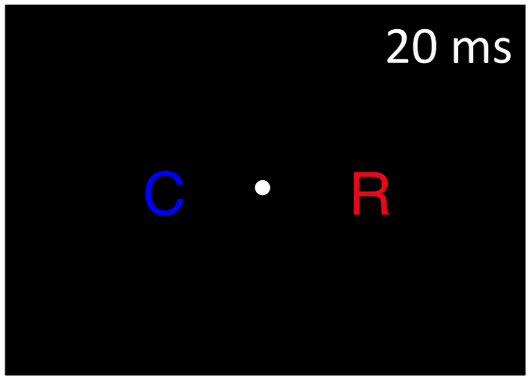


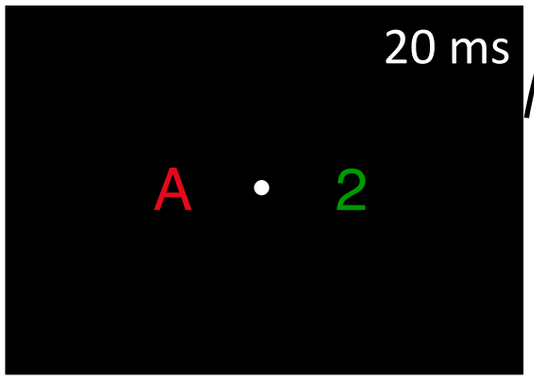
- N2pc: ~200 ms attentional capture (target selection)

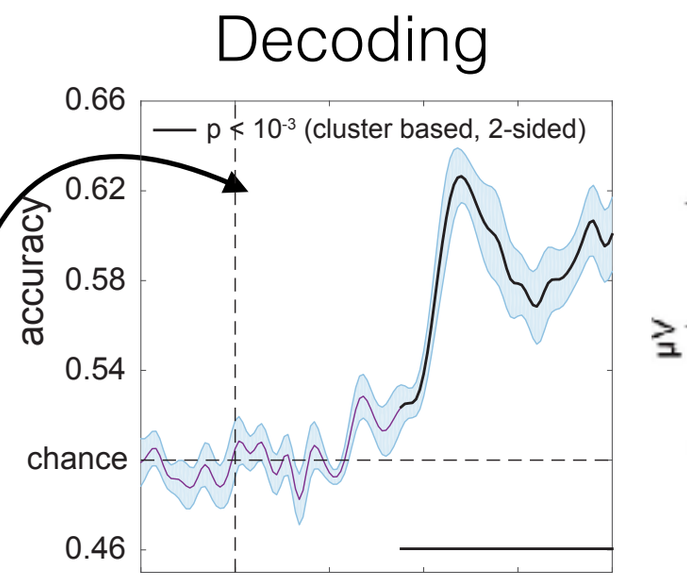
Task: search for gap
in red item while maintaining fixation

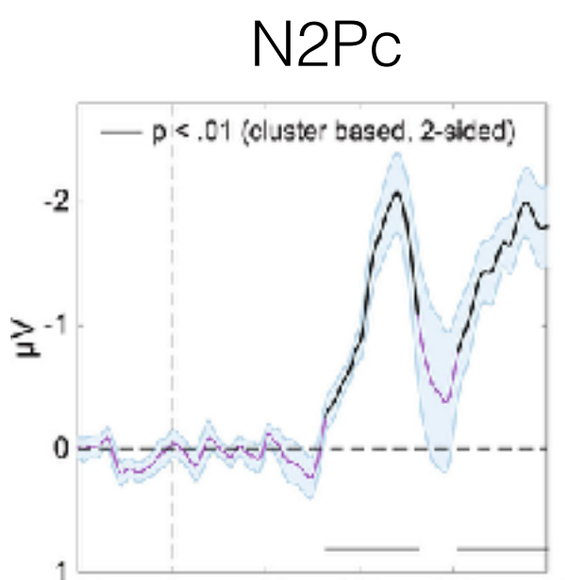
MVPA:classify where the target is

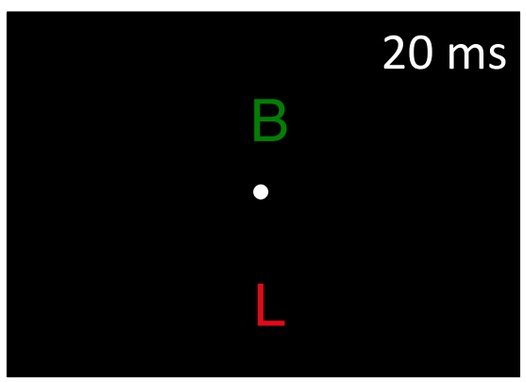
Task: detect whether red item is digit or letter

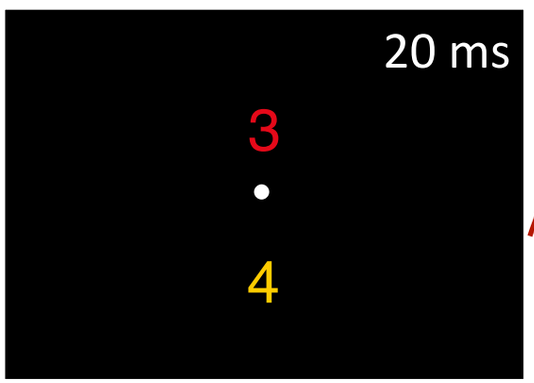


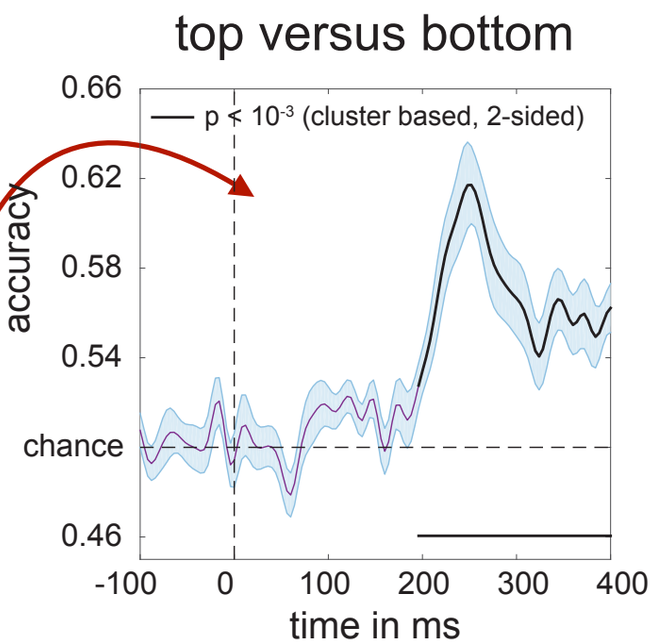








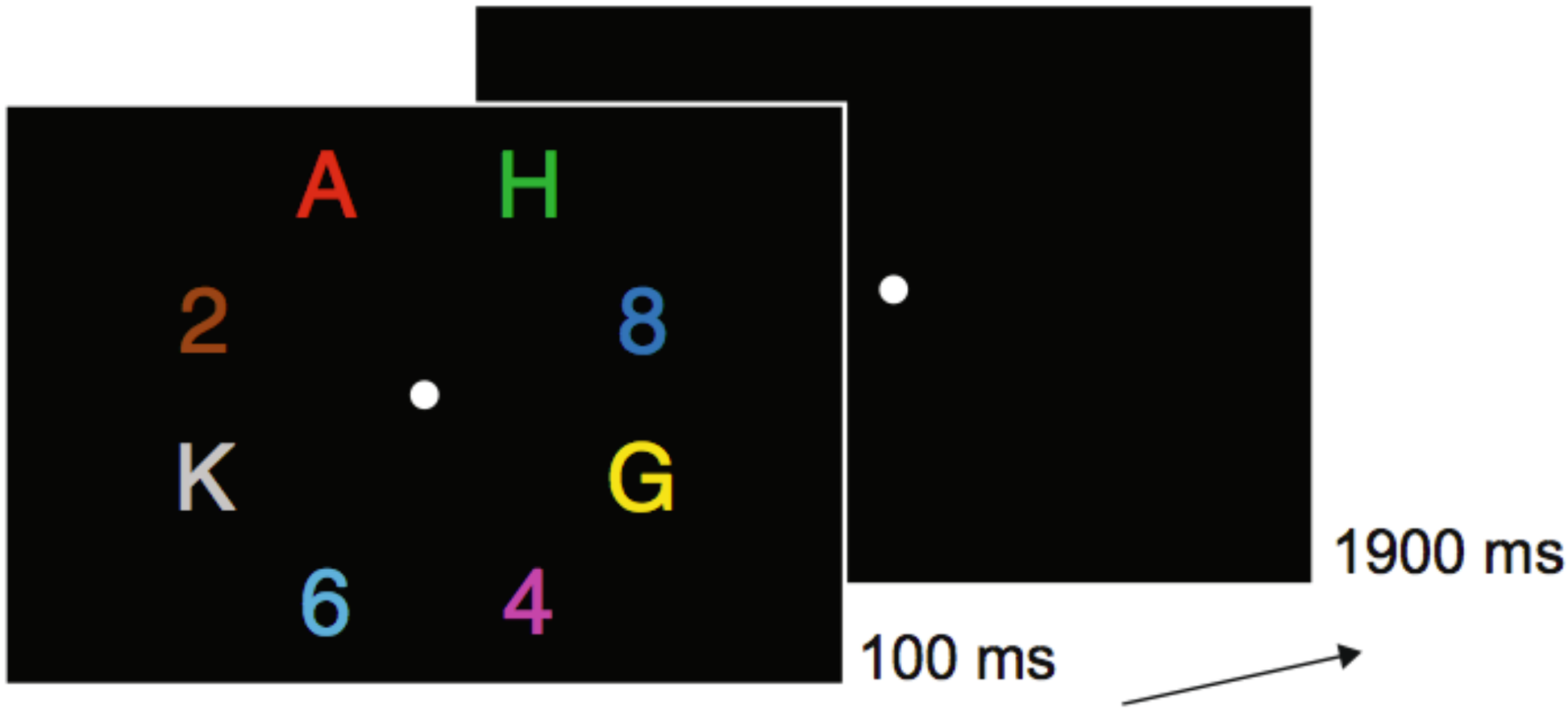




Also for
top vs bottom!

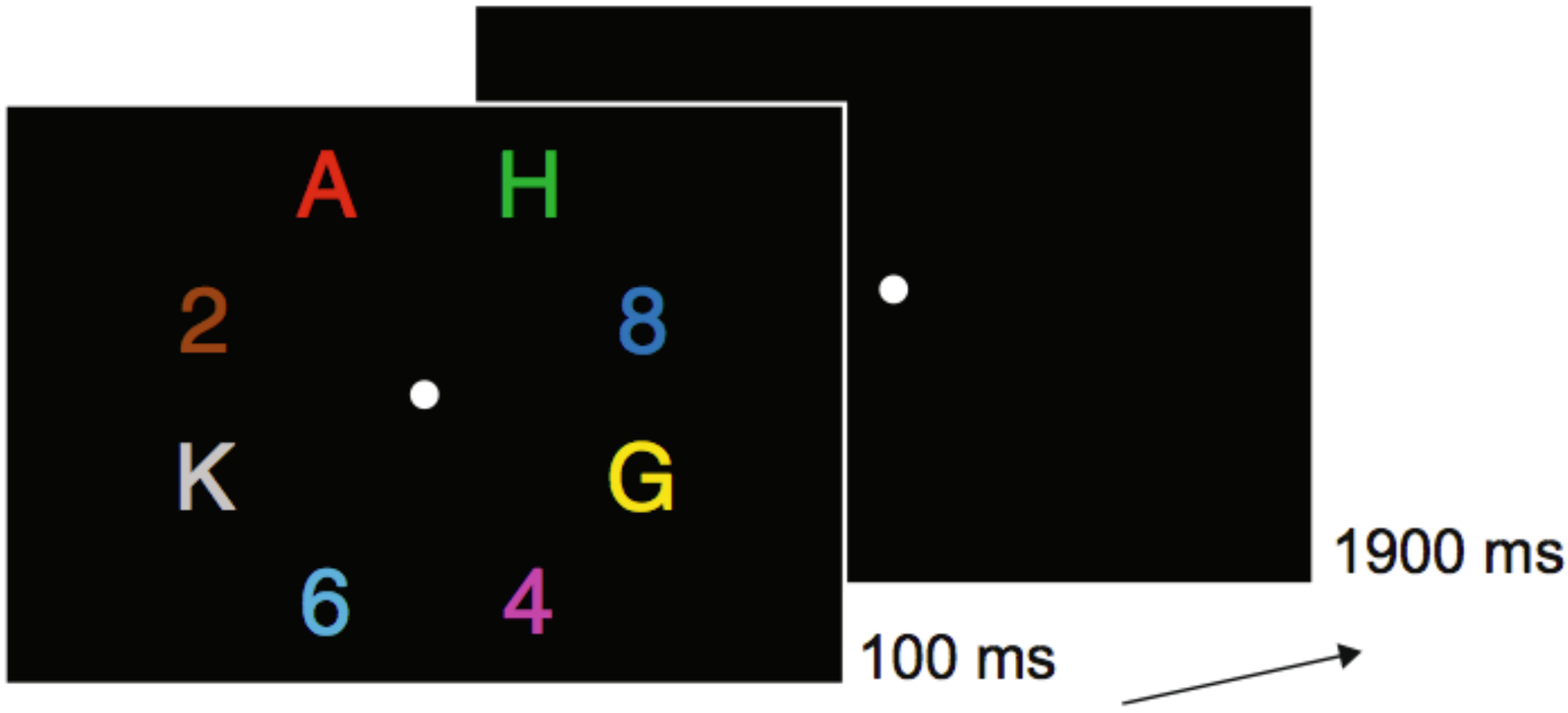
MVPA extracts ***any*** pattern from the data (does not have to be lateralized)

Experiment 2:
attention at a finer resolution!

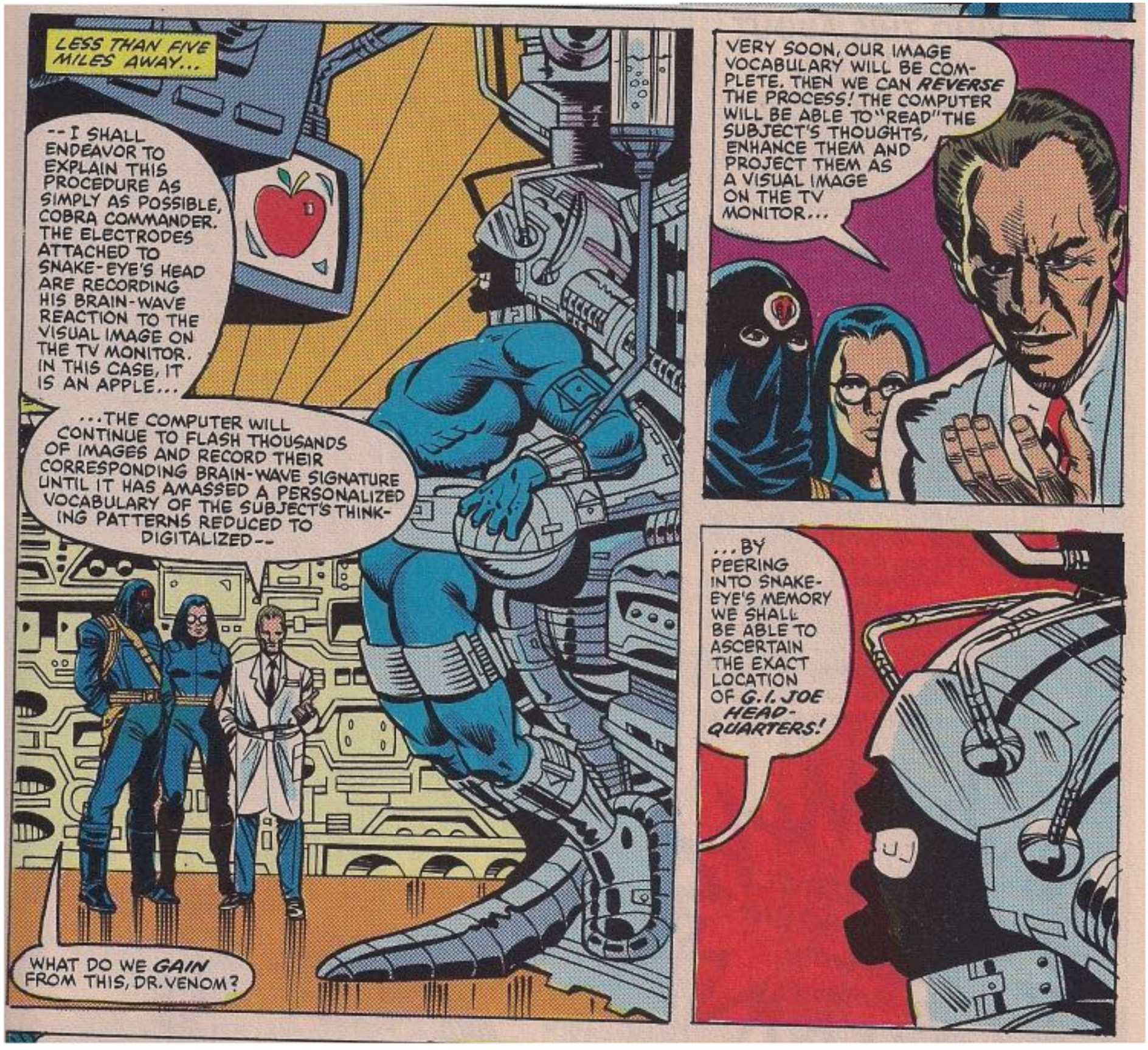


report the red item

Or for every part of the visual field!



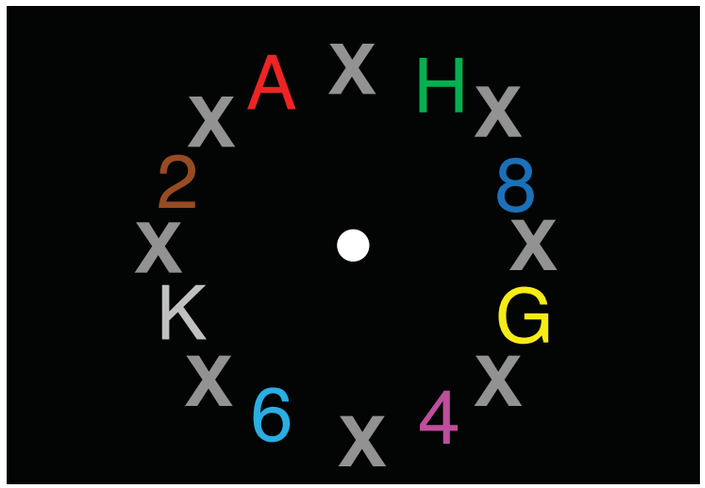
From a 1983 comic book



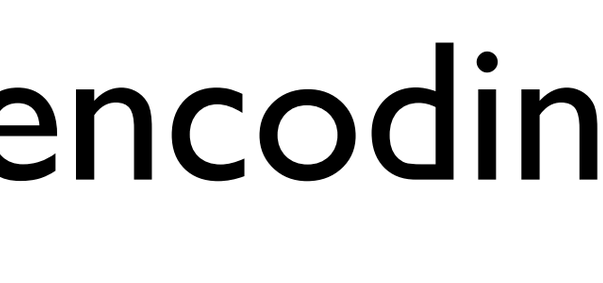
Forward Encoding Models: FEMs

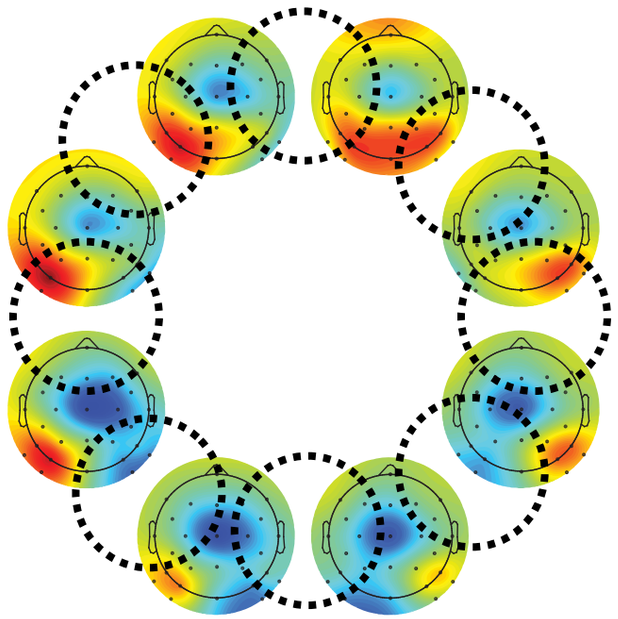
Make new predictions for stimuli that were never shown!

Forward encoding models



Forward encoding models (continuous)

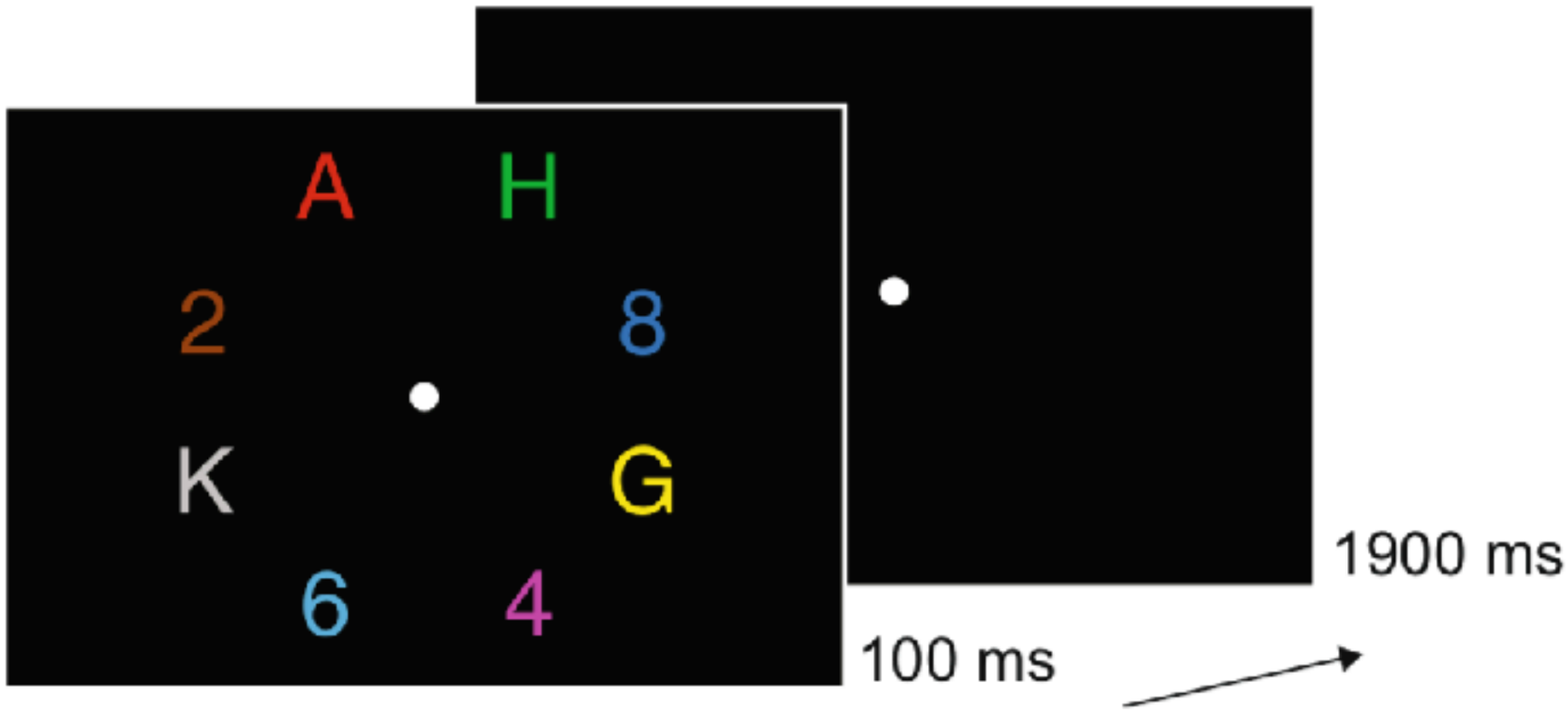




Continuous relationship between stimulus space and neural data

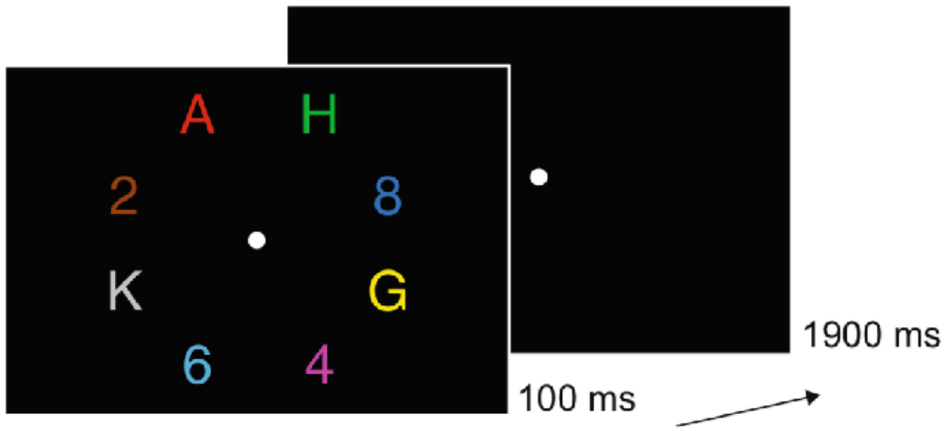
Predict neural patterns for stimuli that the model was never trained on!

Experiment 2:

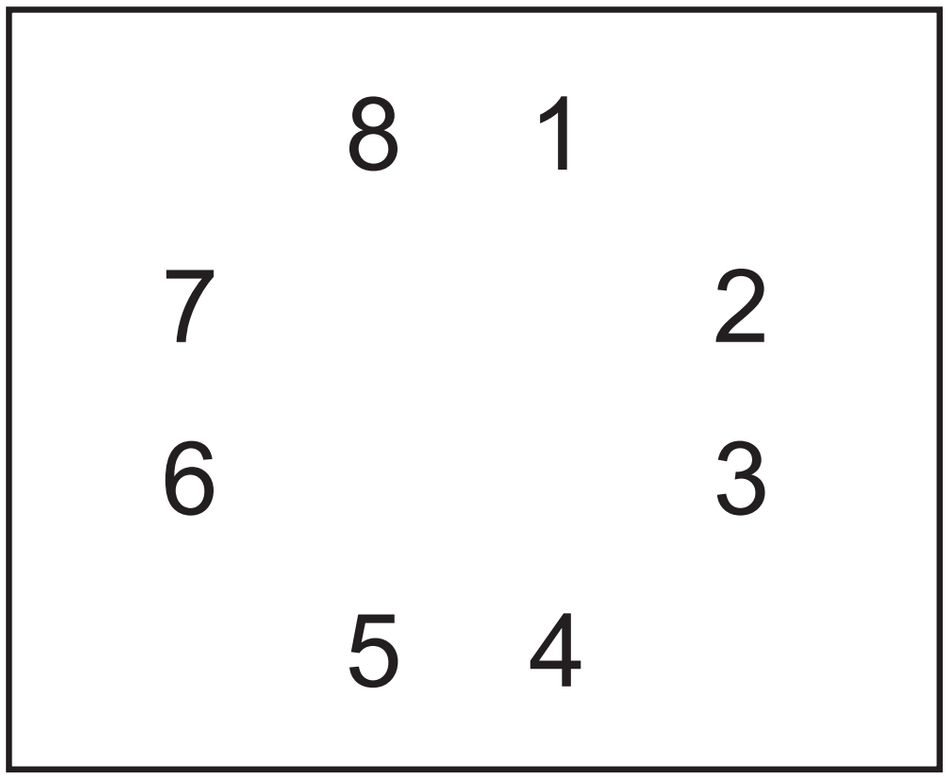


report the red item

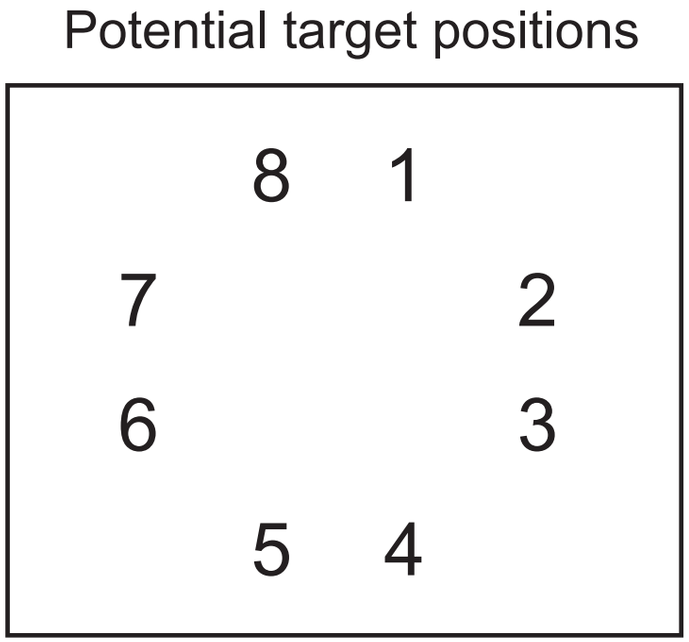
Fahrenfort, J. J., Grubert, A., Olivers, C. N. L., & Eimer, M. (2017). Multivariate EEG analyses support high-resolution tracking of feature-based attentional selection. *Scientific Reports*, 7(1), 1886.

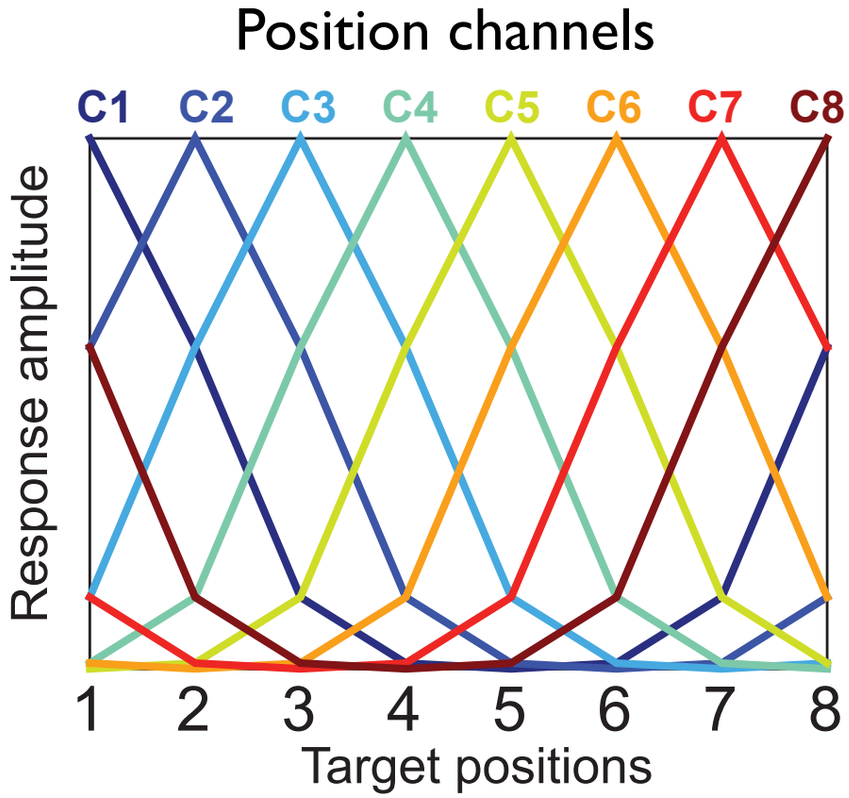


Potential target positions



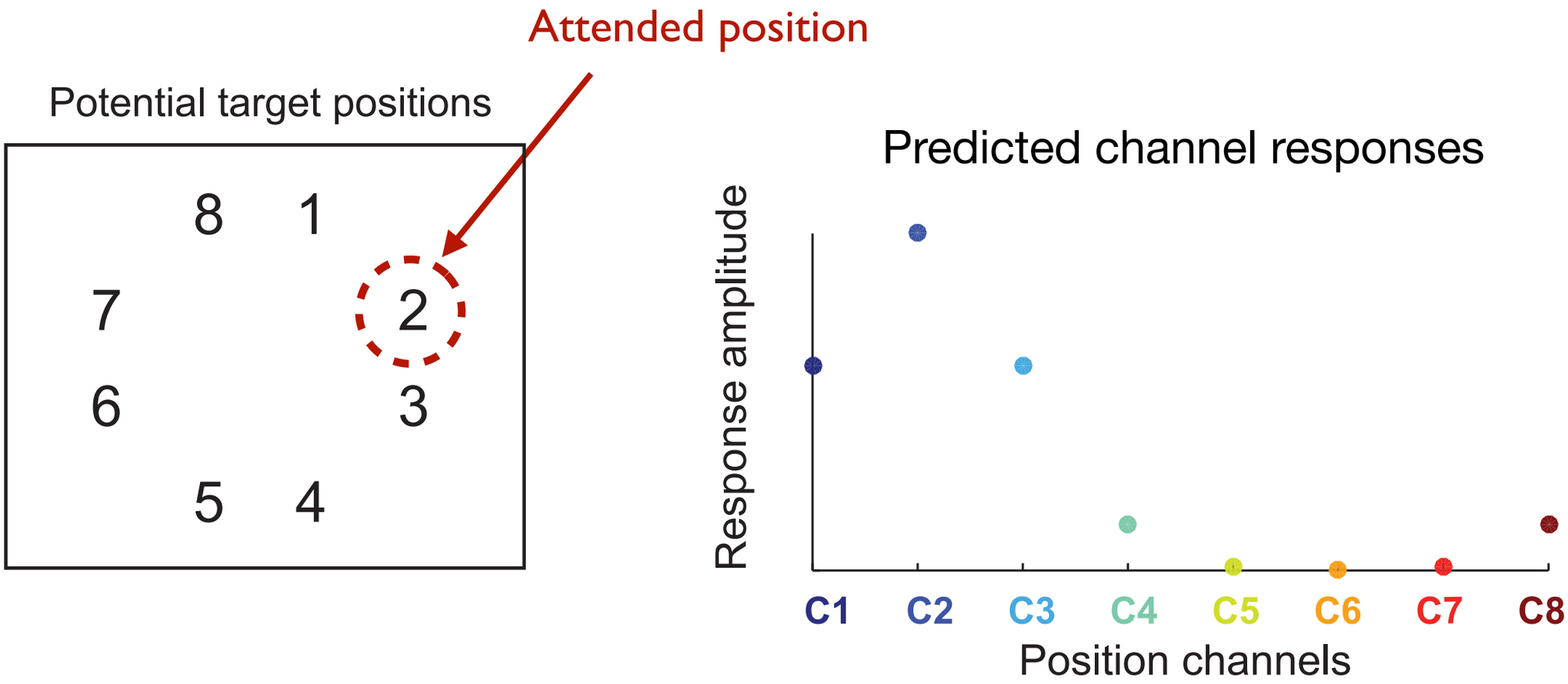
Create eight hypothetical position channels





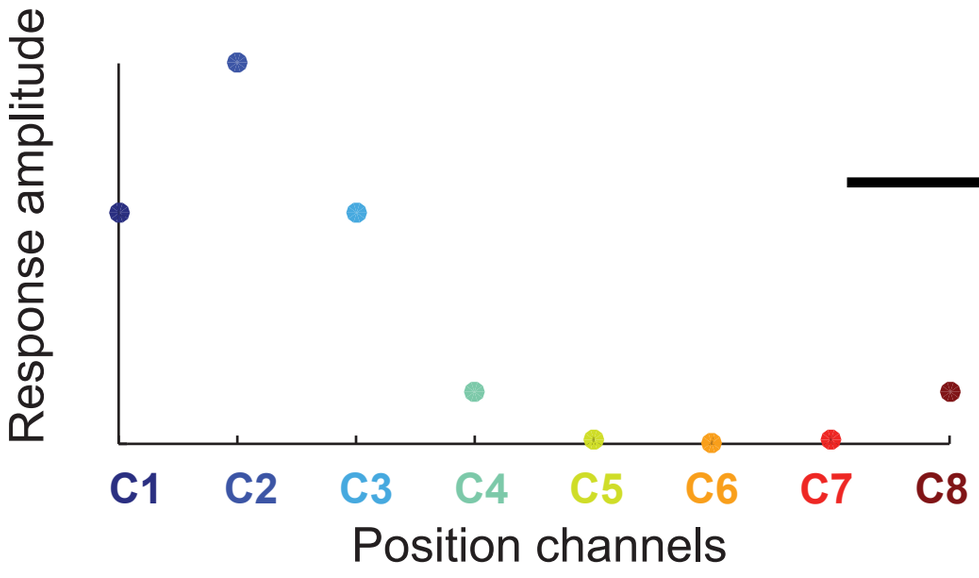
The "position channels" specify the hypothesized relationship between attended position and EEG response amplitudes

Model prediction for each attended position



Training phase

Predicted channel responses



Estimate channel weights

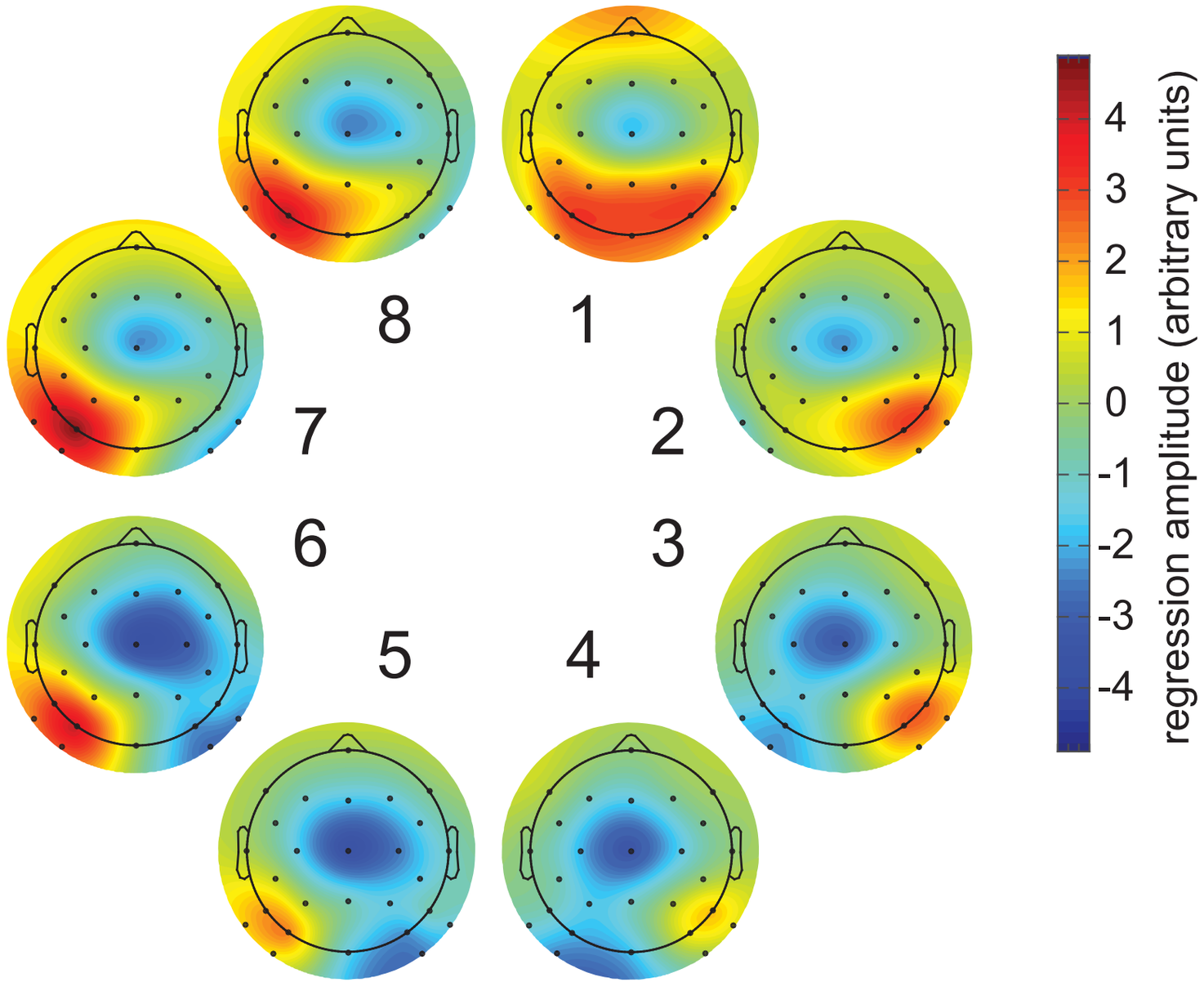
Weights

Observed EEG patterns in training set

Estimate ***weights*** (electrodes × channels)

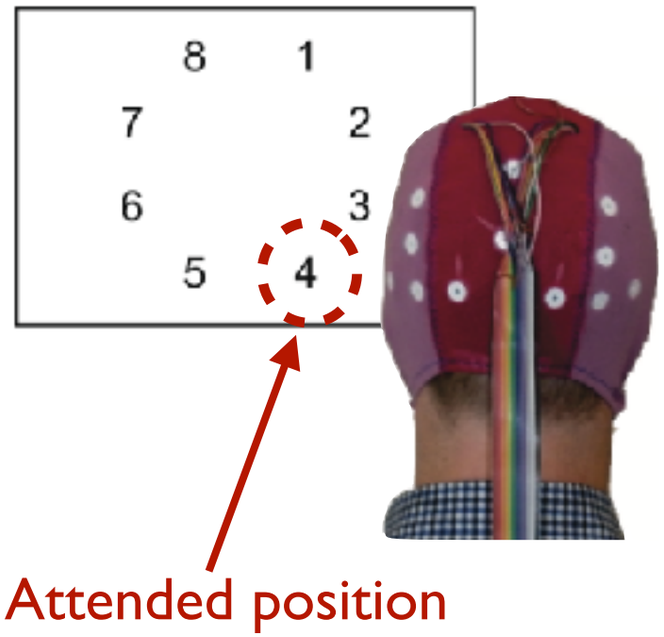
Weights characterize the mapping from channel to electrode space

Channel weights from training phase



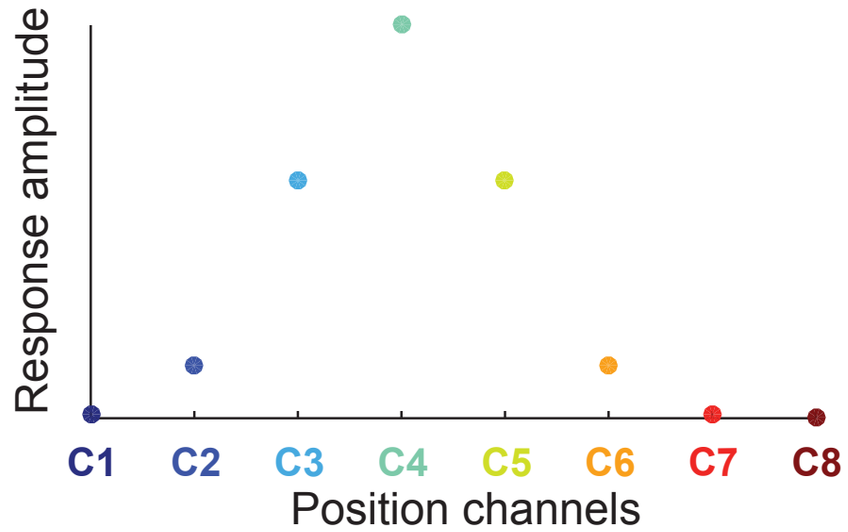
Testing phase: obtain Channel Tuning Functions (CTFs)

Combine observed EEG patterns from testing set



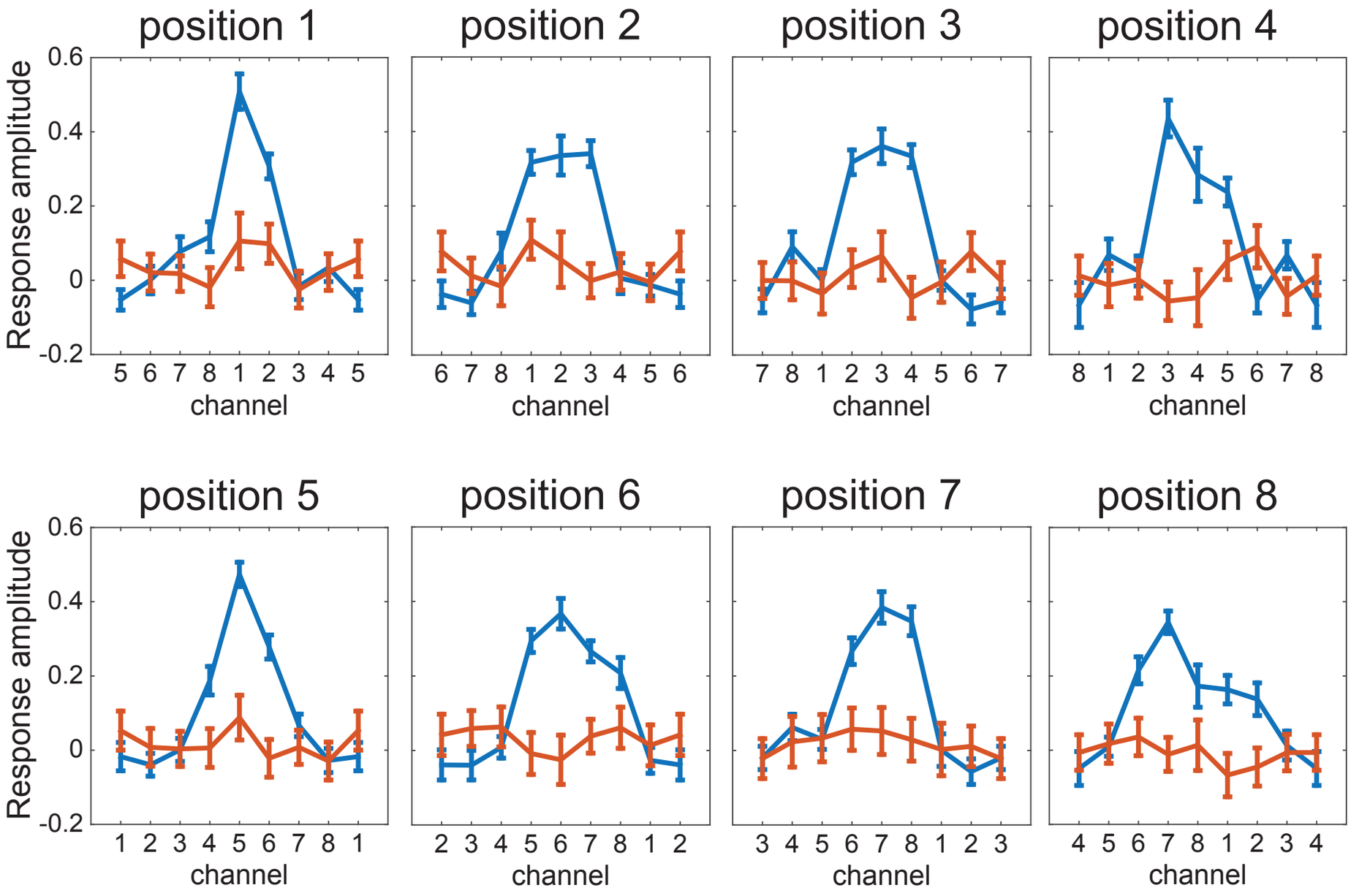
with channel weights to estimate channel responses

Channel Tuning Function (CTF)

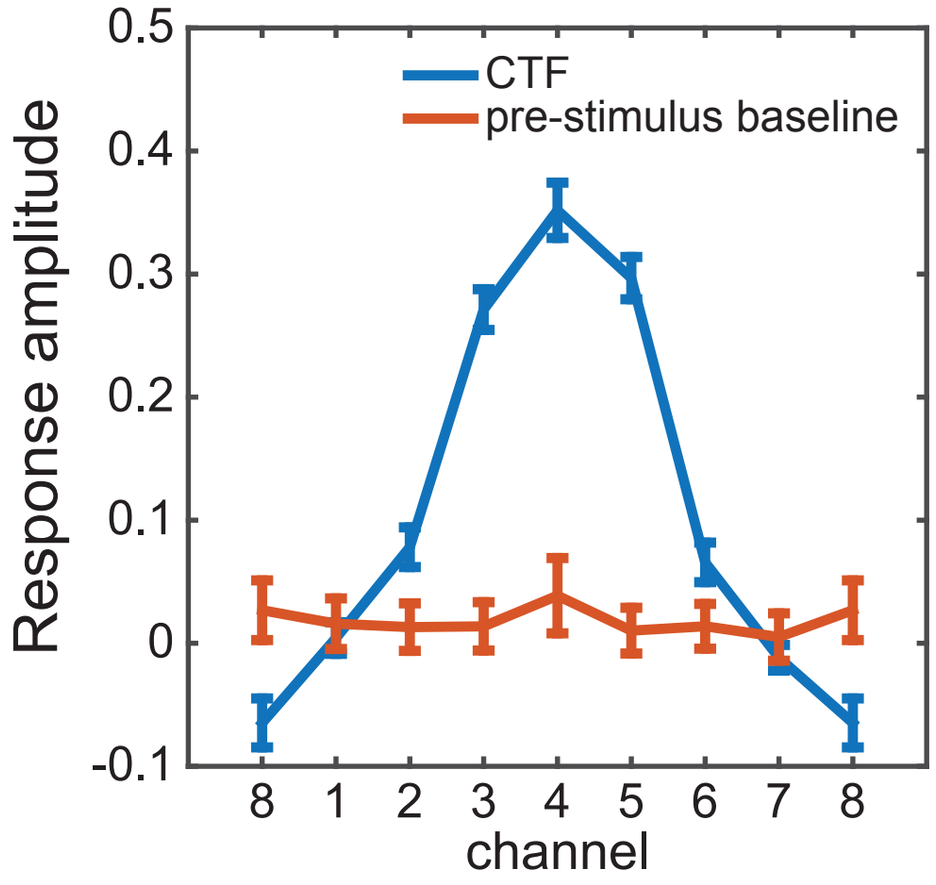


Estimate CTF for every for every attended position

CTF for each attended position

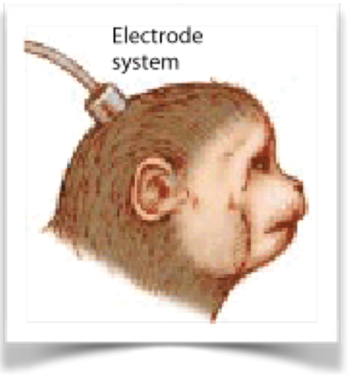


Average CTFs across positions to obtain single CTF

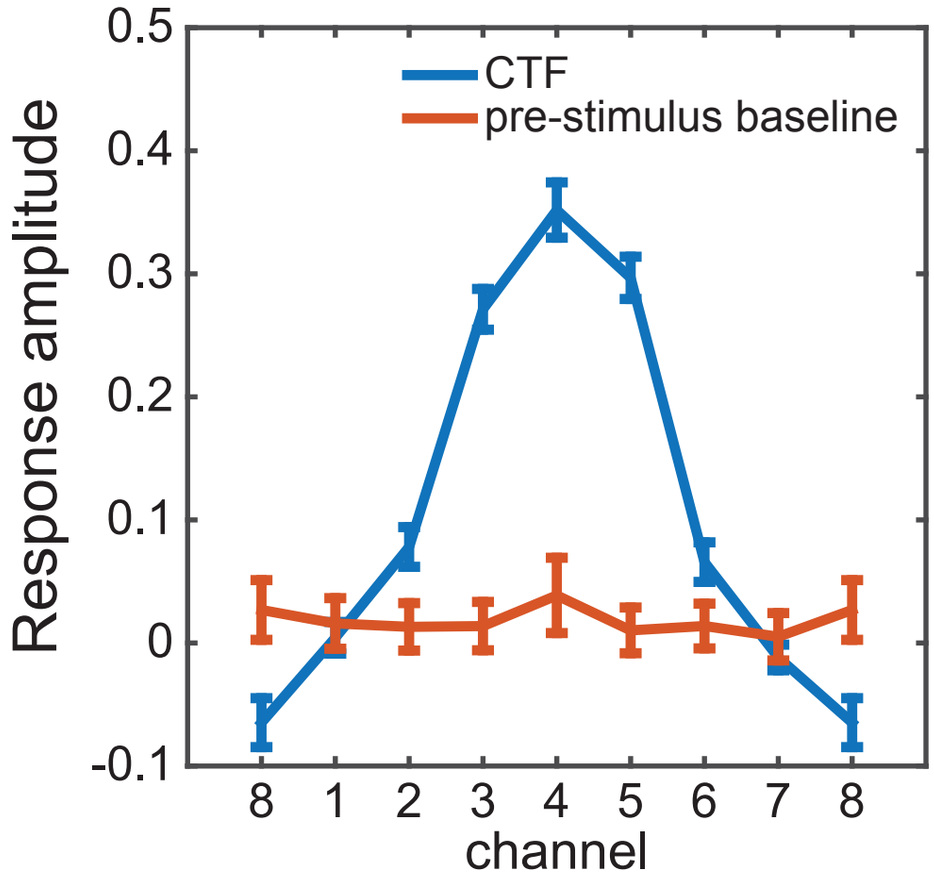


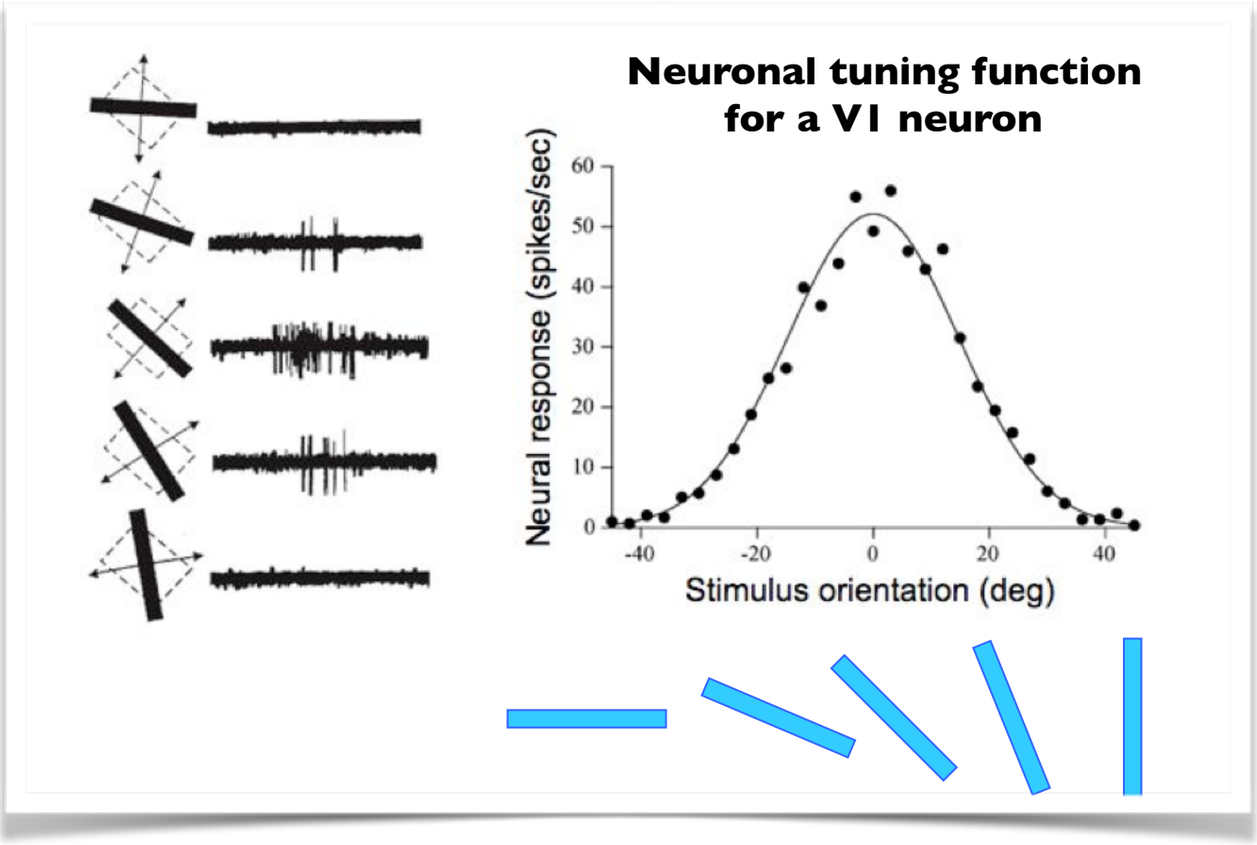
CTF: a continuous relationship between attended position and neural response patterns

Reminiscent of neuronal receptive field tuning functions

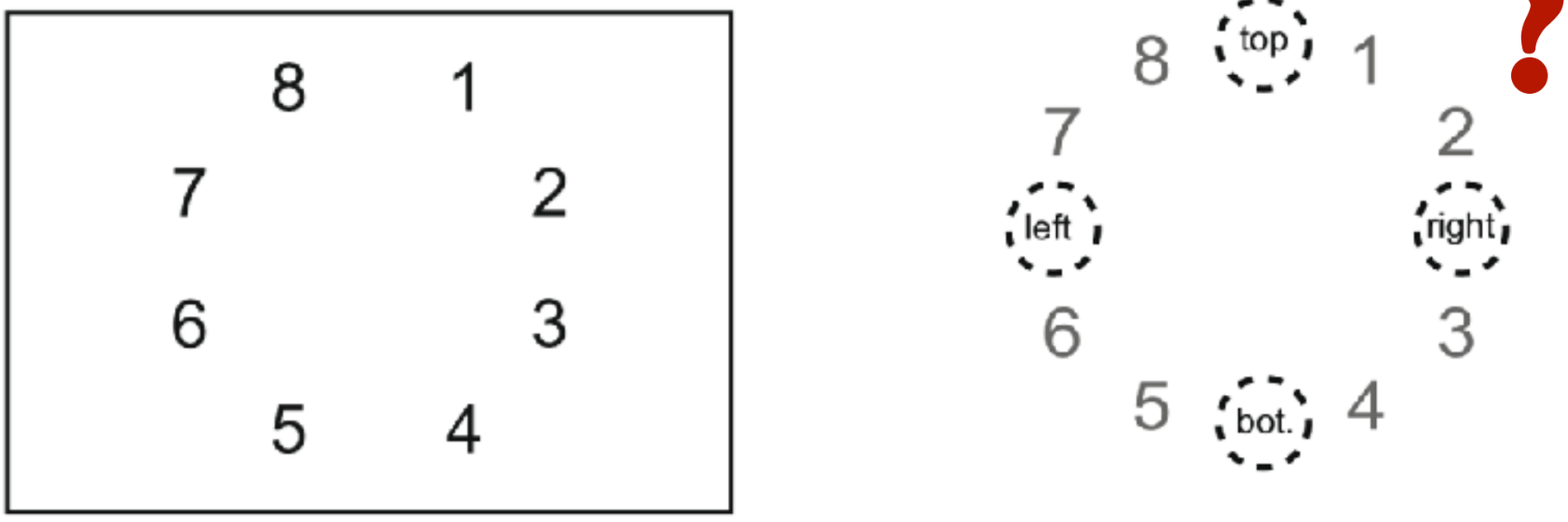


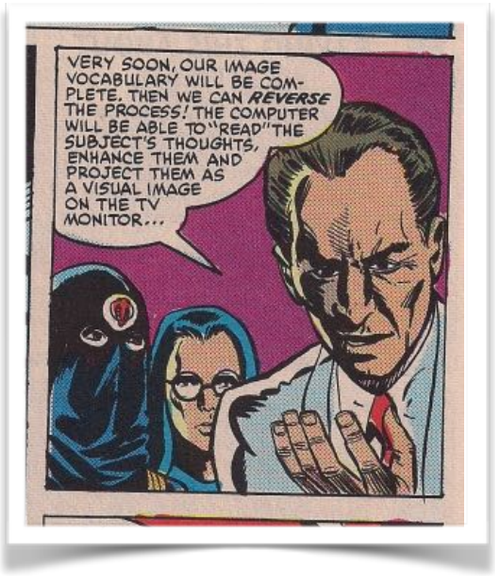
neurophysiology





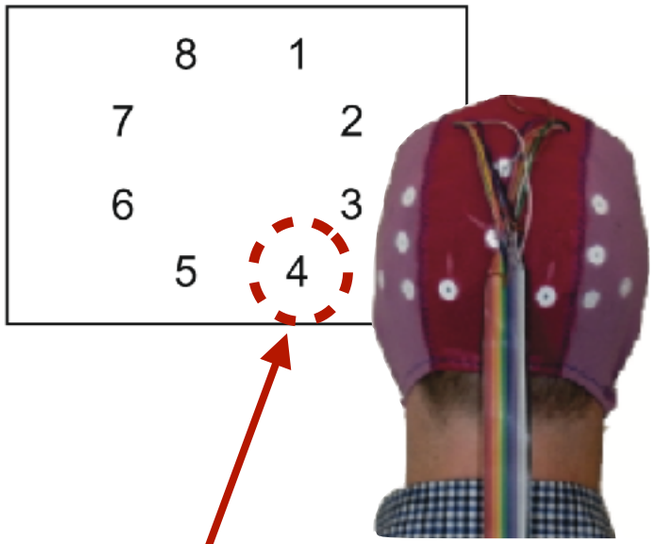
Predict locations that were never attended when generating the model





The procedure from patterns to CTFs is invertible!

Combine observed EEG patterns from testing set

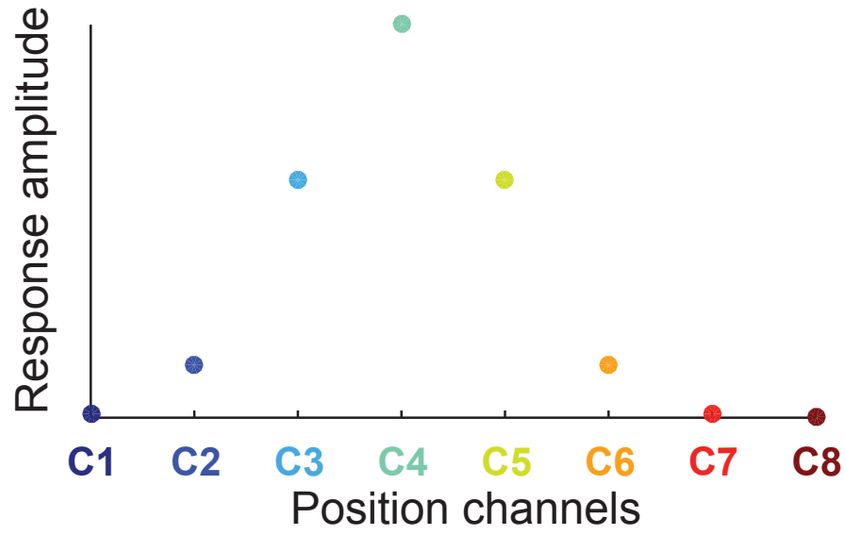


Attended position

with channel weights to estimate channel responses

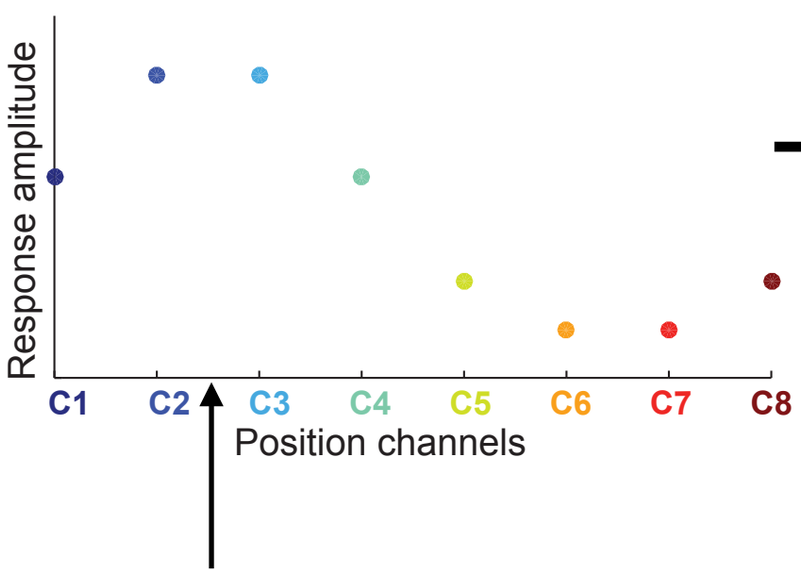
Weights^{-1}

Channel Tuning Function (CTF)



Construct neural patterns for positions that were never attended

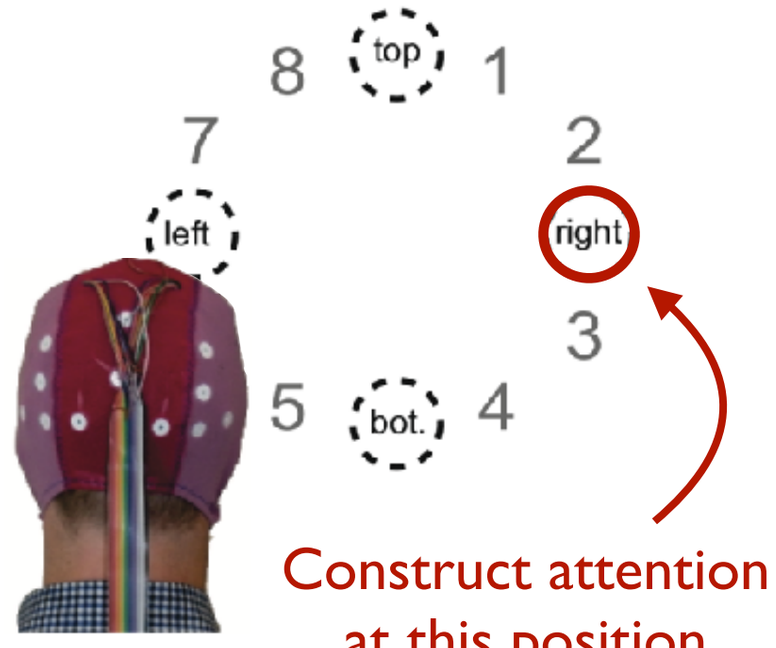
C2/C3 constructed CTF



Channel weights
obtained during training

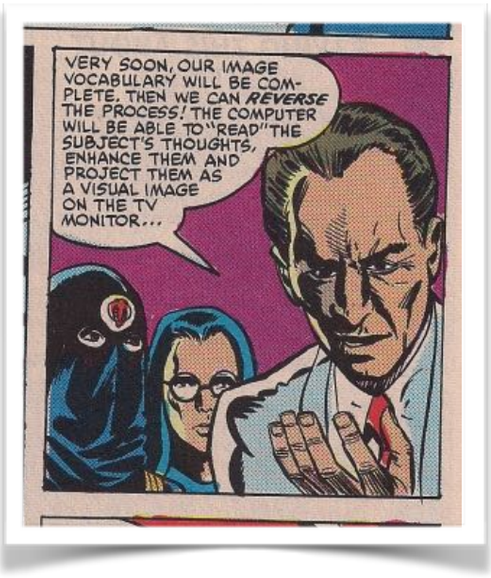
Weights

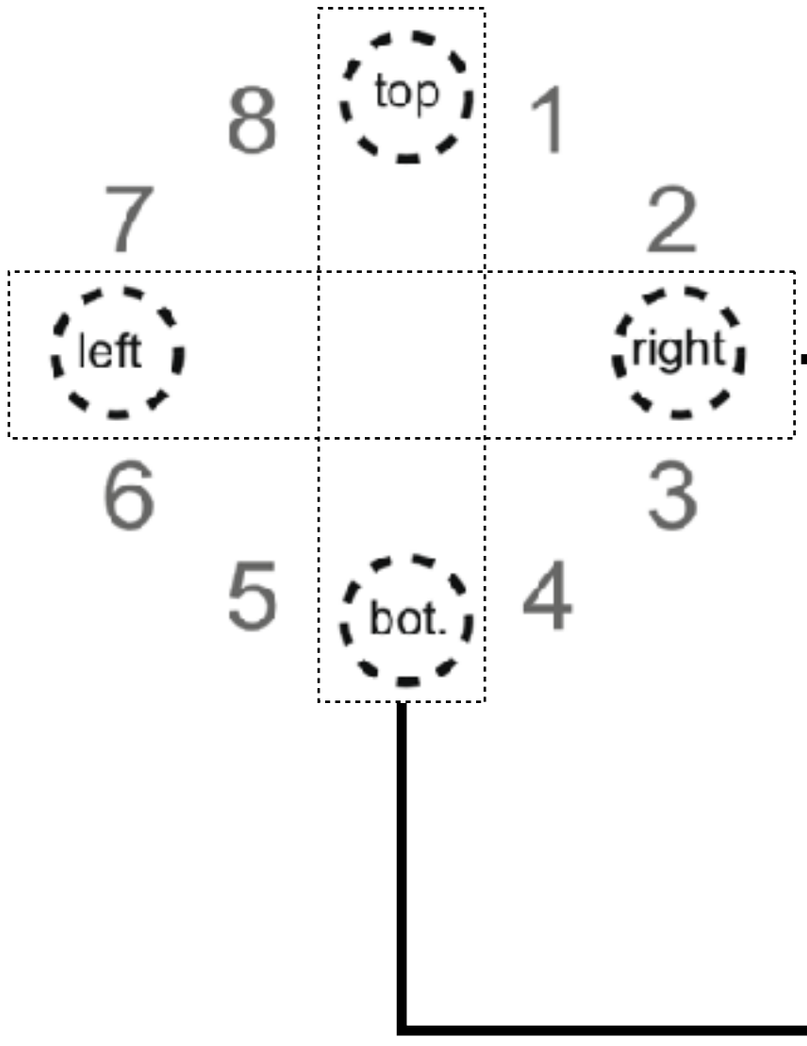
Constructed data pattern

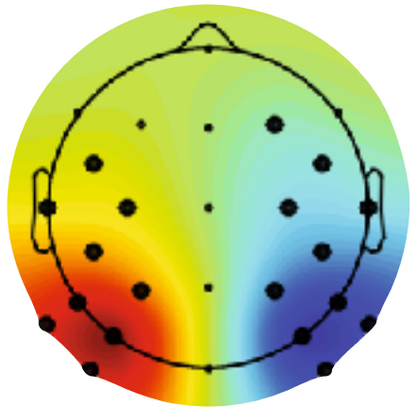


Right position (between 2 and 3)
was never attended during experiment!

Construct neural patterns for positions that were never attended

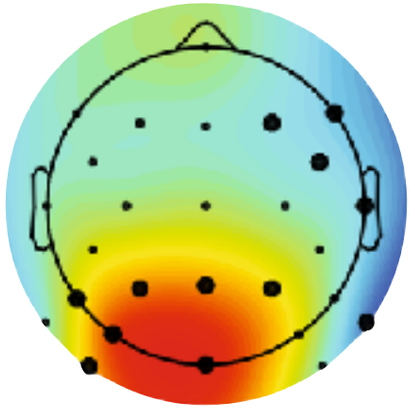






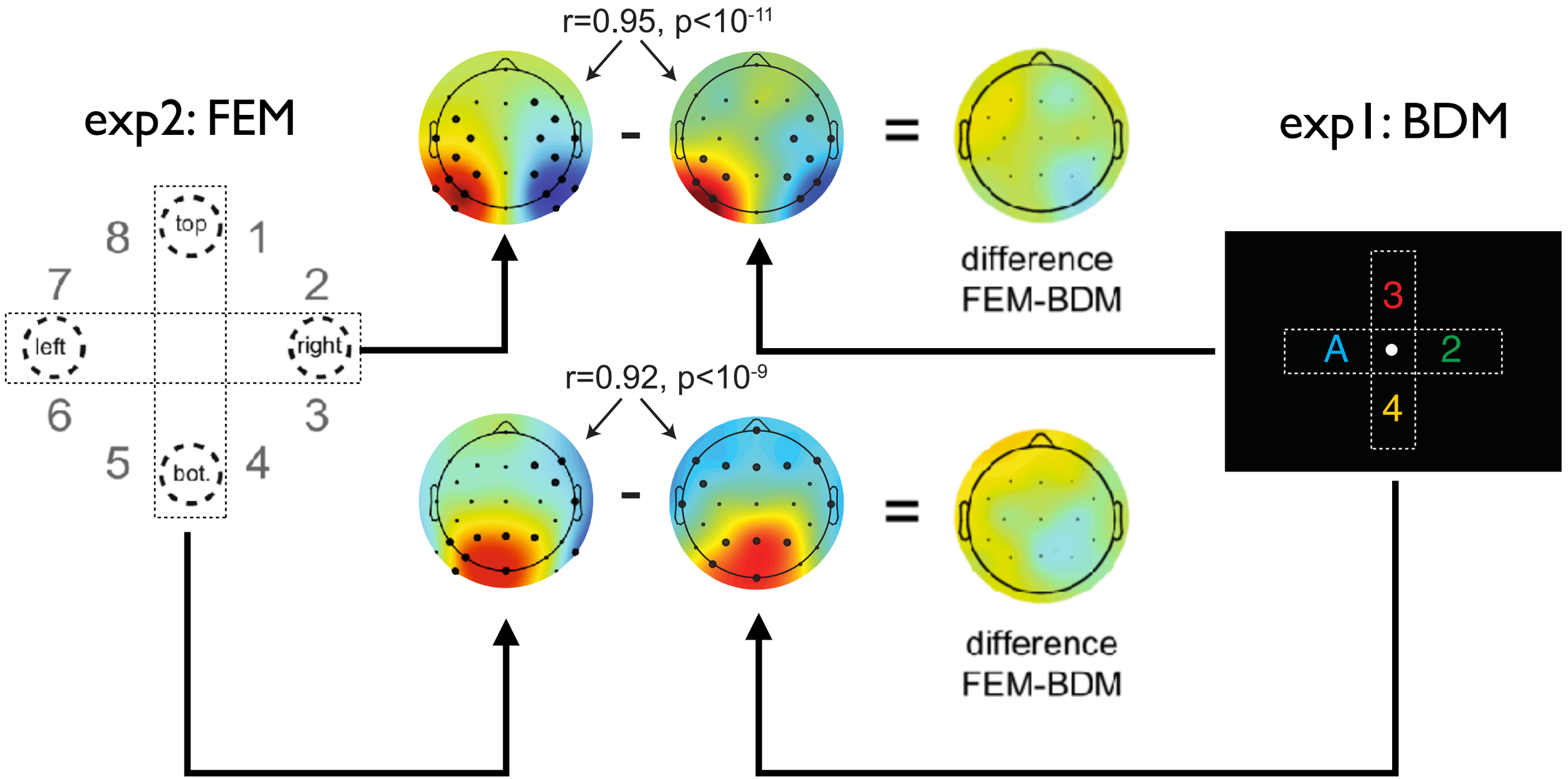
left versus right

Top, bottom, left and right were never attended during the experiment!



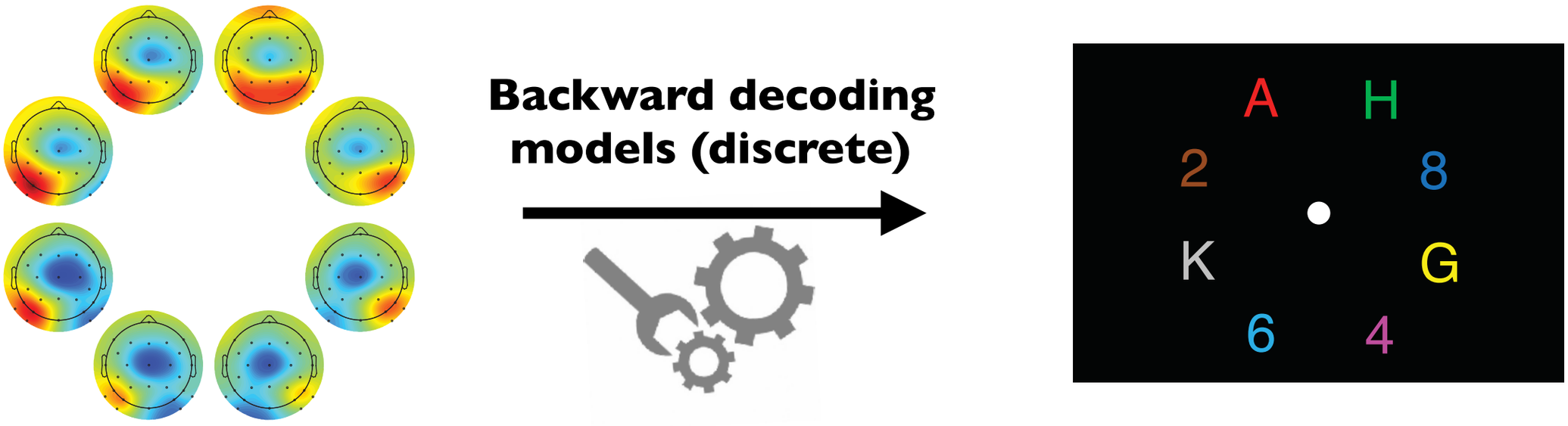
top versus bottom

Correspondence between prediction of forward encoding model (FEM) and actually observed data

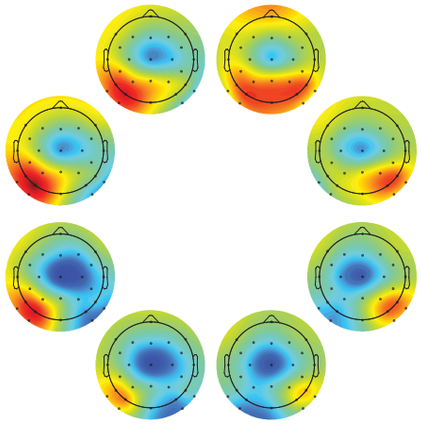


Fahrenfort, J. J., Grubert, A., Olivers, C. N. L., & Eimer, M. (2017). Multivariate EEG analyses support high-resolution tracking of feature-based attentional selection. *Scientific Reports*, 7(1), 1886.

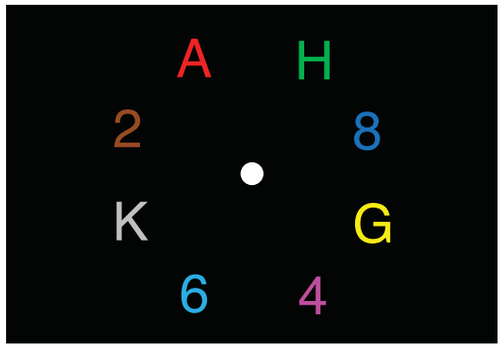
Backward decoding models (BDMs)

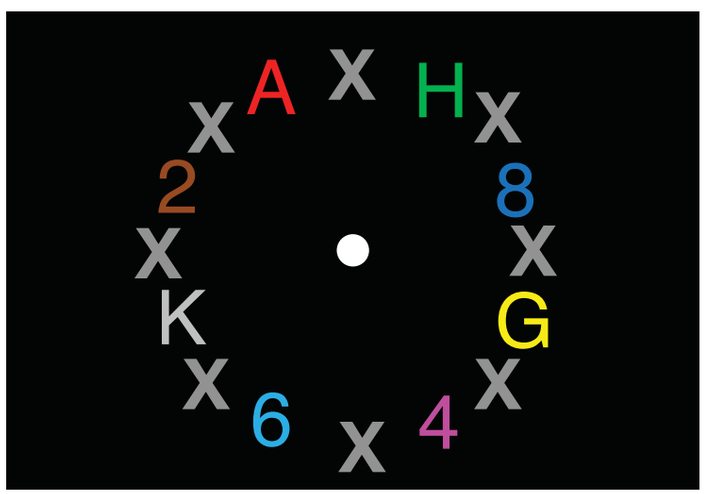


Backward decoding models (BDMs)
versus
Forward encoding models (FEMs)

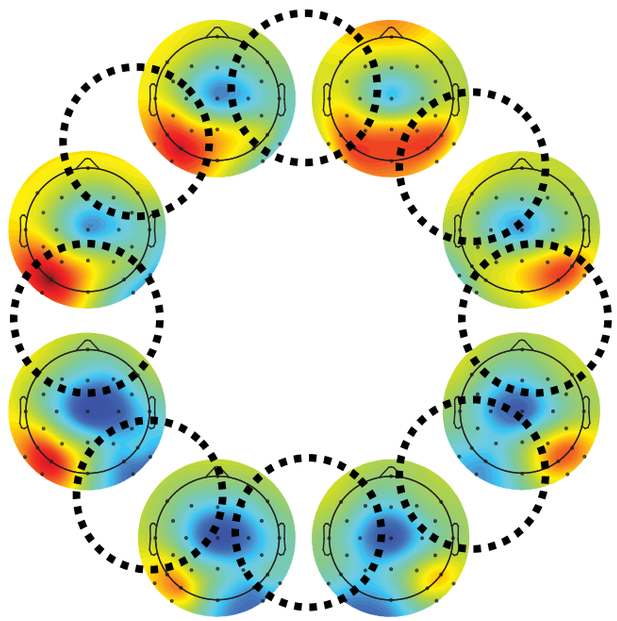


Backward decoding models (discrete)



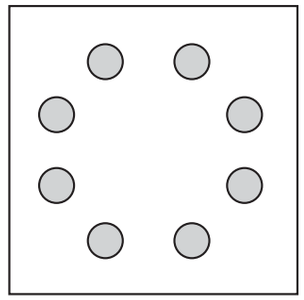


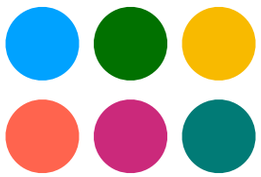
Forward encoding models (continuous)



Backward decoding versus Forward encoding models

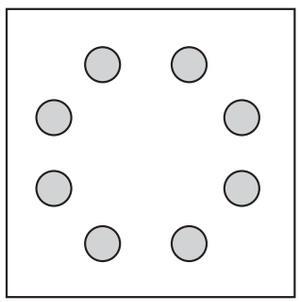
- Backward decoding models work with *discrete* stimulus classes (e.g. object categories, but classes that are continuous can also be treated as discrete).

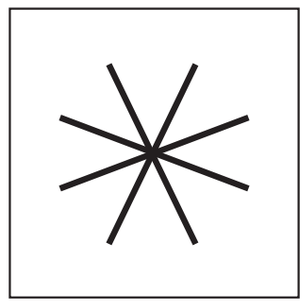




- The model *cannot* make new predictions for stimuli that were never used to train the model

- Forward encoding models only make sense when using *continuous* stimulus classes (e.g. position on a circle, orientation of a bar, color etc)



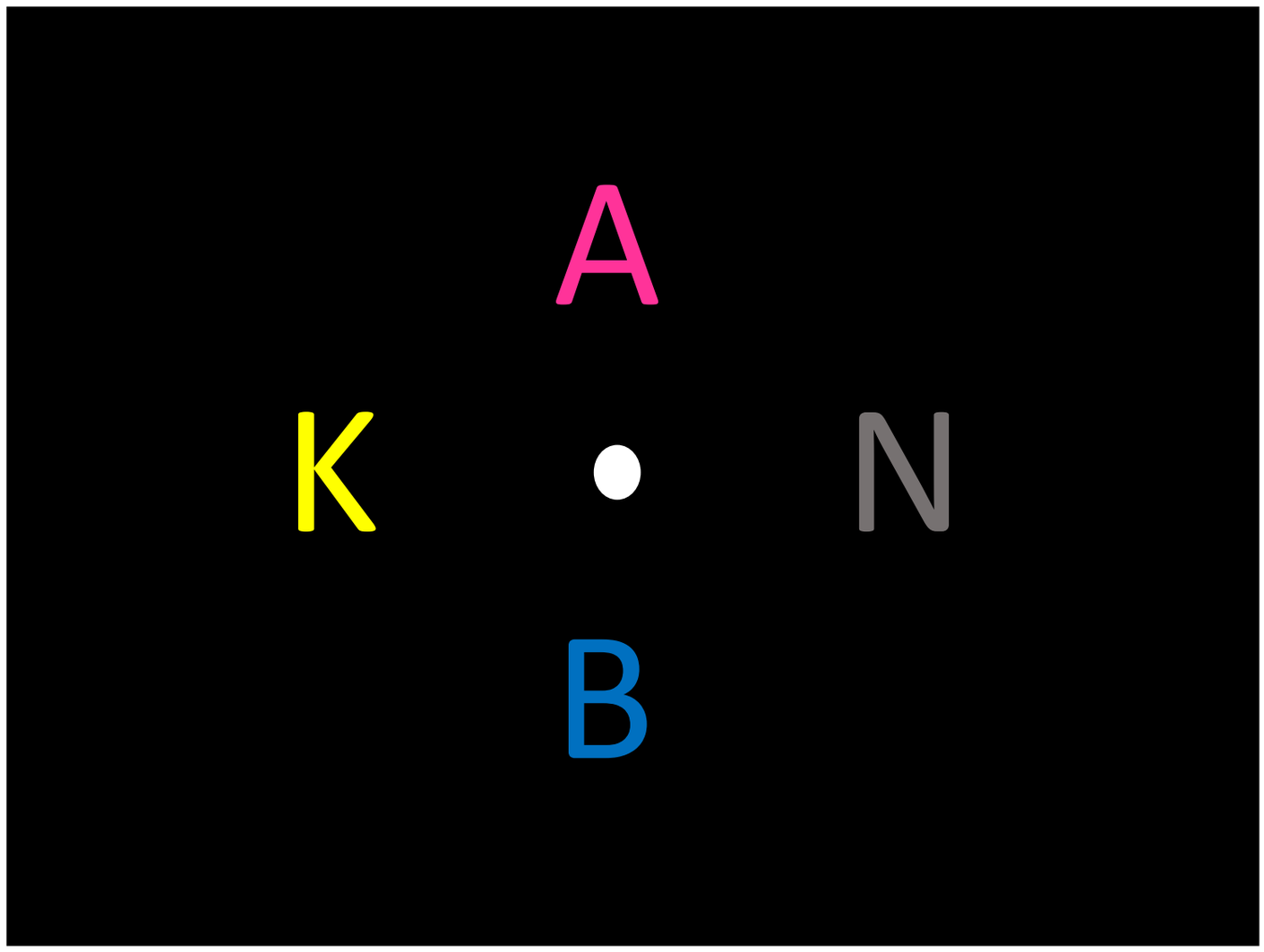


- Allow you to make *new predictions* about cortical responses for stimuli that were never used to create the model

What can we now do?

- Decode attention for any part of the visual field, as long as the condition has been encoded in the experiment (BDMs)
- Identify attention for positions that were not explicitly encoded in the experiment (FEMs)

We are now using these methods to determine whether attention is able to operate in parallel

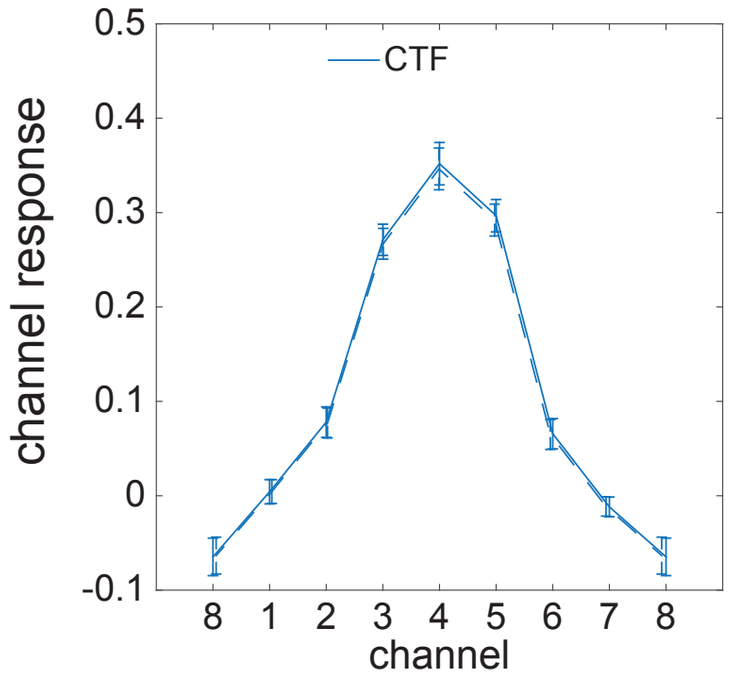


Task: report the identity of the blue and the yellow letter

And to characterize the effect of spatial attention on input over time

attend left

attend right



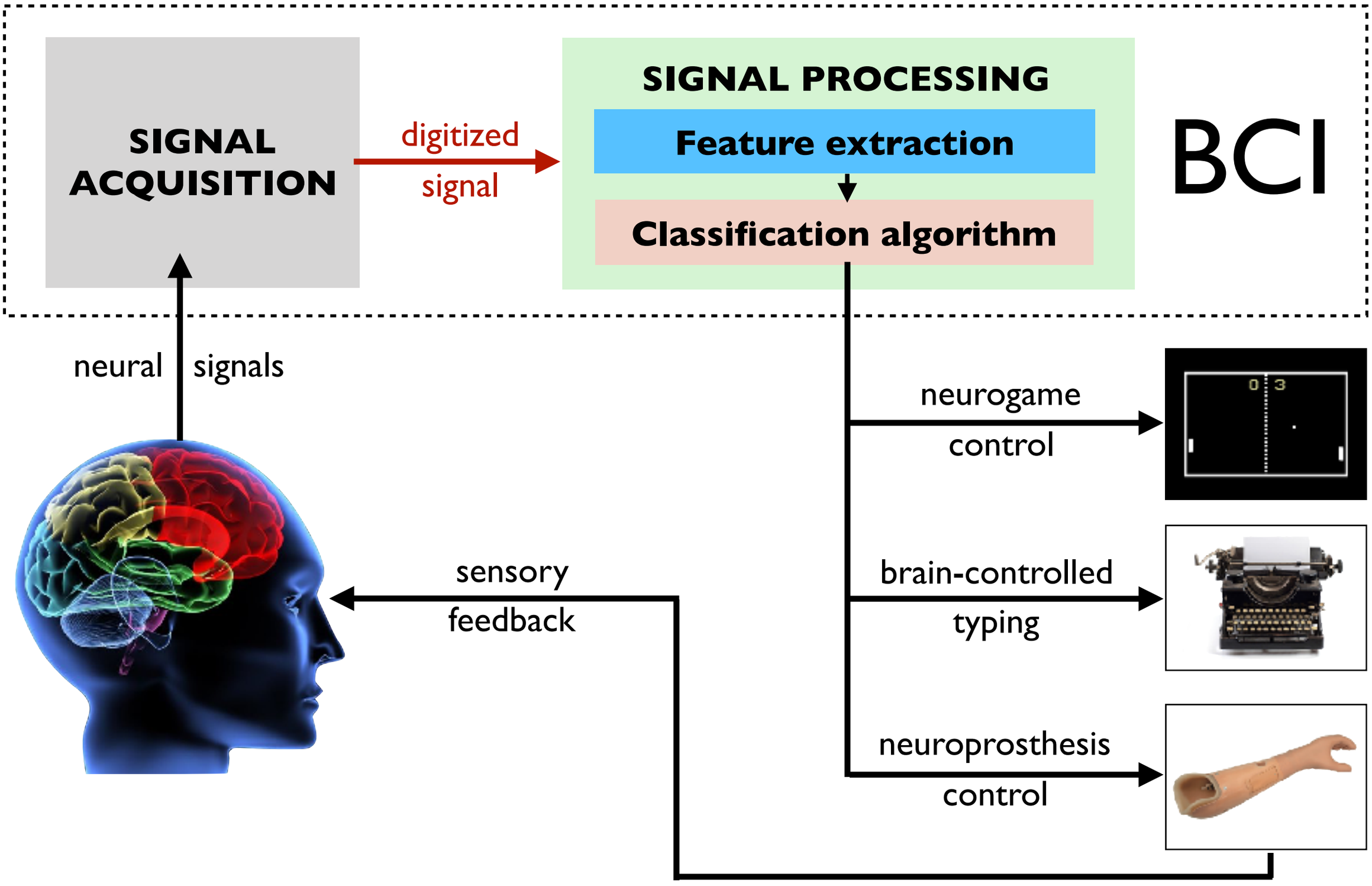
Brain Computer Interfaces

- Multivariate decoding/encoding approaches in EEG and MEG are a relatively recent addition in cognitive neuroscience (in fMRI they've been around for some time)

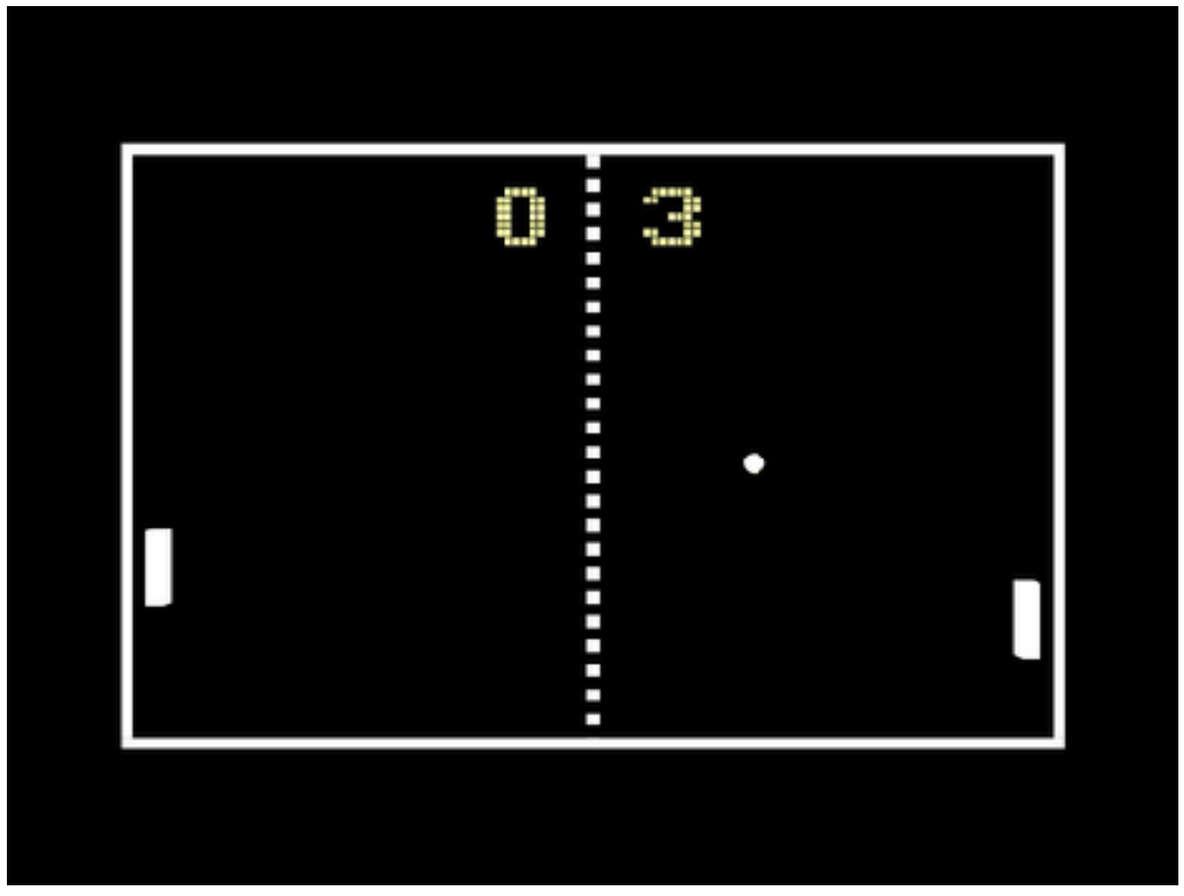
- This is quite odd, as BCIs that utilize multivariate data have been around for quite a bit longer, especially on EEG

Peters, B. O., Pfurtscheller, G., & Flyvbjerg, H. (1998). Mining multi-channel EEG for its information content: an ANN-based method for a brain–computer interface. *Neural Networks*, 11(7), 1429–1433.

Brain Computer Interfaces

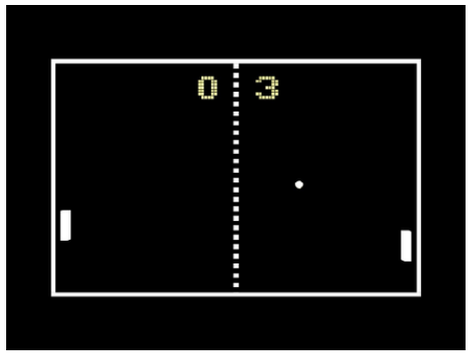


Pong



ATARI®

GAMES



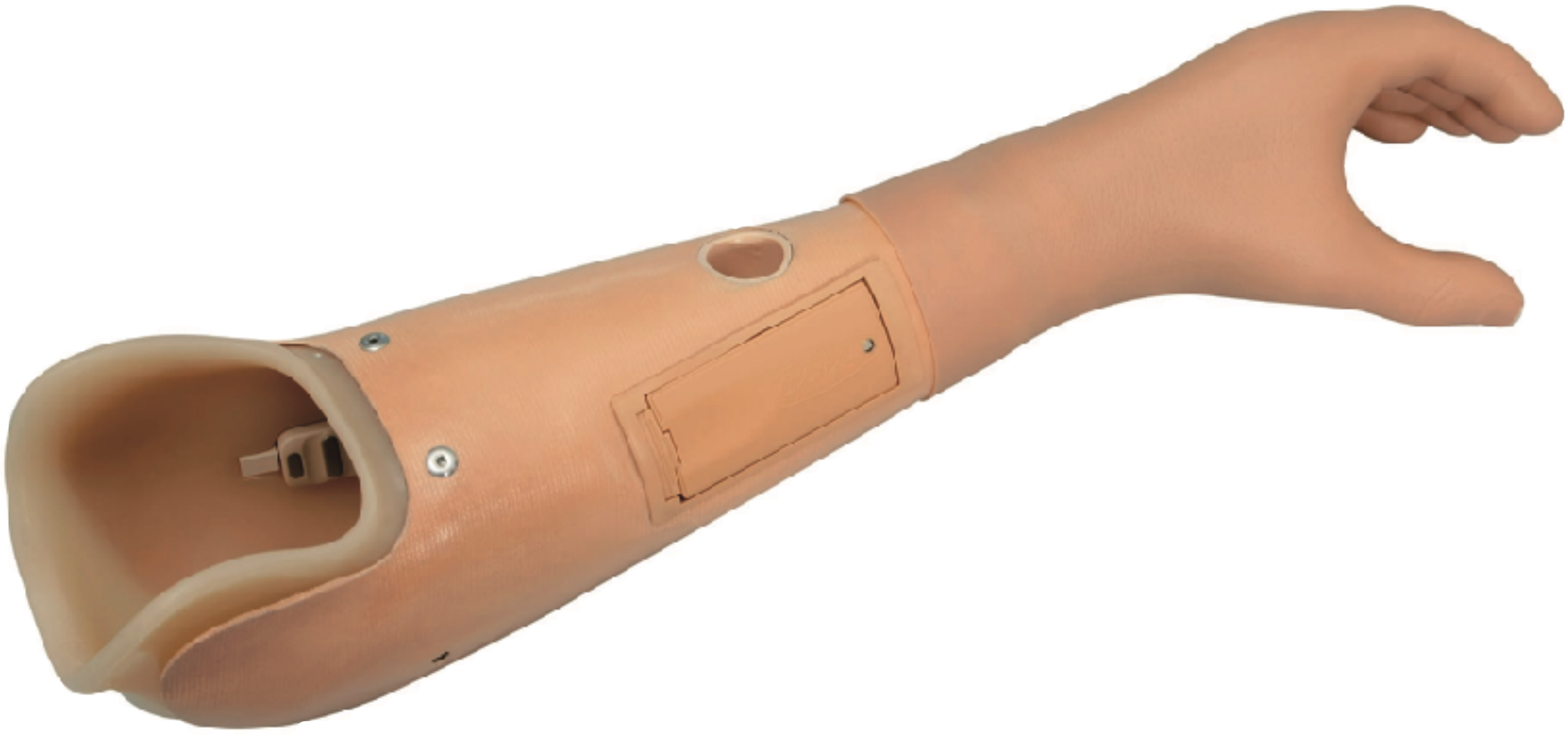
Brain pong

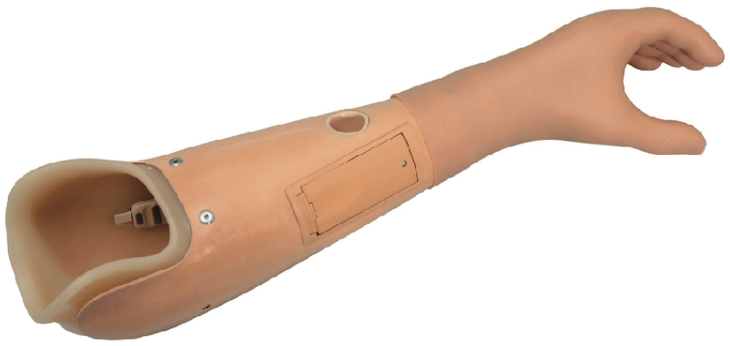
Typing

Electronic Typing

Brain Typing

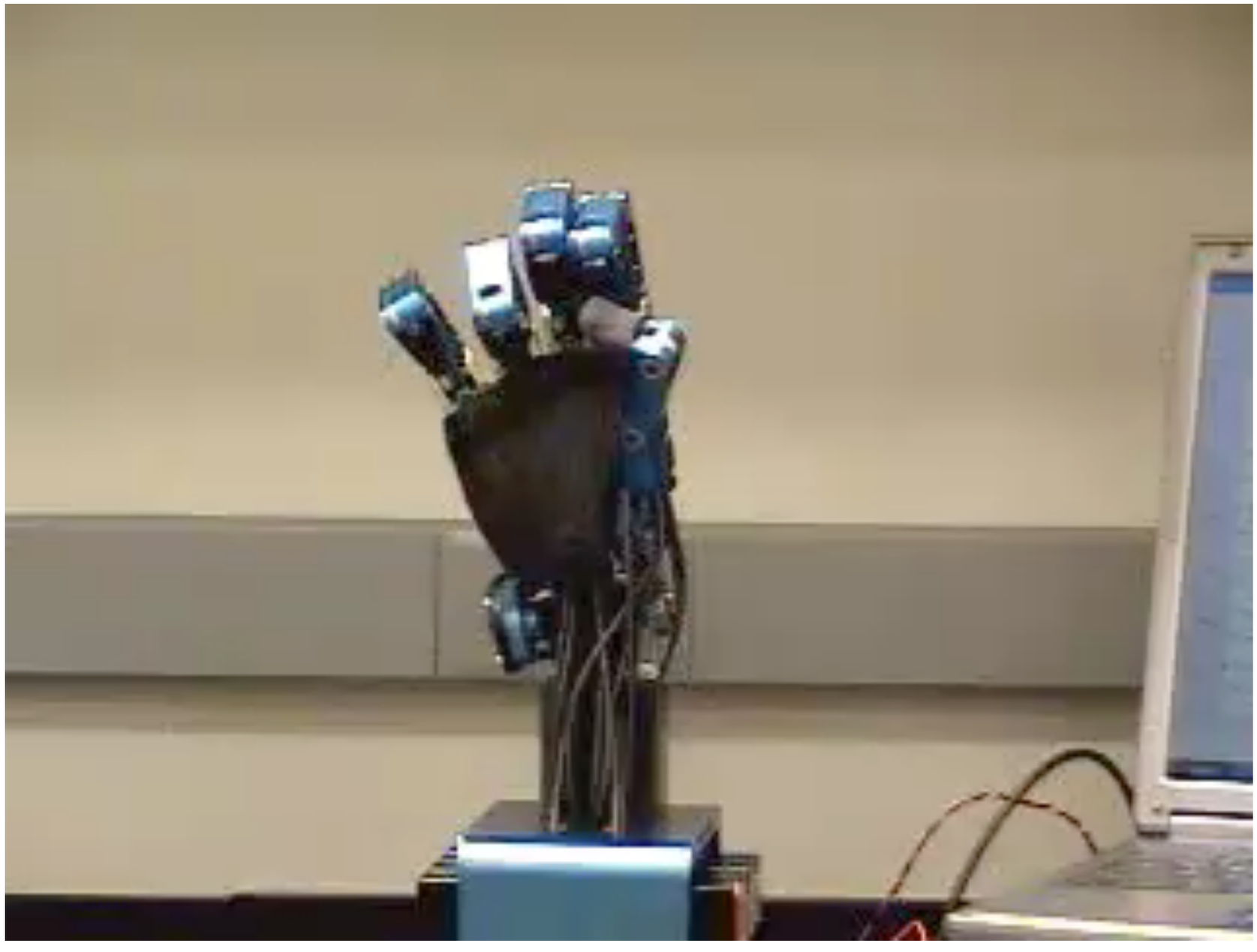
Prosthetic hand





Robotic hand

Brain-controlled hand



Questions?

Course overview

- **Day 1 (introduction)**
  - lecture 1: History and electrophysiological basis of EEG
  - lecture 2: Backward decoding models in MVPA: concepts and analytical approach
  - lecture 3: Advantages of MVPA, the temporal generalization method
  - lecture 4: Architecture of the ADAM toolbox and explaining the experiment of the practical
  - *Afternoon: practical*
- **Day 2 (advanced)**
  - lecture 1: BCI's, MVPA experimental design and various ways to map brain to brain or to behavior
  - lecture 2: Forward encoding models in MVPA: how do they work, concepts and analytical approach
  - *Afternoon: practical, analyze your own data and/or a supplied dataset*