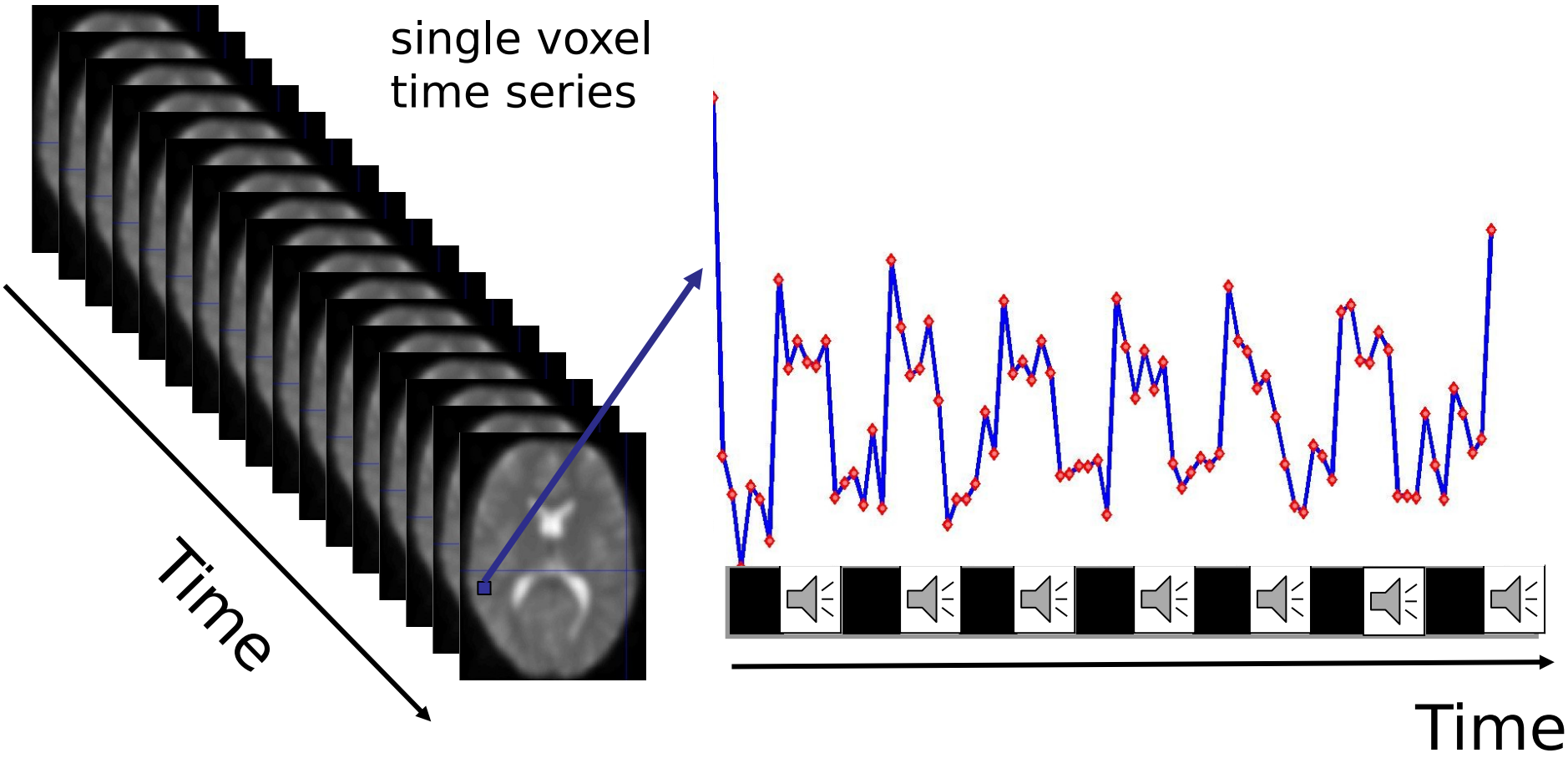
Functional MRI and data analysis

Florent MEYNIEL

Neurospin, CEA, France

Starting with a example: noisy timeseries in fMRI data



fMRI timeseries are noisy, and require a statistical approach:
- Use many data points from a single voxel (or the average value in a region of interest) and test effects of interest. This analysis can be repeated in each voxel. (***the univariate approached***)
- Use many data points from several voxels and test if they collectively convey information about effects of interest. This analysis can be repeated in different regions. (***the multivariate approached***)

Adapted from: SPM courses

I/ From neural activity to BOLD signal: the power of a forward model

The origin of the BOLD signal
The heomodynamic response function
Convolution model for BOLD data

II/ Mass univariate analysis and the General Linear Model

Intuitions for regression
General linear model (GLM) and design matrix
Contrast testing the effect of interest
Adding confounds in the design matrix
Subject and Group level analyses
Categorical and parametric regressors
Model-based approaches: strengths and limits

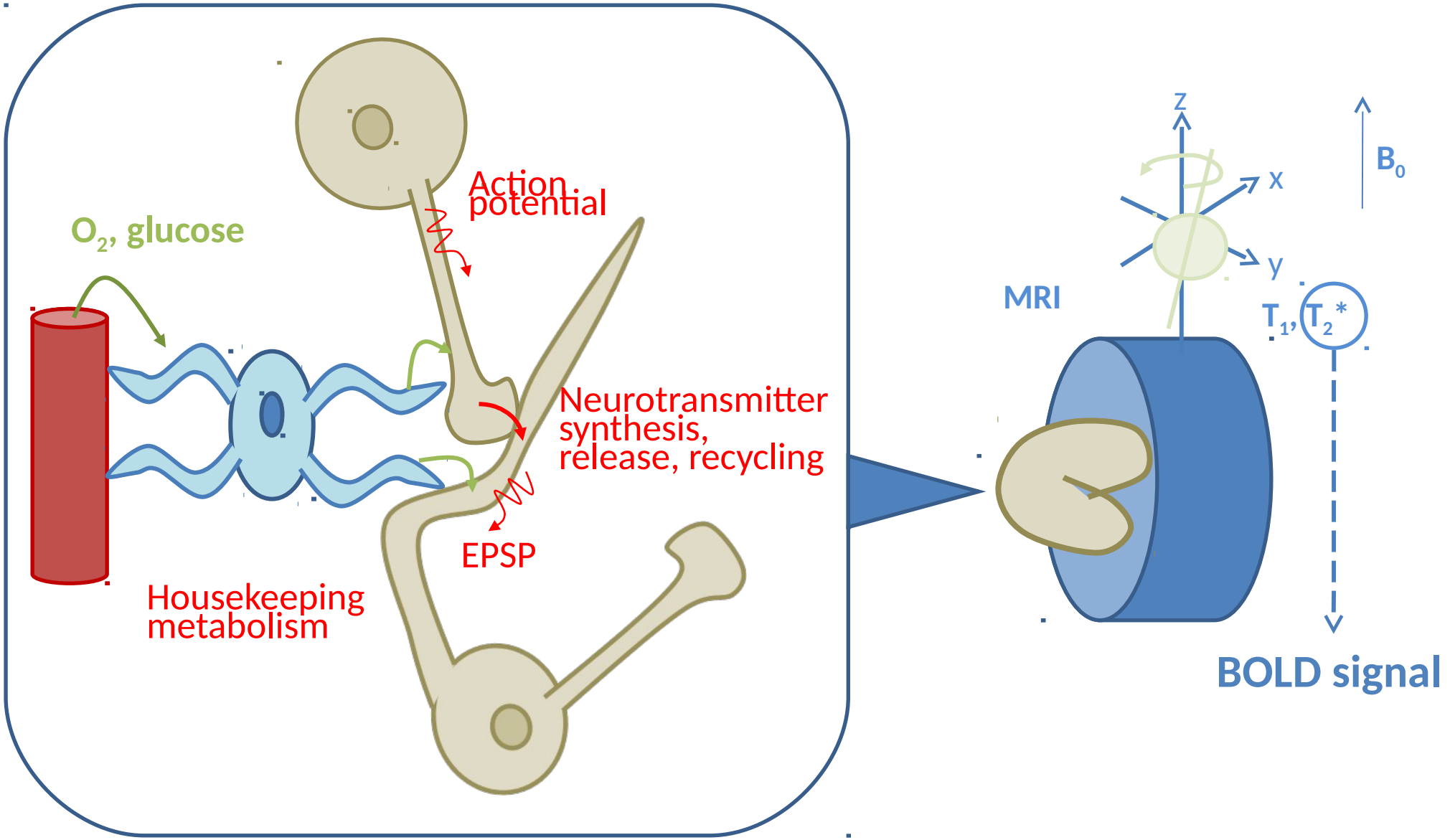
III/ Multivariate analysis

Multivariate information in distributed population codes
Estimation of a classification accuracy
Supervised learning algorithms for classification
Origins of the multivariate information
Comparison of univariate and multivariate approaches

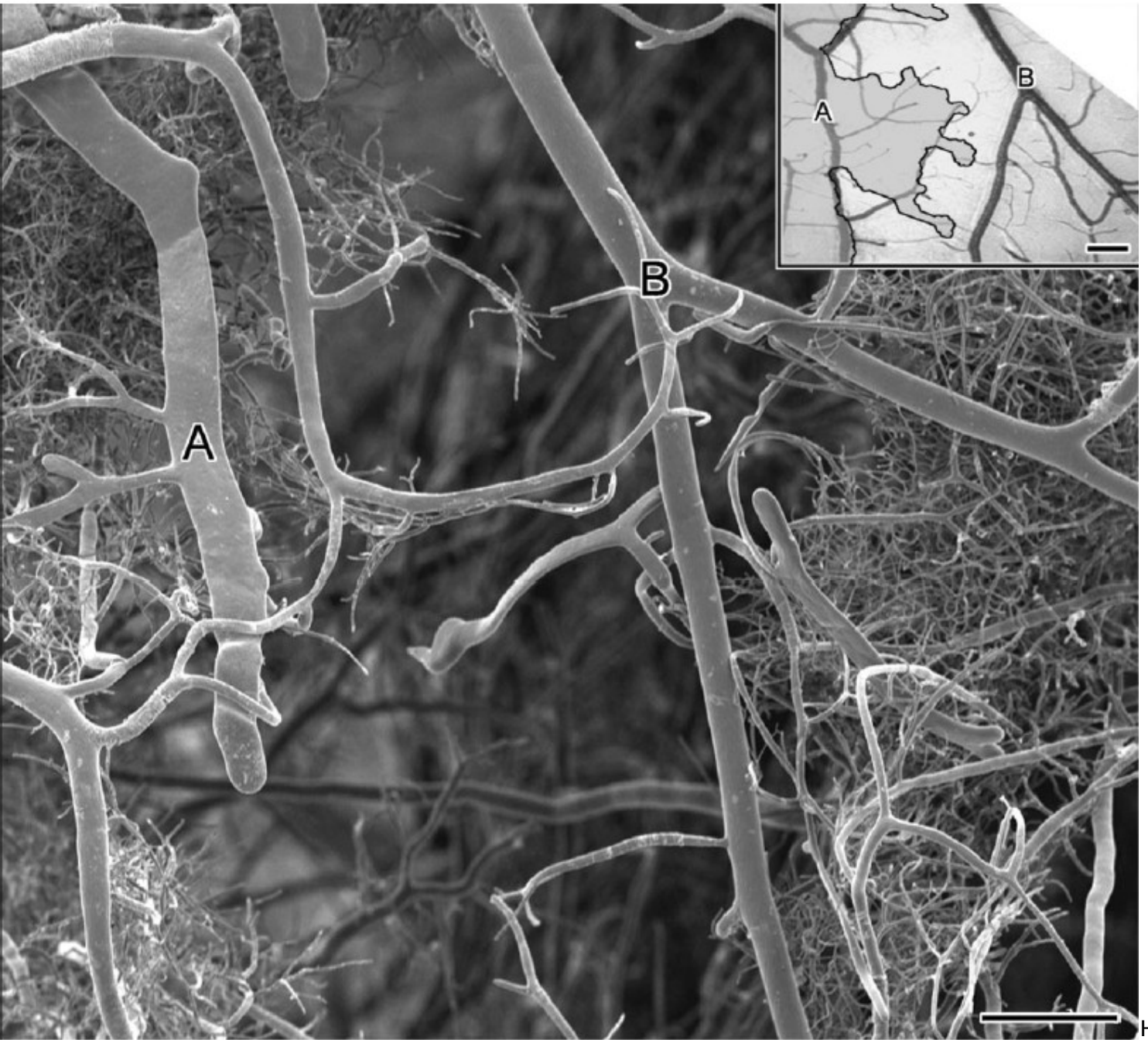
VI/ The problem with multiple comparisons

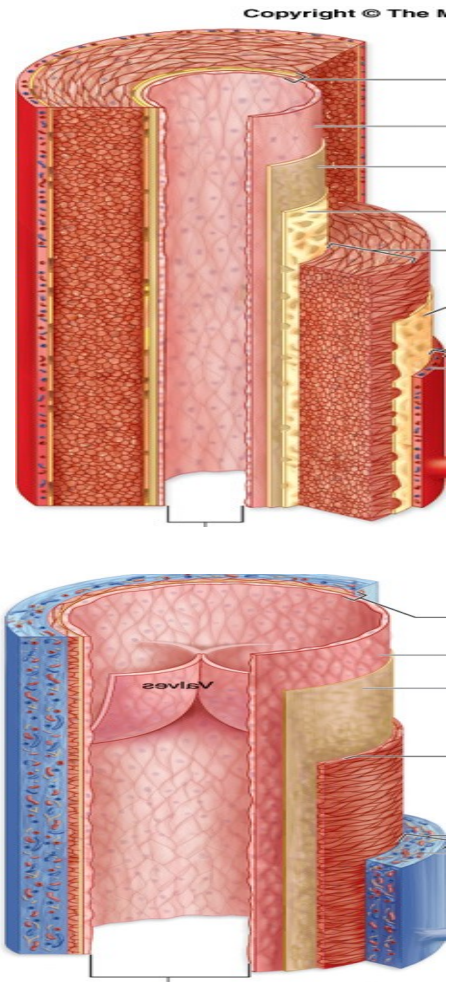
Multiple testing inflates the risk of having a false positive
Statistical methods to correct for multiple comparisons

The BOLD signal reflects changes in the vascular system



The BOLD signal reflects changes in the vascular system

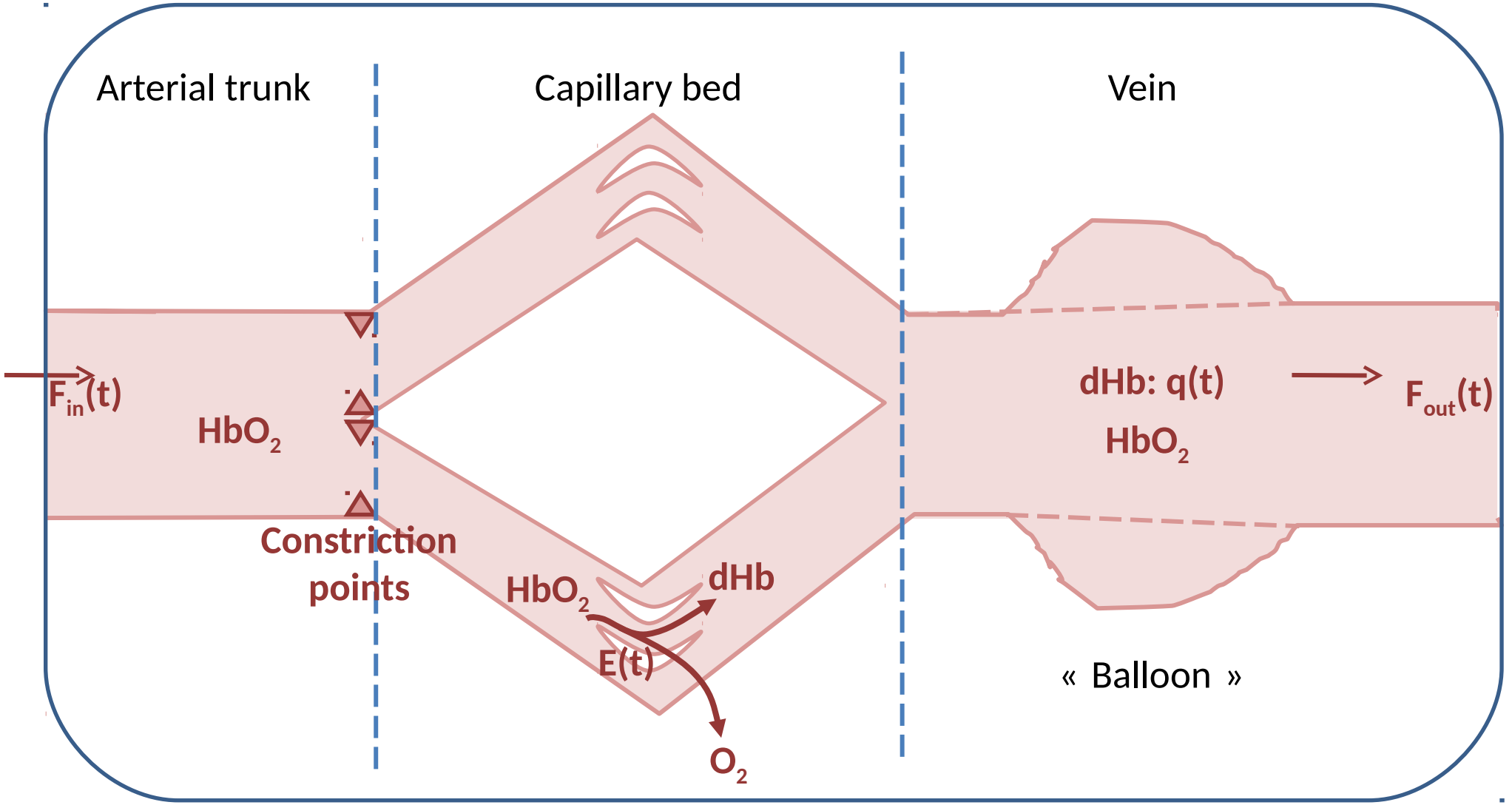




The McGraw-Hill Companies, Inc.

Harrisson, Cereb. Cort. 2002

Models of the BOLD effect, e.g. the “ballon model”



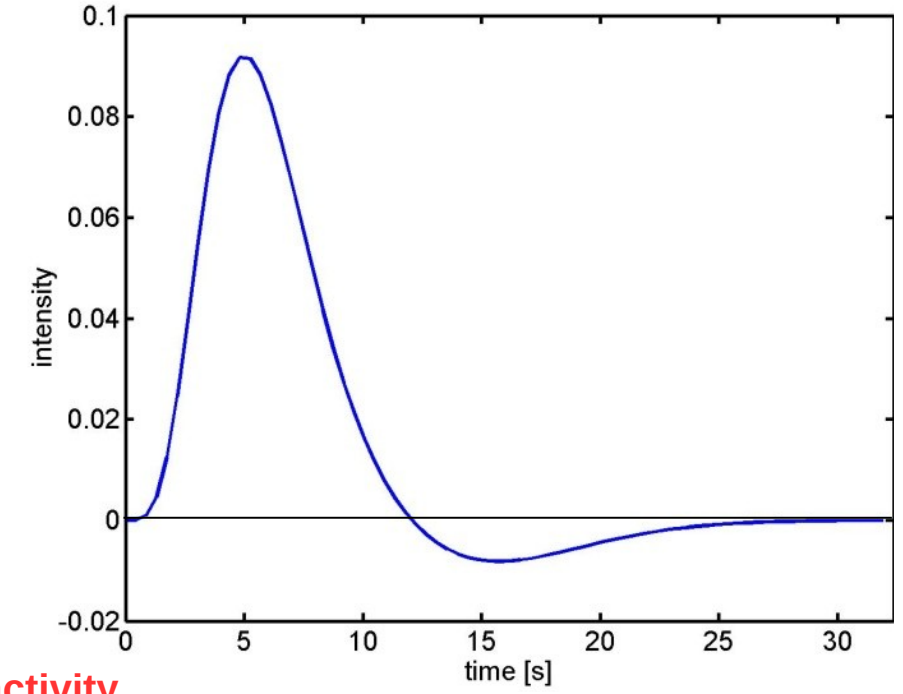
non-blood volume 1-V(t)

blood volume V(t)

Buxton et al, 1998, Mag Res Med

The heamodynamic response function (HRF)

BOLD response

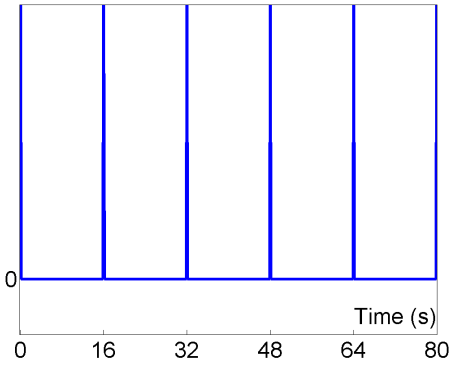


Neural activity

Predicted BOLD responses in a simple experiment

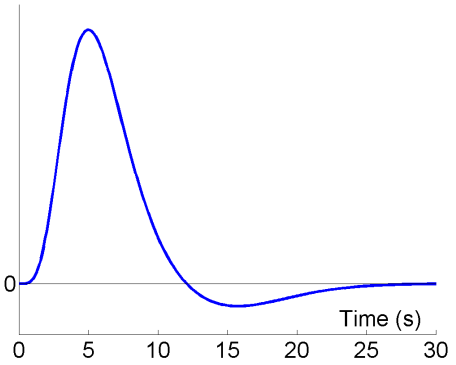
Task: every 16 s, a sound is played.

Expected neuronal activity in the auditory cortex



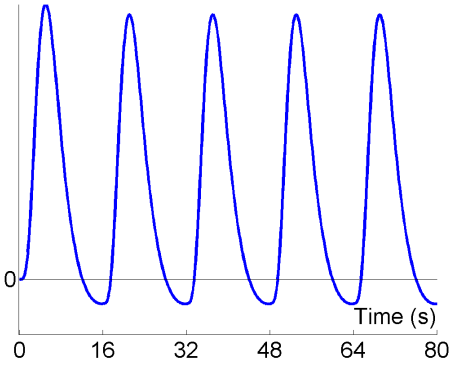
⊗

Expected BOLD response for a single, transient neural event



=

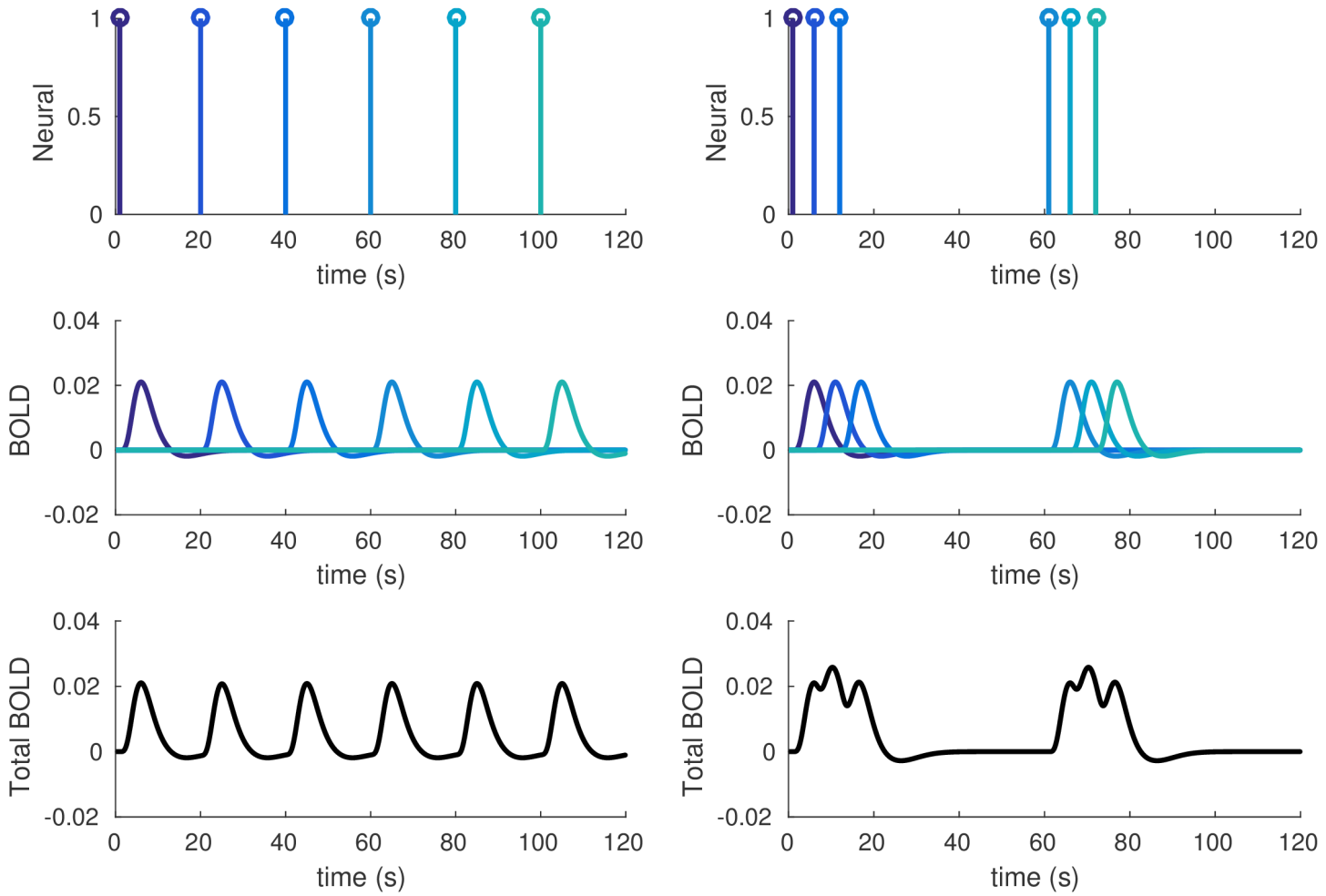
Expected BOLD timeseries in the auditory cortex



In standard fMRI analyses, we assume that the observed BOLD signal is the superposition of (= the sum of) the BOLD responses evoked by every single neural event.

Mathematically, the expected BOLD signal is therefore the timeseries of neural activity convolved with the hemodynamic response function.

Overlap of BOLD responses in fast designs



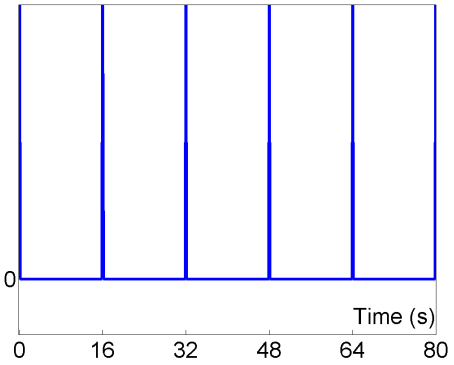
Implications:

- some experimental designs are better than other (e.g. some neuronal effects may be completely smoothed out at the BOLD level)
- Even in very simple designs, events of interest generate time series of observations (sampled by each repetition of the fMRI measurement) that cannot be compared directly.

From neuronal effects to timeseries of BOLD data: The power of forward modelling

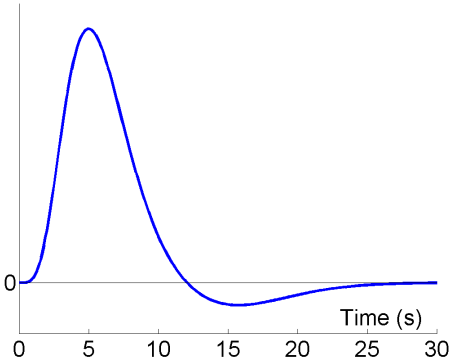
FORWARD MODEL OF OBSERVATIONS

Expected neuronal activity in the auditory cortex



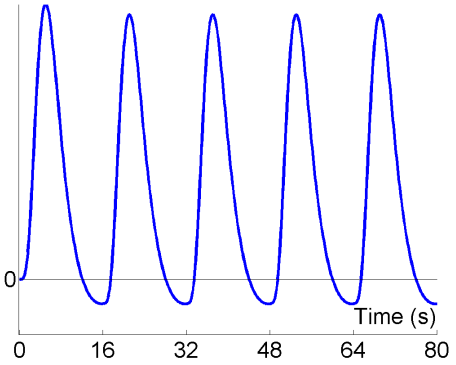
⊗

Expected BOLD response for a single, transient neural event



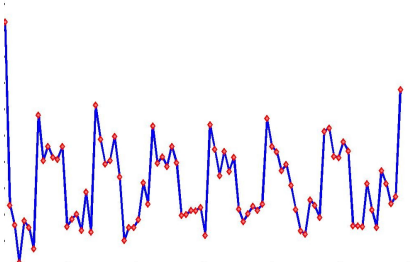
=

Expected BOLD timeseries in the auditory cortex



Match ???

Actual data measured



I/ From neural activity to BOLD signal: the power of a forward model

- The origin of the BOLD signal
- The heomodynamic response function
- Convolution model for BOLD data

II/ Mass univariate analysis and the General Linear Model

- Intuitions for regression
- General linear model (GLM) and design matrix
- Contrast testing the effect of interest
- Adding confounds in the design matrix
- Subject and Group level analyses
- Categorical and parametric regressors
- Model-based approaches: strengths and limits

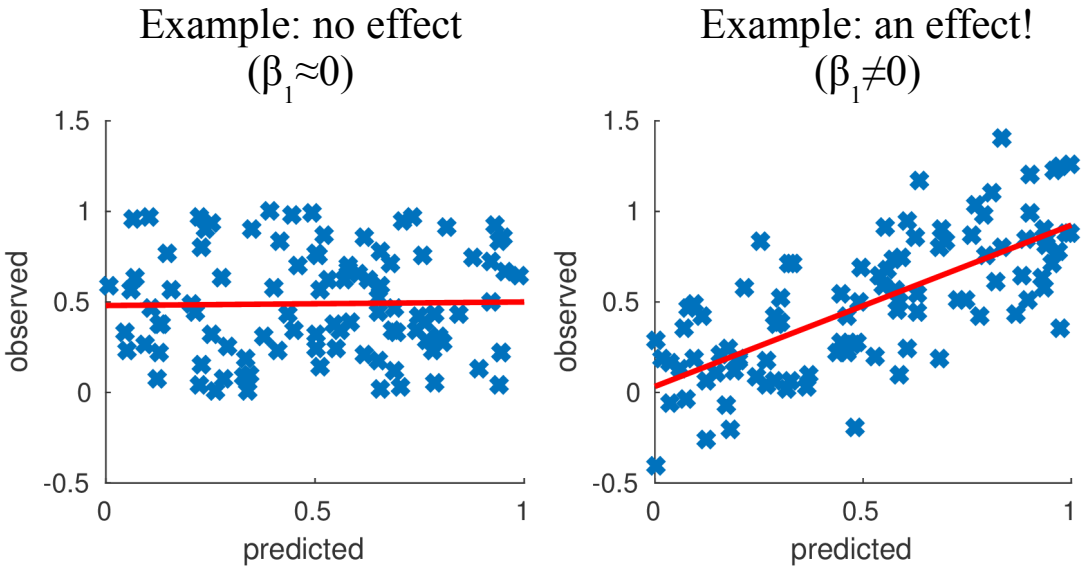
III/ Multivariate analysis

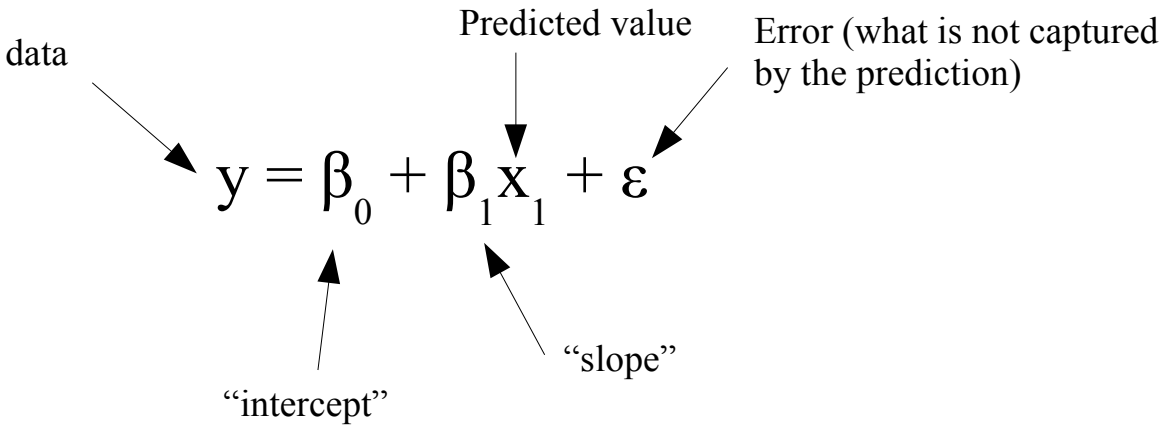
- Multivariate information in distributed population codes
- Estimation of a classification accuracy
- Supervised learning algorithms for classification
- Origins of the multivariate information
- Comparison of univariate and multivariate approaches

VI/ The problem with multiple comparison

- Multiple testing inflates the risk of having a false positive
- Statistical methods to correct for multiple comparisons

From neuronal effects to timeseries of BOLD data: The regression approach





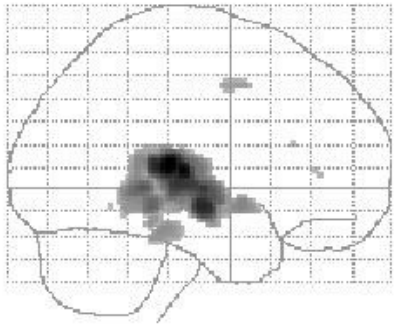
From neuronal effects to timeseries of BOLD data: The regression approach

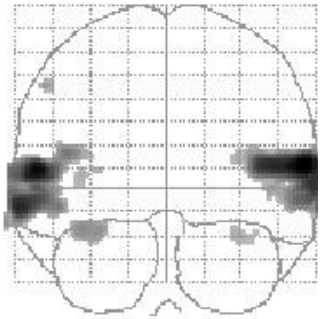
Is $\beta_1 \neq 0$?

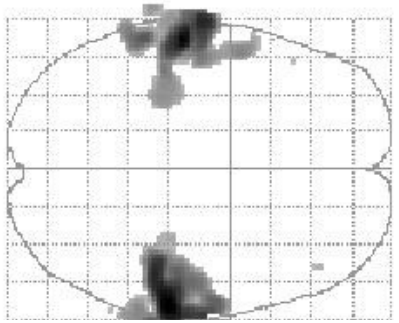
Is β_1 significantly different from 0?

→ Use a Student T-test (a signal-to-noise measure).

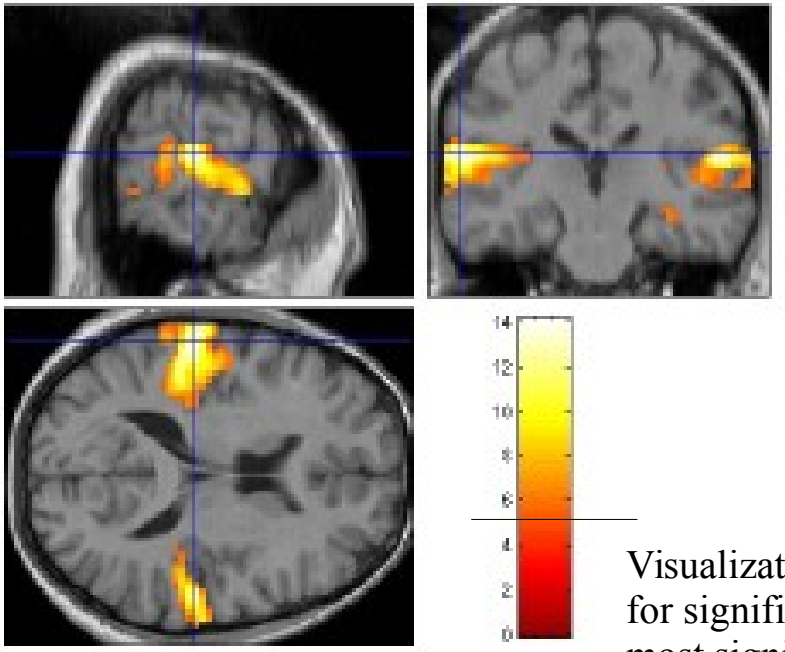
Mass-univariate analysis: compute the Student T-value for every voxel



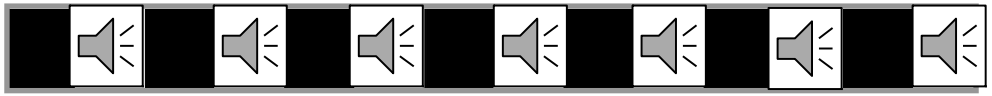




SPM{T_{73}}

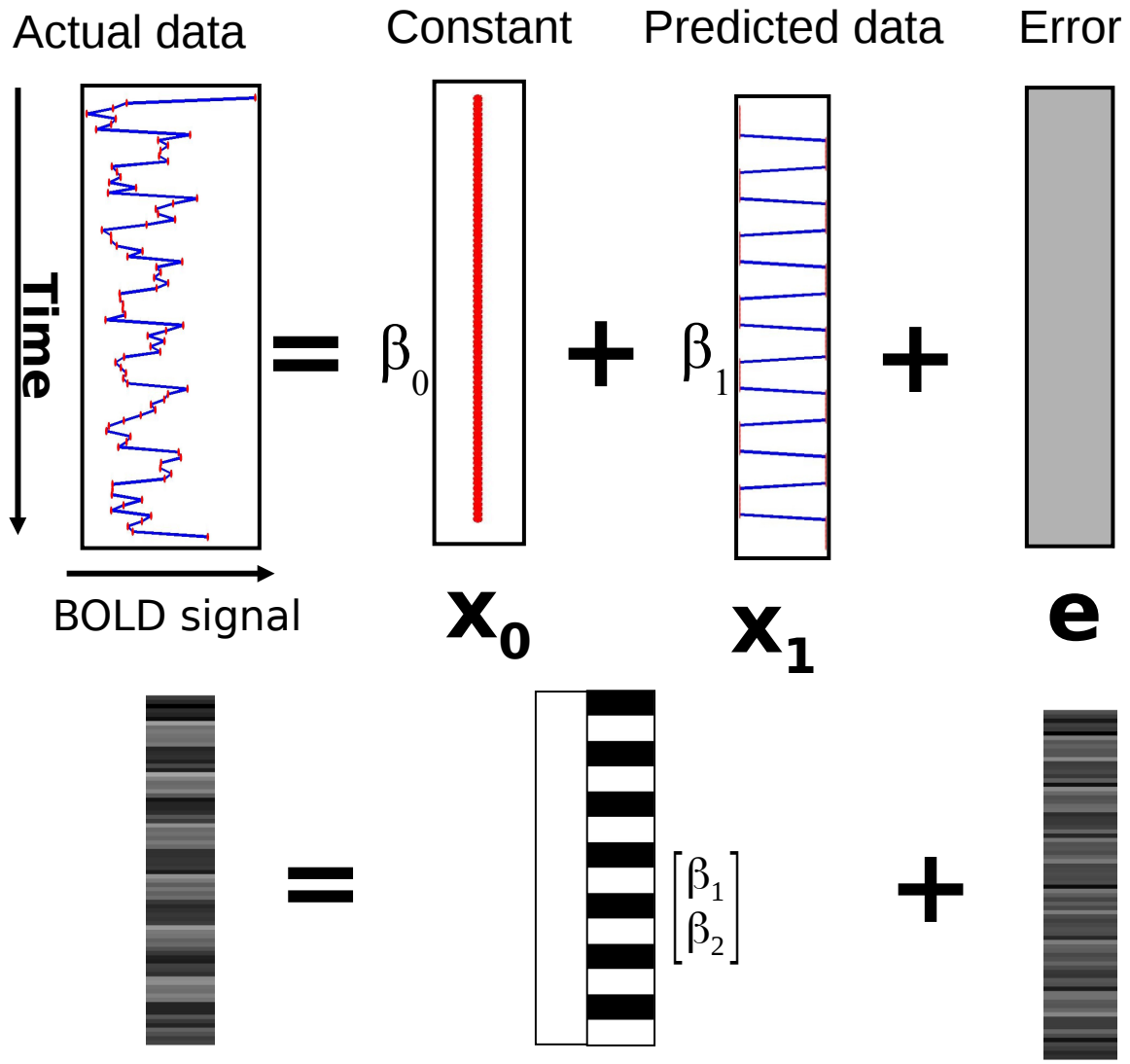


Visualization is thresholded for significance (only the most significant voxels are shown)



Adapted from: SPM courses

From neuronal effects to timeseries of BOLD data: Matrix notations for regression with the General Linear Model (GLM)

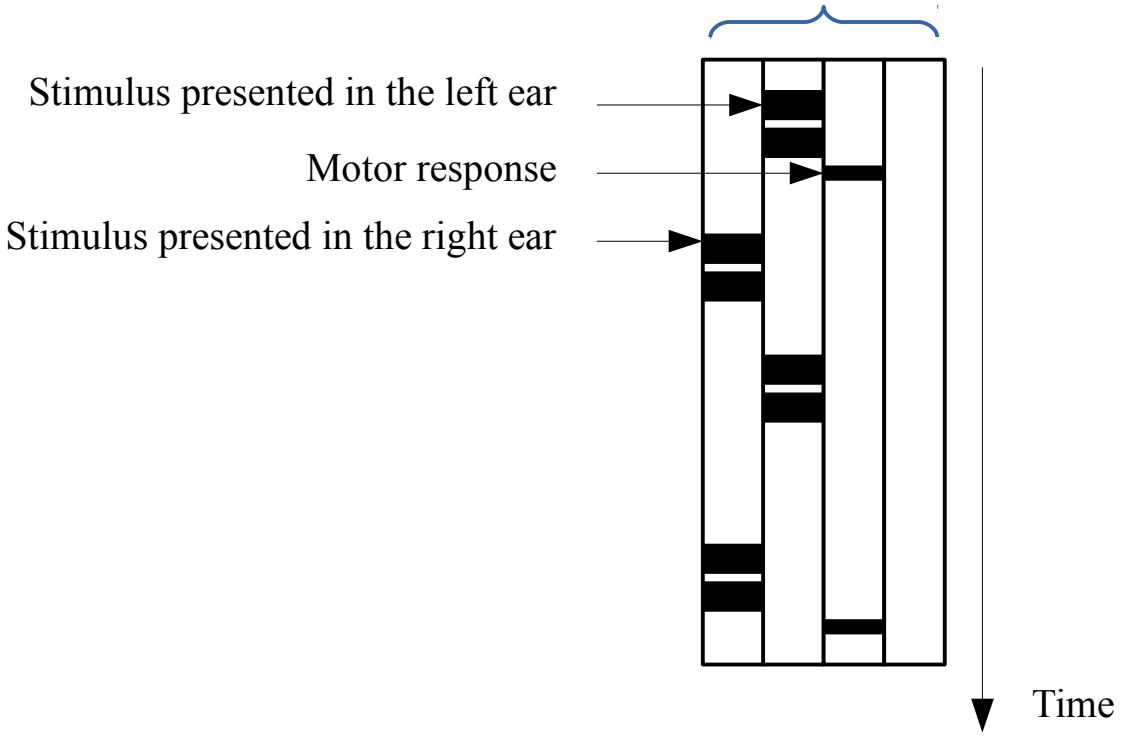


$$Y = X\beta + \varepsilon$$

Adapted from: SPM courses

From neuronal effects to timeseries of BOLD data: The "design matrix" specifies a model of observations as a linear combination of factors (multiple regression)

$$Y = X\beta + \varepsilon$$



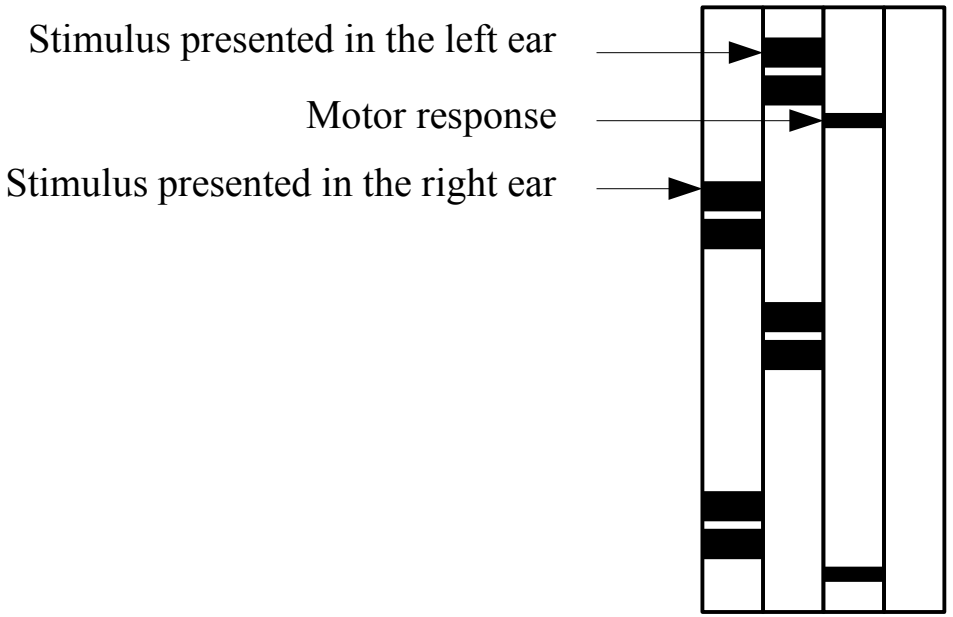
The model of observation is estimated by finding the best-fitting values for the parameters β.
The least-square estimates (those that minimize the residual sum-of-square):

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

NB: the error should be normally, identically and independently distributed
→ data are spatially smoothed (it improves many aspects, including the issue of normal errors)
→ data must be "whitened" to remove the temporal autocorrelation of the data (which is inherent given the BOLD response)

From neuronal effects to timeseries of BOLD data: Testing effects with linear contrast

$$Y = X\beta + \varepsilon$$



C = [0 0 1 0] Contrast testing for more activity when there is a motor response (is $\beta_3 > 0$?).

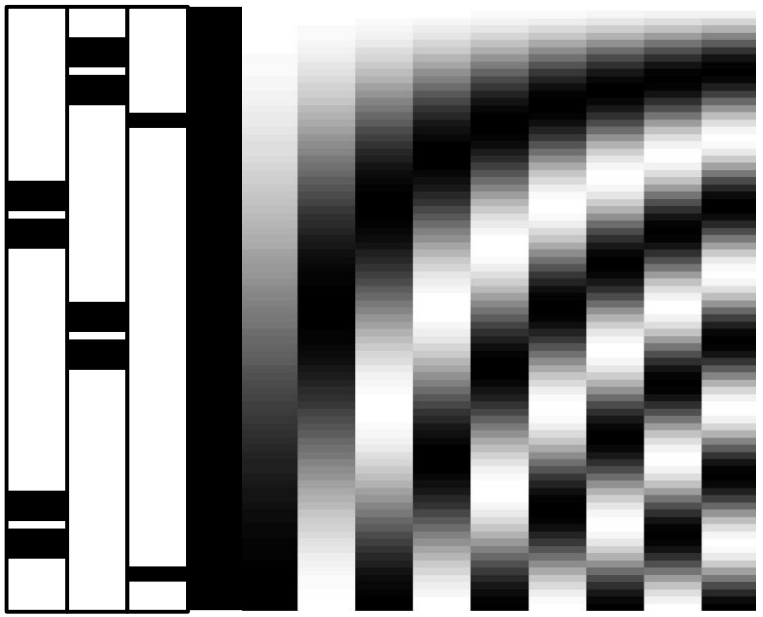
C = [1 -1 0 0] Contrast testing for more activity when the stimulus in presented on the right compared the to left (is $\beta_1 > \beta_2$?)

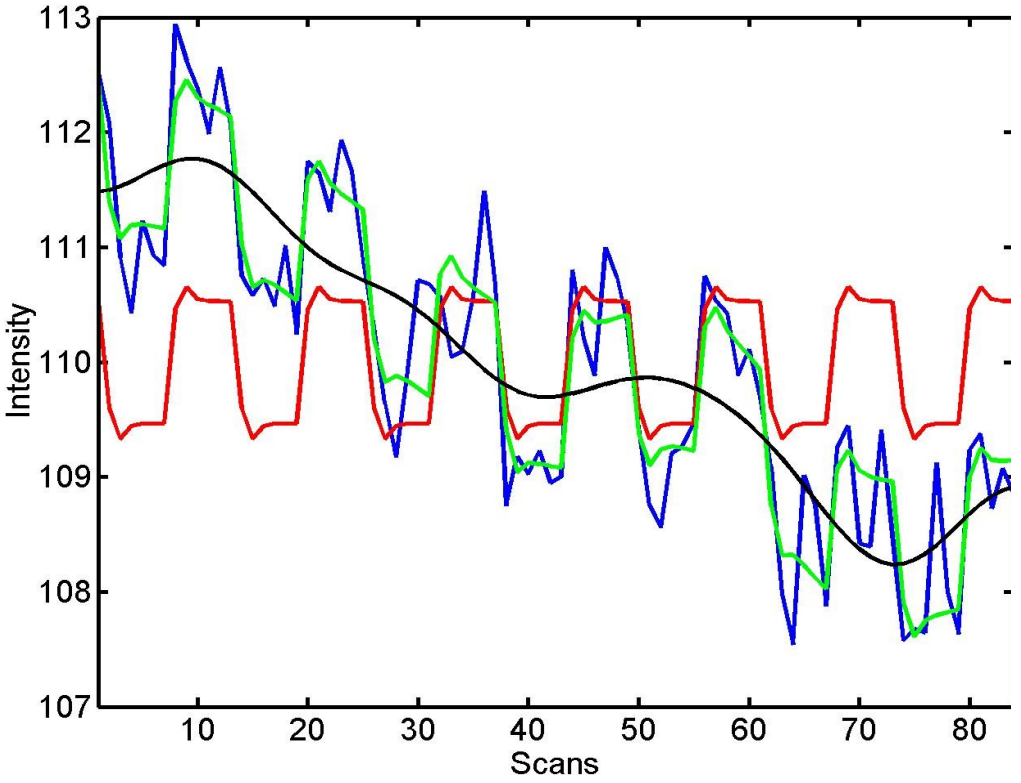
Statistical test: is cβ≠0 ?
→ Use a Student T-test.

$$T = \frac{c\hat{\beta}}{\sqrt{\hat{\sigma}^2 c\left(X^T X\right)^{-1} c^T}}$$

NB: σ^2 is the residual variance
→ any effect that can be accounted for should be included in the design matrix

From neuronal effects to timeseries of BOLD data: Include covariate in the design matrix

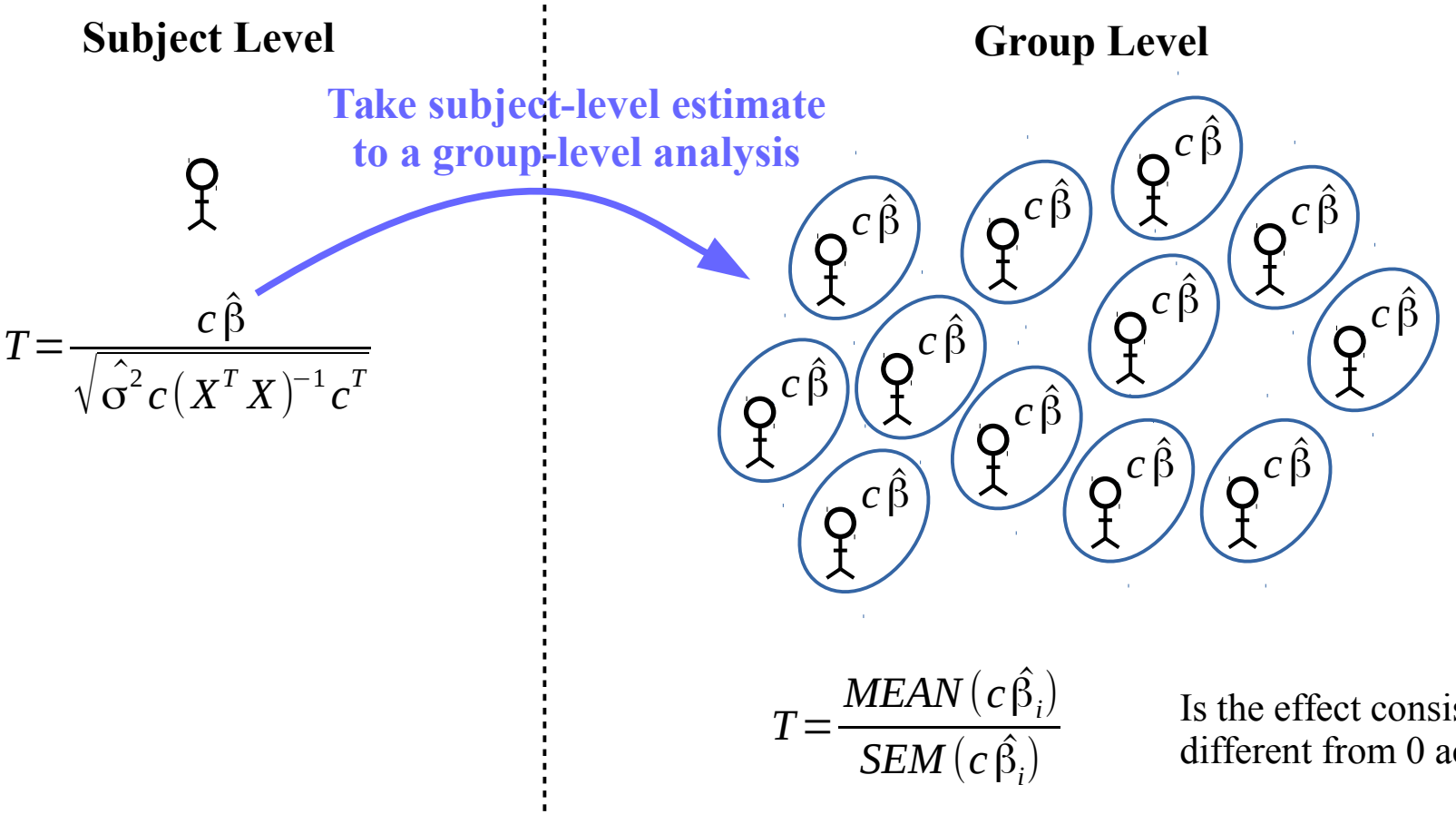




The fMRI signal is often corrupted by slow drifts (instability of the scanner)

Other “confound” variables typically included: subject's motion parameters

Subject-level and group-level analyses



NB1: this method ignores the variance (σ) of the parameters from different subjects. Mixed-models and hierarchical models can take into account the subject-level variance for the group-level inference.

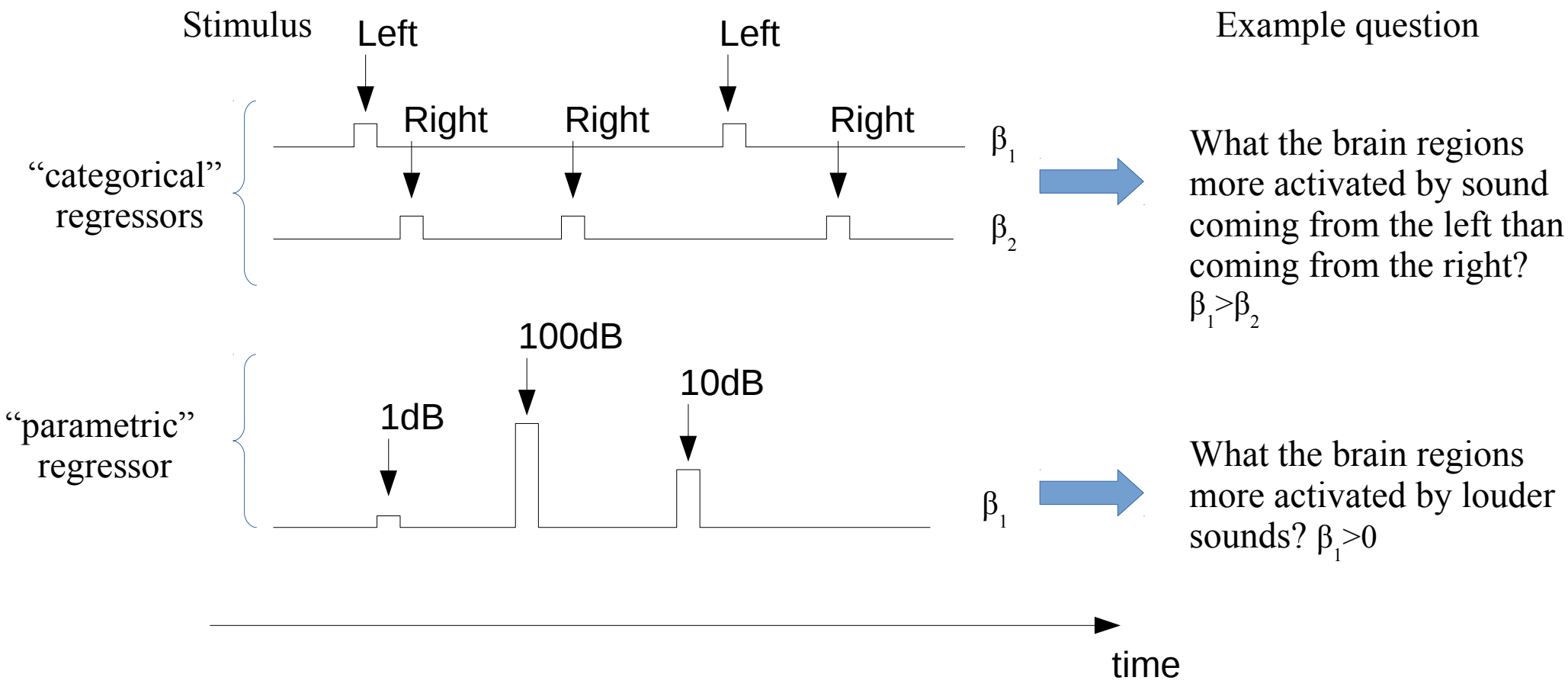
NB2: to compare different subjects voxel-wise, the anatomy first needs to be "normalized", i.e. aligned with another. Usually, they are realigned to a standard anatomical space (e.g. MNI) so that a given voxel can be compared across studies.

The flexibility of General Linear Models

The GLM approach allows different statistical methods:

- Student T-test: Is my effect $E_1 \neq 0$? Is effect $E_1 > E_2$?
- F-test (ANOVA): Is there a difference between any level of factor E_1 (one-way ANOVA)? Is there a difference between any level of factor E_1 will controlling for the effects of E_2 , … E_N (N-way ANOVA)
- T-test and F-test can be performed at the subject-level and at the group-level.

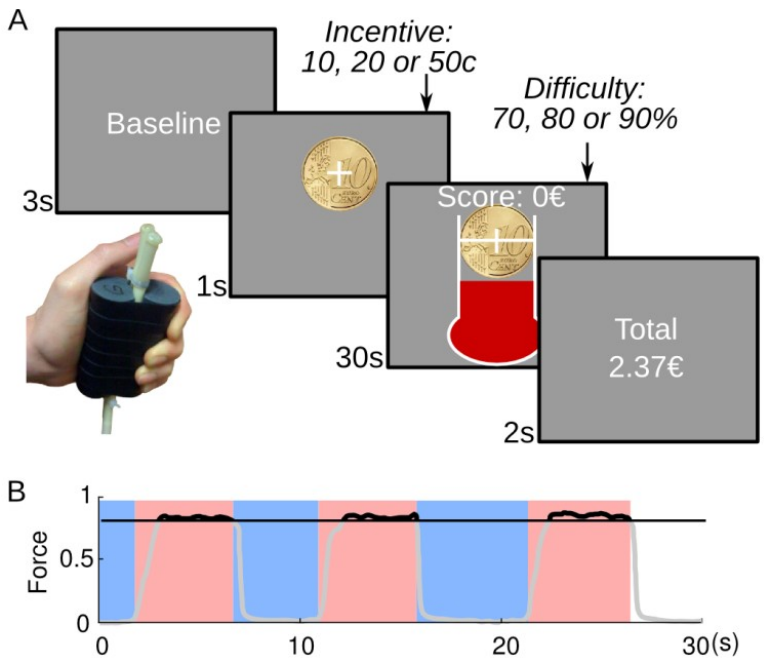
GLM: A conceptual distinction between categorical regressors and parametric regressors

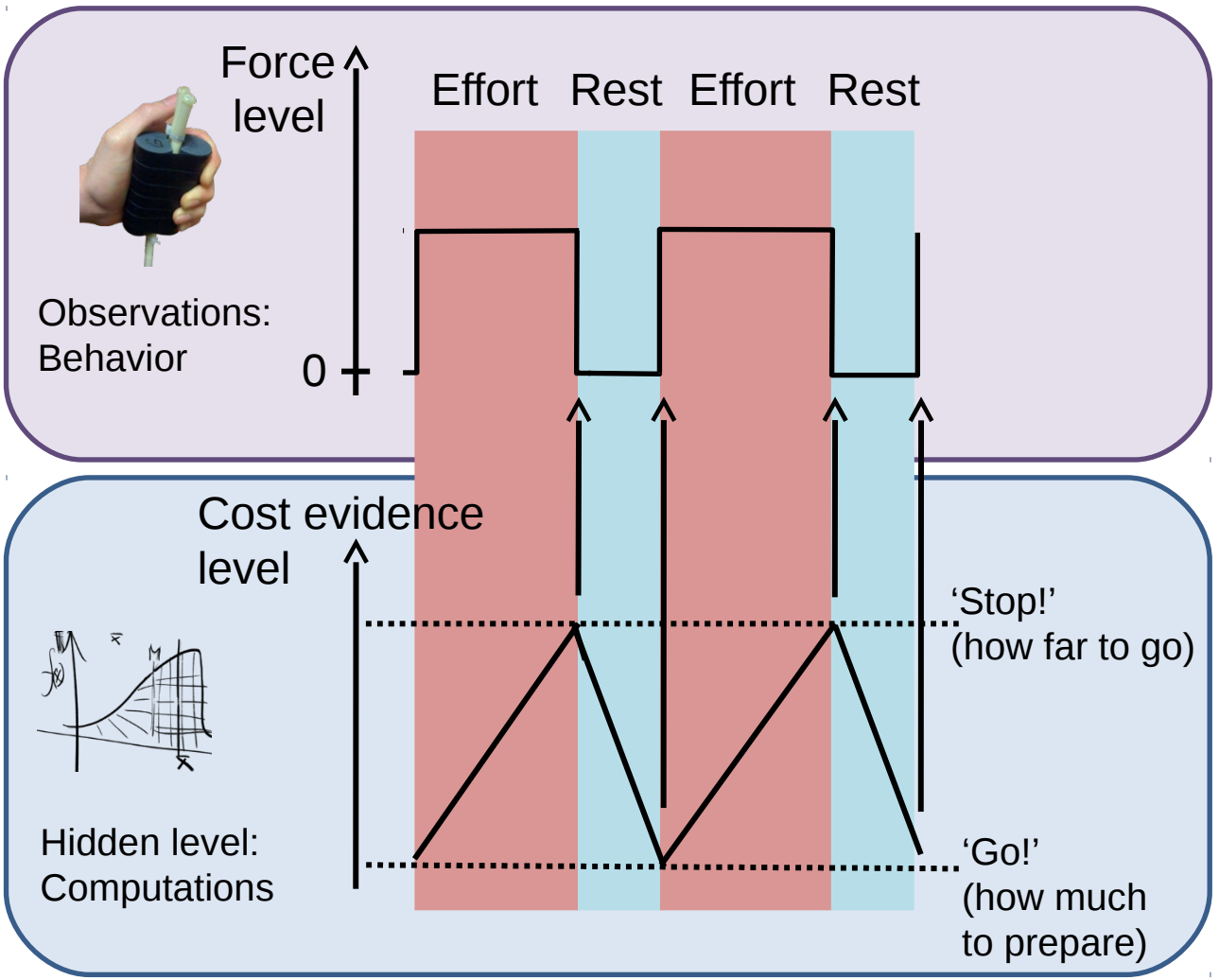


In the end, both types boil down to a regression of time-series in the GLM...

Model-based approaches to design GLM

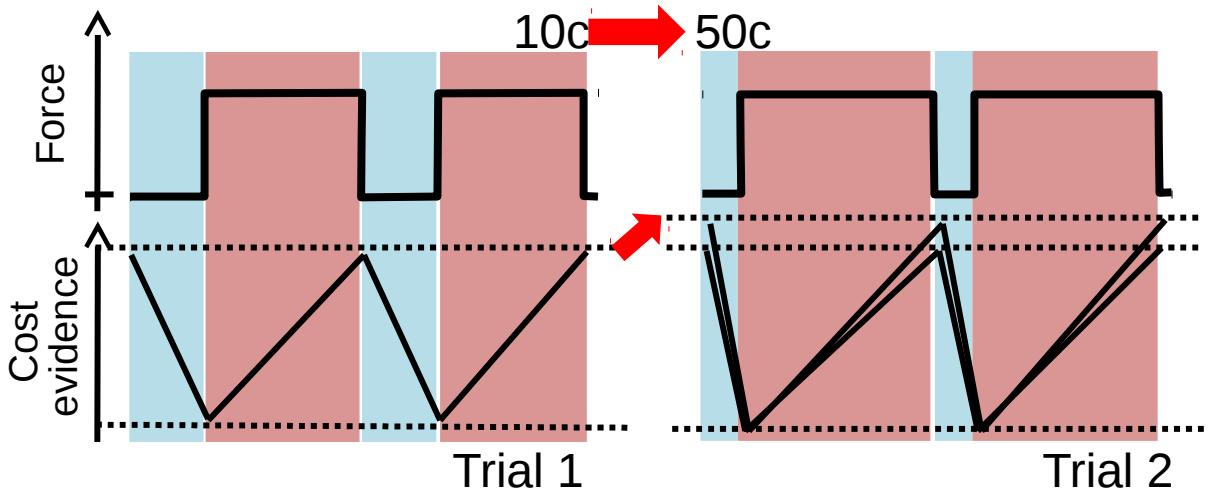
If one has a computational model (a mathematical description) of mental processes, then values computed from this model can be entered in the GLM!





Meyniel et al. PNAS 2013

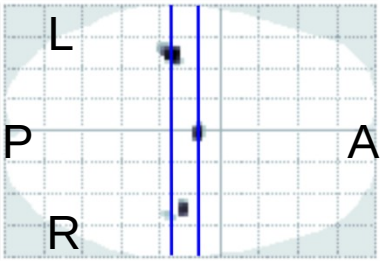
Model-based approaches to design GLM



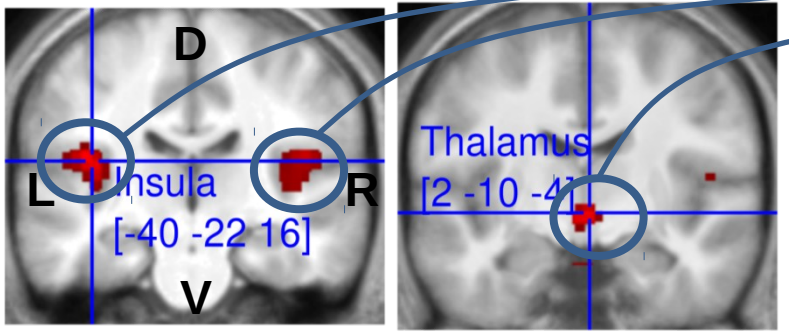
1) A neural correlate?
2) Modulation by incentive?

… Bayesian model selection

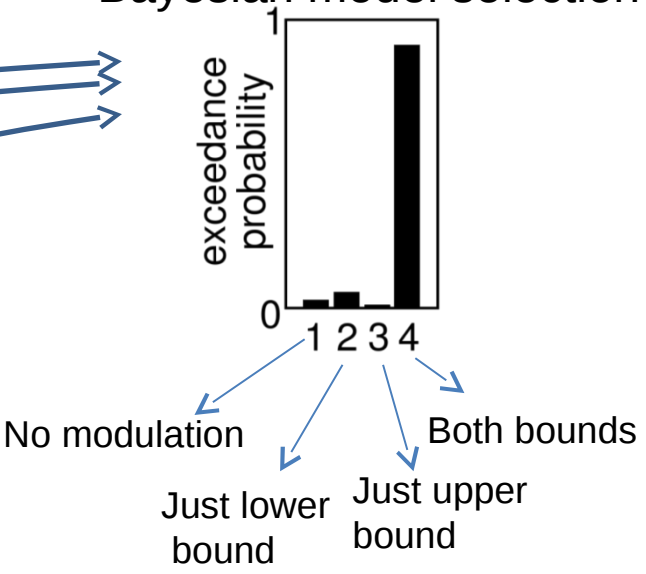
Statistical maps:



Axial projection
P<0.05 FWE
corrected over the
brain

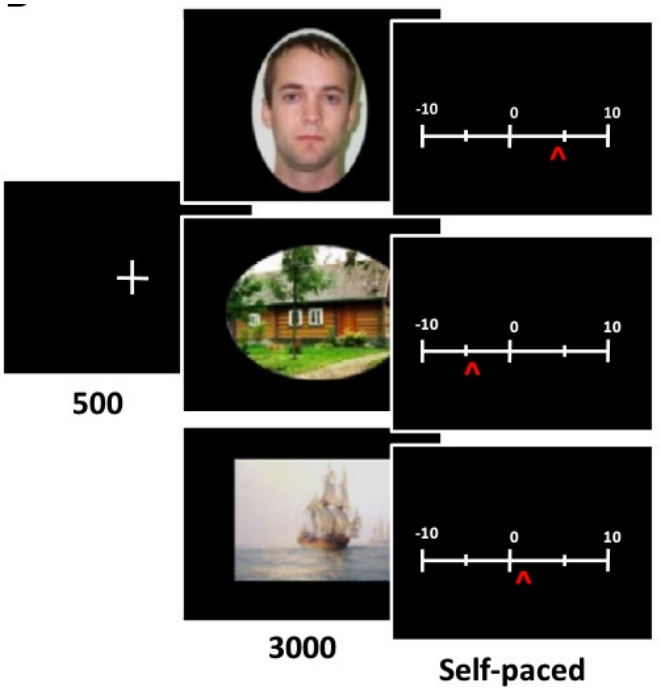


Coronal slice
P<0.001 uncor.

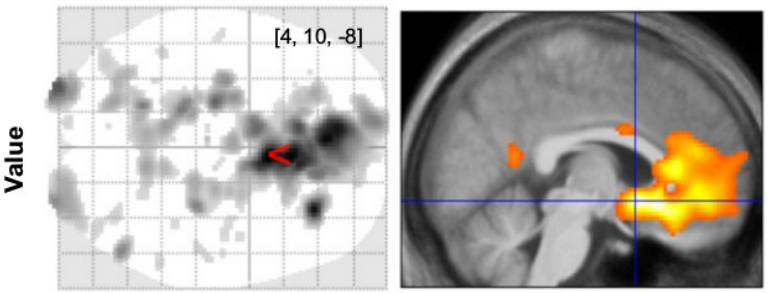


Meyniel et al. PNAS 2013

Limits in the General Linear Model

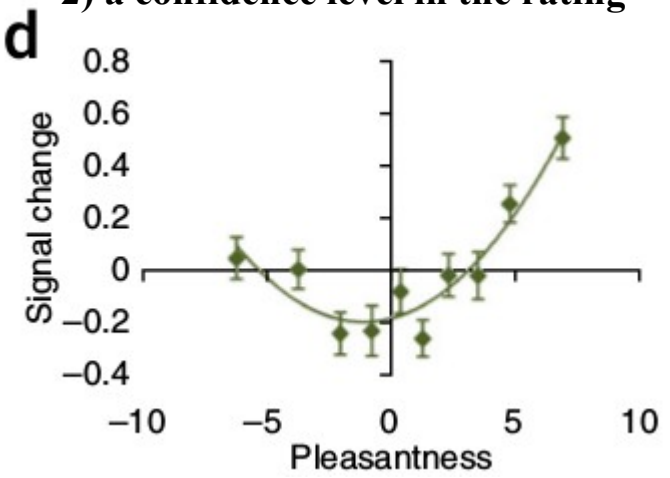


Correlation with pleasantness ratings



Lebreton et al *Neuron* 2009

Re-analysis: the significant correlation actual masks 2 effects!
1) an effect of pleasantness
2) a confidence level in the rating



Lebreton et al *Nat Neurosci* 2015

- A correlation may capture only a fraction of the effects
- The fraction captured may not be diagnostic of your model or hypothesis
- The effect you find may be driven by another mechanism than the one you think of!
- In particular, the BOLD effect smooth transient signals!

See: Wilson & Niv, Plos Comp Biol 2014

I/ From neural activity to BOLD signal: the power of a forward model

- The origin of the BOLD signal
- The heomodynamic response function
- Convolution model for BOLD data

II/ Mass univariate analysis and the General Linear Model

- Intuitions for regression
- General linear model (GLM) and design matrix
- Contrast testing the effect of interest
- Adding confounds in the design matrix
- Subject and Group level analyses
- Categorical and parametric regressors
- Model-based approaches: strengths and limits

III/ Multivariate analysis

- Multivariate information in distributed population codes
- Estimation of a classification accuracy
- Supervised learning algorithms for classification
- Origins of the multivariate information
- Comparison of univariate and multivariate approaches

VI/ The problem with multiple comparison

- Multiple testing inflates the risk of having a false positive
- Statistical methods to correct for multiple comparisons

Beyond the GLM and mass-univariate approaches

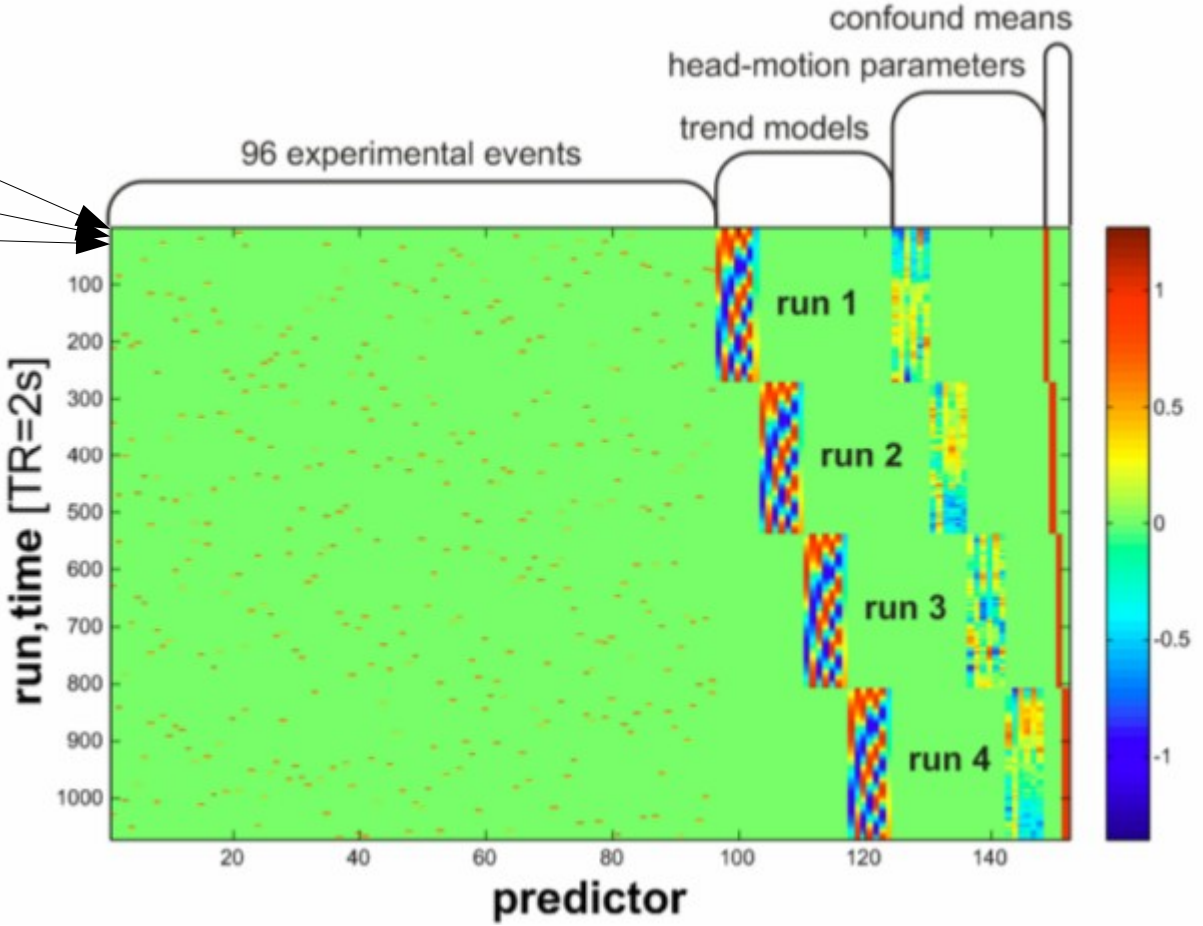
Stimulation: serial presentation of images (dog / cat)



...

time

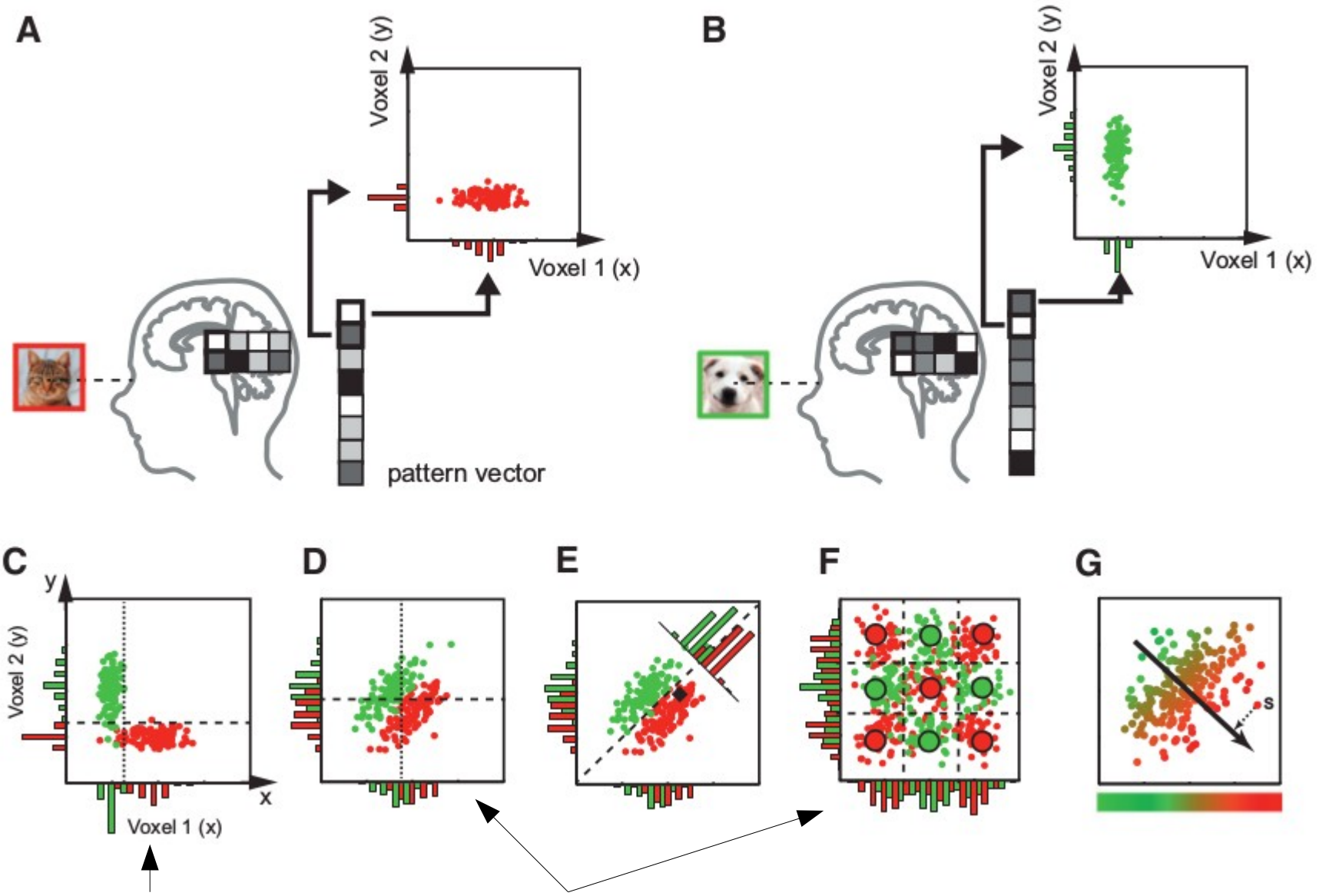
Image #1: cat
Image #2: cat
Image #3: dog
...



Question:
Does the BOLD response to a single image convey information about the stimulus?
Can we guess correctly if the image is a dog or cat based on the BOLD response only?

Image from: N. Kriegeskorte.

Beyond the GLM and mass-univariate approachs: multivariate analysis

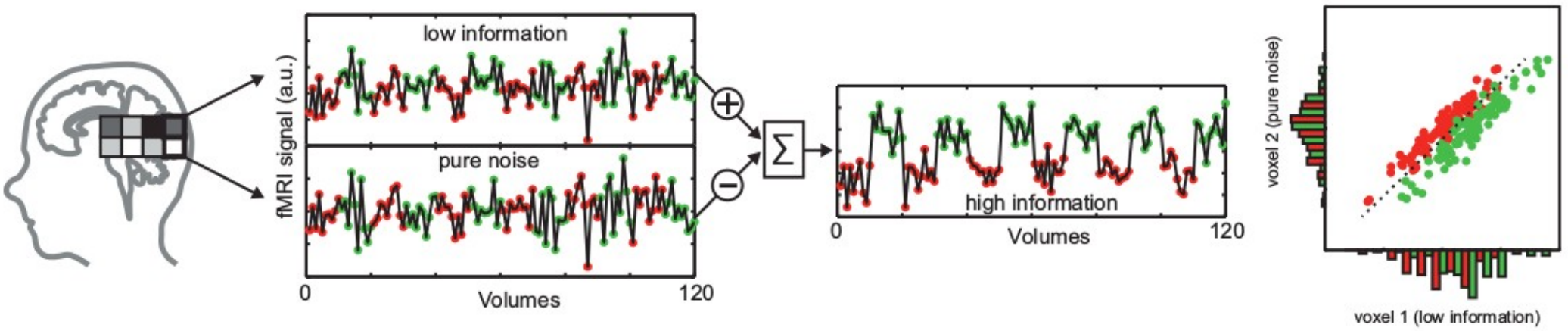


One voxel contains enough information about the the stimulus

The information about the stimulus is distributed, and hence inherently multivariate (a single voxel is not enough to decide)

Haynes, Neuron 2015

Multivariate analysis: why 2 voxels may be better than one to decode information



"Noise" is often shared between neighboring voxels due to:

- Common physiological artifacts (heart beat, respiration)
- Common scanner artifacts (drift, thermal noise)
- Common cognitive processes unrelated to the task.

Multivariate analysis can leverage on these common sources and automatically subtract them.

Haynes, Neuron 2015

Estimation of the “decoding accuracy” with a cross-validation approach

Find the best way to classify trials

Test for generalization with new trial

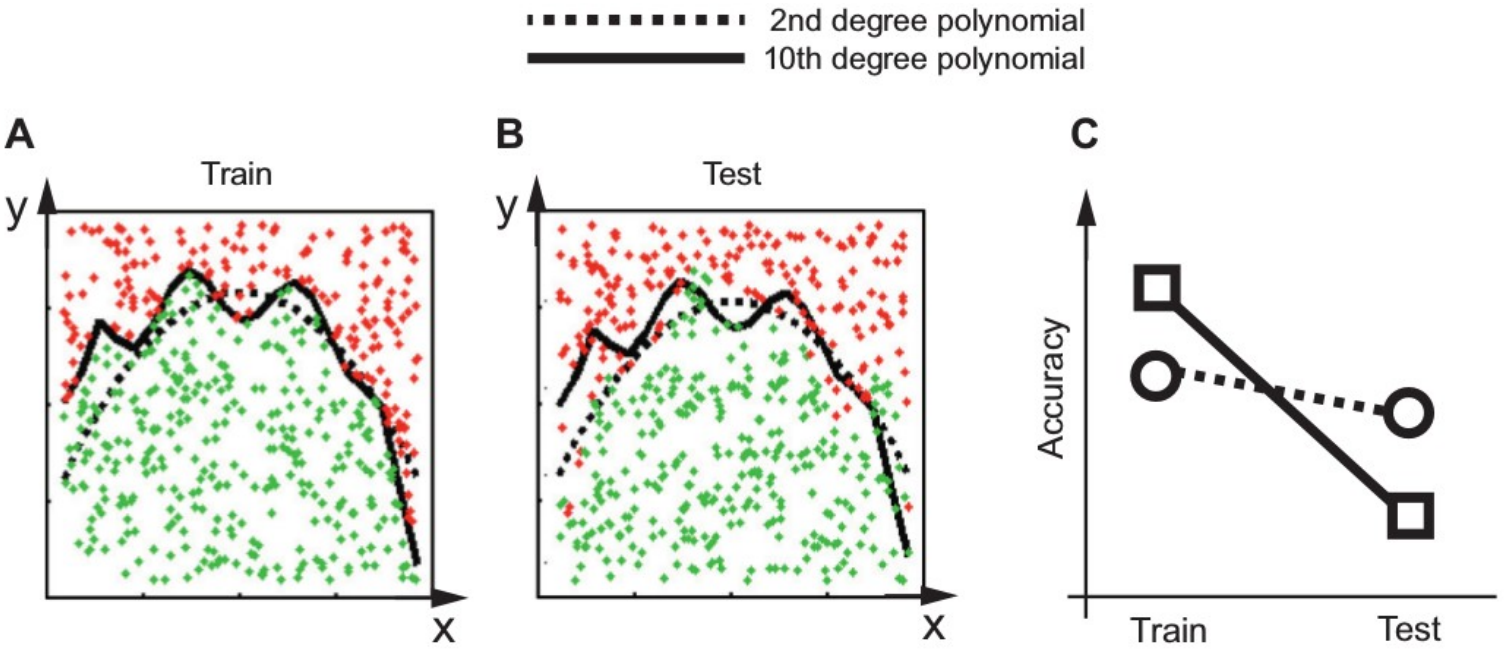
Correct guesses: 8/10
Chance: 5/10
Binomial test: p=0.01 (prob of observing value equal or higher under the null)
NB: permutation test can also be used, and are actually better since they don't assume a specific distribution.

Haynes, Neuron 2015

Estimating the "decoding accuracy" with distinct training and testing sets (cross-validation) penalizes for un-necessary complexity in the classification

Classification algorithms automatically find the features in the data across voxels that are informative about the stimulus.
Note that the information about the stimulus ("cat", "dog") is provided by the experimenter: the algorithms is therefore ***supervised***.

Example algorithms for supervised learning: linear discriminant analysis, linear support vector machine (SVM), non-linear SVM, …



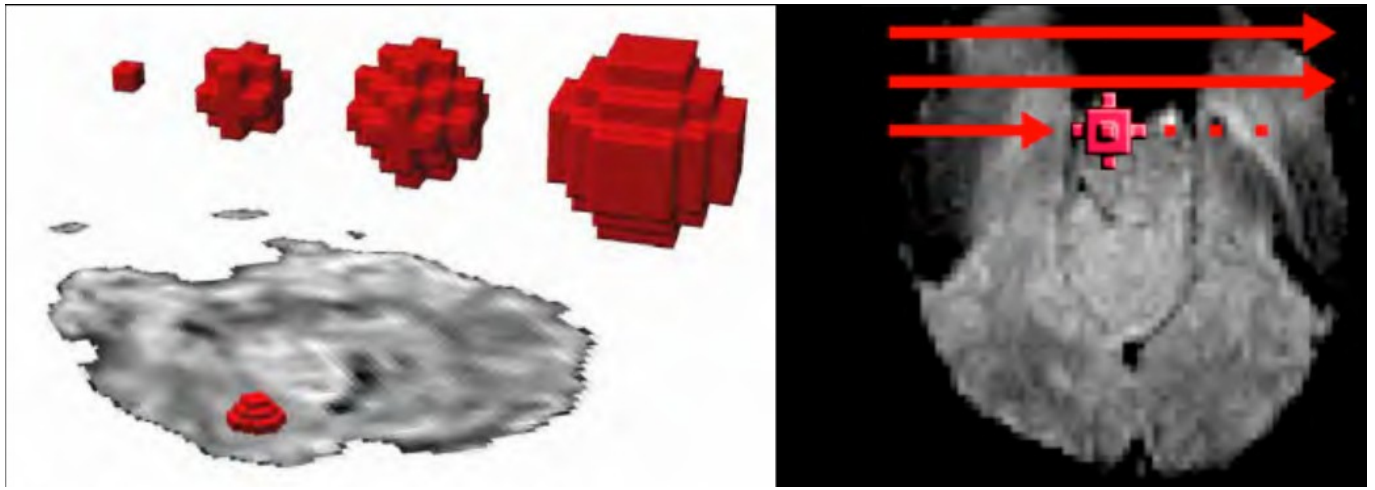
Haynes, Neuron 2015

Multivariate analysis: regions of interest and “search light” through the brain with >2 voxels for classification

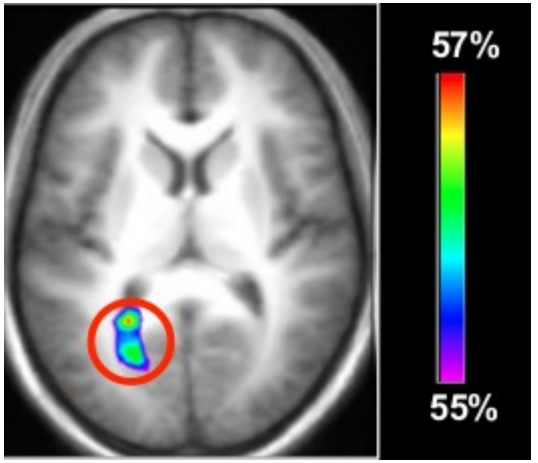
Analysis per Region Of Interest (ROI):
All voxels from a pre-defined ROI (e.g. V1) are used in the classification analysis.

Alternative: “Search ligth”
A local volume of voxels is used for the classification, and the analysis is repeated after moving the center of the volume so as to cover the entire brain.

Search light classification



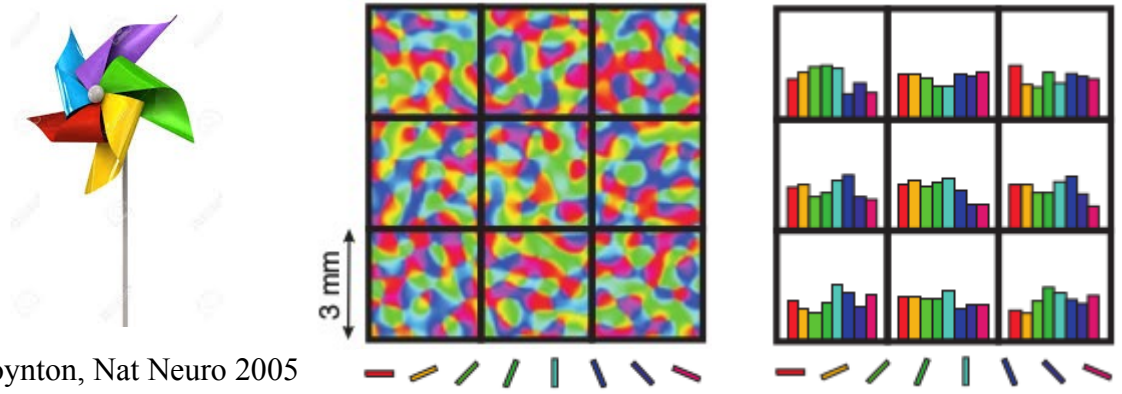
Decoding accuracy



Adapted from Jesse Rissman

Multivariate analysis: Why many voxels may be jointly informative

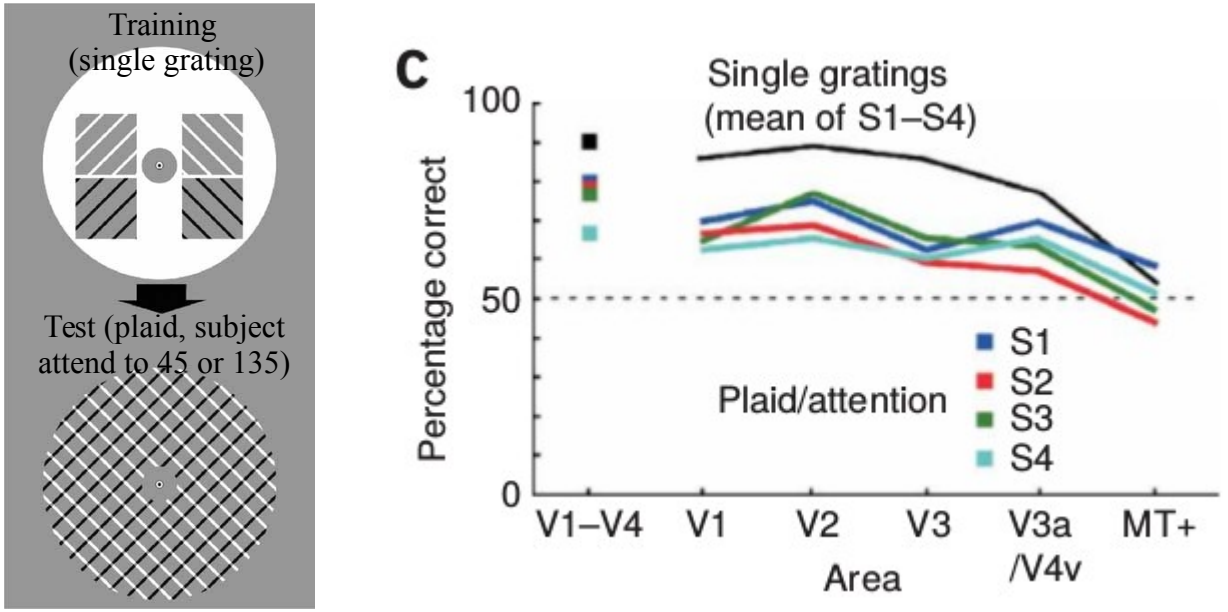
Pinwheel organization of cortical columns in the primary visual cortex (V1)



Boynton, Nat Neuro 2005

If the sampling of cortical columns tuned to different orientations by the MRI voxels is slightly biased, then differences will appear at the voxel level. An alternative explanation is that there are macroscopic biases.

By capitalizing on the level of activity of different voxels tuned to different orientations, one can decode the orientation of the stimulus.



Kamitani and Tong, Nat Neuro 2005

Decoding of fMRI data in the visual cortex identifies the orientation of the stimulus. The classification procedure extends to decoding the perceived (or attended) orientation *even when the stimulus is actually keep constant*!

Comparison of the multivariate and univariate approaches

- Look at voxels independently from one another
- Look for spatially smoothed signal
- Based on a regression approach
- One must fully specify the type of representation looked for
- Statistics: T-test, F-test (parametric or not)
- Computationally cheap. Parametric tests suffice

- Look at the information conveyed jointly by multiple voxels
- Look for spatially structured signals
- Based on a classification approach
- The classification automatically extracts the relevant features
- Statistics: classification accuracy
- Computationally expensive. Requires permutation, cross-validation

I/ From neural activity to BOLD signal: the power of a forward model

The origin of the BOLD signal
The heomodynamic response function
Convolution model for BOLD data

II/ Mass univariate analysis and the General Linear Model

Intuitions for regression
General linear model (GLM) and design matrix
Contrast testing the effect of interest
Adding confounds in the design matrix
Subject and Group level analyses
Categorical and parametric regressors
Model-based approaches: strengths and limits

III/ Multivariate analysis

Multivariate information in distributed population codes
Estimation of a classification accuracy
Supervised learning algorithms for classification
Origins of the multivariate information
Comparison of univariate and multivariate approaches

VI/ The problem with multiple comparison

Multiple testing inflates the risk of having a false positive
Statistical methods to correct for multiple comparisons

It is likely that a rare event occurs if you try multiple times

Bet 1

You throw a pair of dice.
I give you 10€ if you have a double 6.
You give me 10€ otherwise.

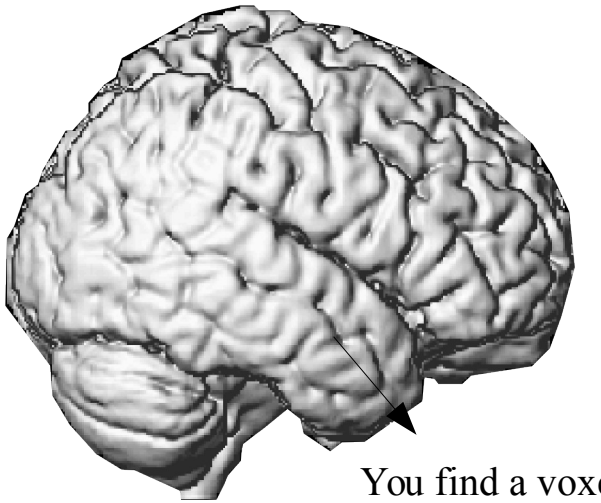
Probability that I win:
$1-(1/6)^2=97\%$

Bet 2

You throw a pair of dice 100 times.
I give you 10€ if you have a double 6 at least once. You give me 10€ otherwise.

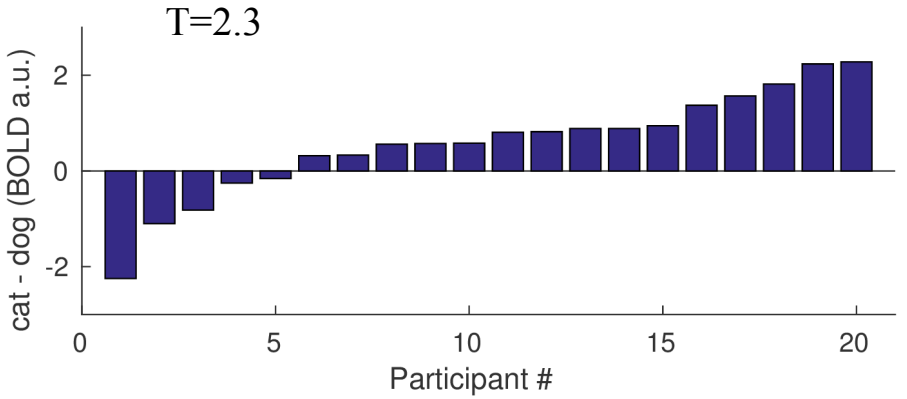
Probability that I win:
$[1-(1/6)^2]^{100}=6\%$

The problem of multiple testing across voxels



Number of voxels that can be considered as grey matter after smoothing, at a resolution of 1,5 mm: ~ 500 000.

You find a voxel with different activity for "cat" and "dogs" in 20 subjects, at p=0.016 (one-tail t-test)!

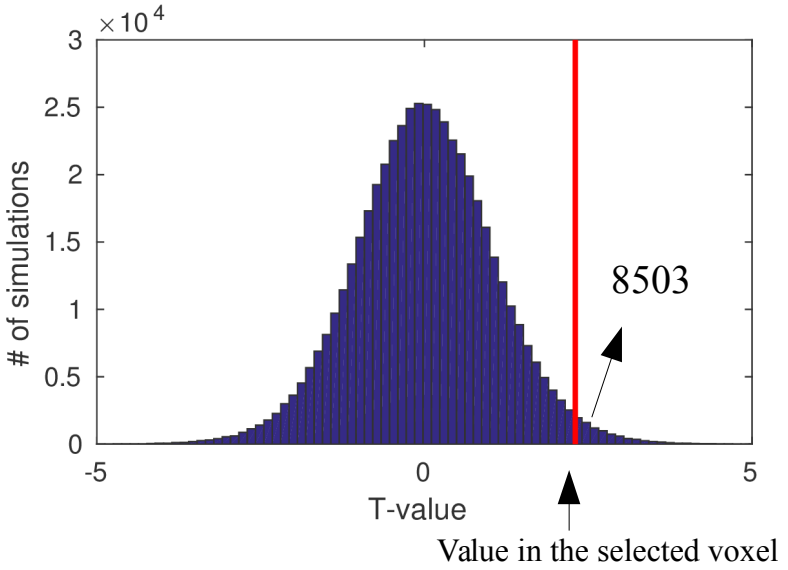


Simulation of the chance level (null effect) in that experiment

Compute the probability of a "null" effect: imagine that the cat vs. dog distinction is based on chance, rather than on a real distinction.
→ in subject #1: cats are indeed cats.
→ in subject #2: cats are actually dogs!
→ …

=> randomly re-assign labels across participant and compute the t-value.

Repeat 500 000 times.



8503 simulations out of 500000 simulations yield the same (or a higher) t-value as the one found.
NB: using the parametric distribution:
500000*p(T≥2,3)=8194.

→ we need a significance level that is protected against the inflation of false positives.

Solution: correct for multiple comparisons

The standard p-value is the probability of obtaining a test statistic (T) at least as extreme as the one that is observed (u), assuming that the null hypothesis is true:

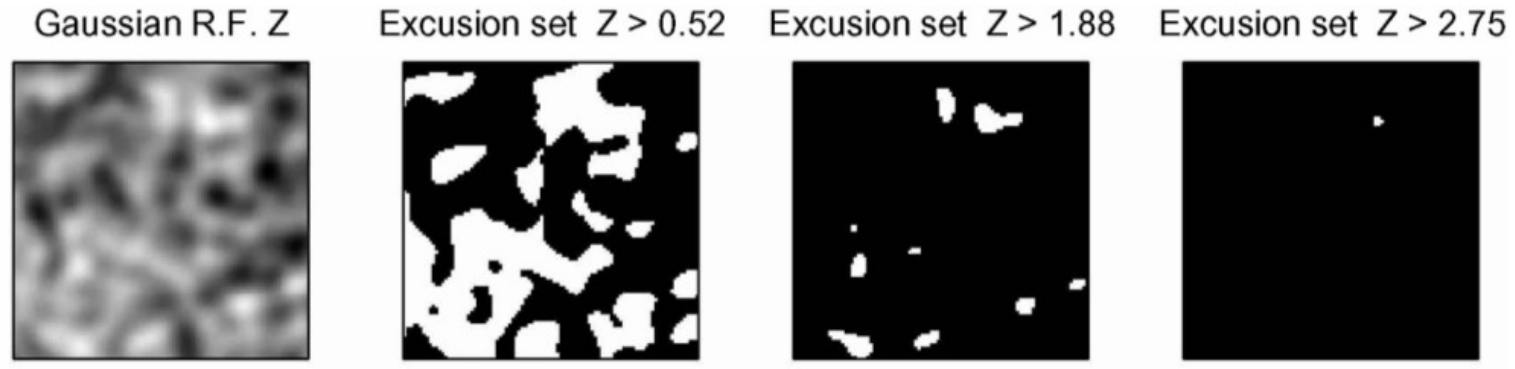
$$p(T>u|H_0)$$

The family-wise error rate: The probability of obtaining a least one test statistic (T) in a given voxel that is equal to, or more extreme than, the one that is observed (u), assuming that the null hypothesis is true. (More details in Thomas Nichols 2003 – review)

$$p\left(\bigcup_{i\in v}\{T_i\geq u\}|H_0\right)$$

Method 1: Bonferroni correction: divide the usual p-value threshold by the number of test. This is correct when all test are exactly independent, which is not the case of neighouring voxels.

Method 2: Gaussian Random Field theory. It takes smoothness into account and compute statistic with assumption about the distribution of data. (More details in Thomas Nichols 2003 – review)



Method 3: permutation approach. It automatically takes smoothness into account and does not rely on assumptions about the distribution of the data. (see Smith & Nichols 2009 Neuroimage)

And more: Holm, Sidak, Hochberg, Simes (False descovery rate), …