Bioinformatics II
Winter Term 2016/17

universitätbonn
Rheinische
Friedrich-Wilhelms-
Universität Bonn

Chapter 9:
Statistical Image Analysis

Jun.-Prof. Dr.-Ing. Thomas Schultz

URL: http://cg.cs.uni-bonn.de/iaan/
E-Mail: schultz@cs.uni-bonn.de
Office: Friedrich-Ebert-Allee 144, 53113 Bonn

January 24, 2017

9.1 Statistical Hypothesis Tests

Goal of Hypothesis Tests

- Now that we know about MR image acquisition and segmentation, we would like to **draw scientific conclusions** from our data
  - Answer questions such as:
    - Can we observe changes in the thickness of the cortex associated with learning how to juggle?
    - Which brain regions are activated when counting?

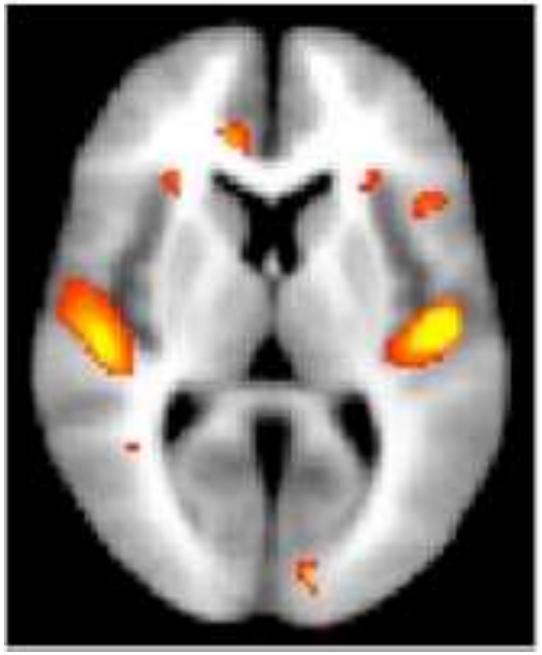
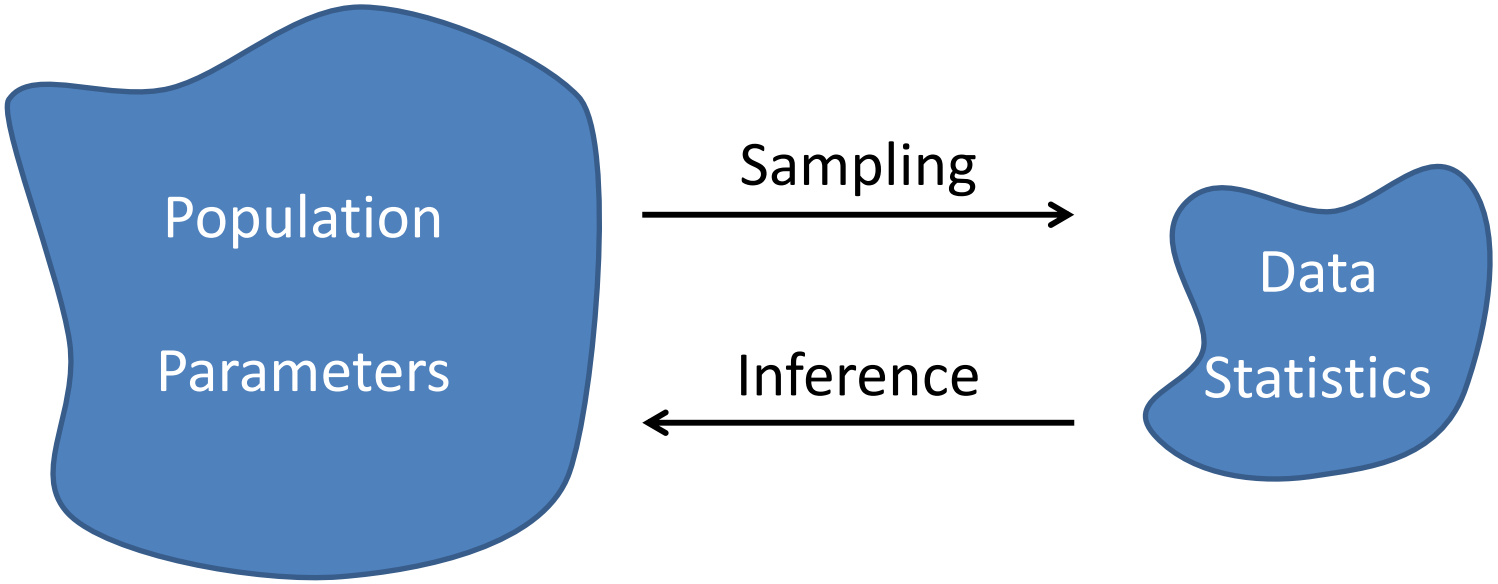


Image from Oxford brain fMRI lab

Population vs. Sample

- **Population:** Set about which we would like to make a statement (e.g., all humans)
  - (Assumed to be) governed by some statistical distribution (e.g., Gaussian)
  - Described by certain parameters (e.g., mean and variance)
- **Sample:** Set of individuals for which we have data
  - Can compute statistics on measured data
  - Would like to make statements about the parameters *of the partially observed population* that hold with a given probability



Hypothesis Testing: Basic Concept

- **Basic principle:** Reject a statement about the population parameters by showing that the observed data is highly unlikely if it holds
- **Null Hypothesis (H_0):** Statement that we are trying to reject
  - The opposite of our true research hypothesis (**alternative hypothesis** H_A: "H_0 is false")
  - Often states that some factor *does not have an effect* on the data (e.g., "The changes from learning how to juggle are not visible in MRI.")

Type I / Type II Errors

- **Type I error** (α): ***Reject*** a ***true*** null hypothesis
  - In science: reporting a false finding
  - Controlled by design of statistical test
  - Widely accepted level: α=5%

H_0	True	False
Reject	Type I	Correct
Reject	Correct	Type II

- **Type II error** (β): ***Fail to reject*** a ***false*** null hypothesis
  - In science: Not being able to report a finding even though the alternate (research) hypothesis was correct
  - "Power" of a test

Widely Used Tests

- Widely used hypothesis tests in neuroimaging include:
  - Single-sample and two-sample t-test
  - Paired t-test
  - One- and two-sided tests
  - Analysis of variance
- I assume you are already familiar with many of them

Definition of *p* value

- **Definition:** The ***p* value** is the *conditional* probability of observing the computed value of the test statistic or a more extreme value *if* H_0 *is true*
  - Often needs to be computed using software or using pre-computed lookup tables
  - Smaller *p*: Stronger evidence against H_0
  - Obtain a test with type I error α if we reject H_0 if $p<\alpha$
    - In neuroimaging, $p<0.05$ is usually taken to be "significant"

Summary: Hypothesis Testing

- Hypothesis testing rejects a null hypothesis H_0 based on the fact that the observed data is unlikely given the hypothesis.
  1. Formalize null hypothesis H_0
  2. Compute test statistic from data (e.g., t score)
  3. Compute *p* value
     - Conditional probability of test statistic taking on the computed or a more extreme value given H_0
  4. Reject H_0 if $p<\alpha$
     - Specify *p* as a measure of evidence against H_0
     - *p* does **not** provide a bound on the type I error, nor the posterior probability of H_0

9.2 Family-Wise Error (FWE) Correction

Motivation: Family-Wise Errors

- **Quiz:** Suppose you have N=10 coins in your wallet and throw away all for which a statistical test with type I error level α=0.05 rejects the null hypothesis that it is legitimate. If none of the coins is counterfeit, what is the probability of still discarding at least one of them?

$$p = 1 - (1 - 0.05)^{10} = 0.40$$

- **Lesson:** If we perform multiple tests and would like to control the rate of performing a type I error in any of them (i.e., the **family-wise error**), we have to account for the number of comparisons!
  - In brain imaging: N≈10^5 voxels, practically certain to detect false differences when using α=0.05 per-voxel

Šidàk and Bonferroni Correction

- **Šidàk correction:** An obvious way to correct for multiple comparisons is to demand a stricter α_{IND}
  - Assuming independent tests, we obtain:

$$1-(1-\alpha_{IND})^{N}=\alpha_{FW} \text{ iff } \alpha_{IND}=1-\sqrt[N]{(1-\alpha_{FW})}$$

- **Bonferroni correction:** A computationally more convenient, conservative, and quite close approximation is given by:

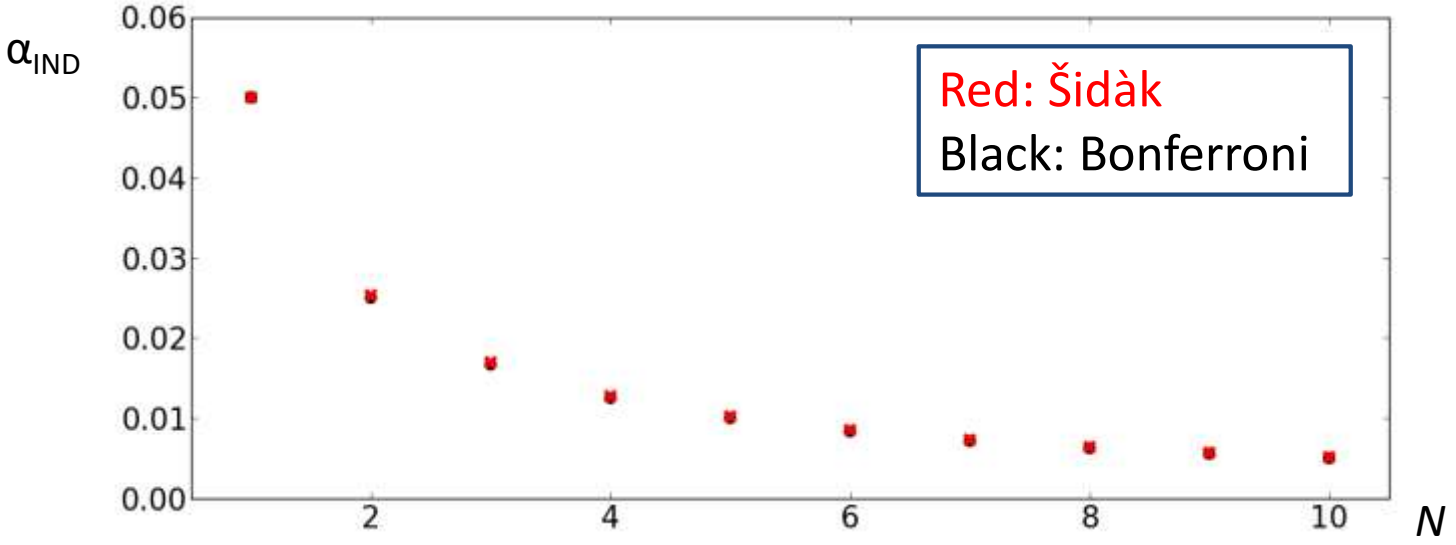
$$\alpha_{IND}=\alpha_{FW}/N$$

Bonferroni Correction: Proof

- Consider N hypotheses $H_1,\ldots,H_N$ with associated p values $p_1,\ldots,p_N$. Let T be the set of N_T true hypotheses. Let $\alpha_{IND}:=\alpha_{FW}/N$. Then,

$$FWE = P\left(\bigcup_{i\in T}\left(p_i \le \frac{\alpha_{FW}}{N}\right)\right) \le \sum_{i\in T} P\left(p_i \le \frac{\alpha_{FW}}{N}\right)$$

$$= N_T \frac{\alpha_{FW}}{N} \le \alpha_{FW}$$



Ad-hoc Correction

- In neuroimaging, tests in neighboring voxels are often strongly correlated. Thus, Šidàk and Bonferroni correction are **far too conservative** and dramatically increase type II errors.
- **Ad-hoc alternative:** Adapt threshold on test statistic until map "looks right" (e.g., $t>3$)
  - Some labs have a quasi-consensus based on their setup and experiments with predictable outcome
  - Can be useful for quick initial impression
  - More principled solution is clearly desirable to draw reproducible and reliable conclusions

Random Field Theory

- **Random Field Theory** (RFT) is a principled way to account for spatial correlations
  - Less conservative than Bonferroni correction
- **Main result:** In 3D and for sufficiently large Z (Gaussianized t),

$$p_{FWE}(Z) \approx R \times \frac{(4 \ln 2)^{\frac{3}{2}}}{(2\pi)^2} e^{-\frac{Z^2}{2}} (Z^2 - 1)$$

with number of resolution elements (RESEL)

$$R = \frac{V}{FWHM_x \times FWHM_y \times FWHM_z}$$

where the full-width half-maximum (FWHM) refers to the hypothetical Gaussian kernel that *could have been used* to achieve observed smoothness from white noise

Notes on Random Field Theory

- **Amount of smoothness** results from combined effects of intrinsic smoothness (due to physiology, PSF of acquisition) and explicit image smoothing
  - In practice, is estimated from spatial derivatives in normalized residual images
- RFT is **different** from Bonferroni correction with RESELs instead of voxels!
  - RESELS only used to make parameters in the approach more intuitive, derivation does *not* assume squared exponential decay of correlation

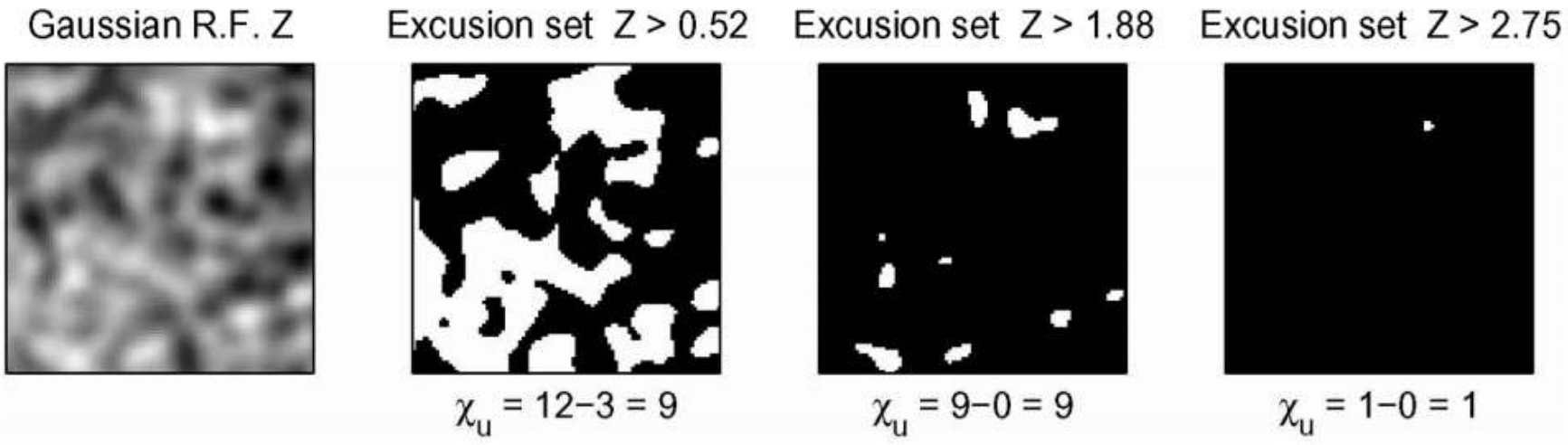
Basic Idea Behind Random Field Theory

- Random Field Theory considers the **Euler characteristic** χ_u of the excursion set

$$\{\mathbf{x} \in \Omega \,|Z(\mathbf{x}) > u\}$$

$$FWE = P\left(\bigcup_{i \in T} Z_i \geq u\right) = P\left(\max_i Z_i \geq u\right)$$

$$\approx P(\chi_u > 0) \approx E[\chi_u] \approx \frac{V\sqrt{|\Lambda|}}{(2\pi)^2} e^{-\frac{Z^2}{2}}(Z^2 - 1)$$



Images from [Nichols/Hayasaka 2003]

Cluster-based Testing

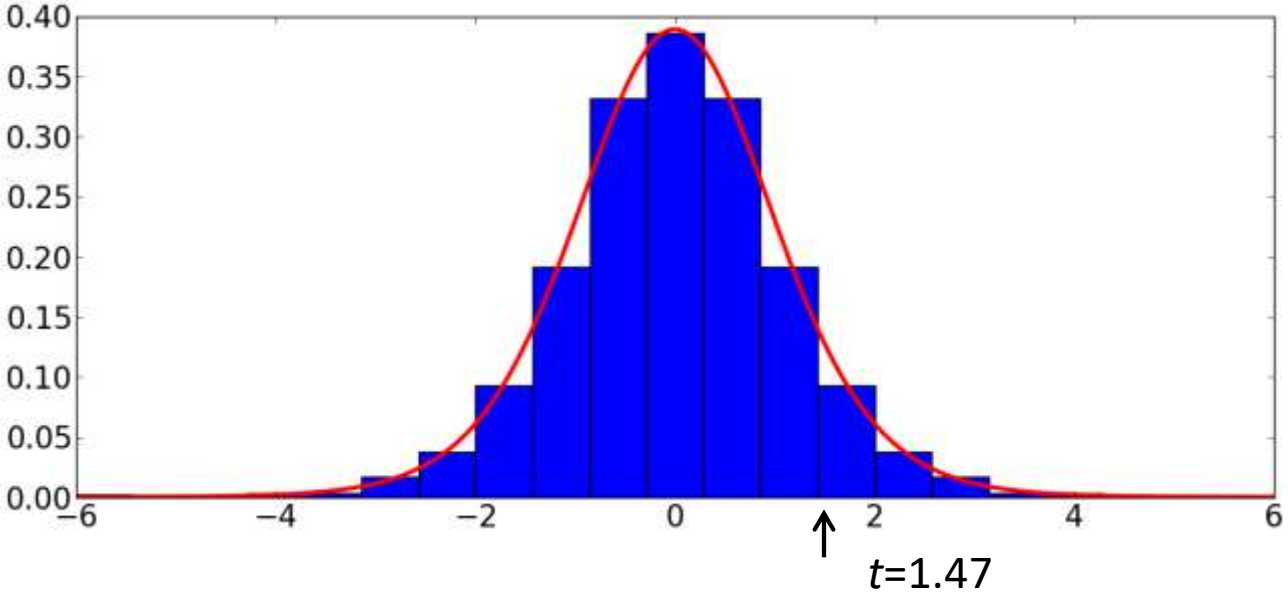
- So far, Random Field Theory still works voxel-wise
- **Alternative:** Consider size of connected regions
  - Most true effects extend over some spatial region, isolated voxels are likely to be false positives
- **Approach:**
  1. Form clusters by heuristically thresholding statistical maps at some level (e.g., $t>3$)
  2. Formal testing can be done **cluster-wise**
     - Size (number of voxels) or mass (sum of scores) serve as test statistics
     - **But:** How to determine null distributions?

Permutation-based Testing

- **Permutation methods** are widely used in neuroimaging to convert test statistics into *p* values
  - **Null hypothesis:** The assignment of individuals to groups (e.g., healthy vs. patient) is unrelated to the value of the test statistic
  - Null distribution can be built by **permuting the group labels** w.r.t. the measurements and recomputing the test statistic for each configuration
  - *p* value given by location of test statistic for true label assignment w.r.t. null distribution

Illustration: Permutation-Based *p* Value

- Example measurements for two-sample t-test:
  - $v_i^C = \{1400, 1220, 1280, 1360, 1290, 1350\}$
  - $v_i^{NC} = \{1190, 1210, 1310, 1370, 1250, 1230\}$
- The 12 measurements can be labelled in $\binom{12}{6} = 924$ ways, leading to the following distribution of *t* scores:



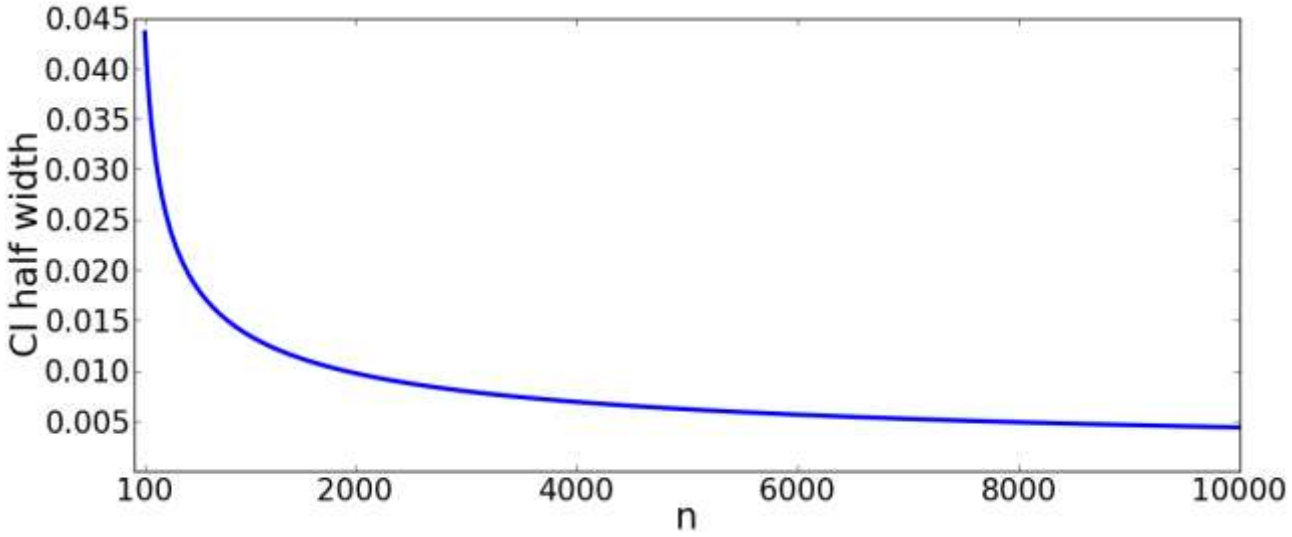
Red: Student's t-Distribution
p=0.086 (one-sided)
Blue: Permutation-based Distribution
p=0.091

- **Note:** Based on 924 possible permutations, no *p* values smaller than 1/924≈0.001 will be computed

Required Number of Permutations

- **In practice:** With more realistic sample sizes, total number of permutations too large to enumerate, instead perform a random subsample
  - **Consequence:** Repeating the analysis of exactly the same data can lead to a different *p* value!
  - Randomly performing *n* out of a sufficiently large number of possible permutations produces the following 95% confidence interval (CI) of *p*:

$$p \pm 2\sqrt{p(1-p)/n}$$



Permutation-/Cluster-Based FWE Correction

- **For each permutation** of group labels:
  - Compute per-voxel test statistic (e.g., t-Test)
  - Threshold map at some level, compute clusters
  - Store the largest / heaviest cluster in the full brain
    - This is the relevant cluster to control FWE rate!

- **For each cluster** found using the true labels:
  - Compute per-cluster “FWE-corrected” *p* value by comparing size / mass to the null distribution

Reporting Cluster-Based Results

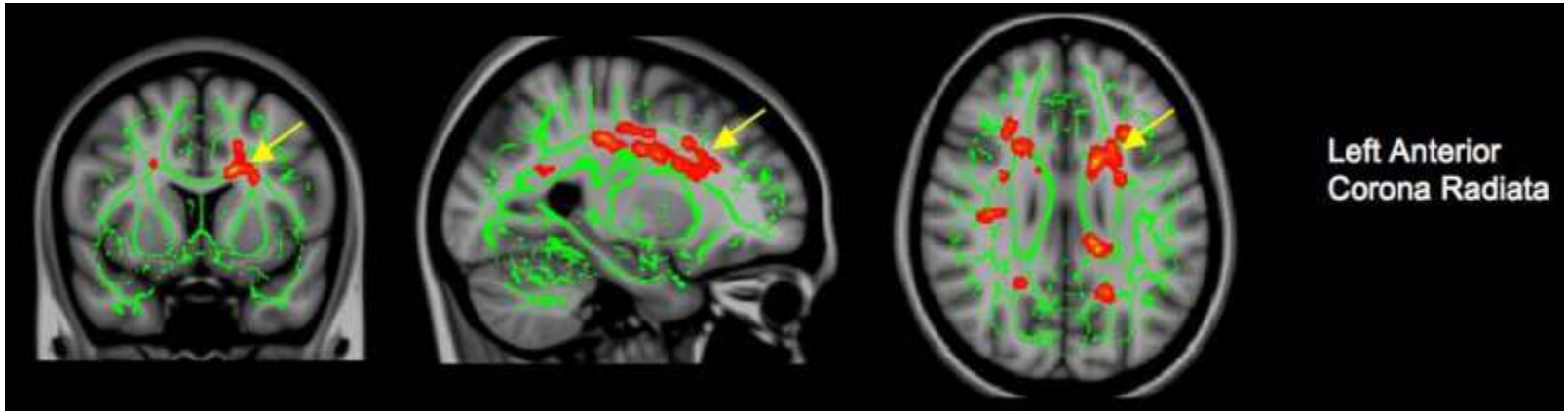
- Example: Table from [Jung et al. 2010]:

Table 3 DTI differences between NPSLE patients and controls

FA Controls > NPSLE

Voxels	p-value	Max-t	MNI X	MNI Y	MNI Z	Approximate white matter tract
266	0.001	3.135	30	17	18	Right superior longitudinal fasciculus
223	0.001	3.242	11	-34	25	Splenium of CC
172	0.003	3.408	-8	5	25	Body of CC
165	0.003	3.028	-26	28	12	Left anterior corona radiata
120	0.006	3.006	34	-41	28	Right superior longitudinal fasciculus

- Brain map from [Jung et al. 2010]:
  - Map *t* values, restricted to significant clusters



Threshold-Free Cluster Enhancement

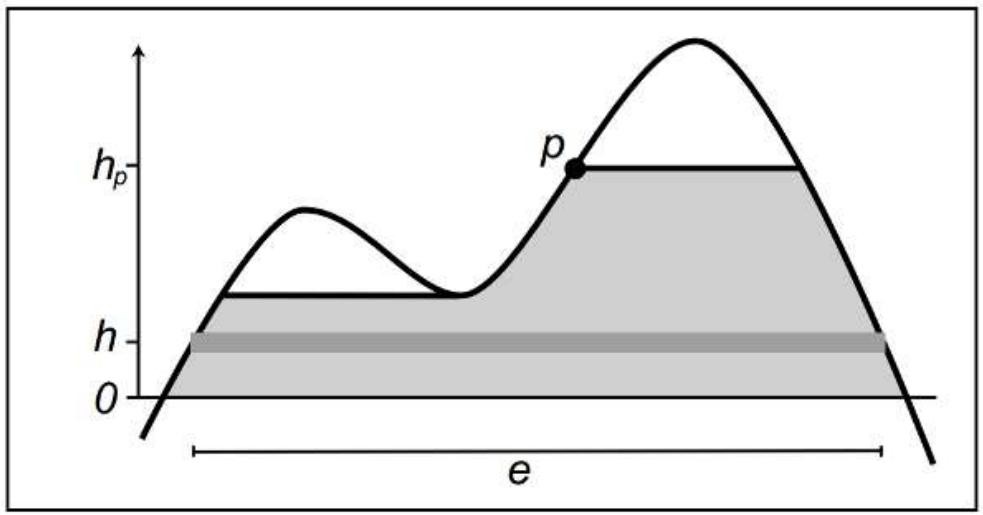
- **Remaining drawbacks** of cluster-based testing:
  - No principled way to select cluster-forming threshold
  - Results may change considerably based on threshold
  - A few voxels might cause clusters to split or merge
  - Pre-smoothing commonly applied, but what bandwidth?
  - Center of gravity no longer an adequate description of localization if clusters are large
- Goals of **threshold-free cluster enhancement (TFCE)**:
  - Boost belief based on agreement in spatial neighborhoods
  - Avoid having to set an arbitrary threshold
  - Still perform testing per-voxel

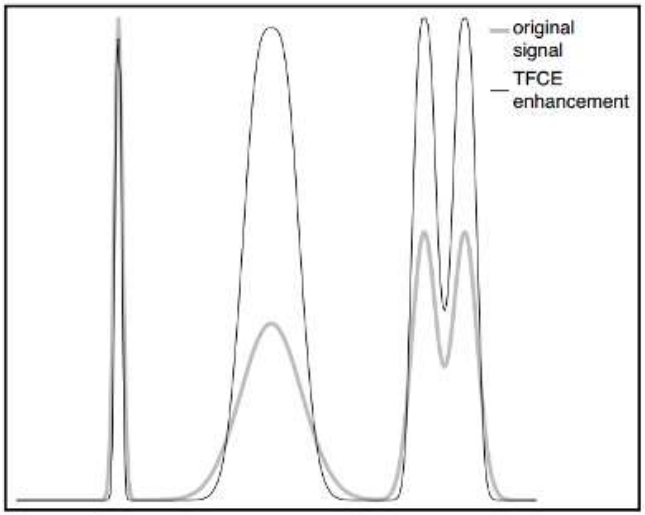
Basic Idea of TFCE

- **"Trick" in TFCE:** Perform image transformation that boosts values (e.g., *t* scores) that are surrounded by other large values

$$TFCE(p) = \int_{h=h_0}^{h_p} e(h)^E h^H dh$$

  - Recommended: E=0.5, H=2





Images from [Smith/Nichols 2009]

Properties of TFCE

- TFCE only enhances **positive values,** sets negative ones to zero
  - Corresponds to one-sided t-Test
  - If desired, repeat test with negated t scores (account for multiple comparisons!)
- Unlike Gaussian smoothing, TFCE **preserves locations of local maxima**
- **Permutation-based testing** using TFCE:
  - Compute per-voxel test statistic (e.g., t-Test)
  - Apply TFCE to the map of t values
  - Store the largest value in the full brain
  - Testing compares per-voxel TFCE value to the null distribution

Example: Cluster-based vs. TFCE

- *Example:* Systemic Lupus Erythematosus (SLE)
  - Autoimmune disease; would like to better understand difference between
    - NP-SLE: Neuropsychiatric variant, neural symptoms such as seizures or cognitive decline
    - non-NP-SLE: SLE without such symptoms
  - Using cluster-based analysis with cluster-forming threshold t>3, [Jung et al. 2010] conclude that NP-SLE (but not non-NP-SLE) leads to a significant difference in brain structure
    - Might suggest that non-NP-SLE "does not affect the brain" (even though they did not test/claim this!)

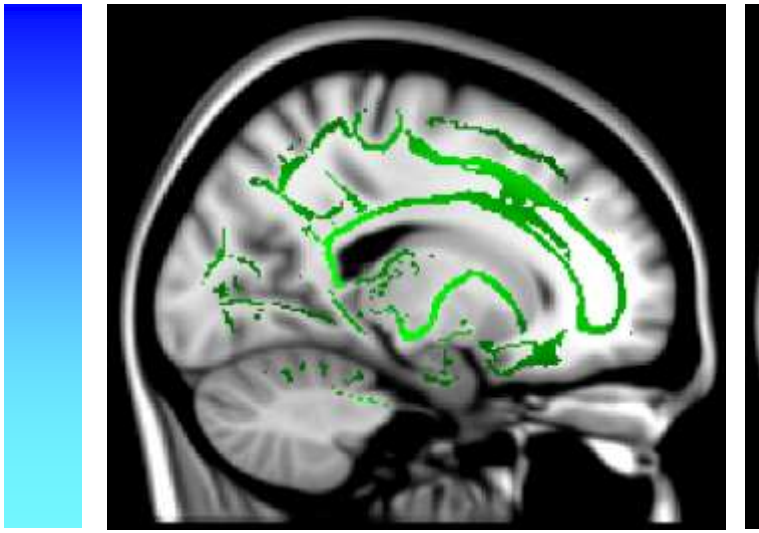
Cluster-based vs. TFCE in SLE

- Own results (based on different data):
  - Cluster-based analysis replicates Jung et al.
  - TFCE finds significant difference between healthy controls and non-NP-SLE!
    - Extent of affected regions much smaller
    - Suggests that even non-NP-SLE affects the brain, but small enough damage does not (yet) lead to neural symptoms

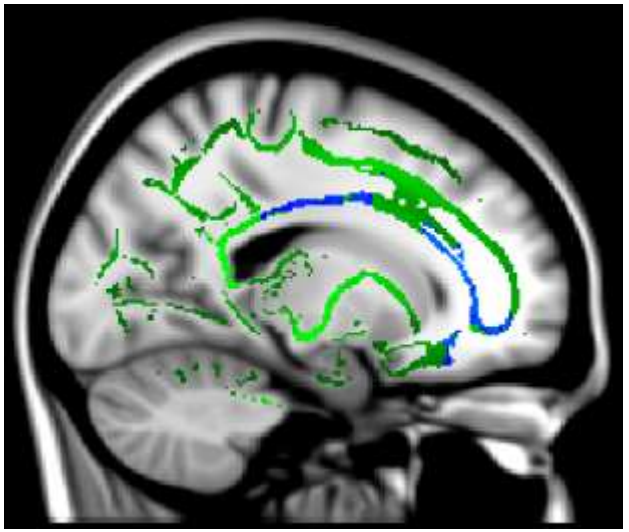
p=0.05

FWE-corrected
One-sided:
non-NP-SLE FA < healthy FA

p=0



Cluster-based



TFCE-based

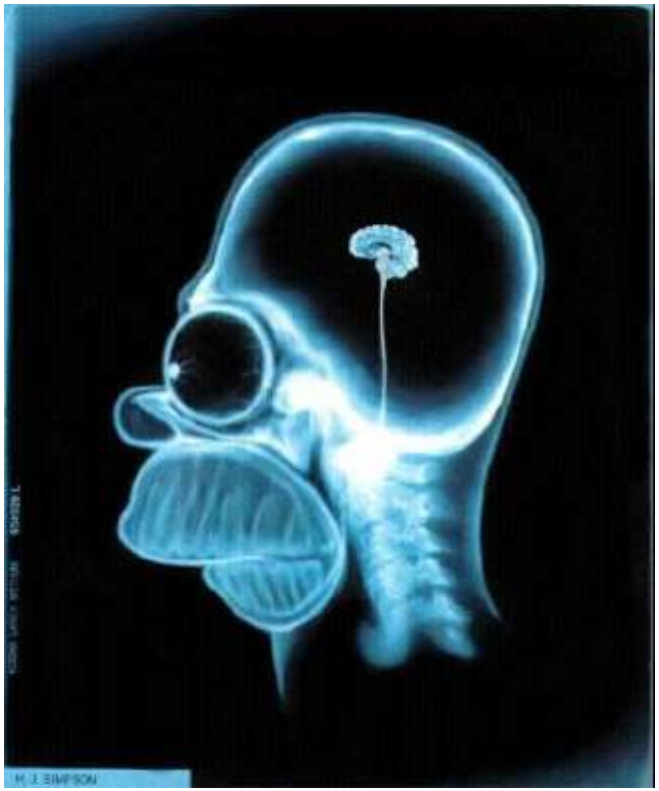
[Schmidt-Wilcke et al., NeuroImage: Clinical 2014]

Summary: FWE Correction

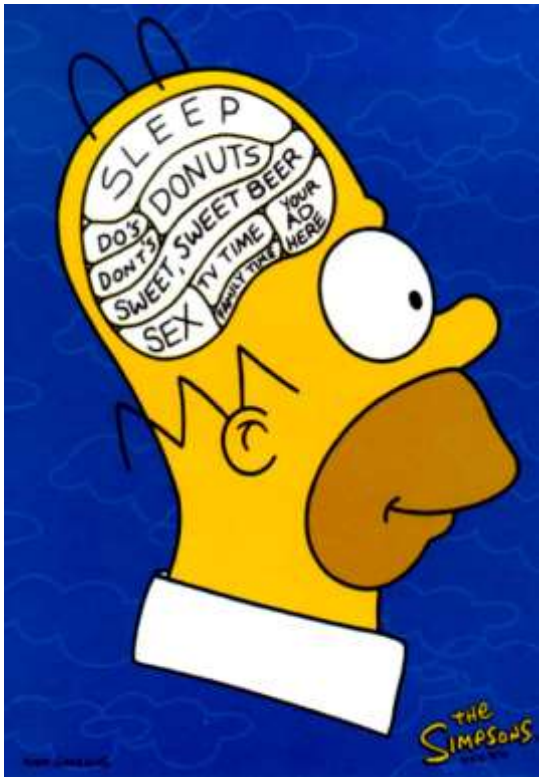
- **Type-I errors accumulate** when performing multiple hypothesis tests
- **Random Field Theory** provides a way to correct for this while accounting for smoothness
  - Less conservative than Bonferroni correction
- Another solution is to boost confidence using **information from neighboring voxels**
  - Cluster-based
    - Cluster-forming threshold, permutation-based *per-cluster* FWE corrected *p* value
    - Random Field Theory for clusters should be avoided
  - Threshold-Free Cluster Enhancement (TFCE)
    - Nonlinear peak enhancement, permutation-based *per-voxel* FWE corrected *p* value

9.3 Functional MRI

Structural vs. Functional MRI



Structural MRI studies brain anatomy

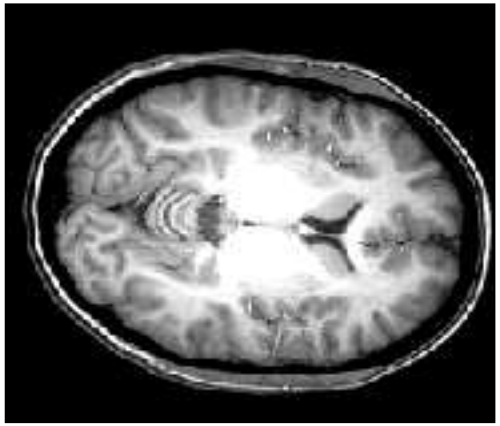


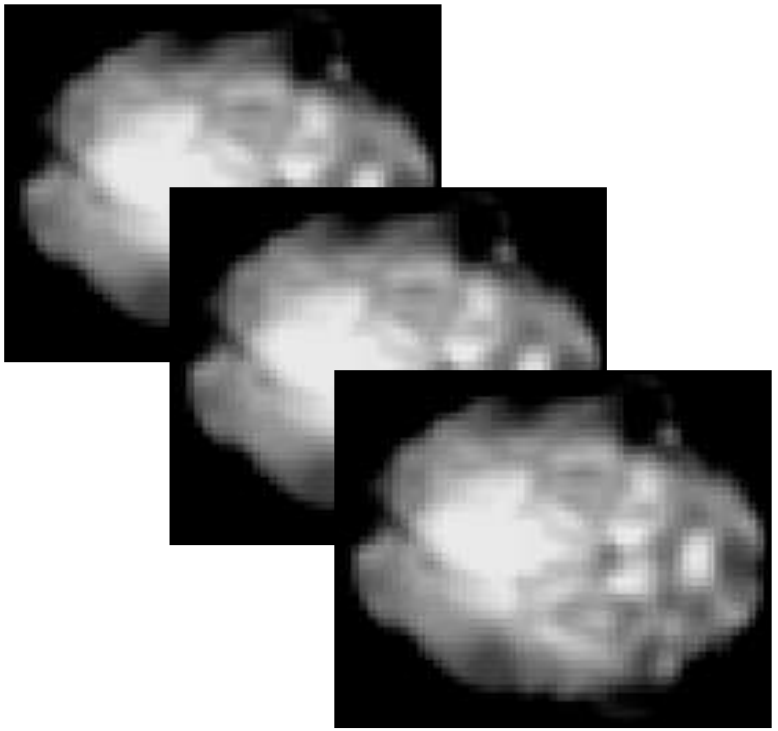
Functional MRI studies brain function

Slide from Jody Culham

fMRI: The Raw Data

- In **structural MRI,** we take one high-resolution image (e.g., 1x1x1 mm^3)
- In **functional MRI,** we repeatedly (e.g., every 2 sec) take low-resolution images (e.g., 3x3x5 mm^3)

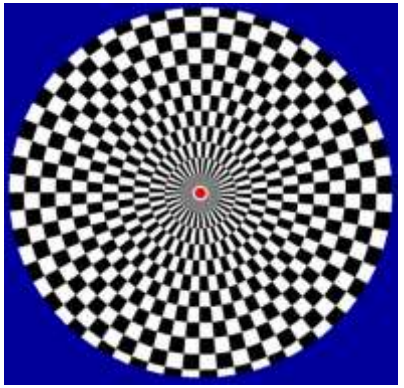


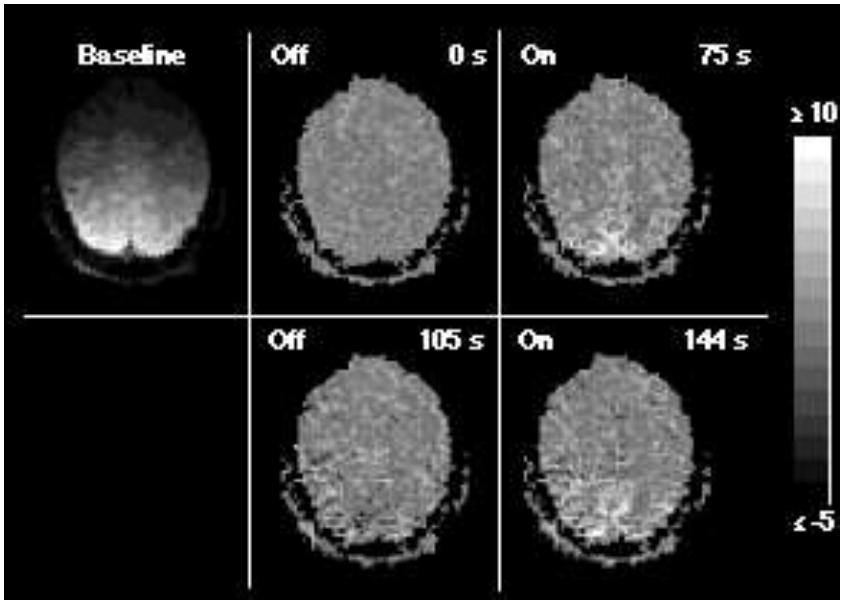


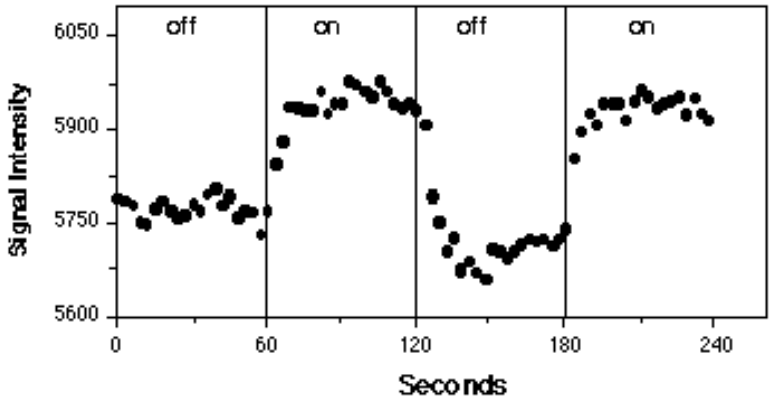
Images from Jody Cullham

fMRI: Activations

- In fMRI, an **activation** is measured by taking the difference between the signal for different tasks
- **Example stimulus:** Checkerboard
  OFF (60 sec) – ON (60 sec) – OFF …
- **Example MR response:**



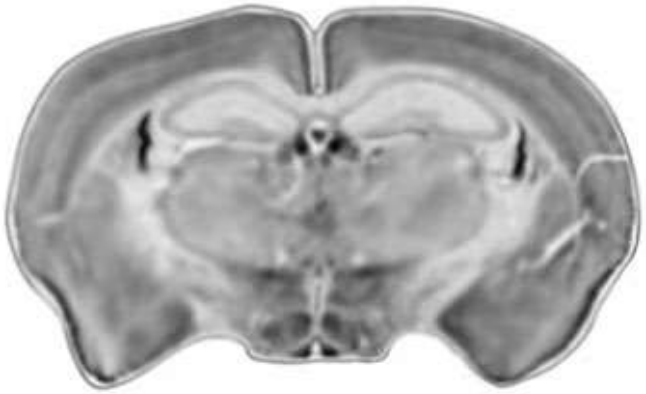




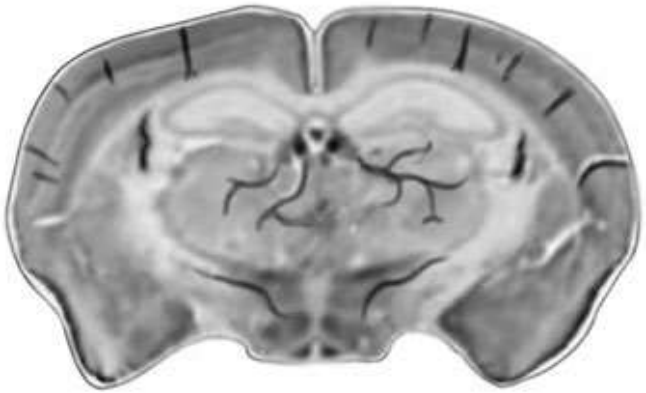
Images from Kwong et al., 1992

Where Does the Signal Come From?

- **BOLD effect:** MR intensity is **B**lood **O**xygenation **L**evel **D**ependent
- **Oxygenated Blood**
  - **Diamagnetic:** (weakly) counteracts the local magnetic field
  - Effect similar to water
  - (Almost) no change in signal
- **Deoxygenated Blood**
  - **Paramagnetic:** (slightly) enhances the field
  - Decreases T_2^*
  - Attenuates the MR signal



rat breathing pure oxygen



rat breathing normal air

Images from Ogawa et al., 1990

Temporal Complexity of BOLD Response

MR response to flashing visual stimulus

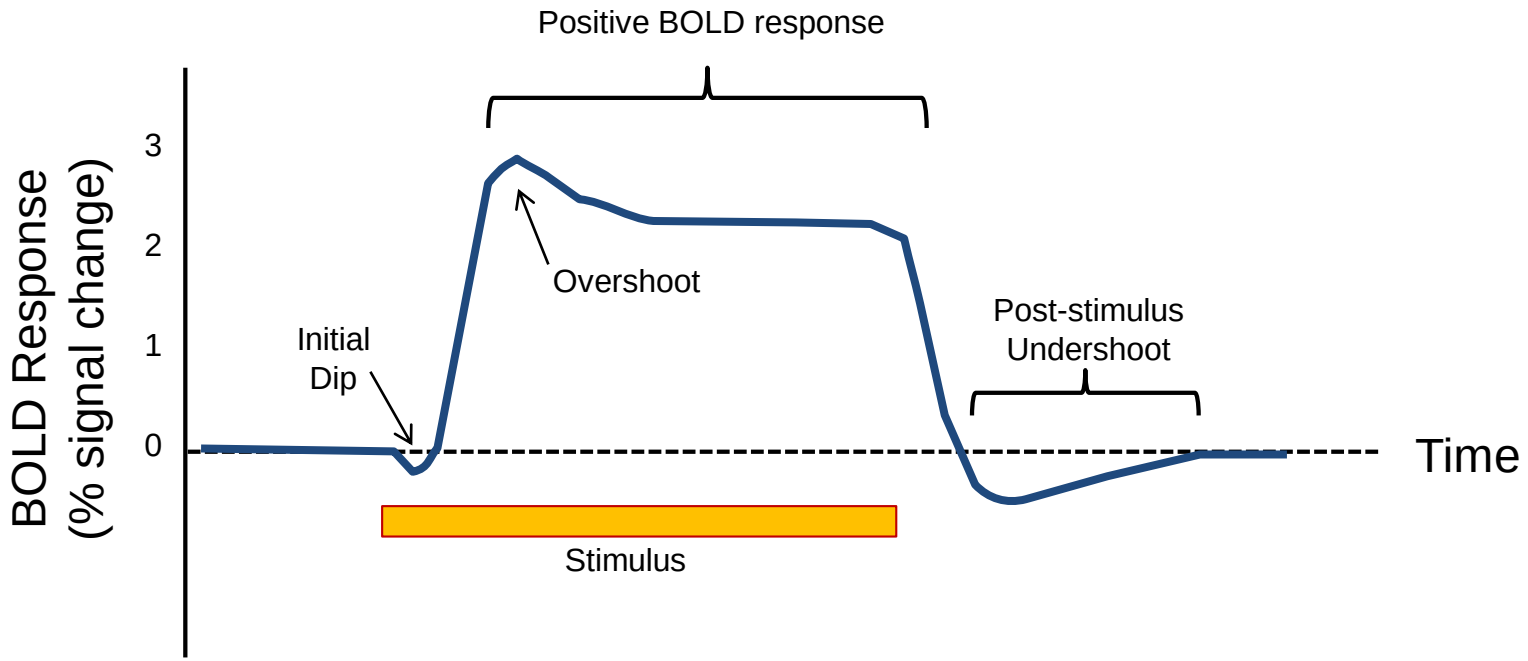
Red: Increased MR Signal

Blue: Decreased MR Signal

Images from Hu et al., 1997

How Does BOLD Relate to Neuronal Activity?

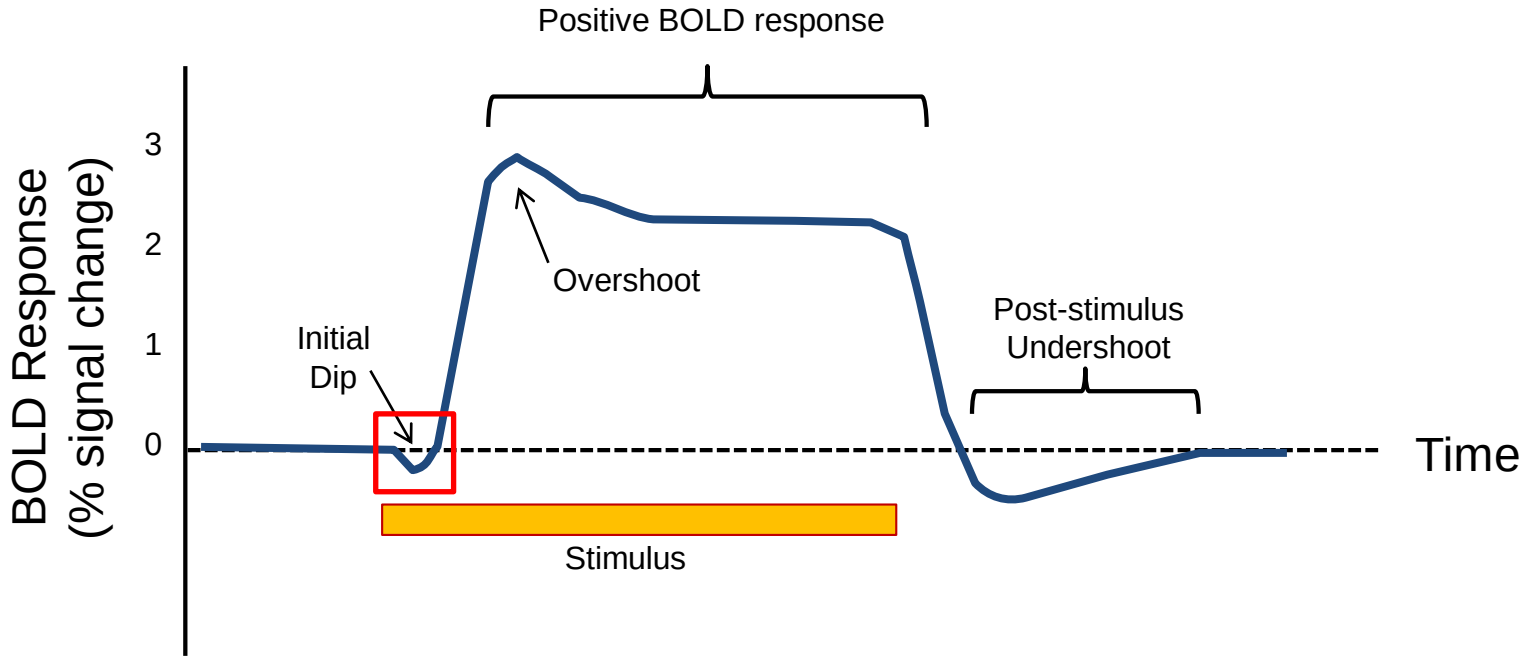
- Typical example of the time course of the MR signal change in response to an extended stimulus:



Plot from Jody Culham

Mechanism: Initial Dip

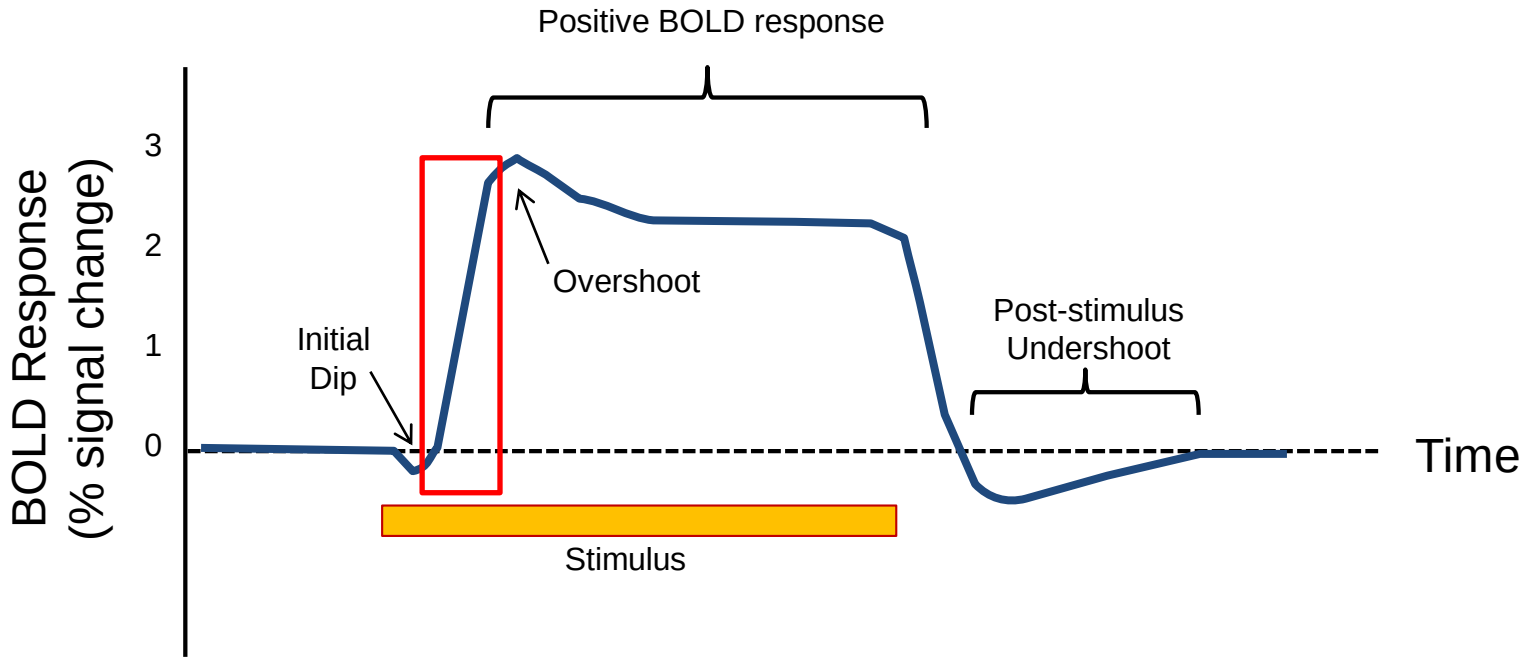
- Stimulus **increases neuronal activity**, which requires more oxygen and increases the amount of deoxygenated blood
  - Leads to a decrease in MR signal intensity
- **"Initial dip"** rather weak and not observed in all studies
  - If observed, often spatially more restricted than main peak



Plot from Jody Culham

Mechanism: Rise

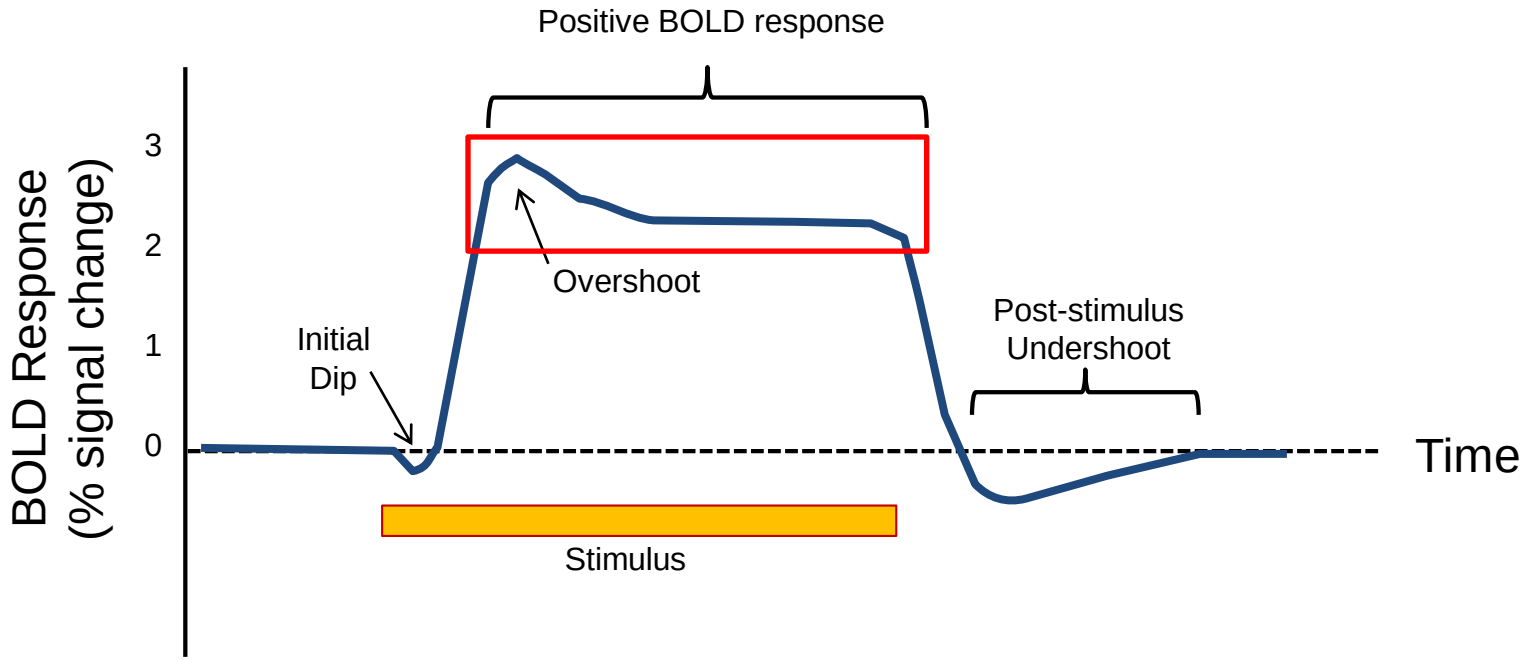
- **Local blood flow increases,** providing more oxygenated blood and **overcompensating** the additional need from neural activity
  - Change in signal strength between 5% (primary sensory stimulation) and 0.1-0.5% (cognitive tasks)
  - Basis of most fMRI studies



Plot from Jody Culham

Mechanism: Overshoot / Plateau

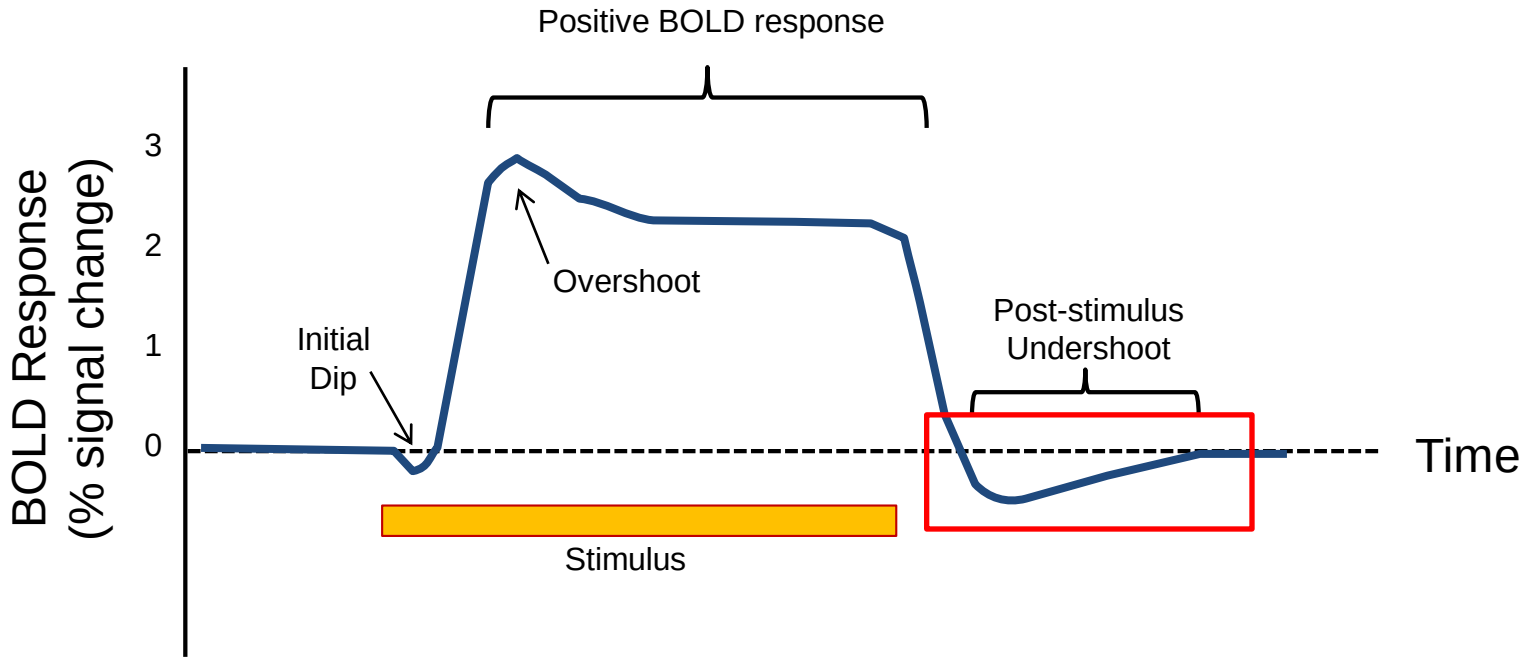
- While neuronal activity persists, increased oxygen uptake continues to be overcompensated
  - In “blocked” fMRI design with extended phases of stimulation, a plateau follows an initial overshoot



Plot from Jody Culham

Mechanism: Post-Stimulus Undershoot

- Many studies observe a **post-stimulus undershoot** in MR intensity that often persists for tens of seconds
  - Mechanism still not fully agreed upon [van Zijl et al., 2012]
  - May indicate uncoupling of metabolic and blood flow response:
    - Blood flow returns to baseline earlier than need for oxygen
    - Much oxygen used while "cleaning up" after activity



Plot from Jody Culham

Mechanism: Overview

- fMRI provides a rather **indirect measure** of neuronal activity

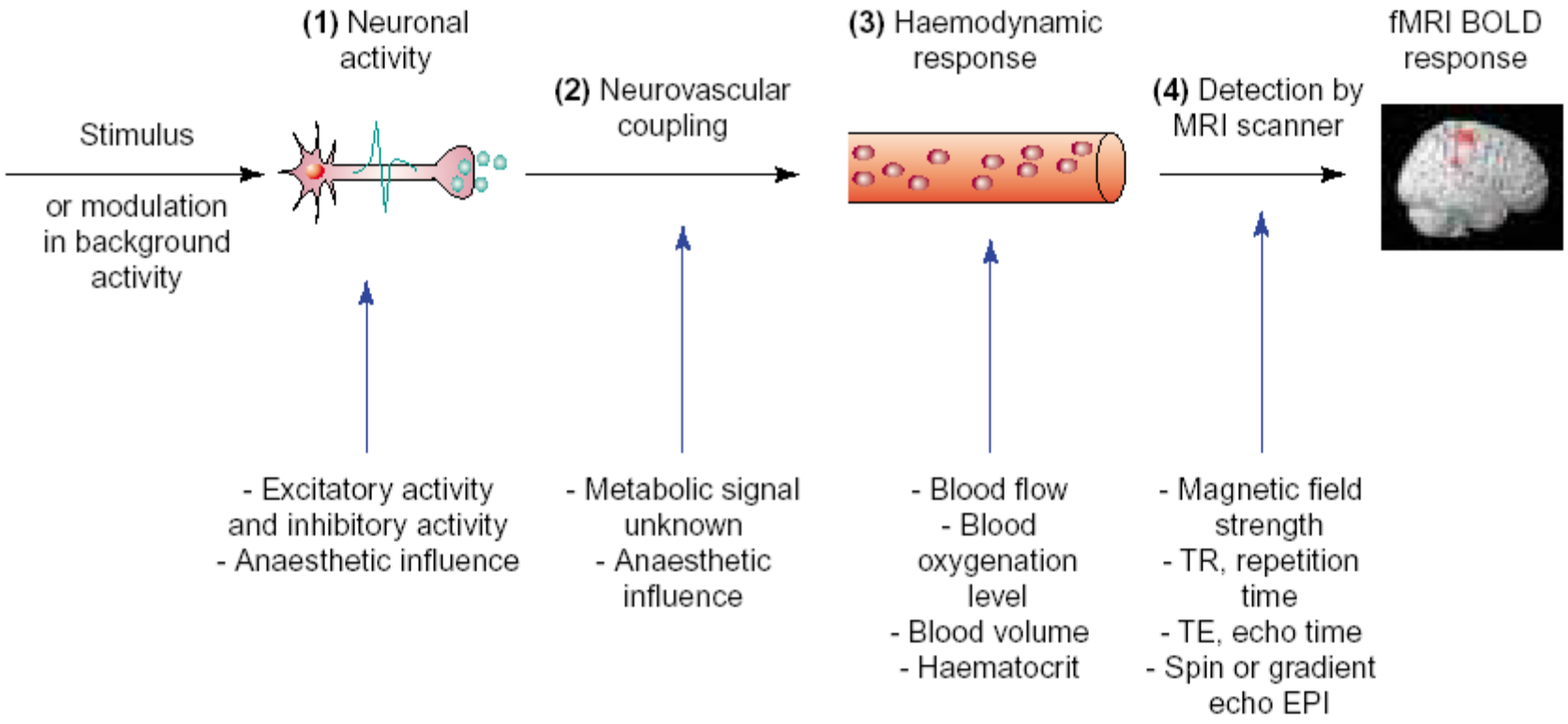
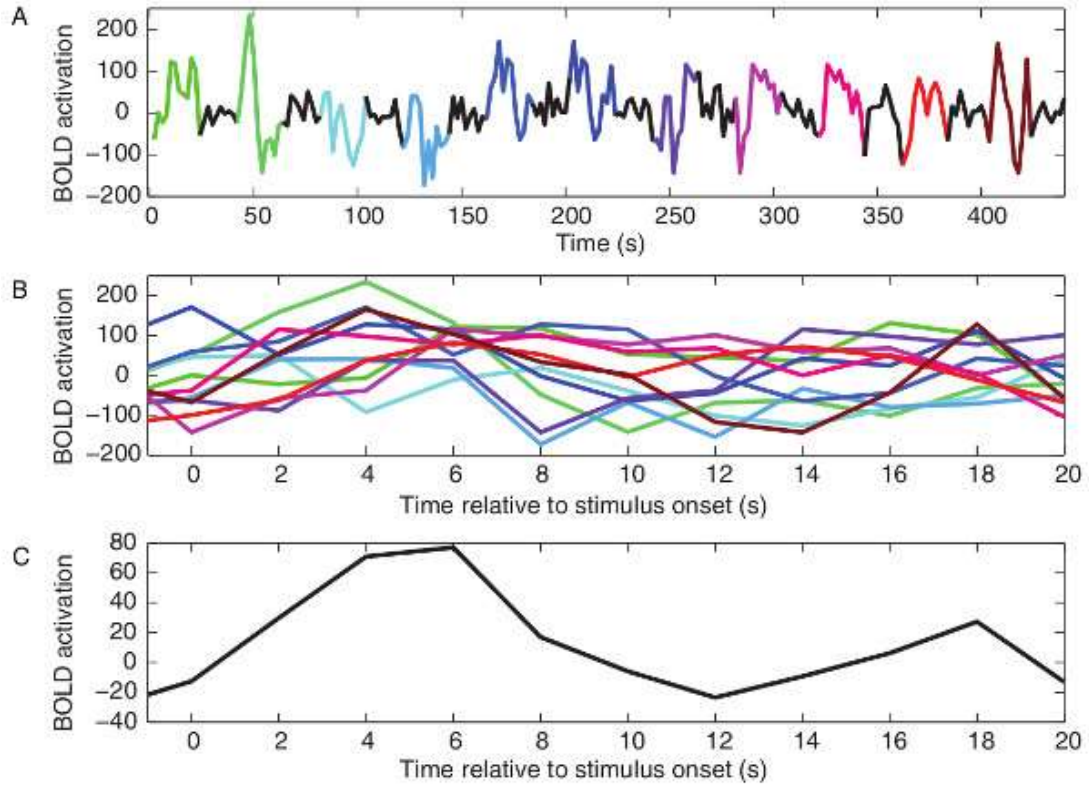


Image from Arthurs / Boniface 2002

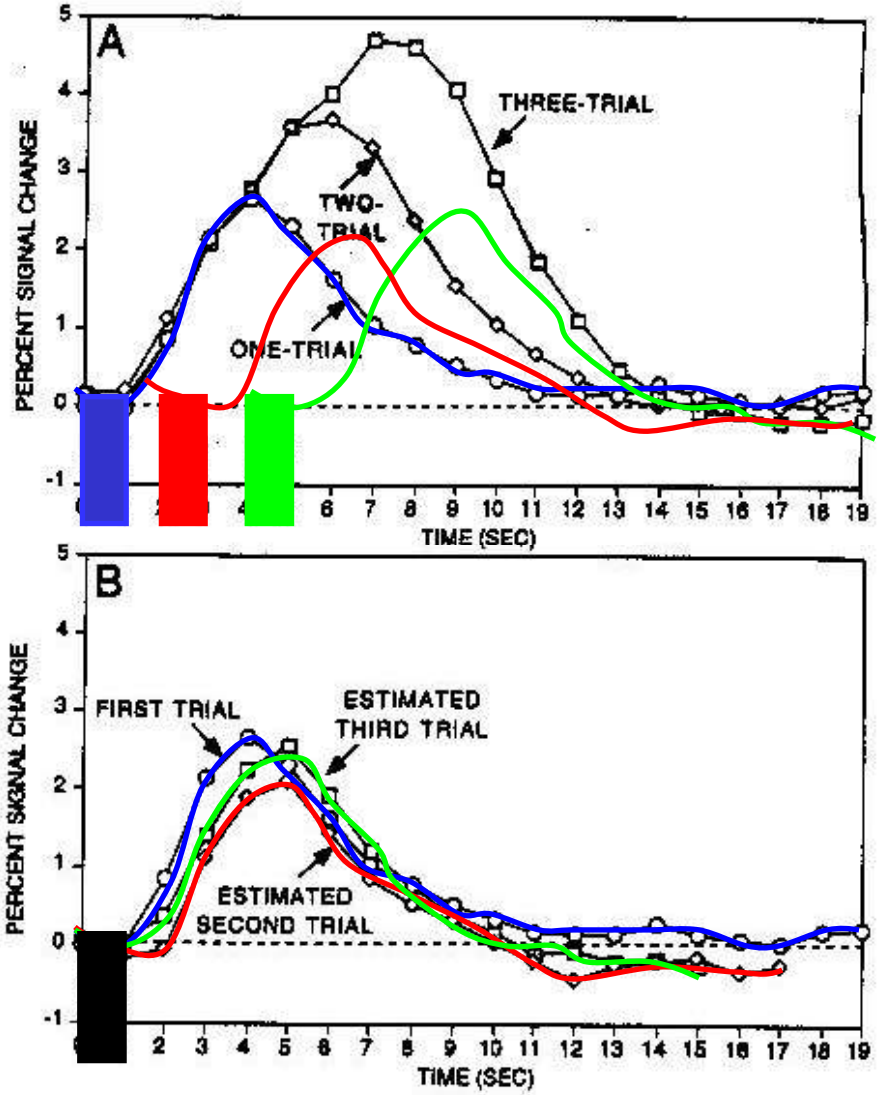
Hemodynamic Response Function

- The **Hemodynamic Response Function (HRF)** describes the time course of MR signal change in response to a short stimulus
  - Can be estimated by repeatedly presenting a short stimulus and averaging the measurements



Plots from Poldrack et al. 2011

Linearity of Hemodynamic Response



red = 2 - 1

green = 3 - 2

Sync each trial response to start of trial

Not quite linear
but good enough!

Plots from Dale & Buckner, 1997

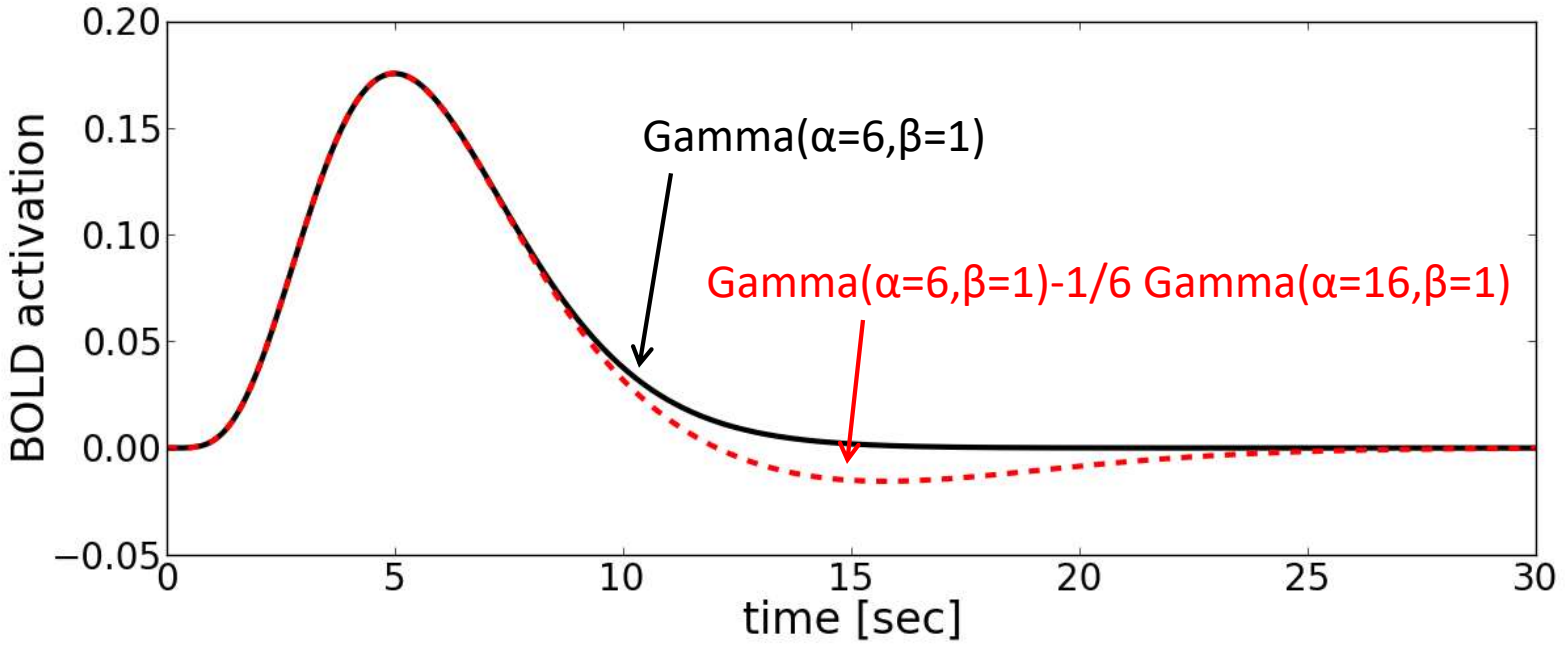
Slide from Jody Culham

Limits to Linearity of Hemodynamic Response

- Linearity is widely assumed in fMRI data analysis. Caveats of this assumption include:
  - **Spacing:** Responses to stimuli less than two seconds apart are slightly smaller than expected
  - **Duration:** Very short stimuli have a much larger response than would be expected from longer stimuli
    - Yesilyurt et al. 2008: Response to 5 ms visual stimulus half as large as to 1000 ms stimulus
- **Solution:** Quick succession of stimuli and ultra-short stimuli are rarely used in practice

Canonical HRF

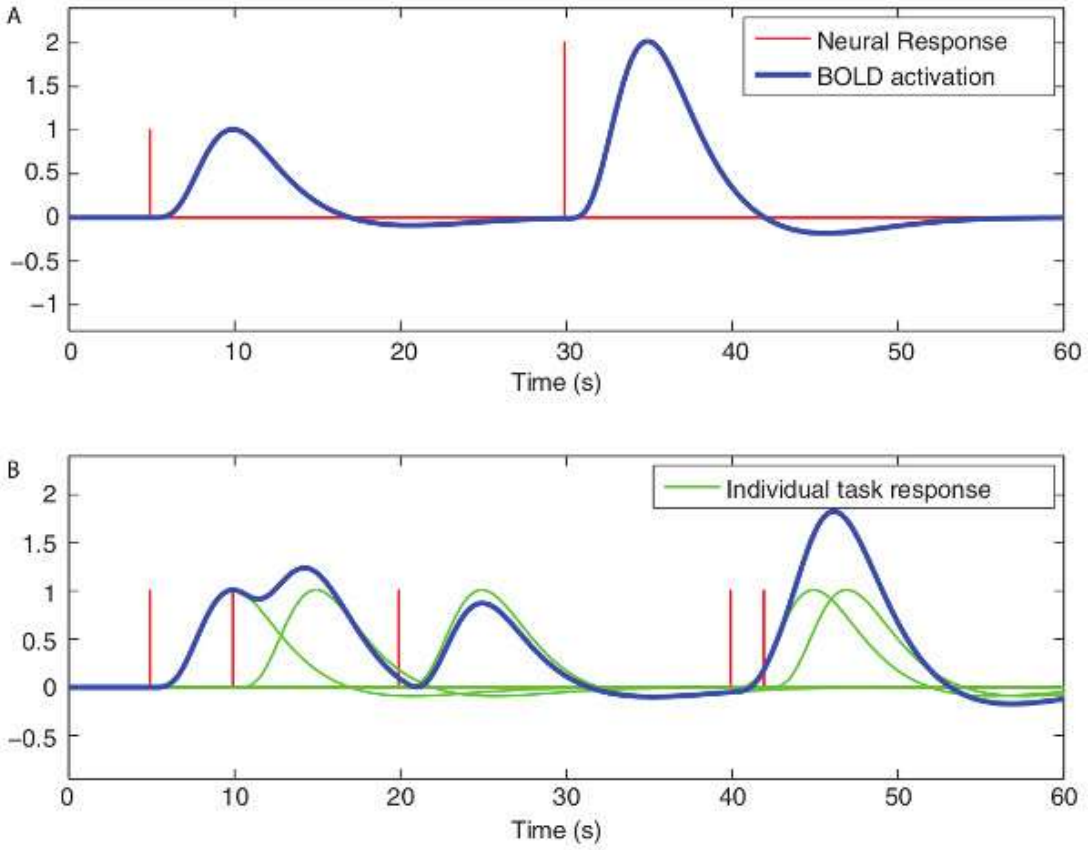
- It has been found empirically that the shape of the **Gamma distribution** (for certain parameters) closely resembles the HRF
  - Often used as a **"canonical HRF"** in analysis
  - Usually (imprecisely) called **"Gamma function"**
  - **"Double Gamma"** includes poststimulus undershoot



Convolution

- If hemodynamic response is linear, the BOLD signal is given by convolving the neural response *f* with the HRF *h*

$$(h * f)(t) = \int h(\tau) f(t - \tau) d\tau$$



Plots from Poldrack et al. 2011

Summary: Functional MRI

- **BOLD effect:** MR intensity reduced by presence of deoxygenated blood
- Amount of deoxygenated blood changes during **neuronal activity**
  - After initial dip, additional need for oxygen is overcompensated, leading to a stronger signal
  - Mechanism still not known in detail, but found to correlate mostly with local field potentials
- **Observing signal change over time** allows us to draw conclusions about brain activity
  - Tradeoff between temporal and spatial resolution

9.4 General Linear Model (GLM)

Motivation: General Linear Model

- **General Linear Model** provides a unified framework for different statistical tests
  - Provides additional flexibility that we will need for fMRI analysis
  - Standard for statistical analysis in neuroimaging
  - Expresses vector **y** of *N* measurements as the product of an *N*x(*p*+1) design matrix **X** and a vector **β** of *p+1* parameters, plus an i.i.d. Gaussian noise vector **ε**:

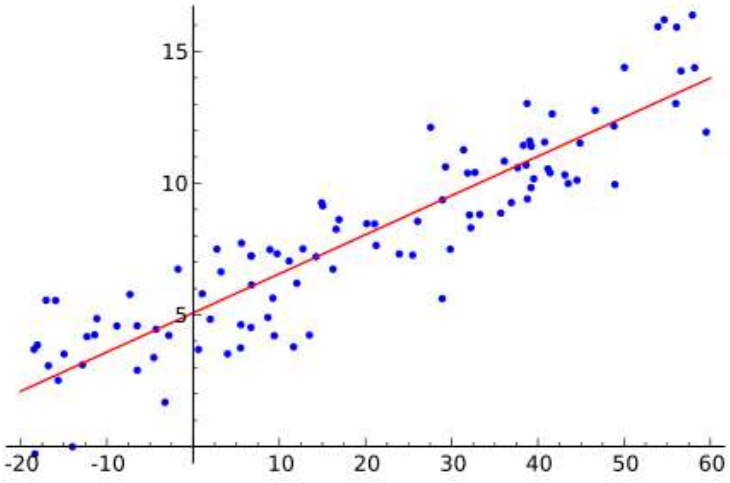
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Linear Regression

- The GLM can be considered a generalization of simple linear regression:

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \boldsymbol{\epsilon}$$

intercept slope



- Given data $\{(x_{1,1}, y_1), \ldots, (x_{1,N}, y_N)\}$, the model can be written as $\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\epsilon}$ with $\boldsymbol{\beta} = [\beta_0, \beta_1]^T$ and

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} \\ \vdots & \vdots \\ 1 & x_{1,N} \end{bmatrix}$$

Multiple Linear Regression

- Linear regression with p>1 independent variables is called **multiple regression**:

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \cdots + \beta_p \mathbf{x}_p + \boldsymbol{\epsilon}$$

- Same GLM $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with longer $\boldsymbol{\beta}$ vector
- Estimation of $\boldsymbol{\beta}$:
  - Since number of measurements typically exceeds p+1, we cannot simply invert $\mathbf{X}$ to obtain $\boldsymbol{\beta} = \mathbf{X}^{-1}\,\mathbf{y}$
  - Instead, multiply both sides by $\mathbf{X}^T$ to obtain normal equations: $\mathbf{X}^T\,\mathbf{y} = \mathbf{X}^T\mathbf{X}\,\boldsymbol{\beta}$
  - If $\mathbf{X}$ has *full column rank,* we can solve for

$$\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\,\mathbf{y}$$

Two-Sample t-Test in the GLM

- *Example:* Two-sample t test (chess player vs. non-chess player) in the GLM framework: **β**=[μ^C, μ^{NC}]$^\top$

$$\begin{pmatrix} v_1^C \\ \vdots \\ v_n^C \\ v_1^{NC} \\ \vdots \\ v_n^{NC} \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \boldsymbol{\beta} + \boldsymbol{\epsilon}$$

- **Null hypothesis** $\beta_1 = \beta_2$ can be written in the general form $\mathbf{c}\boldsymbol{\beta} = 0$ using the (row) contrast vector $\mathbf{c} = (1 \ -1)$
- For any **c**, the *t* score (with $\nu = N - (p+1)$) can be shown to be

$$t = \frac{\mathbf{c}\boldsymbol{\beta}}{\sqrt{\mathbf{c}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}^T\sigma^2}} \quad \text{with} \quad \sigma^2 = \frac{\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}}{N-(p+1)}$$

Regressing Out Nuisance Variables

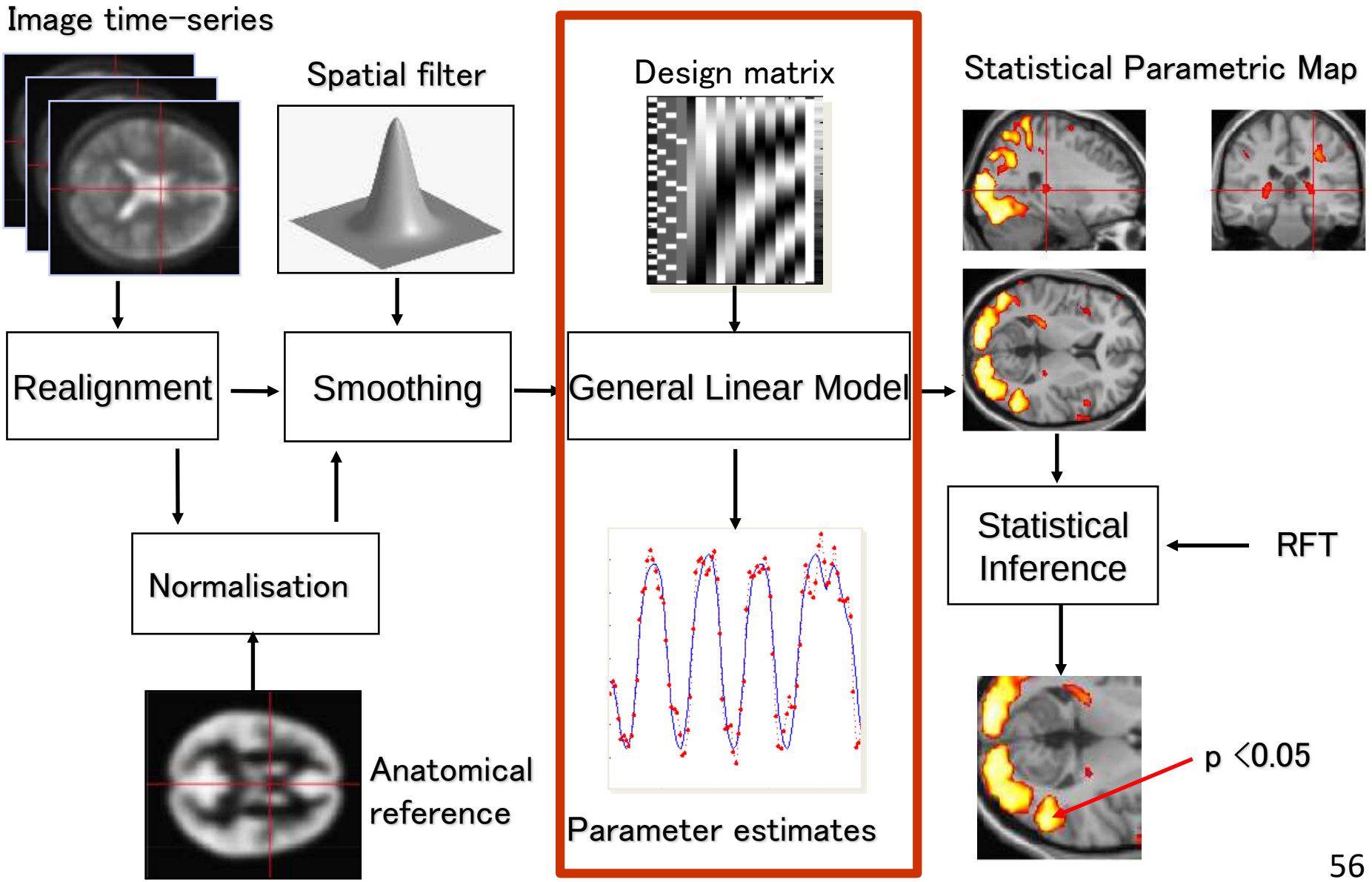
- **Additional benefit of GLM:**
  - Can insert βs which are ignored by the contrast to "regress out" nuisance variables of no interest
  - *Example:* Our subject pool is not age-matched. Inserting a β for the effect of age (corresponding to a column of ages in **X**) implies an "age-correction"
    - *Caveat:* If you model the mean of a group and include a regressor for age, your estimated mean will be at age zero, unless you mean-center the age column!
  - *Limitation:* Assumes that effect of age is linear. It's still preferable to use matched subjects.
  - *Caveat:* In permutation-based testing, nuisance variables are *not* interchangeable under the null
    - Often regressed out separately before permutation

Summary: General Linear Model

- The General Linear Model $\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$ unifies all statistical tests that are widely used in neuroimaging
  - Additional flexibility will be used in next section
  - Specification using design matrix **X** and contrast vector **c** or contrast matrix **C**
  - Can be used to compute *t* scores, *F* scores, or combined with permutation-based testing
  - Allows us to "regress out" variables of no interest

9.5 Statistical Analysis of fMRI Data

Overview: fMRI Analysis



Slide taken from SPM course

Example: Auditory Stimulus

- **Early proof-of-concept experiment** performed at Functional Imaging Laboratory, University College London ("mother of all experiments")
  - Full-brain EPI scan with 64x64x64 voxels (3x3x3 mm^3 resolution), repeated at TR=7 sec
  - Alternating between 42 sec blocks of rest and auditory stimulation (spoken words)
  - 96 acquisitions overall
    - Initial 12 discarded to reach steady state
  - Data publicly available:
    - http://www.fil.ion.ucl.ac.uk/spm/data/auditory

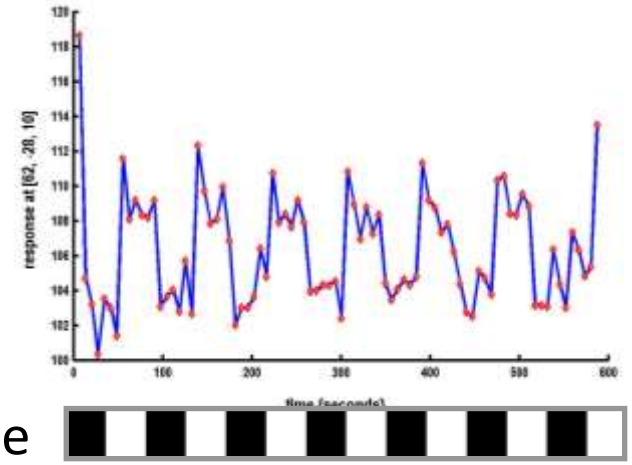
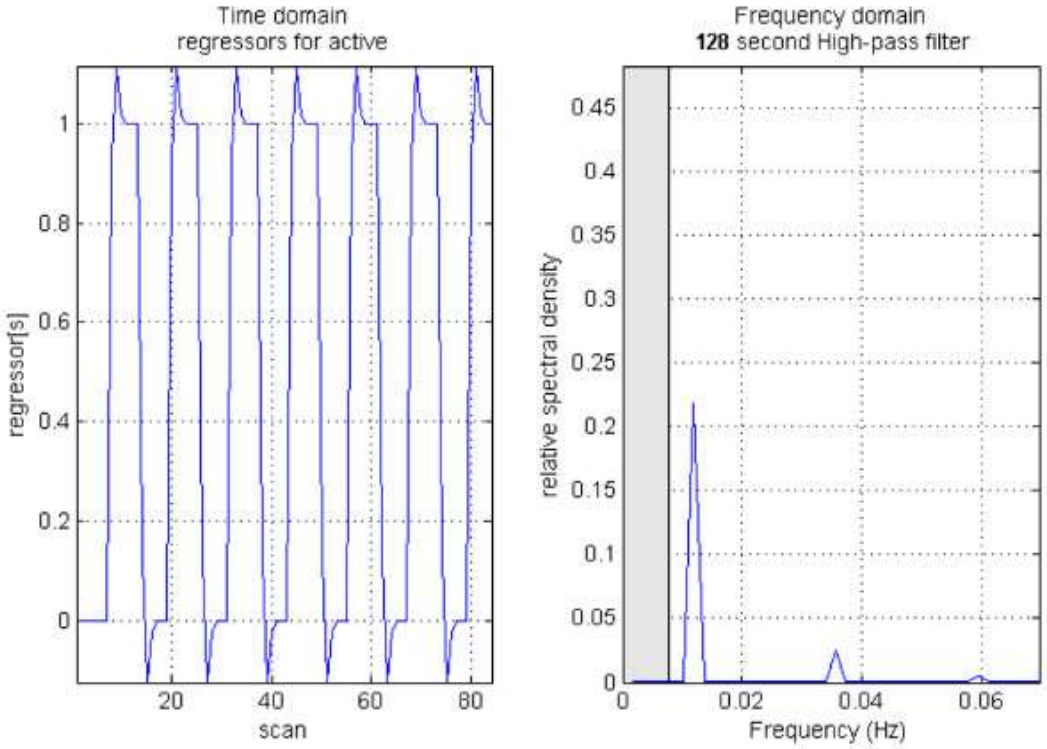


Image taken from SPM course

Statistical Modeling

- In principle, we could use a mass-univariate **two-sample t-test** to compare MR images during stimulation to images at rest
- **But:** That would ignore HRF
  - Convolving stimulus with HRF predicts signal ("regressor")
  - Use regressor as a column in General Linear Model



Plots from SPM8 manual, Ashburner et al. 2013

A Typical Design Matrix

- It is common to visually inspect the design matrix for the General Linear Model:

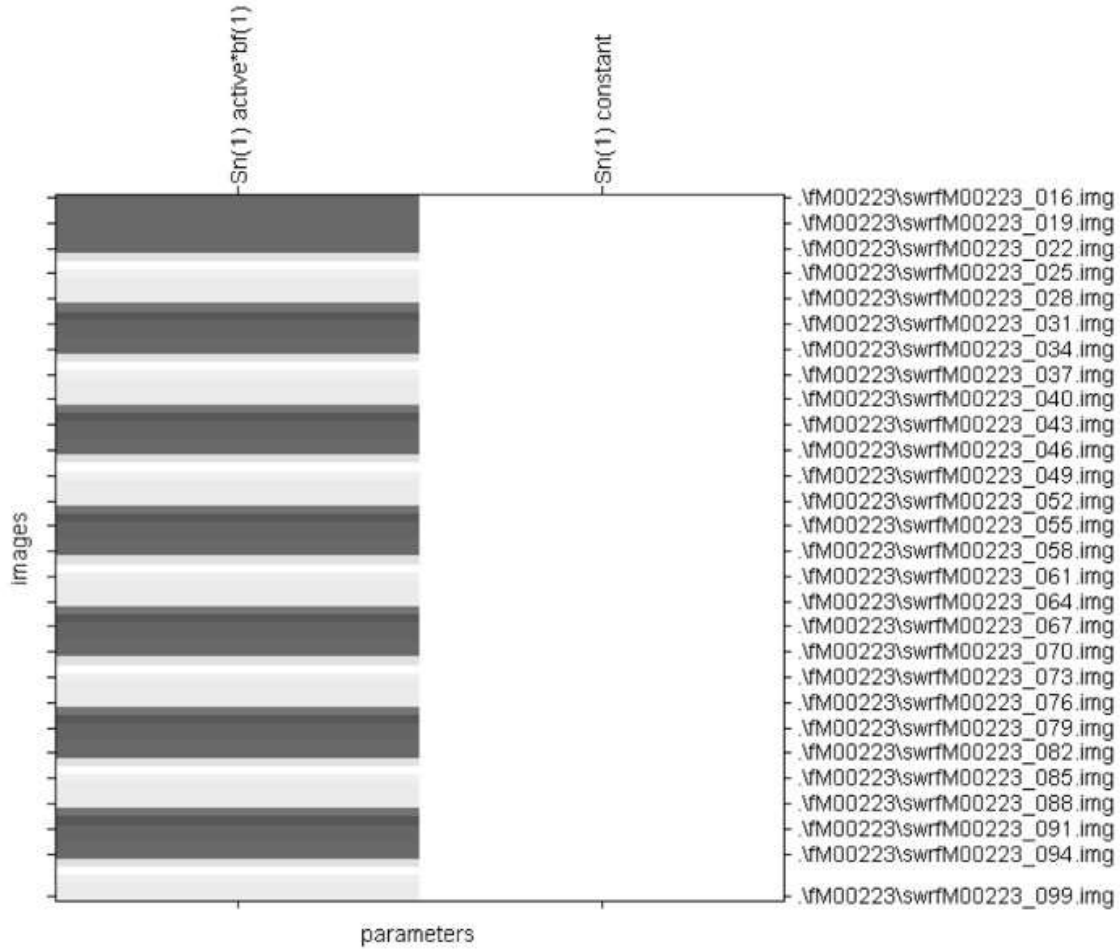
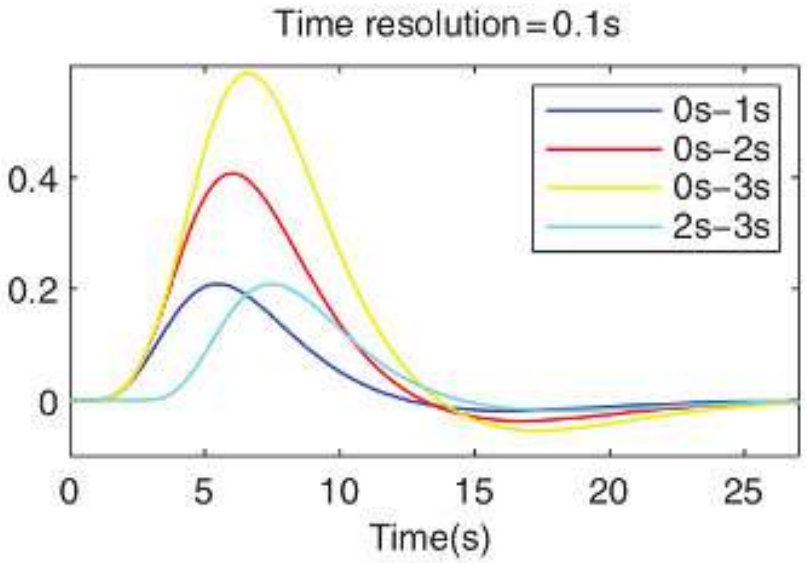


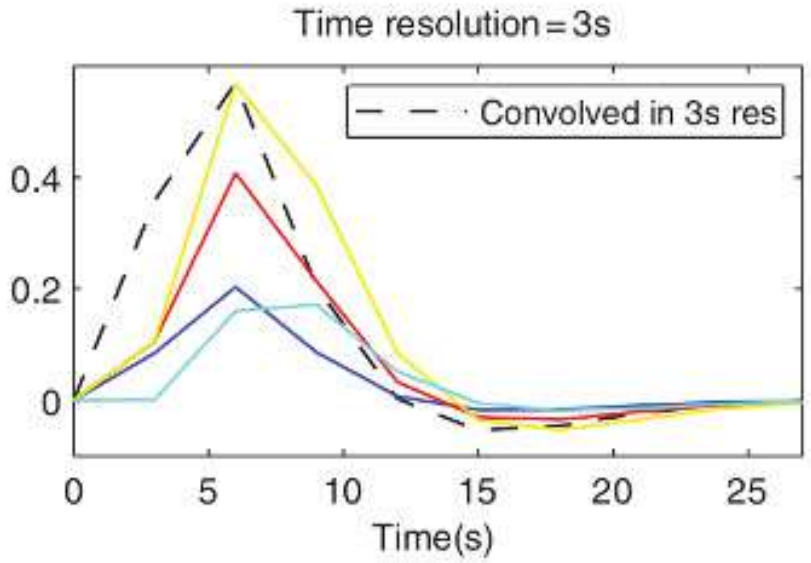
Image from SPM8 manual, Ashburner et al. 2013

Perform Convolution with High Resolution!

- At small temporal resolution (e.g., TR=3 s), short stimuli are undersampled ☹
- Convolution with HRF smoothes out stimulus ☺
- But: Have to perform convolution in high temporal resolution to avoid aliasing



Four different stimuli convolved with HRF at 0.1s res



Subsampled results (color) vs. convolution at 3s res

Plots from Poldrack et al. 2011

Maximum Intensity Projection

- Perform statistical testing as in voxel-based morphometry
- **Maximum Intensity Projection:**
  - 2D map showing maximum across third dimension

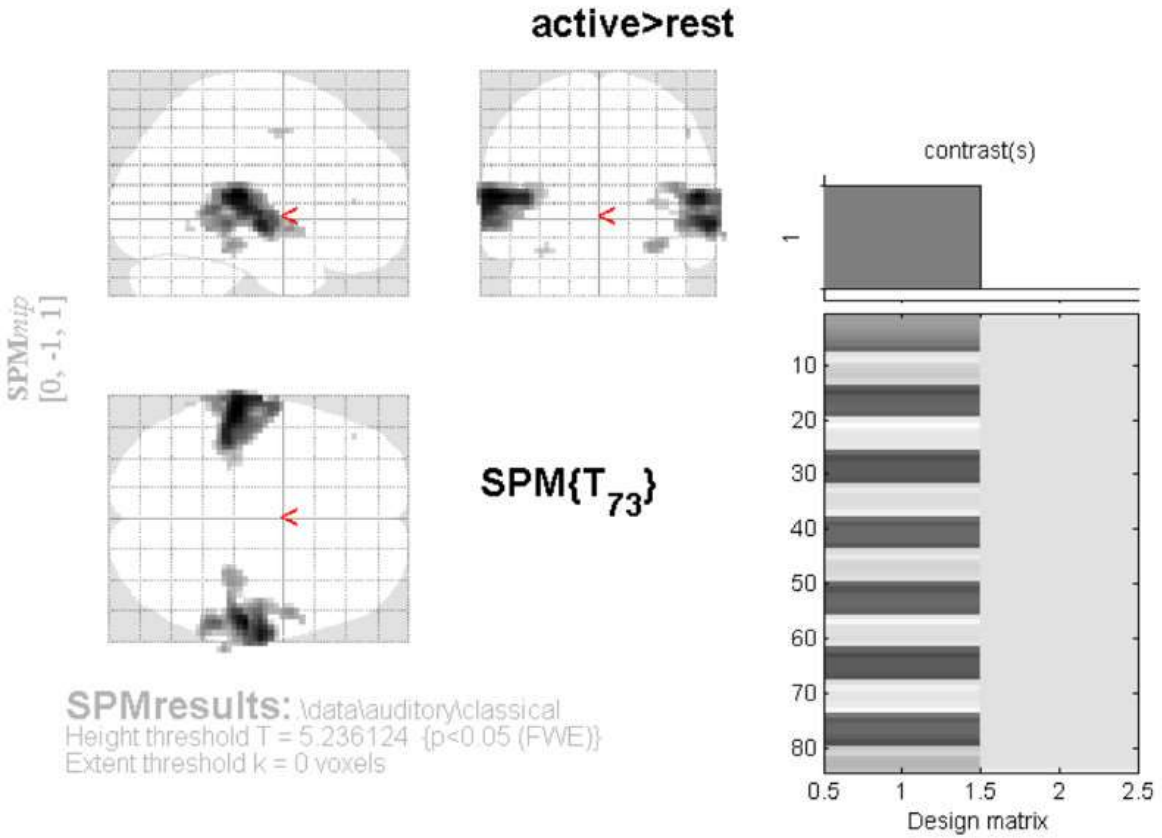


Image from SPM8 manual, Ashburner et al. 2013

Overlay after Coregistration

- After coregistration of functional and high-resolution structural data, activations can be overlaid

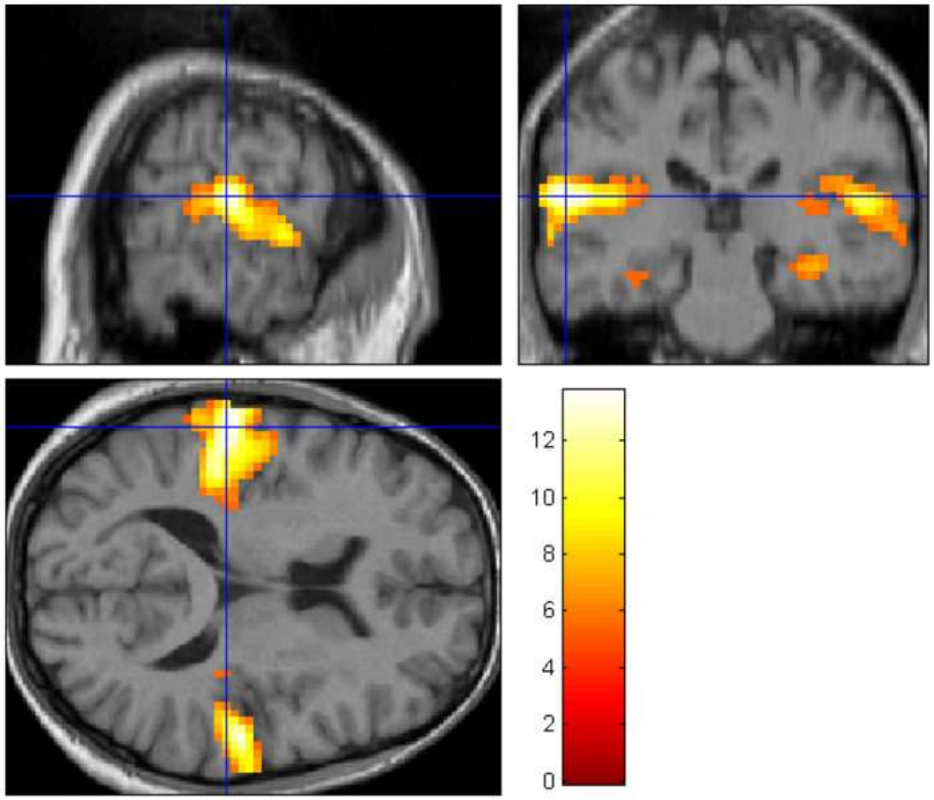


Image from SPM8 manual, Ashburner et al. 2013

Overlay after Normalization

- After non-linear normalization, activations can be overlaid on a generic brain template
  - **Two-step:** Functional to structural scan of same subject, structural scan to template
  - Can use **generic models of gray matter surface**
    - Ignores variability in brain anatomy

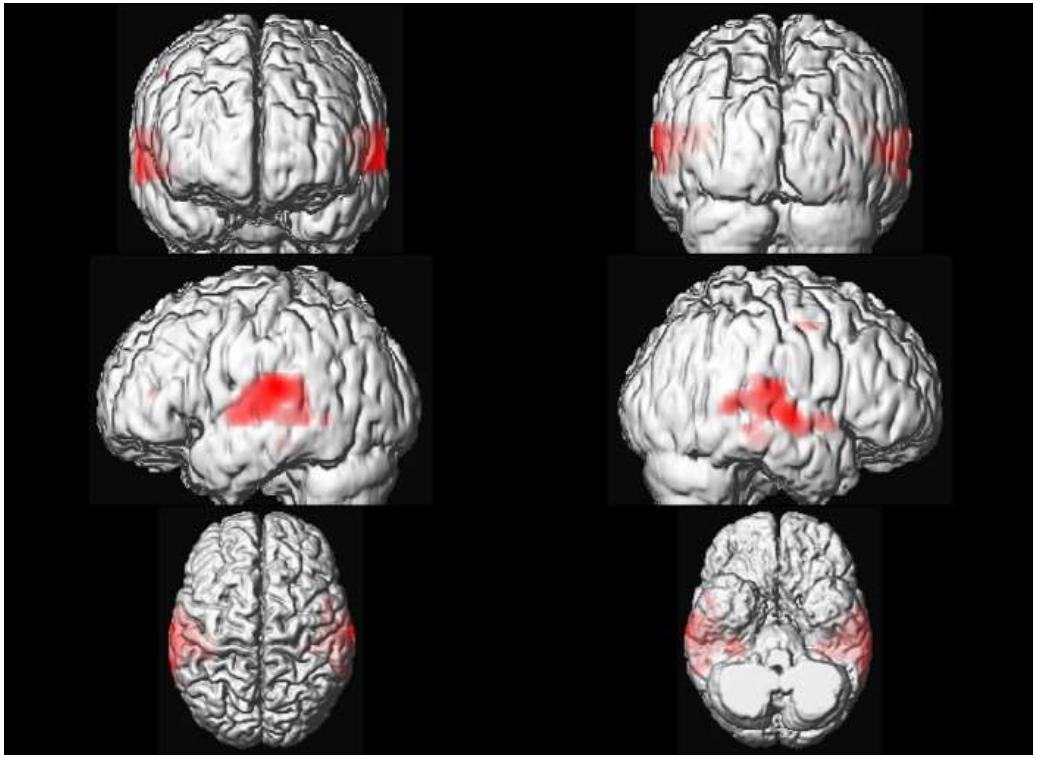


Image from SPM8 manual, Ashburner et al. 2013

Inspecting Individual Time Series

- Selecting individual voxels ("voxel surfing") allows us to compare their **individual time series** with the model fit
  - Check for substantial model misspecification

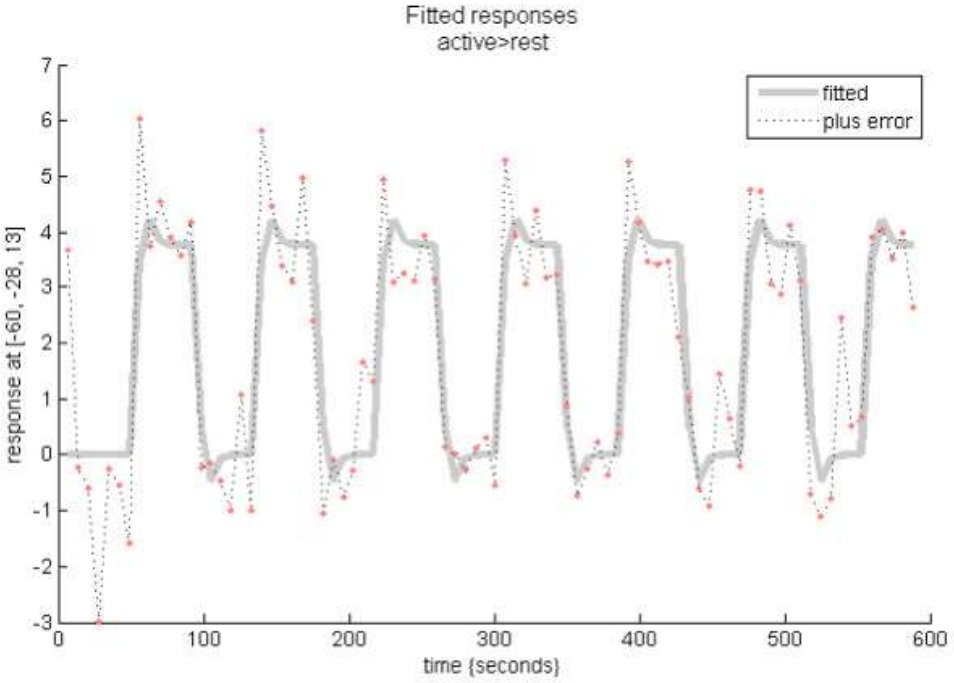


Image from SPM8 manual, Ashburner et al. 2013

fMRI Group Analysis

- How to combine fMRI data **from several subjects** to answer questions such as „is activation stronger in women than in men“?
- Tempting to simply normalize and temporally concatenate all scans
  - Ignores the fact that the subjects themselves are **samples from a larger population**
    - Fails to distinguish variances and degrees of freedom w.r.t. number of fMRI time steps vs. number of subjects
  - Statistical results will **not generalize** to the wider population!
  - fMRI literature calls this „fixed“ vs. „mixed“ effects

Two-Level Analysis

- „Second-level“ models are commonly estimated in two steps, each of them using a GLM:
  1. Estimate activation per-subject
  2. Compare contrasts between groups

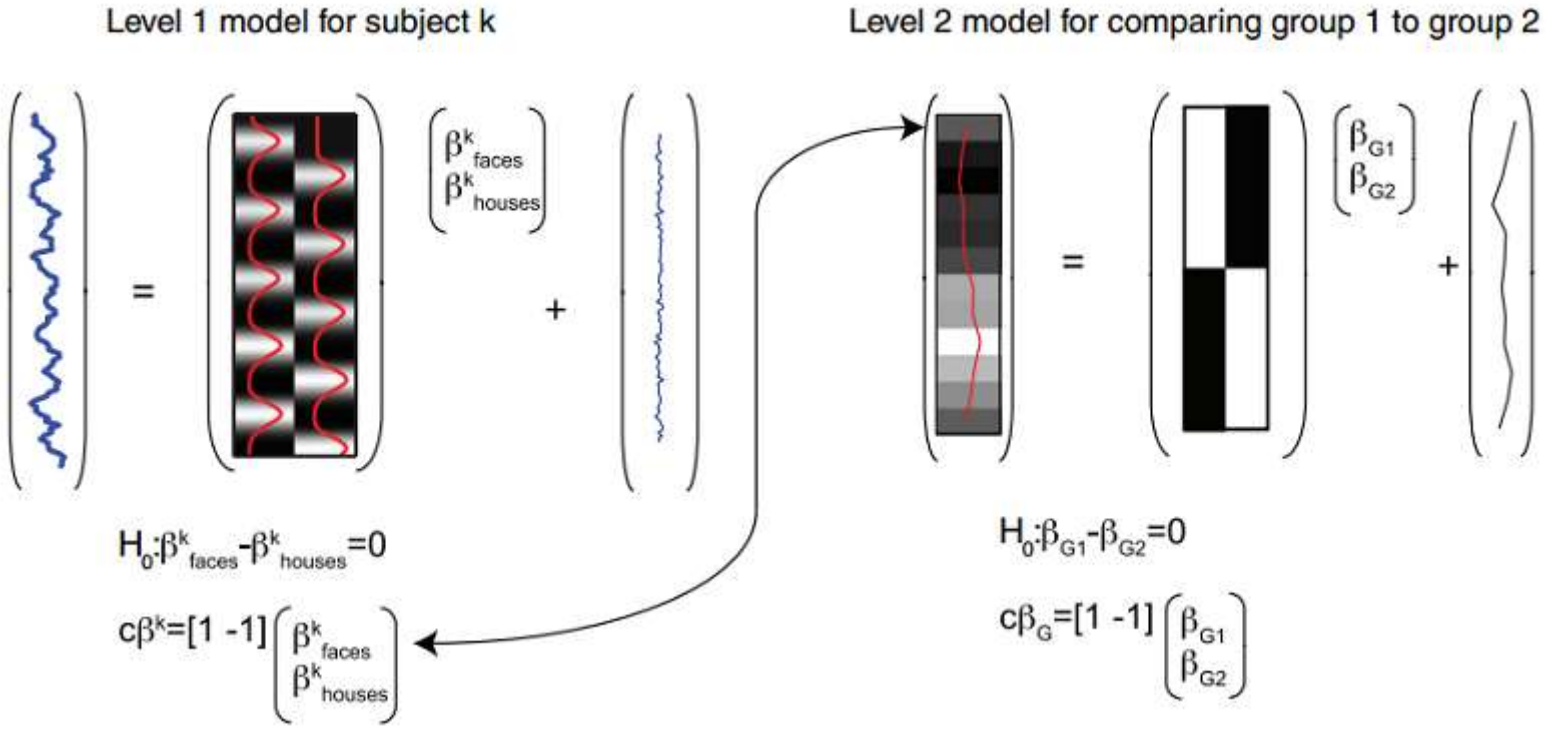


Illustration from Poldrack et al. 2011

Model Unequal Within-Subject Variance?

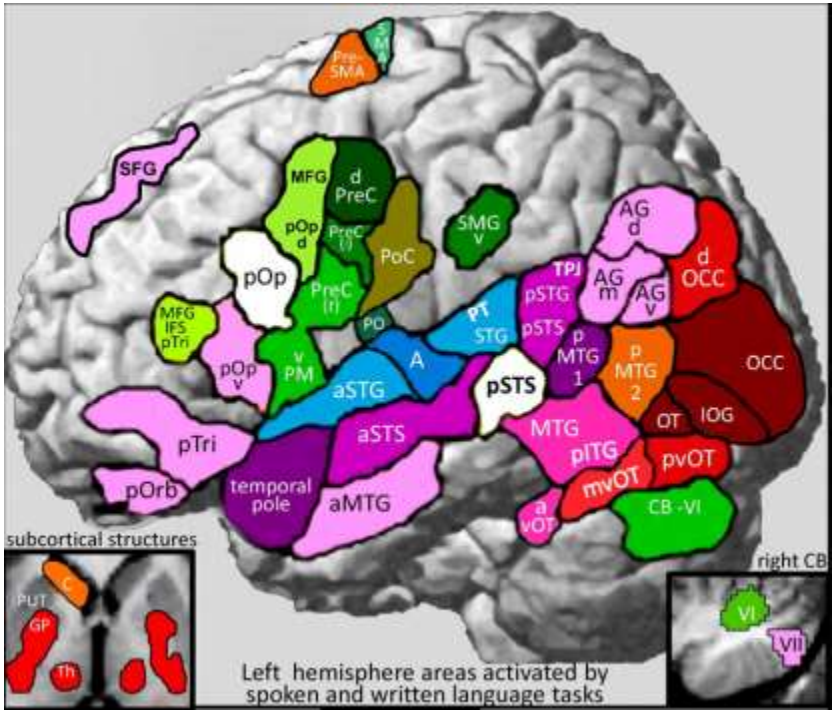
- In case variances differ between subjects, we need to perform a **weighted least squares fit** for the second-level GLM:

$$\boldsymbol{\beta} = \left(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

  - Reduces the impact of highly uncertain subjects
- Some software packages **do not propagate** individual within-subject variances to the second level
  - Estimates of variances that have been obtained from few samples are themselves quite uncertain!

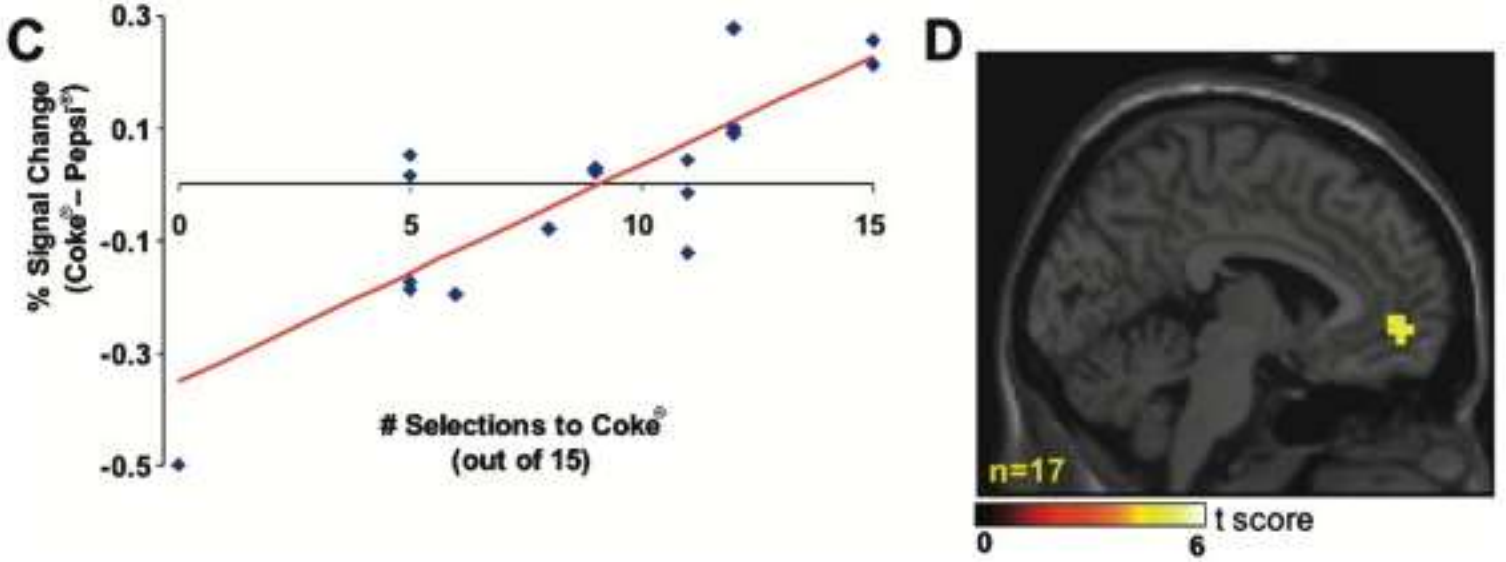
fMRI Example: Speech

- **Price 2012** summarizes hundreds of studies that provide converging evidence on which brain regions are involved in hearing speech, producing speech, or reading
  - Reading meaningful vs. meaningless words, generating vs. repeating words, processing simple vs. complex grammar, etc.
  - Concludes that next big challenge is to understand how brain regions interact



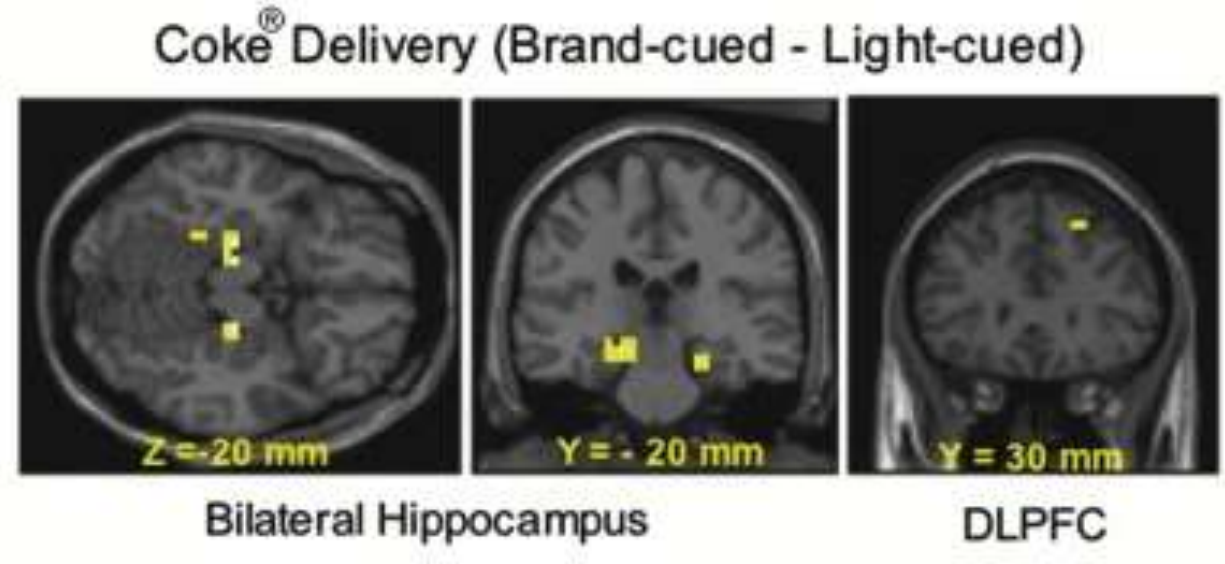
fMRI Headline: Coke vs. Pepsi

- **McClure et al. 2004** report results on behavioral preference and fMRI response to cola with and without knowing the brand
  - People split evenly between Coke vs. Pepsi in a blind taste test
  - Prefrontal activity correlated with preference



fMRI Headline: Coke vs. Pepsi

- Most subjects preferred labeled cup of coke over unlabeled cup of coke (despite balanced distribution during blind taste test)
- Brand-cued vs. light-cued delivery of coke led to activations in several other brain regions
  - No such effects for Pepsi



Summary: Statistical fMRI Analysis

- After applying pre-processing...
  - Correlate observed with predicted BOLD response using the General Linear Model
  - Everything you learned about family-wise error correction applies!
  - Group analysis done in two steps to correctly account for mixed effects
- For visual presentation...
  - slice through data or present MIP
  - overlay on top of anatomical image

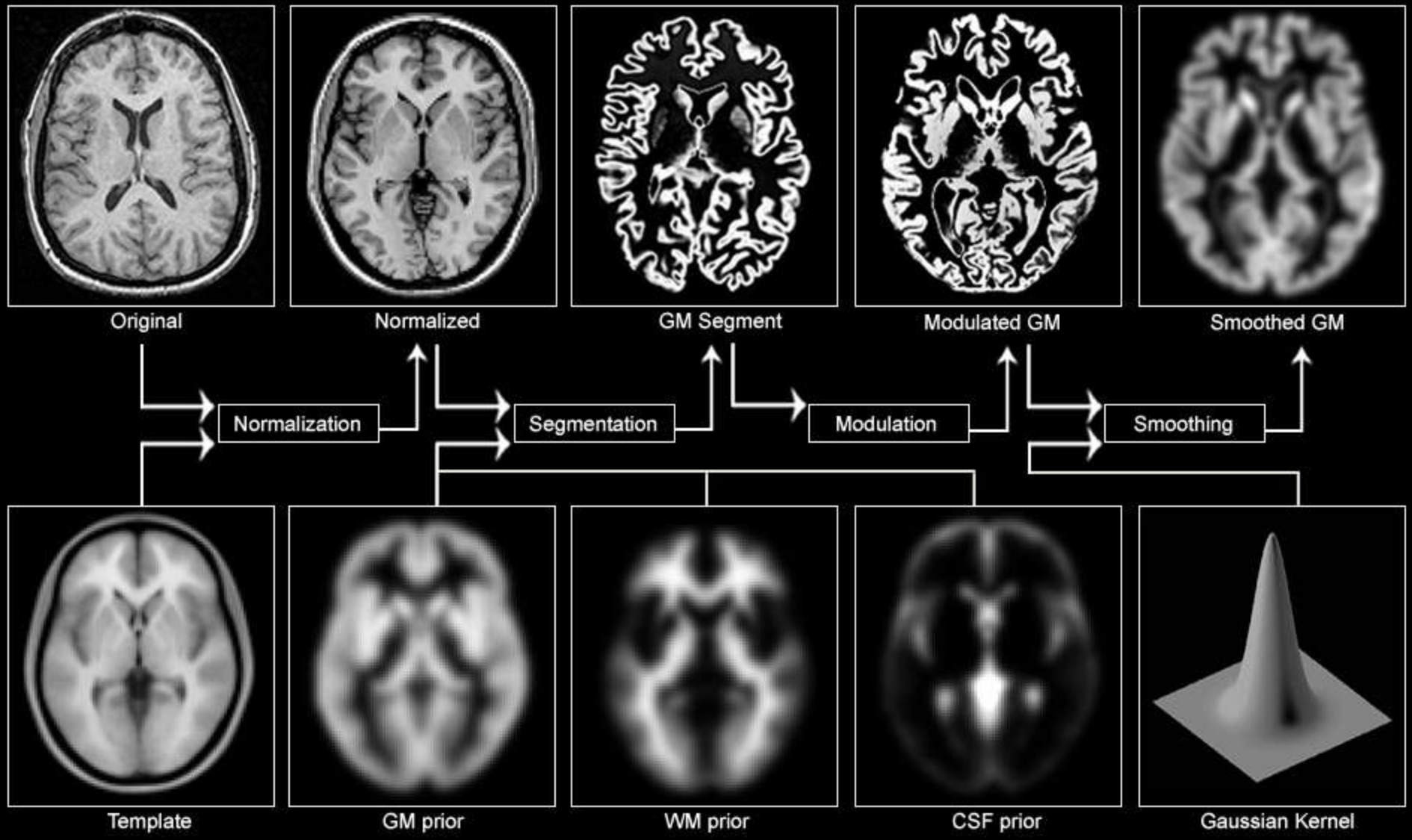
9.6 Voxel-Based Morphometry (VBM)

Goal of Voxel-Based Morphometry

- We just learned how to perform mass univariate tests, but how to derive images for which it makes sense to apply them?
- **Voxel-Based Morphometry** is one option
  - **Goal:** Would like to compare the size of specific brain areas between subjects
  - **But:** Segmenting specific regions of interest is difficult and limits our analysis to those (few) regions
  - **Idea:** Normalize subjects so that anatomical structures are aligned and compare regional gray matter volume *at each point of the brain*

Image Source: brainmaps.org

Voxel-Based Morphometry
Pre-processing Overview



lide from Nicola Hobbs & Marianne Novak

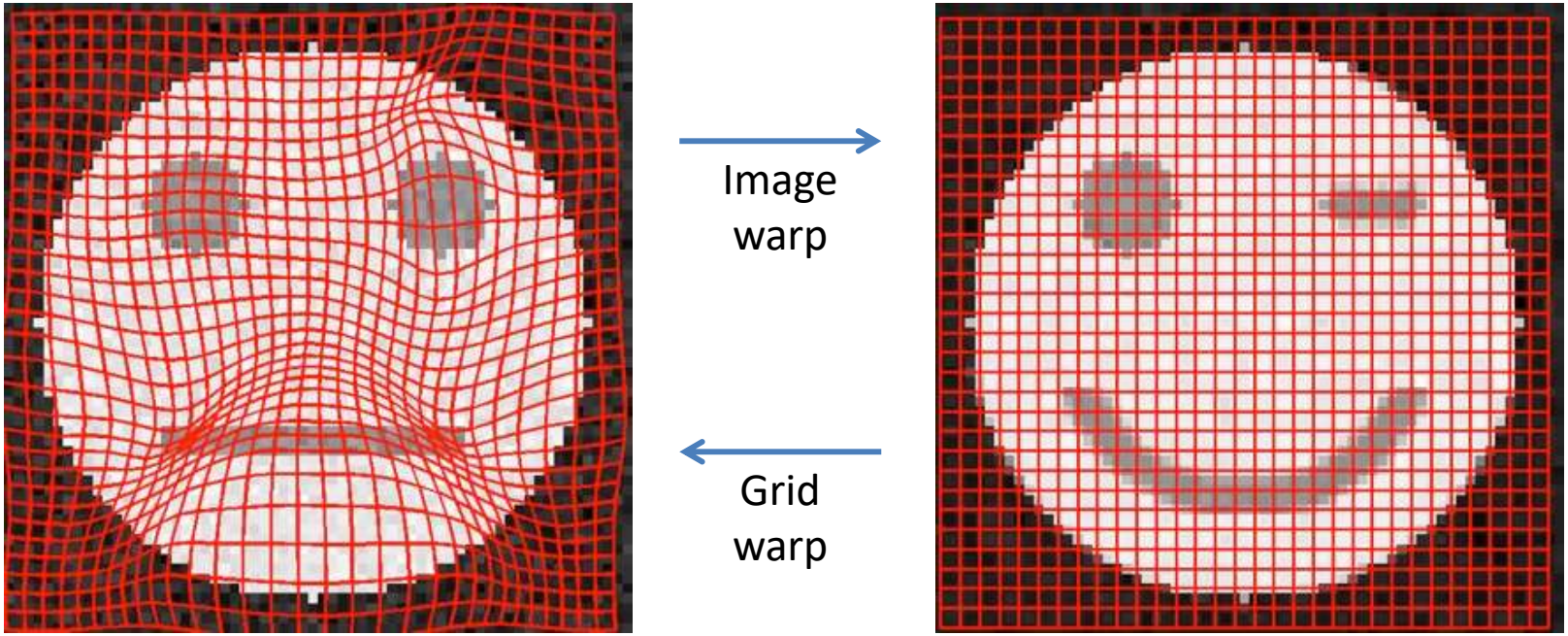
Pre-Processing in VBM

1. Normalization
   - Rigid, then affine, then non-linear registration
2. Segmentation
   - Tissue classification into GM / WM / CSF
   - MRF to deal with noise and bias fields

- **Optimized VBM:**
  - Performs automated brain extraction first
  - Builds a study-specific atlas
  - Achieves improved gray matter alignment by segmenting first, normalizing gray matter maps, applying transformation to original image, and segmenting again (including prior maps)

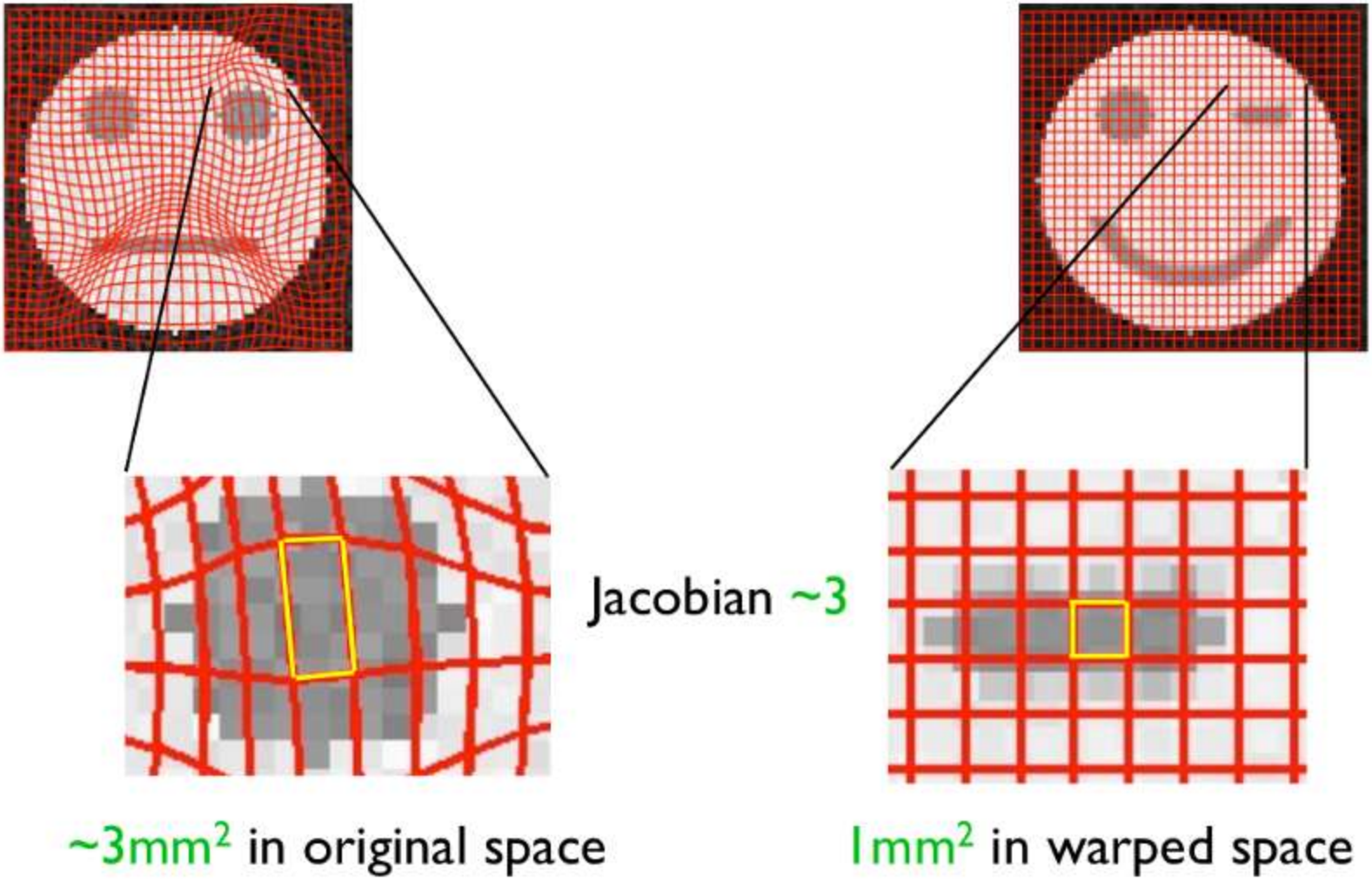
Modulation

- Desired interpretation is in terms of **gray matter volume,** but nonlinear warping distorts volumes
- **Modulation** corrects for this by multiplying normalized gray matter map with the Jacobian determinant



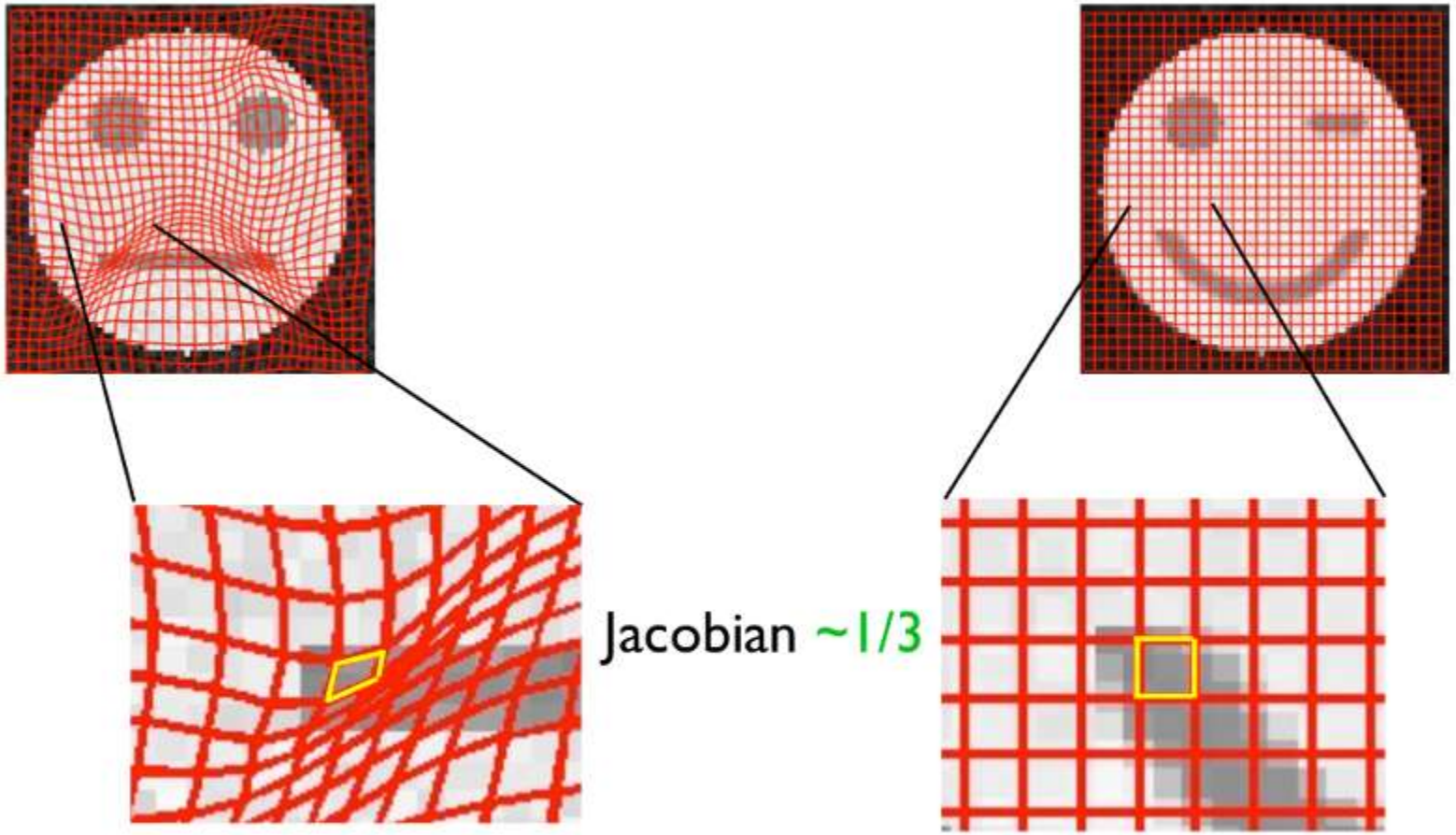
Images from Oxford brain fMRI lab

Illustration: Contraction



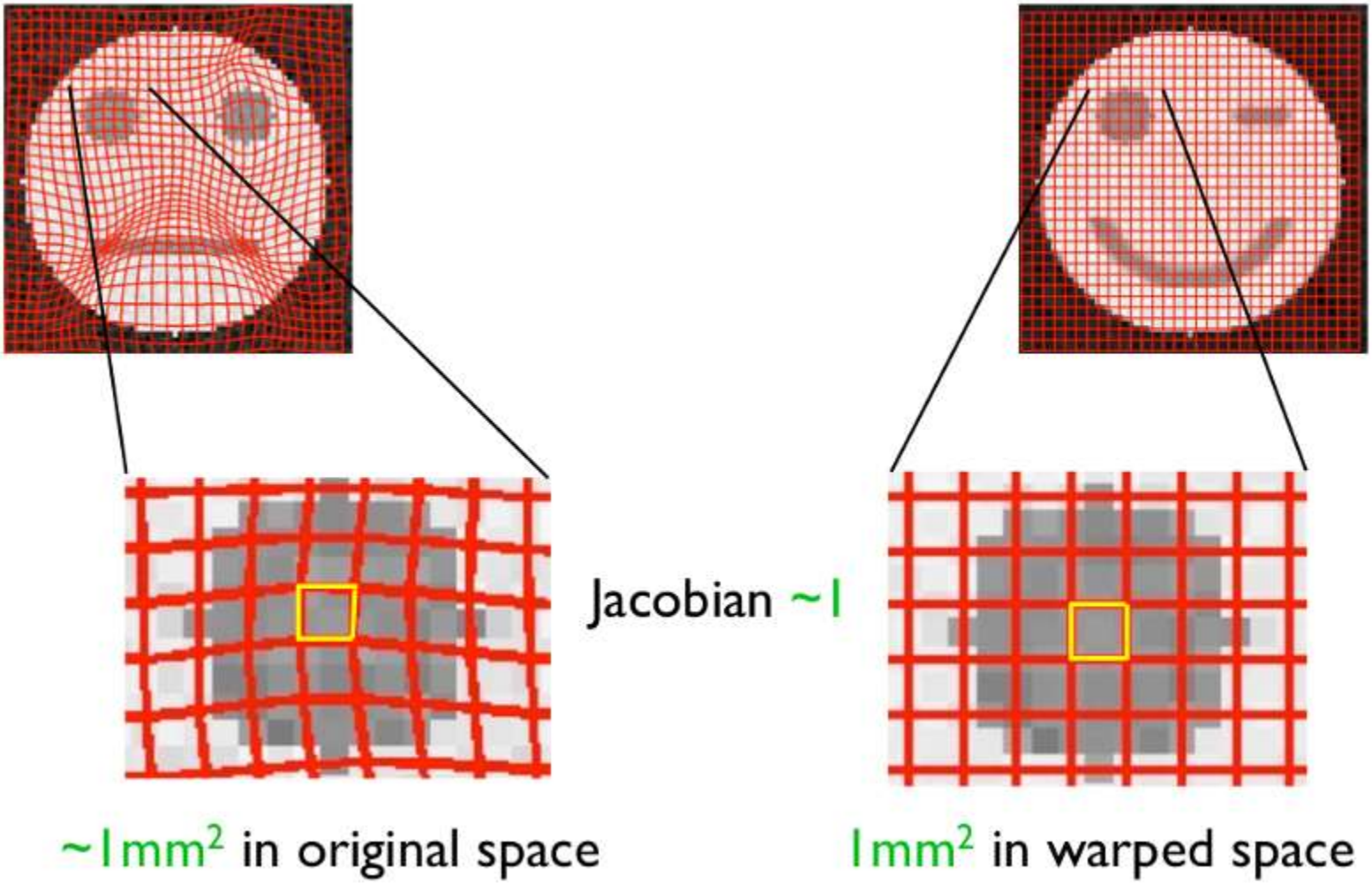
Slide from Oxford brain fMRI lab

Illustration: Expansion



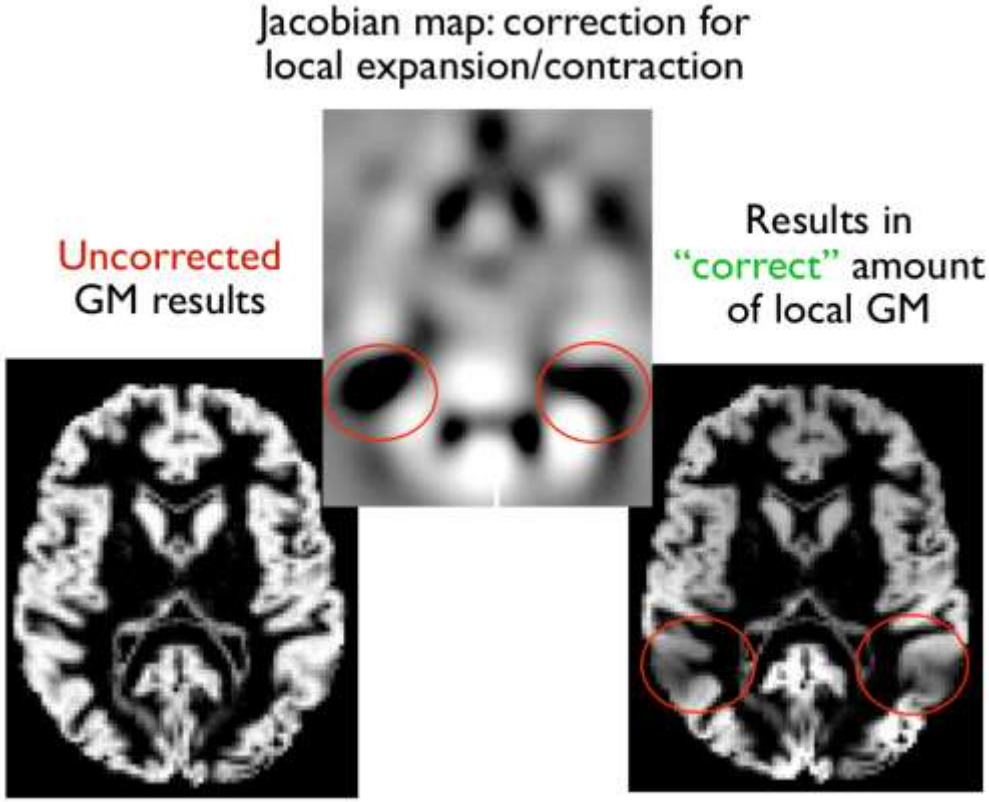
Slide from Oxford brain fMRI lab

Illustration: Volume Preservation



Slide from Oxford brain fMRI lab

Applying the Modulation

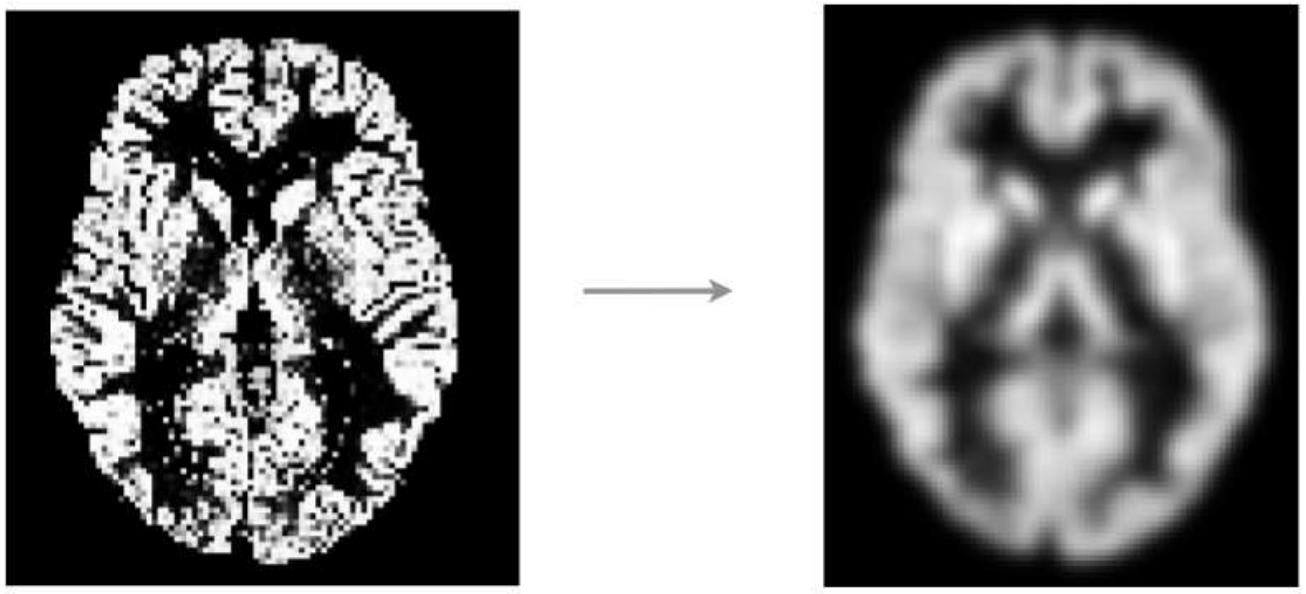


Images from Oxford brain fMRI lab

- Sometimes, modulation is omitted and results are interpreted as "gray matter density" (ratio of gray matter vs. white matter and CSF)
  - Result will depend strongly on flexibility of normalization!

Smoothing

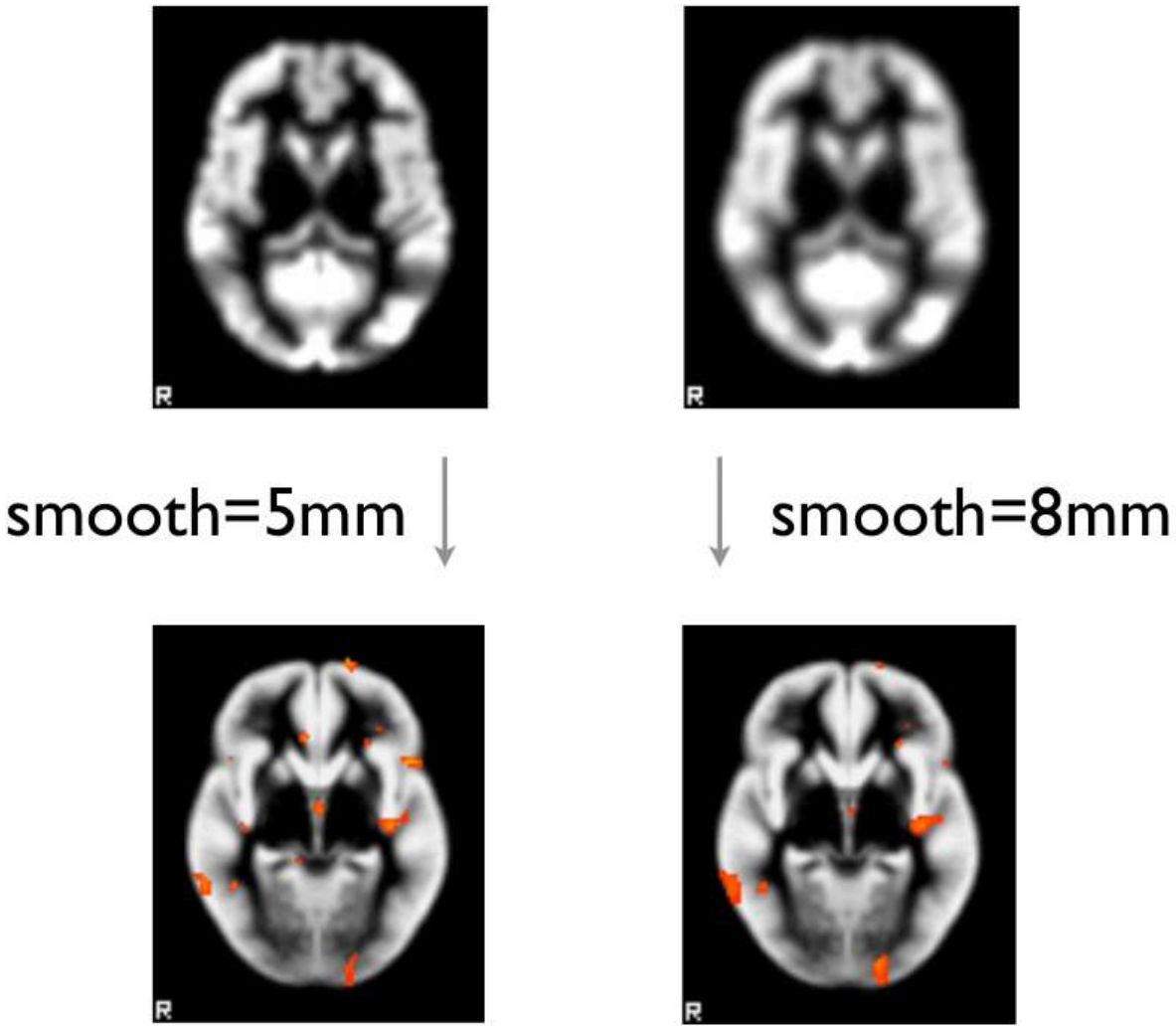
- Apply some amount of **Gaussian smoothing**
  - Interpretation as “regional” (rather than strictly local) gray matter volume / density
  - Compensates for slight inaccuracies in normalization
  - Makes the data more normally distributed (central limit theorem)



Images from Oxford brain fMRI lab

How Much Smoothing?

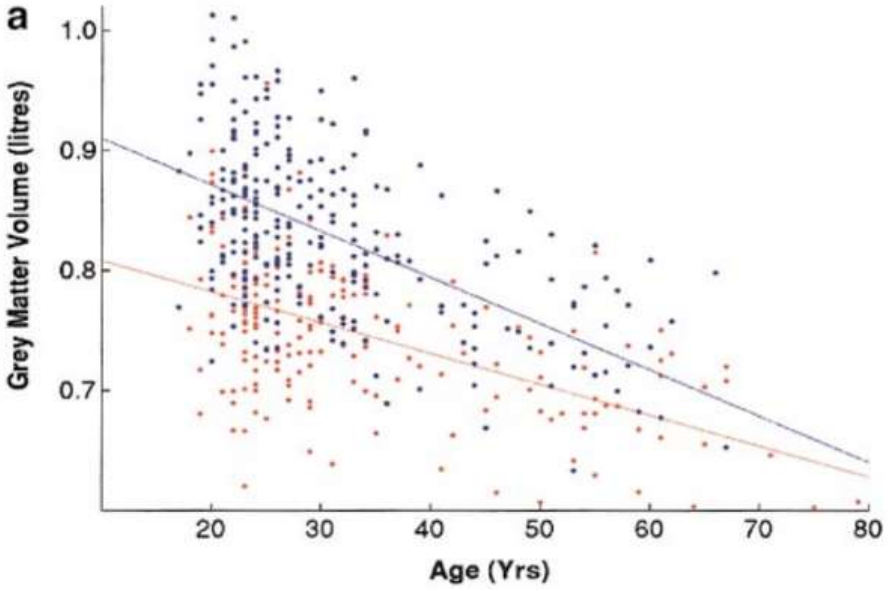
- Extent of smoothing will influence result; ideally, bandwidth should be tuned to expected size of effect

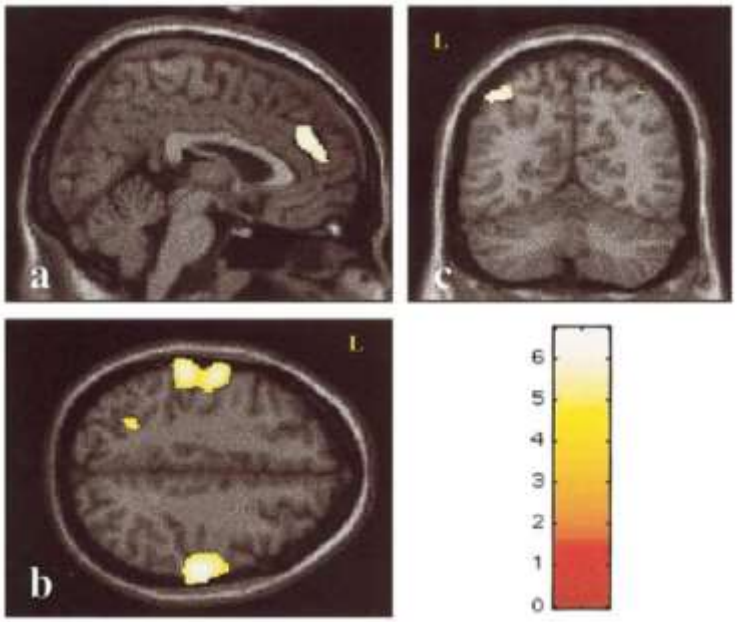


Images from Oxford brain fMRI lab

A VBM Aging Study

- [Good et al. 2001] used VBM to study the effect of age on global and local gray and white matter volumes in 465 healthy adults (200 female, 265 male)
  - Found global GM loss with age, accelerated in some regions, including pre-/postcentral gyrus

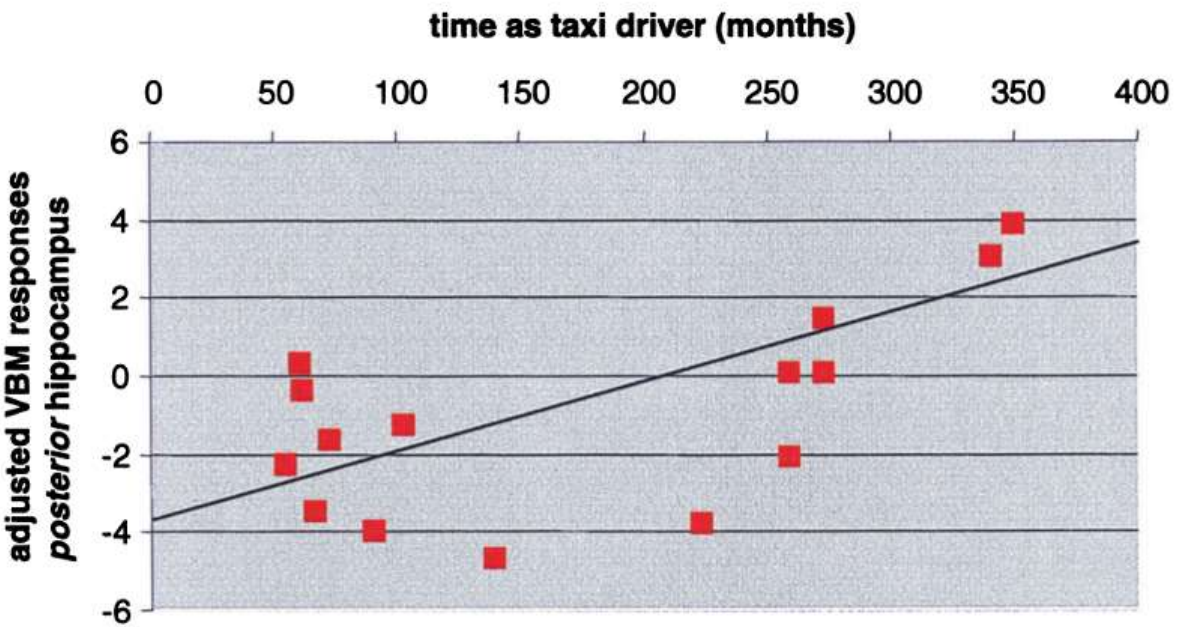


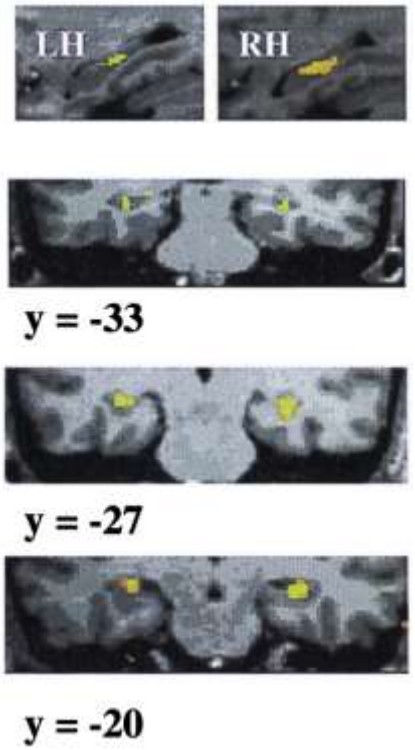


Accelerated loss of GM volume

The VBM Taxi Driver Study

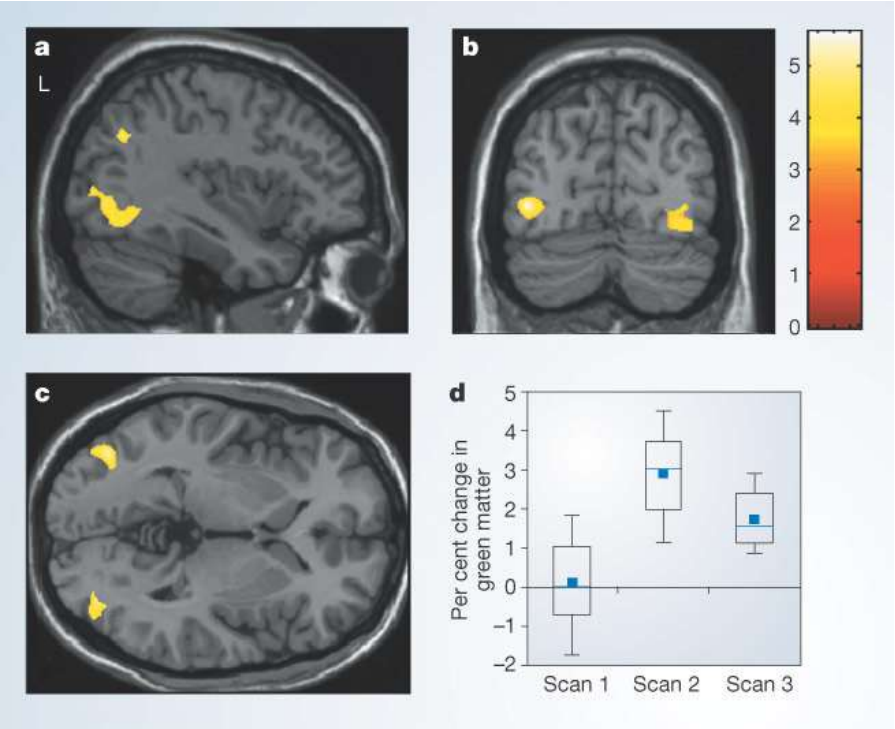
- [Maguire et al. 2000] used VBM to compare the brains of 16 taxi drivers to those of 50 gender- and age-matched non-taxi driving controls
  - Found enlarged posterior hippocampus in taxi drivers, correlated with time spent as a driver (age-corrected)





A Longitudinal VBM Juggling Study

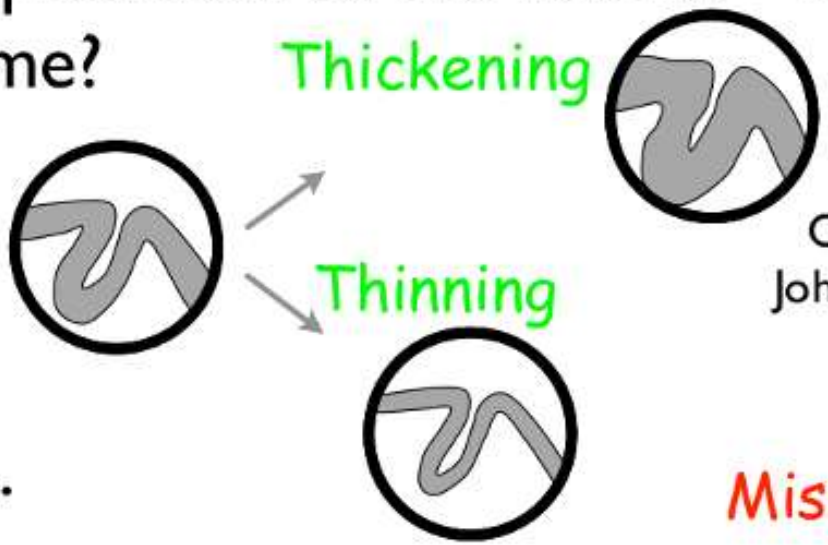
- [Draganski et al. 2004] used VBM to visualize transient structural changes in the brain associated with learning a new task (i.e., juggling)
  - Random assignment of subjects to groups, no significant differences initially
  - Changes localized in vision- and motor-related areas



Issues with Voxel-Based Morphometry

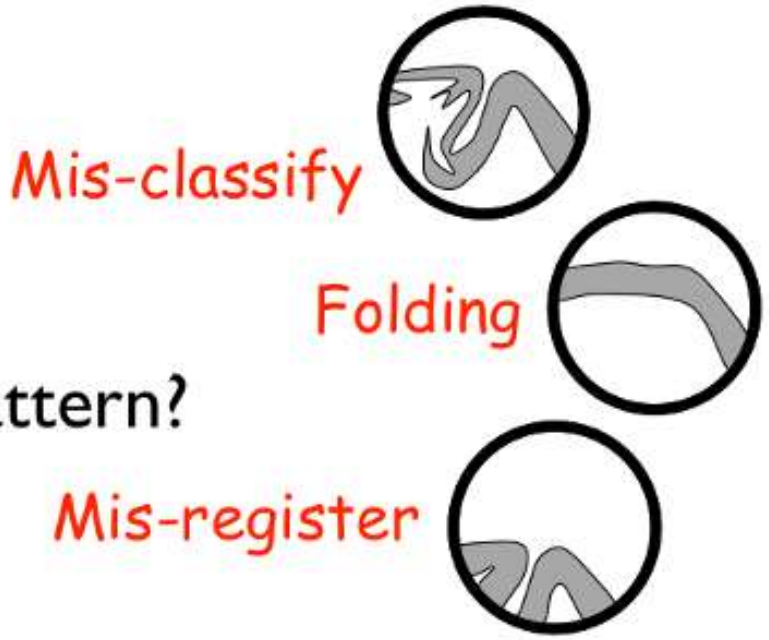
- Controversial approach - back to the issues:

1) Interpretation of the results - real loss/increase of volume?



Courtesy of John Ashburner

Or ...



- Difference in the contrast?
- Difference in gyrification pattern?
- Problem with registration?

Slide from Oxford brain fMRI lab

VBM: Summary

- **Voxel-Based Morphometry (VBM)** provides a tool to study regional changes in gray or white matter volume (or “density”) based on structural MRI scans. It involves:
  - Brain Extraction
  - Normalization
  - Segmentation
  - Modulation
  - Smoothing
  - Statistical Testing
- VBM is used widely and successfully
- Points of criticism include dependence on accuracy of registration and smoothing parameters

Further Reading

- Nicole A. Lazar: *The Statistical Analysis of Functional MRI Data.* Springer, 2008
- Russell A. Poldrack, Jeanette A. Mumford, Thomas E. Nichols: *Handbook of Functional MRI Data Analysis.* Cambridge University Press, 2011
- J. Ashburner, K.J. Friston: *Voxel-Based Morphometry – The Methods.* NeuroImage 11:805-821, 2000 [original VBM paper]
- C.D. Good et al.: *A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains.* NeuroImage 14:21-36, 2001 [optimized VBM paper]