

## All you wanted to know about fMRI\*: \*but were afraid to ask

**Vince D. Calhoun, Ph.D.**  
 Director, Image Analysis & MR Research  
 The MIND Research Network  
 Associate Professor, Electrical and Computer Engineering,  
 The University of New Mexico

## MRI vs. fMRI

**MRI studies brain anatomy.**      **Functional MRI (fMRI) studies brain function.**

Source: Jody Culham's fMRI for Dummies web site

### Hemodynamic Measure of Brain Function (1881)

Arm →  
Brain →

Pressure Traces      "Bertino"

Angelo Mosso

### Blood Oxygen Level Dependent (BOLD)

- Neural activity **increases**
- Blood flow **increases** ("reactive hyperemia")
- Deoxyhemoglobin concentration **decreases**
- Magnetic field homogeneity **increases**
- Gradient echo EPI signal **increases**

## MRI vs. fMRI

**MRI**      **fMRI**

high resolution (1 mm)      low resolution (~3 mm but can be better)

**fMRI**  
 Blood Oxygenation Level Dependent (BOLD) signal  
 indirect measure of neural activity

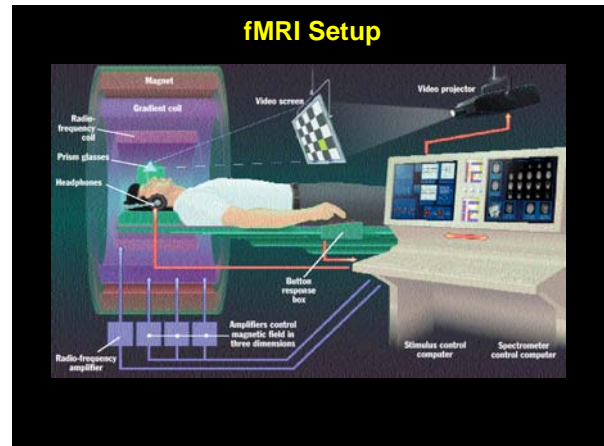
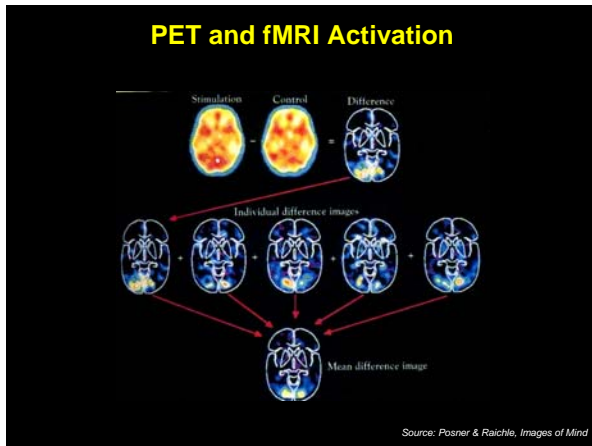
↑ neural activity → ↑ blood oxygen → ↑ fMRI signal

Source: Jody Culham's fMRI for Dummies web site

## fMRI Activation

Flickering Checkerboard  
 OFF (60 s) - ON (60 s) - OFF (60 s) - ON (60 s) - OFF (60 s)

Source: Kwong et al., 1992



### fMRI Experiment Stages: Prep

- 1) Prepare subject**
  - Consent form
  - Safety screening
  - Instructions
- 2) Shimming**
  - putting body in magnetic field makes it non-uniform
  - adjust 3 orthogonal weak magnets to make magnetic field as homogenous as possible
- 3) Sagittals** *Note: That's one g, two t's*
  - Take images along the midline to use to plan slices

Source: Jody Culham's [fMRI for Dummies](#) web site

### fMRI Experiment Stages: Anatomicals

- 4) Take anatomical (T1) images**
  - high-resolution images (e.g., 1x1x2.5 mm)
  - **3D data:** 3 spatial dimensions, sampled at one point in time
  - 64 anatomical slices takes ~5 minutes

Source: Jody Culham's [fMRI for Dummies](#) web site

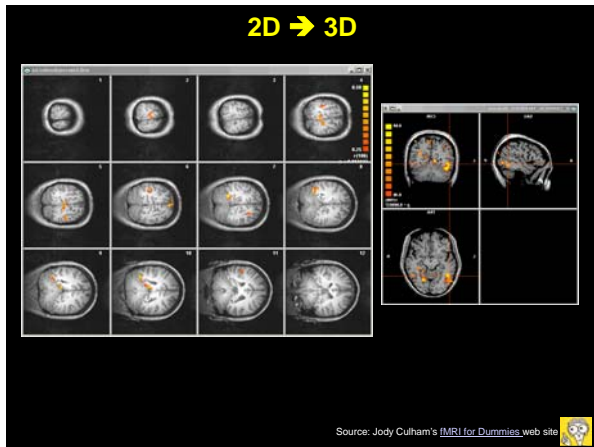
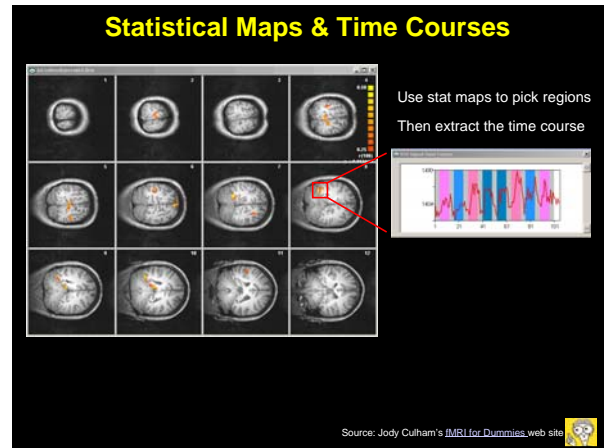
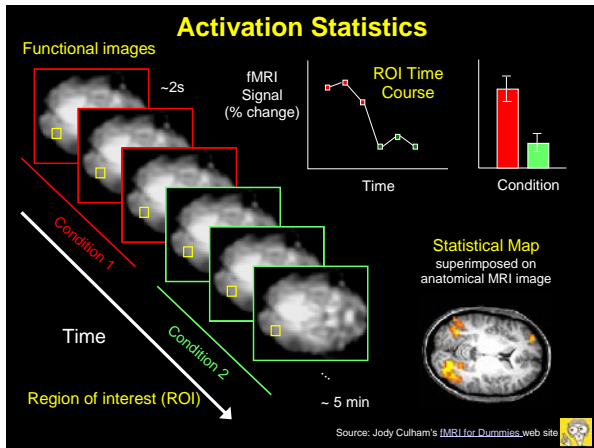
### Slice Terminology

Source: Jody Culham's [fMRI for Dummies](#) web site

### fMRI Experiment Stages: Functionals

- 5) Take functional (T2\*) images**
  - images are indirectly related to neural activity
  - usually low resolution images (3x3x5 mm)
  - all slices at one time = a **volume** (sometimes also called an **image**)
  - sample many volumes (time points) (e.g., 1 volume every 2 seconds for 150 volumes = 300 sec = 5 minutes)
  - **4D data:** 3 spatial, 1 temporal

Source: Jody Culham's [fMRI for Dummies](#) web site



### Design Jargon: Runs

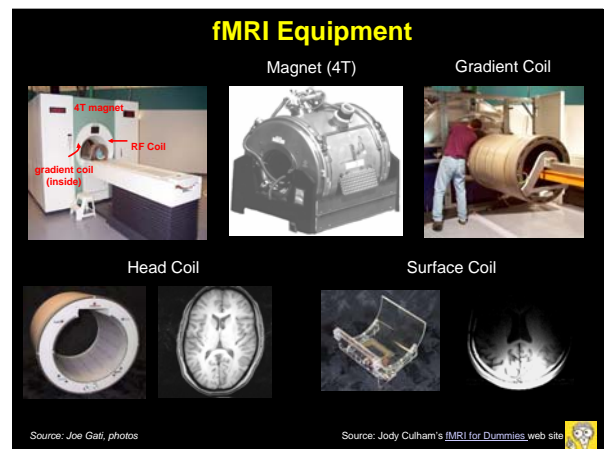
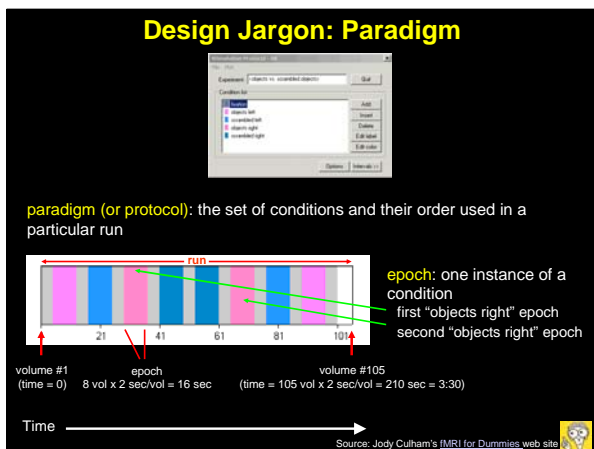
**session:** all of the scans collected from one subject in one day  
**run (or scan):** one continuous period of fMRI scanning (~5-7 min)  
**experiment:** a set of conditions you want to compare to each other  
**condition:** one set of stimuli or one task

Note: Terminology can vary from one fMRI site to another (e.g., some places use "scan" to refer to what we've called a volume).

4 stimulus conditions  
+ 1 baseline condition (fixation)

A session consists of one or more experiments.  
 Each experiment consists of several (e.g., 1-3) runs.  
 More runs/expt are needed when signal:noise is low or the effect is weak.  
 Thus each session consists of numerous (e.g., 5-20) runs (e.g., 0.5 – 3 hours)

Source: Jody Culham's fMRI for Dummies web site

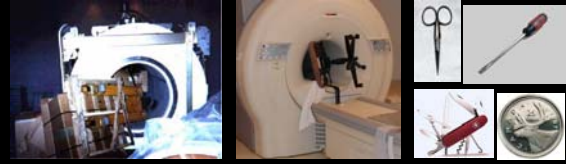


## What Does fMRI Measure?

- **Big magnetic field**
  - protons (hydrogen molecules) in body become aligned to field
- **RF (radio frequency) coil**
  - radio frequency pulse
  - knocks protons over
  - as protons realign with field, they emit energy that coil receives (like an antenna)
- **Gradient coils**
  - make it possible to encode spatial information
- **MR signal differs depending on**
  - concentration of **hydrogen** in an area (anatomical MRI)
  - amount of **oxy- vs. deoxyhemoglobin** in an area (functional MRI)

## Magnet Safety

The whopping strength of the magnet makes safety **essential**.  
Things fly – Even big things!



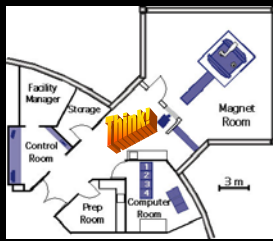
Source: [www.houstonffworks.com](http://www.houstonffworks.com)

Source: <http://www.simgphysics.com/>  
<http://www.abcnews.com/>

Source: Jody Culham's fMRI for Dummys web site 

## Magnet Safety

1. Principal Investigators should be sure all lab members are aware of hazards.
2. Make sure that anyone who is about to enter the magnet room has been filled out consent and screening forms (subjects, lab members, visitors).
3. Remove all metal, coins, credit cards etc. as soon as you enter the magnet area.
4. Think! Train yourself to mini-screen yourself every time you approach the magnet room.
5. Do not enter the magnet room with any tools (e.g., scissors). Use only magnet-friendly tools in the toolbox in the magnet room.



Do the metal macarena!

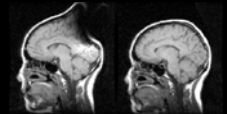
Source: Jody Culham's fMRI for Dummys web site 

## Subject Safety

Anyone going near the magnet – subjects, staff and visitors – must be thoroughly screened:

Subjects must have **no metal in their bodies**:

- pacemaker
- aneurysm clips
- metal implants (e.g., cochlear implants)
- interuterine devices (IUDs)
- some dental work (fillings okay)




This subject was wearing a hair band with a ~2 mm copper clamp. Left: with hair band. Right: without.  
Source: Jorge Jovicich

Subjects must **remove metal from their bodies**

- jewellery, watch, piercings
- coins, etc.
- wallet
- any metal that may distort the field (e.g., underwire bra)

Subjects must be given **ear plugs** (acoustic noise can reach 120 dB)

Source: Jody Culham's fMRI for Dummys web site 

fMRI Basic Experimental Design – block design.

## So you want to do an fMRI study?

Average cost of performing an fMRI experiment in 1998:

**US\$463/hour!!!**



Average cost of performing a thought experiment:

Your Salary



## Thought Experiments

- What do you hope to find? Which brain regions relative to which tasks?
- What would that tell you about the cognitive process involved?
- Would it add anything to what is already known from other techniques?
- Could the same question be asked more easily & cheaply with other techniques?
- Would fMRI add enough to justify the immense expense and effort?
- What would be the alternative outcomes (and/or null hypothesis)?
- Or is there not really any plausible alternative (in which case the experiment may not be worth doing)?
- If the alternative outcome occurred, would the study still be interesting?
- If the alternative outcome is not interesting, is the hoped-for outcome likely enough to justify the attempt?
- What would the headline be if it worked?
- What are the possible confounds? What else could it be?
- Can you control for those confounds? Are appropriate baselines and contrasts available?
- Are there behavioral experimental precedents for your study?
- Has the experiment already been done?

Source: Jody Culham's web slides

## Three Stages of an Experiment

### Sledgehammer Approach

- brute force experiment
- powerful stimulus
- don't try to control for everything
- run a couple of subjects -- see if it looks promising
- if it doesn't look great, tweak the stimulus or task
- try to be a subject yourself so you can notice any problems with stimuli or subject strategies

### Real Experiment

- no more tweaking! Stick with your chosen paradigm (piloted in the sledgehammer approach) and collect more subjects
- incorporate appropriate control conditions
- there is some debate on how many subjects you need
  - some psychophysical studies test two or three subjects
  - many studies test 6-10 subjects
  - random effects analysis requires at least 10 subjects
- can run all subjects in one or two days
  - pro: minimize setup and variability
  - con: "bad magnet day" means a lot of wasted time

### Whipped Cream

- after the real experiment works, then think about a "whipped cream" version
- going straight to whipped cream is a huge endeavor, especially if you're new to imaging

Source: Jody Culham's web slides

## Parameters for Neuroimaging

### You decide:

- number of slices -- how much of the brain do you want to get?
- slice orientation -- ties in with how much but also considers where you want to image?
- slice thickness -- fewer slices with greater thickness may get more brain but will give lower spatial resolution!
- in-plane resolution (field of view and matrix size)
- volume acquisition time -- within the limits of the system -- can't do 4 shot EPI with a TR of 0.5 over 32 slices!!!!
- length of a run (volume time x no. of volumes) -- consider subject comfort here!
- number of runs -- basic power issue, do you have enough to get a reliable effect -- this differs for block vs. event-related fMRI
  - duration and sequence of epochs within each run
    - counterbalancing within or between subjects
    - usually you want to use epoch lengths that are integer multiples of your volume time (e.g., if your volume time is 2 sec (TR=2), make blocks an even number of seconds, such as 16 not 15) -- for event-related fMRI you want your events to be multiples of the TR.

### Your physicist can help you decide:

- pulse sequence (e.g., gradient echo vs. spin echo)
- number of shots
- TR, TE, flip angle, etc.

Source: Jody Culham's web slides

## Tradeoffs



"fMRI is like trying to assemble a ship in a bottle -- every which way you try to move, you encounter a constraint" -- Mel Goodale

"That's on a good day. On a bad day, it's like trying to assemble a ship in a bottle blindfolded, drunk and with one hand tied behind your back" -- Jody Culham

### Number of slices vs. volume acquisition time

- the more slices you take, the longer you need to acquire them
- e.g., 15 slices (4 shot) in 2 sec vs. 20 slices (4 shot) in 3 sec

### Number of slices vs. number of shots

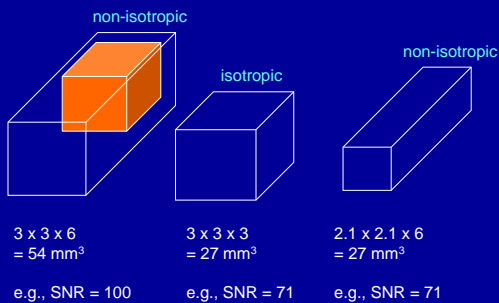
- 4 shot slices take longer than 2 shot slices
- e.g., 15 slices (4 shot) in 2 sec vs. 18 slices (2 shot) in 2 sec

### Number of slices vs. in-plane resolution

- the higher your in-plane resolution, the fewer slices you can acquire in a constant volume acquisition time
- e.g., 128x128 matrix for 1 slice vs. a 64x64 matrix for 4 slices

Source: Jody Culham's web slides

## Voxel size



In general, larger voxels buy you more SNR.

EXCEPT when the activated region does not fill the voxel (partial voluming)

Source: Jody Culham's web slides

## Power issues.

### Statistical Power

- the probability of rejecting the null hypothesis when it is actually false
- "if there's an effect, how likely are you to find it?"

### Effect size

- bigger effects, more power
  - e.g., motor or somatosensory tasks (move vs. rest or touch vs rest) -- signal is usually so strong that 1 run is usually enough
  - looking for activation during imagined movements might require many more runs
  - similarly, single event designs may need more runs even for motor tasks!

### Sample size

- larger n, more power
  - more subjects
  - longer runs
  - more runs

### Signal:Noise Ratio

- better SNR, more power
  - stronger, cleaner magnet
  - more focal coil
  - fewer artifacts
  - more filtering

Source: Jody Culham's web slides

### Subtraction Logic

Cognitive subtraction originated with reaction time experiments (F. C. Donders, a Dutch physiologist).



Measure the time for a process to occur by comparing two reaction times, one which has the same components as the other + the process of interest.

Example:

- T1: Hit a button when you see a light
- T2: Hit a button when the light is green but not red
- T3: Hit the left button when the light is green and the right button when the light is red

T2 – T1 = time to make discrimination between light color

T3 – T2 = time to make a decision

**Assumption of pure insertion:** You can insert a component process into a task without disrupting the other components.

Source: *Jody Culham's web slides*

### Subtraction Logic: Brain Imaging

Example: Simple motor task

- T1: tap fingers alternately
- T2: rest

T2 – T1 = "finger motor" areas

Possible factors added

- motor control
- monitoring of performance

Possible factors removed

- cognition! This is a boring task!

Possible confounds

- motion artifacts in one condition but not the other
- rest possibly not the best baseline

You must always consider the possible components you could be adding or affecting. More sophisticated designs (e.g., parametric designs, conjunction designs) may better address true contribution of components.

Source: *Jody Culham's web slides*

### Subtraction Logic: Brain Imaging

Example: Simple vs complex motor task

- T1: tap fingers alternately
- T2: tap fingers in sequence

T2 – T1 = possible "praxis" areas

Possible factors added

- motor control
- sequential motor programming – praxis?

Possible factors removed

- attentional load – T2 is harder than T1

Source: *Jody Culham's web slides*

### Dealing with Attentional Confounds

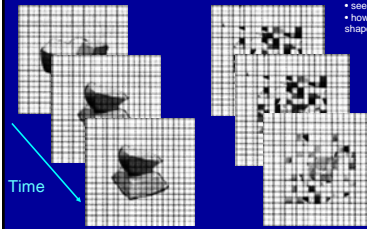
fMRI data seem highly susceptible to the amount of attention drawn to the stimulus or devoted to the task.

How can you ensure that activation is not simply due to an attentional confound?

Add an attentional requirement to all stimuli or tasks.

One-back tasks – A Kanwisher lab eg...

- Basic experiment:
- compare intact shapes to scrambled shapes during passive viewing
  - see activation in lateral occipital complex (LOC)
  - how can we be sure it's not just that the intact shapes are more attentionally engaging?



- Add a "one back" task
- subject must hit a button whenever a stimulus repeats
  - the repetition detection is much harder for the scrambled shapes
  - any activation for the intact shapes cannot be due only to attention

- Other common confounds that reviewers love to hate:
- eye movements
  - motor movements

### Is concurrent behavioral data necessary?

"Ideally, a concurrent, observable and measureable behavioral response, such as a yes or no bar-press response, measuring accuracy or reaction time, should verify task performance."

– Mark Cohen & Susan Bookheimer, *TINS*, 1994

"I wonder whether PET research so far has taken the methods of experimental psychology too seriously. In standard psychology we need to have the subject do some task with an externalizable yes-or-no answer so that we have some reaction times and error rates to analyze – those are our only data. But with neuroimaging you're looking at the brain directly so you literally don't need the button press... I wonder whether we can be more clever in figuring out how to get subjects to think certain kinds of thoughts silently, without forcing them to do some arbitrary classification task as well. I suspect that when you have people do some artificial task and look at their brains, the strongest activity you'll see is in the parts of the brain that are responsible for doing artificial tasks."

– Steve Pinker, interview in the *Journal of Cognitive Neuroscience*, 1994

But what about measuring component processes?

Source: *Nancy Kanwisher*

### Change only one thing between conditions!

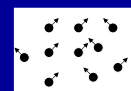
As in Donders' method, in functional imaging studies, two paired conditions should differ by the inclusion/exclusion of a *single* mental process

How do we control the mental operations that subjects carry out in the scanner?

- i) Manipulate the stimulus
  - works best for automatic mental processes



- ii) Manipulate the task
  - works best for controlled mental processes



**DON'T DO BOTH AT ONCE!!!**

Source: *Nancy Kanwisher*

## And put your conditions in the same run!

As far as possible, put the two conditions you want to compare within the same run.

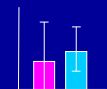
Why?

- subjects get drowsy and bored
- magnet may have different amounts of noise from one run to another (e.g., spike)

Common flawed logic:

Run1: A – baseline  
Run2: B – baseline

"A – 0 was significant, B – 0 was not, ∴ Area X is activated by A more than B"



Faces Places  
Error bars = 95% confidence limits

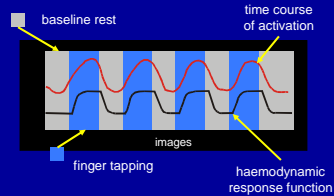
If you do this, you can get a situation where A is significantly > 0 but B is not, yet the difference between A and B is not significant

Bottom line: If you want to compare A vs. B, compare A vs. B! Simple, eh?

Source: *Jody Culham's web slides*

## Block Design Sequences

Consider the simplest case, a block design with two conditions e.g., alternate tapping of two fingers vs. rest let's assume 2 sec/volume



How long should a run be?

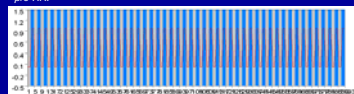
- Short enough that the subject can remain comfortable without moving or swallowing.
- Long enough that you're not wasting a lot of time restarting the scanner.
- Ideal is  $-5 \pm 2$  minutes

Source: *Jody Culham's web slides*

## Block Design Sequences

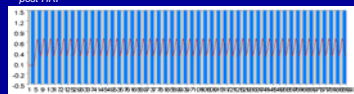
How fast should the conditions cycle?

pre-HRF



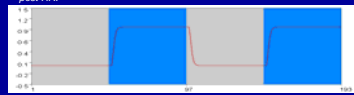
Every 4 sec (2 images)  
• signal amplitude is weakened by HRF  
• not too far from range of breathing frequency (every 4-10 sec) → could lead to respiratory artifacts

post-HRF



• if design is a task manipulation, subject is constantly changing tasks, gets confused

post-HRF

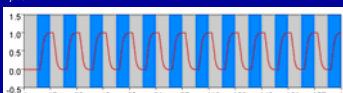


Every 96 sec (48 images)  
• more noise at low frequencies  
• linear trend confound  
• subject will get bored  
• very few repetitions – hard to do eyeball test of significance

Source: *Jody Culham's web slides*

## Block Design Sequences

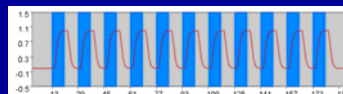
post-HRF



Every 16 sec (8 images)  
• allows enough time for signal to oscillate fully  
• not near artifact frequencies  
• enough repetitions to see cycles by eye  
• a reasonable time for subjects to keep doing the same thing

Other factors:

- symmetric design – some use longer rest vs. activation periods
- add a few extra images at the end to allow the hemodynamic response to catch up
- add extra time at the beginning to allow for the magnet to warm up and the subject to warm up (let the startle response die down)



Source: *Jody Culham's web slides*

## Nancy K's PET (and fMRI) Peeves 1-5

1. "Brain Area X is activated by Task A."

Compared to what? Activations are differences!

2. "Baseline".

Baseline's are not "null activation" conditions – appropriate comparisons are the key.

3. Inferring: Because Region X responded significantly more strongly in Task A than control, but didn't respond significantly more strongly in Task B than control, it is selectively activated by Task A.

A difference in significances is not necessarily a significant difference.

4. Imputing a specific function to a region of cortex from a difference in only two conditions.

Data always underdetermines theory, but reasonable hypotheses about function require multiple tests applied to the same region of cortex.

5. "Gyrus X was active in my comparison of tasks B and C, and in Joe Shmo's comparison of tasks D and E, so the same area must be involved in both tasks B and D."

Gyri can be very big places; need within-subject data.

Source: *Nancy Kanwisher*

## Nancy K's PET (and fMRI) Peeves 6-10

6. Difficulty/attention confounds.

Task manipulations are subject to the former; stimulus manipulations subject to the latter – but don't use "attention" as an easy out!

7. Any study with concludes: "The results of the present study demonstrate that Task A is carried out in a distributed network of cortical areas."

This feels like a copout.

8. Localizing tasks, not processes.

Lexical decision and stem completion are tools to study cognitive processes, not basic cognitive operations.

9. Showing the data from the "best" voxel.

With 20,000 voxels and noisy data this can be very dangerous – as can be showing the best subject.

10. "I did it because it was an option in SPM."


Don't run any analyses you don't understand! Don't assume that what everyone else is doing must be OK.

Source: *Nancy Kanwisher*


### Dealing with frustration

Murphy's law acts with particular vigour in fMRI imaging:

- Number of pieces of equipment required in an fMRI experiment: ~50
- Probability of any one piece of equipment working in a session: 95%
- Probability of everything working in a session:  $0.95^{50} = 7.6\%$



Sign that used to be at the 1.5 T at MCH




Solution for a good imaging session =  
\$4 million magnet  
+ \$3 roll of duct tape

Source: Jody Culham's web slides

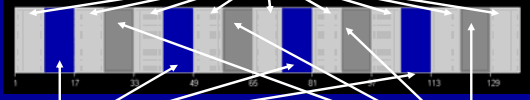
## DATA ANALYSIS

### A Simple Experiment



Lateral Occipital Complex  
• responds when subject views objects

Blank Screen



Intact Objects

Scrambled Objects

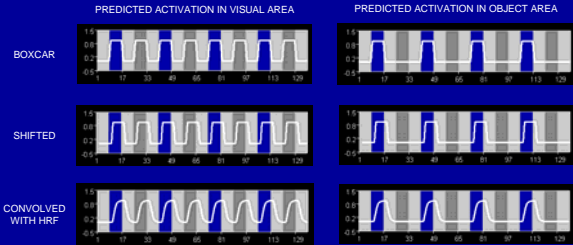
TIME →

One volume (12 slices) every 2 seconds for 272 seconds (4 minutes, 32 seconds)  
Condition changes every 16 seconds (8 volumes)

Source: Jody Culham's fMRI for Dummies web site

### Predicted Responses

- fMRI is based on the Blood Oxygenation Level Dependent (BOLD) response
- It takes about 5 sec for the blood to catch up with the brain
- We can model the predicted activation in one of two ways:
  - shift the boxcar by approximately 5 seconds (2 images x 2 seconds/image = 4 sec, close enough)
  - convolve the boxcar with the hemodynamic response to model the shape of the true function as well as the delay



PREDICTED ACTIVATION IN VISUAL AREA

PREDICTED ACTIVATION IN OBJECT AREA

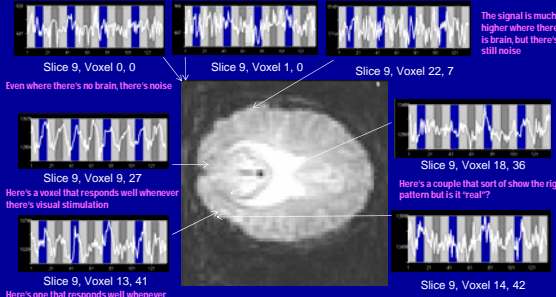
BOXCAR

SHIFTED

CONVOLVED WITH HRF

### Why do we need stats?

- We could, in principle, analyze data by voxel surfing: move the cursor over different areas and see if any of the time courses look interesting



Slice 9, Voxel 0, 0

Slice 9, Voxel 1, 0

Slice 9, Voxel 22, 7

Slice 9, Voxel 9, 27

Slice 9, Voxel 13, 41

Slice 9, Voxel 18, 36

Slice 9, Voxel 14, 42

Even where there's no brain, there's noise

The signal is much higher where there is brain, but there's still noise

Here's a voxel that responds well whenever there's visual stimulation


Here's one that responds well whenever there's intact objects

Here's a couple that sort of show the right pattern but is it "real"?

Source: Jody Culham's fMRI for Dummies web site

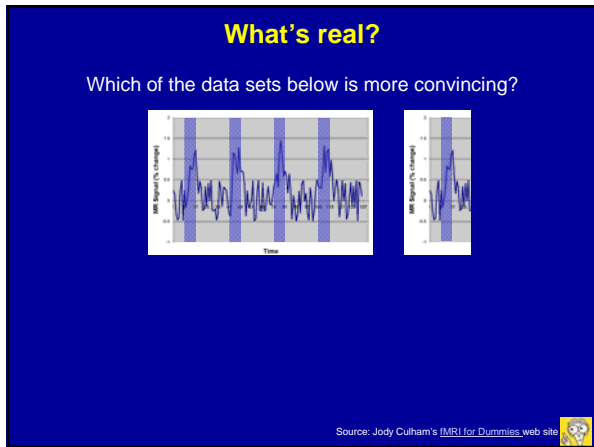
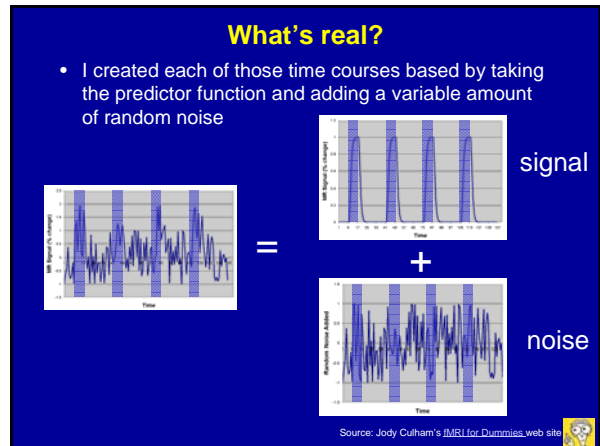
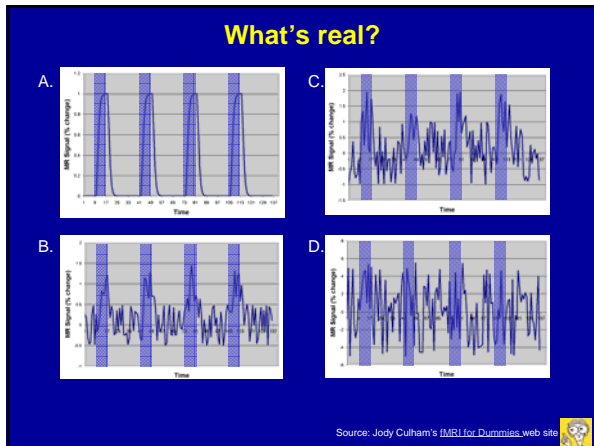
### Why do we need stats?

- Clearly voxel surfing isn't a viable option. We'd have to do it 49,152 times and it would require a lot of subjective decisions about whether activation was real
- This is why we need statistics
- Statistics:
  - tell us where to look for activation that is related to our paradigm
  - help us decide how likely it is that activation is "real"



The lies and damned lies come in when you write the manuscript

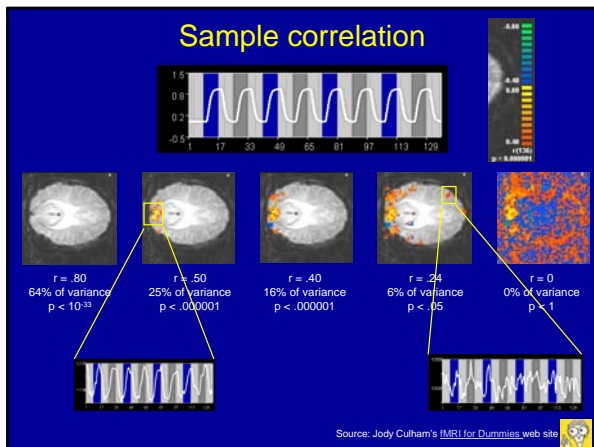
Source: Jody Culham's fMRI for Dummies web site



### Formal Statistics

- Formal statistics are just doing what your eyeball test of significance did
  - Estimate how likely it is that the signal is real given how noisy the data is
- confidence**: how likely is it that the results could occur purely due to chance?
- "p value"** = probability value
  - If "p = .03", that means there is a .03/1 or 3% chance that the results are bogus
- By convention, if the probability that a result could be due to chance is less than 5% ( $p < .05$ ), we say that result is **statistically significant**
- Significance depends on
  - signal (differences between conditions)
  - noise (other variability)
  - sample size (more time points are more convincing)

Source: Jody Culham's [fMRI for Dummies web site](#)



### Complications

- Not only is it hard to determine what's real, but there are all sorts of statistical problems

#### Potential problems

- data may be contaminated by artifacts (e.g., head motion, breathing artifacts)
- $.05 * 49,152 = 2457$  "significant" voxels by chance alone
- many assumptions of statistics (adjacent voxels uncorrelated with each other; adjacent time points uncorrelated with one another) are false

#### What's wrong with these data?

$r = .24$   
6% of variance  
 $p < .05$

Source: Jody Culham's [fMRI for Dummies web site](#)

## Approach #1: Region of interest (ROI) analysis

- If you are looking at a well-established area (such as visual cortex, motor cortex, or the visual motion area MT+), it's fairly easy to activate and identify the area

- Do the stats and play with the threshold till you get something believable in the right vicinity based on anatomical location (e.g., sulcal landmarks) or functional location (e.g., Talairach coordinates from prior studies)
- Once you have found the ROI, do independent experiments, extract the time course information and determine whether activation differences between conditions are significant
  - Because the runs that are used to generate the area are independent from those used to test the hypothesis, liberal statistics ( $p < .05$ ) can be used

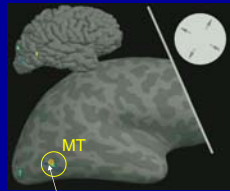
Source: Jody Culham's fMRI for Dummies web site

## Example of ROI approach

Tootell et al, 1995, *Nature*, Motion Aftereffect

Are motion-responsive areas of the brain active during illusory motion?

Run #1: Low contrast moving rings vs. stationary rings

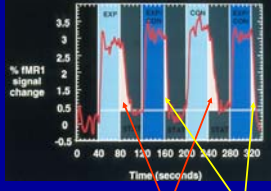


MT

We can be fairly sure this is MT+ because:

- nothing else in that vicinity is activated by low contrast motion
- the Talairach coordinates match previous studies
- the activation is at the junction of the inferior temporal sulcus and lateral occipital sulcus

Runs #2, 3 and 4: Alternate between two new conditions, unidirectional motion (which causes a motion aftereffect) and bidirectional motion (which does not cause an aftereffect). Look at activation during subsequent stationary pattern.



Time (seconds)

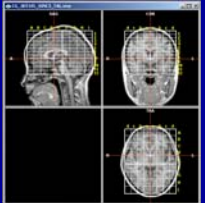
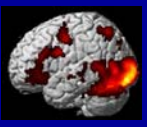
Source: Tootell et al., 1995

Source: Jody Culham's fMRI for Dummies web site

## Approach #2: Whole Brain Stats

- You don't necessarily need a priori hypotheses (though sometimes you can use less conservative stats if you have them)
- Average all of your data together in Talairach space
- Compare two (or more) conditions using precise statistical procedures and assumptions. Anything that passes at a carefully determined threshold is considered real.
- Make a list of these areas and publish it.

This is the tricky part!






Source: Jody Culham's fMRI for Dummies web site

## Example of Whole Brain Stats

Chao & Martin, 2000, Neuroimage

Which areas of the dorsal brain are activated by images of tools?

- Compare images of tools with images of animals
- Standard SPM-style preprocessing (motion correction, spatial smoothing, minimum cluster of 7 contiguous voxels)
- Publish list of significant activations

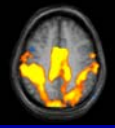
How many foci is this?

Condition	n	Volume (mm <sup>3</sup> )		Number of clusters	
		k = 5	k = 7	k = 5	k = 7
Tools (vs. Animals)	10	100	100	10	10
Animals (vs. Tools)	10	100	100	10	10
Tools (vs. Tools)	10	100	100	10	10
Animals (vs. Animals)	10	100	100	10	10

I chose this paper because the list is short, though sometimes these tables go on for pages!

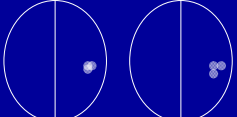
Source: Jody Culham's fMRI for Dummies web site

## Some problems



So what's the Talairach coordinate here?

Both of the data sets below show activation sites in three Talairached subjects superimposed. Will they provide the same whole brain stats?



Source: Jody Culham's fMRI for Dummies web site

## Comparing the two approaches

### Region of Interest (ROI) Analyses

- Gives you more statistical power because you do not have to correct for the number of comparisons
- Hypothesis-driven
- ROI is not smeared due to intersubject averaging
- Easy to analyze and interpret
- Neglects other areas which may play a fundamental role
- Popular in North America

### Whole Brain Analysis

- Requires no prior hypotheses about areas involved
- Includes entire brain
- Often neglects individual differences
- Can lose spatial resolution with intersubject averaging
- Can produce meaningless "laundry lists of areas" that are difficult to interpret
- You have to be fairly stats-savvy
- Popular in Europe

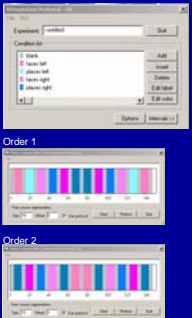
NOTE: Though different experimenters tend to prefer one method over the other, they are NOT mutually exclusive. You can check ROIs you predicted and then check the data for other areas.

Source: Jody Culham's fMRI for Dummies web site

## Example Experiment

### Faces vs. Places in Left or Right or Central Visual Field

- pilot experiment in Subject JC in preparation for hemianopic patient
- Purpose: Can we identify the fusiform face area (FFA) and parahippocampal place area (PPA) with peripheral viewing?
- 7 conditions
- each visual condition has a stream of faces or places (e.g., 1/sec)
- head coil
- 12 quasi-coronal slices
- volume time = 2 sec
- 3 x 3 x 6 voxels
- 155 volumes per run
- 2 orders
  - 2 runs x order 1
  - 2 runs x order 2
- fixation epochs = 8 sec (4 vol)
- image epochs = 16 sec (8 vol)



Source: Jody Culham's fMRI for Dummies web site

## Statistical Analyses: T-test

### T-test

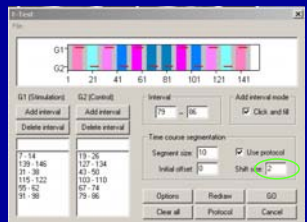
- simple, sometimes seems more reliable than fancy stuff
- compare two conditions (or sets of conditions)
- shift activation to compensate for hemodynamic lag

### Kolmogorov-Smirnov

- non-parametric version of t-test
- sensitive to differences in variance as well as means
- controversial – best to avoid

Source: Jody Culham's fMRI for Dummies web site

## T-test: Setup



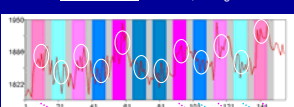
**Sample question:**  
Can we find the FFA and PPA (collapsing across visual field locations)?

**Note:** We must shift the function to compensate for the hemodynamic lag. Assume 4-6 sec (at 2 sec/volume, that's 2-3 images) shift

Source: Jody Culham's fMRI for Dummies web site

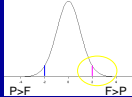
## T-test: Stats

To look for **Faces – Places** Activation, for a given voxel:

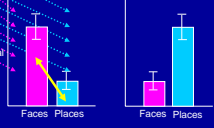


Measure average MR signal and SD for each volume in which **faces** were presented (6 epochs x 8 volumes/epoch = 48 volumes)

Measure average MR signal and SD for each volume in which **places** were presented (48 volumes)



To look for **Places – Faces** Activation, look at the negative tail of the comparison



**Determine if mean difference is statistically significant:**  
Calculate t value.  
Use t to look up p value for that number of degrees of freedom (df = 48 x 2 = 96).

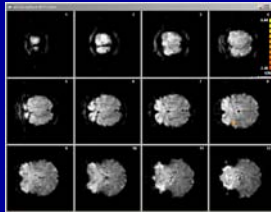
e.g., For -96 df  
 $t > 1.98 \rightarrow p < .05$  (1/20 chance)  
 $t > 3.39 \rightarrow p < .001$  (1/1000 chance)

Repeat this process 49,152 more times (64x64x12), once for each voxel in the volume obtained.

Source: Jody Culham's fMRI for Dummies web site

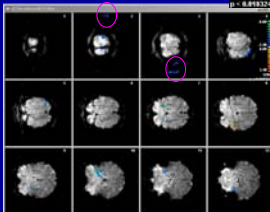
## T-test: Maps

For each voxel in the brain, we can now color code a map based on the computed t and p values:



We can do this for the **positive tail (Faces – Places)**  
 Orange = low significance  
 Yellow = high significance

Schmitz or ESP voxels?



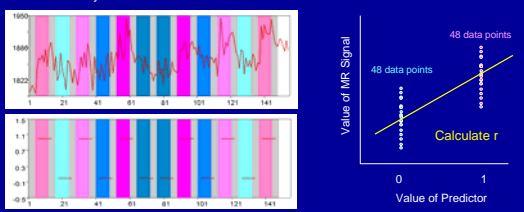
And we can also do this for the **negative tail (Places – Faces)**  
 Blue = low significance  
 Green = high significance

Source: Jody Culham's fMRI for Dummies web site

## Statistical Analyses: Basic Correlation

### Correlation analysis

- voxels with time course correlated with reference function
- can incorporate hemodynamic response function (HRF) to predict time course more accurately



For each voxel:

- Find the correlation between the predictor and the MR signal
- Extract the correlation (r value) and find the corresponding p value.
- Determine whether it is statistically significant
- In this example, similar in spirit to a t-test.

Remember  $r^2$  is the proportion of variance accounted for by our predictor, e.g., if  $r = .7$ ,  $r^2 = .5 = 50\%$

Source: Jody Culham's fMRI for Dummies web site

## Hemodynamic Response Function

**Initial dip**  
-more focal and potentially a better measure  
-somewhat elusive so far, not everyone can find it

**time to rise**  
signal begins to rise soon after stimulus begins

**time to peak**  
signal peaks 4-6 sec after stimulus begins

**post stimulus undershoot**  
signal suppressed after stimulation ends

Source: Jody Culham's fMRI for Dummies web site

## Correlations: Incorporating the HRF

We can model the expected curve of the data by convolving our predictor with the hemodynamic response function.

To find a face-responsive area, we can correlate the convolved face predictor with each voxel time course

Source: Jody Culham's fMRI for Dummies web site

## Correlations: Negative Tail

Note:  
The negative no longer identifies the PPA  
Now it's anything that shows a reduced response during faces (compared to both places and fixation)

Source: Jody Culham's fMRI for Dummies web site

## Problems with t-tests and correlations

- 1) How do we evaluate runs with different orders?  
Right now, we could average our two runs done in Order1 together, and also average our two runs done in Order2 together and then do stats on the two orders separately. There is no way to collapse between orders. If there is an artifact on part of one run, we have to exclude the whole run.
- 2) If we test more subjects, how can we evaluate the subjects together?  
As with the single subject runs, we could average all the subjects together (after morphing them into a common brain space) but that still means we have to run all of them in the same order.
- 2) We can get nice hemodynamic predictors for faces, and for places, but how can we compare them accurately?

If this predictor is significant, we won't know if it's because Faces>Places vs. because Faces>Fixation

**Solution: General Linear Model**

Source: Jody Culham's fMRI for Dummies web site

## General Linear Model (GLM): Logic

Parcel out variance in the voxel's time course to the contributions of six predictors plus residual noise (what the predictors can't account for).

$$\text{fMRI signal} = \beta_1 \times \text{Predictor 1} + \beta_2 \times \text{Predictor 2} + \beta_3 \times \text{Predictor 3} + \beta_4 \times \text{Predictor 4} + \beta_5 \times \text{Predictor 5} + \beta_6 \times \text{Predictor 6} + \text{residuals}$$

Design Matrix

Adapted from Brain Voyager course slides

## GLM Baseline

Here are all our 6 GLM predictors shown together:

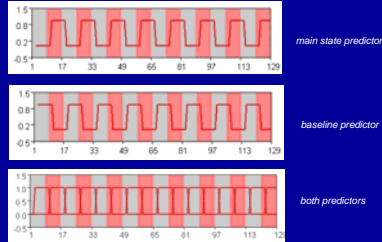
Why is there no "baseline" predictor?

Because if there was, the model would be overdetermined (everything would be high at some point).

Source: Jody Culham's fMRI for Dummies web site

## GLM Baseline

To understand why overdetermination is a problem, consider an example with only two states (e.g., an MT localizer comparing moving rings to stationary rings) and shifted rather than convolved with the HRF:



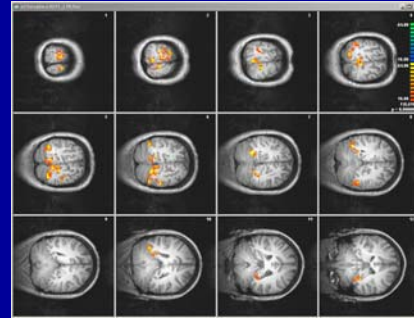
The baseline predictor is exactly the *inverse* of the main state predictor (the negative tail in a t-test or correlation). If we know the strength of the main state predictor, we must know the strength of the baseline predictor (baseline = main \* -1) so the second predictor adds no information to our model. The problem extends to more predictors and HRF-convolved models.

In the GLM, the number of predictors <= number of states - 1

Source: Jody Culham's fMRI for Dummies web site

## GLM Output

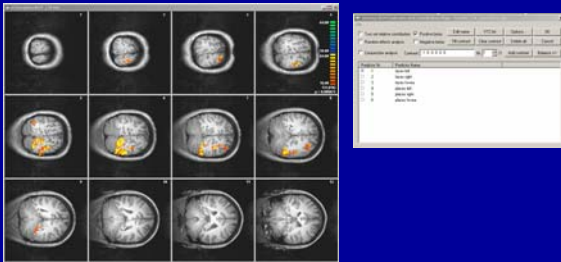
Initially, the output shows us where our model (or any part within it) accounts for a significant amount of variance:



Source: Jody Culham's fMRI for Dummies web site

## Single Predictors

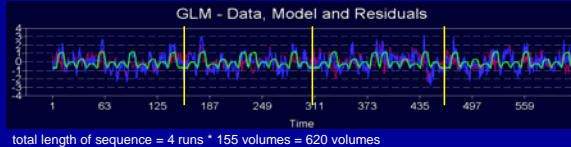
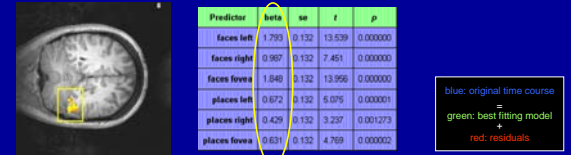
We can look at voxels where a single predictor (e.g., faces left) accounts for a significant amount of variance:



Source: Jody Culham's fMRI for Dummies web site

## GLM Stats

For any given region, we can evaluate the GLM stats



Source: Jody Culham's fMRI for Dummies web site

## GLM Stats

ANOVA					
Source of Variation	df	SS	MS	F	p
Regression	9	208.111	23.123	34.245	0.000000
Model   Confounds	6	208.111	34.685	51.368	0.000000
Residual	610	411.889	0.675		
Total	619	620.000	1.002		

data points = 620 Model | confounds: R = 0.579 Adj R = 0.571 AR(1) = 0.374

Entire model is significant for this region and accounts for 0.579<sup>2</sup> = 33.5% of its variance

Predictor	beta	se	t	p
faces left	1.793	0.132	13.539	0.000000
faces right	0.987	0.132	7.451	0.000000
faces fovea	1.848	0.132	13.956	0.000000
places left	0.672	0.132	5.075	0.000001
places right	0.428	0.132	3.237	0.001273
places fovea	0.631	0.132	4.769	0.000002

beta = weight of predictor in model  
SE = standard error (variability in estimates)  
t = beta/SE (e.g., 1.793/0.132 = 13.58)  
p = probability value for that level of t

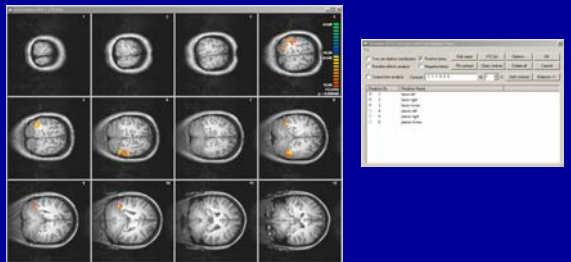
Source of Variation	df	SS	MS	F	p
100000	1	123.781	123.781	183.317	0.000000

$$F = t^2$$

Source: Jody Culham's fMRI for Dummies web site

## Combination Predictors

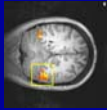
We can look at voxels where a combination of predictors (e.g., all face conditions) accounts for a significant amount of variance:



Source: Jody Culham's fMRI for Dummies web site

## GLM Combo Stats

For any given region, we can evaluate the GLM stats for the combination of predictors:



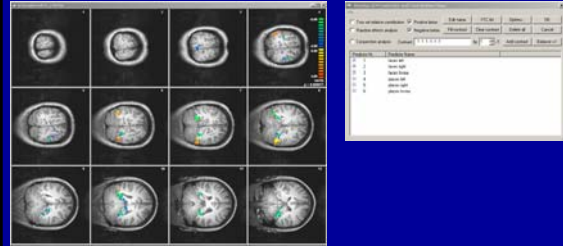
Predictor	beta	se	t	p	
faces left	1.714	0.132	13.002	0.000000	
faces right	1.103	0.132	8.369	0.000000	
faces fovea	1.978	0.132	15.000	0.000000	
places left	0.763	0.132	5.790	0.000000	
places right	0.595	0.132	4.439	0.000011	
places fovea	0.626	0.132	4.767	0.000000	
Contrast / LF	value	se	t	p	
111000	4.795	0.269	16.906	0.000000	
Source of Variation	df	SS	MS	F	p
111000	3	206.741	68.914	102.929	0.000000

The sum of the 3 face predictors ( $1.714 + 1.103 + 1.978 = 4.795$ ) are used in the computation of t (Note: the SE is not computed from the sum of the 3 SEs).

Source: Jody Culham's fMRI for Dummies web site

## Contrasting Predictors

We can look at voxels where a contrast between predictors (e.g., all face conditions vs. all place conditions) accounts for a significant amount of variance:



Source: Jody Culham's fMRI for Dummies web site

## GLM Contrast Stats

For any given region, we can evaluate the GLM stats for the contrast between predictors:

Predictor	beta	se	t	p
faces left	1.677	0.133	12.600	0.000000
faces right	0.929	0.133	6.963	0.000000
faces fovea	1.736	0.133	13.048	0.000000
places left	0.256	0.133	1.924	0.054855
places right	0.243	0.133	1.829	0.068063
places fovea	0.342	0.133	2.570	0.010407
Contrast / LF	value	se	t	p
111-1-1-1-1	3.501	0.273	12.826	0.000000

Sum of the 3 face predictors minus the sum of the 3 place predictors ( $1.677 + 0.929 + 1.736 - 0.256 - 0.243 - 0.342 = 0.841 = 3.501$ ) is used in the computation of t (Note: the SE is not computed from the sum of the 6 SEs).

Source: Jody Culham's fMRI for Dummies web site

## Flexibility of GLM

With our example data, we could ask many more questions such as:

left vs. right field

peripheral vs. foveal stimulation

factorial design:

stimulus (face/place)

field (left/right/foveal)

interaction between visual field and face/place.

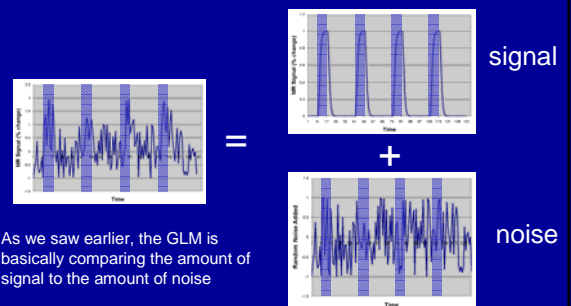
Source: Jody Culham's fMRI for Dummies web site

## Advantages of General Linear Model (GLM)

- can perform data analysis within and between subjects without needing to average the data itself
- allows you to counterbalance orders
- allows you to exclude segments of runs with artifacts
- can perform more sophisticated analyses (e.g., 2 factor ANOVA with interactions)
- easier to work with (do one GLM vs. many t-tests and correlations)

Source: Jody Culham's fMRI for Dummies web site

## Maximizing Your Power



As we saw earlier, the GLM is basically comparing the amount of signal to the amount of noise

How can we improve our stats?

- increase signal
- decrease noise
- increase sample size (keep subject in longer)

Source: Jody Culham's fMRI for Dummies web site

## How to Increase Signal

- Go with a stronger magnet (4T rules!)
- Go with a better coil (surface coil)
  - Tradeoff: lose other areas
- Take bigger voxels
  - Tradeoff: lose spatial resolution, if voxels are too big, you lose signal
- Make sure your subject is paying attention to the stimuli/task
  - e.g., “1 back” task: hit button after every stimulus repetition
  - Tradeoff: Attention itself can distort the activation
  - Caveat: need to be sure attention is comparable across conditions

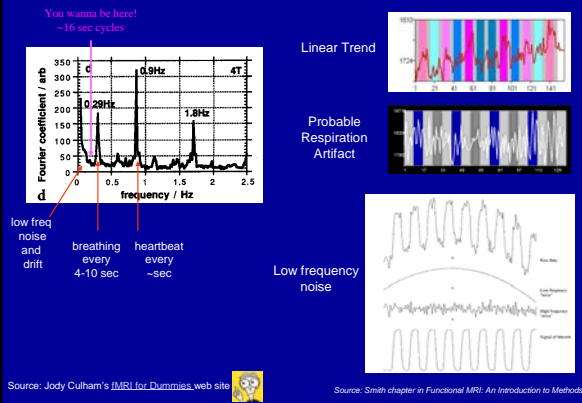
Source: Jody Culham's [fMRI for Dummies web site](http://fMRIforDummies.com)

## How to Reduce Noise

- Head motion artifacts are the worst
  - Test experienced subjects
  - Give your subjects clear instructions (e.g., try not to swallow)
  - Consider motion correction algorithms (Jody's opinion: garbage in, garbage out, when in doubt throw it out; exception: rare subjects such as patients)
- Spatially smooth the data
  - Tradeoff: lose resolution
- Choose your signal power in a frequency with low noise
  - high noise at low frequencies (so don't use long epochs: optimum epoch duration is 16 sec)
  - try to have multiple repetitions of conditions within each run
  - high noise at breathing frequency (once every 4-10 seconds)
- Temporal smoothing
  - At a minimum, perform linear trend removal
  - Consider filtering out low frequencies (high-pass filter)
  - Jody advises against filtering out high frequencies (temporal Gaussian) because it artificially inflates your stats

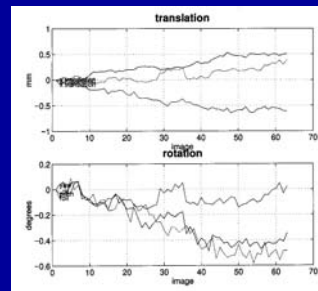
Source: Jody Culham's [fMRI for Dummies web site](http://fMRIforDummies.com)

## How to Reduce Noise



## How to Reduce Noise

- If you can't get rid of an artifact, you can include it as a “predictor of no interest” to soak up variance



Example: Some people include predictors from the outcome of motion correction algorithms

Corollary: Never leave out predictors for conditions that will affect your data

Source: Jody Culham's [fMRI for Dummies web site](http://fMRIforDummies.com)

## Correction for Multiple Comparisons

- If we use a nominal p value of .05 but do ~50,000 times, there's a high probability we'll get some bogus activation. How can we deal with this?

## Correction for Multiple Comparisons

With conventional probability levels (e.g.,  $p < .05$ ) and a huge number of comparisons (e.g.,  $64 \times 64 \times 12 = 49,152$ ), a lot of voxels will be significant purely by chance  
e.g.,  $.05 * 49,152 = 2458$  voxels significant due to chance

How can we avoid this?

### 1) Bonferroni correction

- divide desired p value by number of comparisons  
Example:  
desired p value:  $p < .05$   
number of voxels: 50,000  
required p value:  $p < .05 / 50,000 \rightarrow p < .000001$

- quite conservative
- can use less stringent values
  - \* e.g., Brain Voyager can use the number of voxels in the cortical surface
- some people use different correction values for expected regions and unexpected regions

By this logic, Columbus didn't discover America because he was looking for India!

Source: Jody Culham's [fMRI for Dummies web site](http://fMRIforDummies.com)

## Correction for Multiple Comparisons 2

### 2) Cluster correction

- falsely activated voxels should be randomly dispersed
- set minimum cluster size to be large enough to make it unlikely that a cluster of that size would occur by chance
- assumes that data from adjacent voxels are uncorrelated (not true)

### 3) Test-retest reliability

- Perform statistical tests on each half of the data
- The probability of a given voxel appearing in both purely by chance is the square of the p value used in each half
  - e.g.,  $.001 \times .001 = .000001$
- Alternatively, use the first half to select an ROI and evaluate your hypothesis in the second half.

### 4) Consistency between subjects

- Though seldom discussed, consistency between subjects provides further encouragement that activation is not spurious.

### 5) Screen saver scan

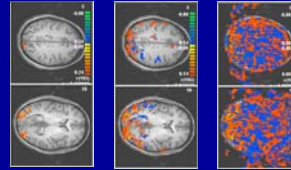
- You can test how many voxels really do light up when really nothing is happening

Source: Jody Culham's fMRI for Dummies web site 

## Correction for Multiple Comparisons 3

### 6) Poor man's Bonferroni

- Jack up the threshold till you get rid of the schmutz (especially in air, ventricles, white matter)
- If you have a comparison where one condition is expected to produce much more activity than the other, turn on both tails of the comparison
- Jody's rule of thumb: "If ya can't trust the negatives, can ya trust the positives?"



Example: MT localizer data

Moving rings > stationary rings (orange)  
Stationary rings > moving rings (blue)

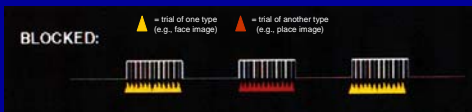
Source: Jody Culham's fMRI for Dummies web site 

## Random Effects Analysis

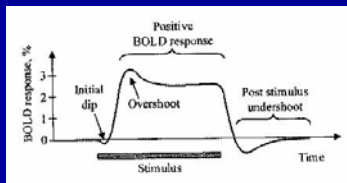
- Typical fMRI stats test whether the differences between conditions are significant in the sample of subjects we have tested
- Often, we want to be able to generalize to the population as a whole including all potential subjects, not just the ones we tested
- **Random effects** analyses allow you to generalize to the population
- Rainer recommends you don't even toy with random effects unless you've got 10 or more subjects (and 50+ is best)
- You don't have to worry about it if you're using the ROI approach because (1) presumably the ROI has already been well-established across multiple labs; and (2) posthoc analyses of results in an ROI approach allow you to generalize to the population (assuming you include individual variance)

## fMRI Basic Experimental Design – event-related fMRI.

### Block Designs



Assumption: Because the hemodynamic response delays and blurs the response to activation, the temporal resolution of fMRI is limited.



This is not entirely true.

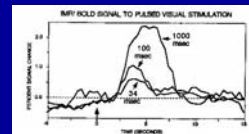
### Assumption of steady-state dynamics.

For block designs we assume that the BOLD effect remains constant across the epoch of interest.

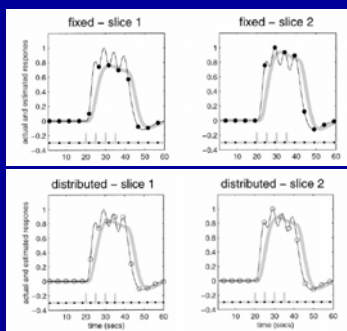
For PET this assumption is valid given the half-life of the tracers used to image the brain.

But the BOLD response is much more transient and more importantly may vary according to brain regions and stimulus durations and maybe even stimulus types.

Price et al. (1999) Neuroimage, 10, 36 – 44.



### Assumption of steady-state dynamics.



Price et al. 1999, Neuroimage, 10, 36 – 44.

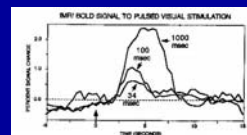
### What are the temporal limits?

What is the briefest stimulus that fMRI can detect?

Blamire et al. (1992) – 2 sec

Bandettini (1993): 0.5 sec

Savoy et al (1995): 34 msec



With enough averaging, anything seems possible.

Assume that the shape of the HRF is predictable.

Event-related potentials (ERPs) are based on averaging small responses over many trials.

Can we do the same thing with fMRI?

## SNR in block vs. ER-fMRI – trade-offs again!

Block Design - widely spaced ER-fMRI ~ 33% loss of SNR (Bandettini and Cox, 2000)

Widely spaced ER-fMRI - rapid ER-fMRI ~ 17% loss of SNR (Miezin et al. 2000)

So from Block Design to rapid ER-fMRI ~ 50% loss of SNR!!

Claim is that the power lost in SNR is made up for by increased numbers of trials for event-related averaging. This may differ across regions and for different tasks – yet to be determined.

## Why do an event-related design?

Pros:

- multiple trial types in one run – randomization becomes possible

- avoid confounds of motor artifacts – haemodynamic lag

- greater temporal control

- can look for activation to a single specific trial types (usually the average of many trials)

Cons:

- smaller SNR means smaller n – ramp up number of trials (~ 50 – 100 per condition is considered reasonable)

- more complex design and analysis (esp. timing and baseline issues)

## Thought Experiments for event-related fMRI.

- What do you hope to find? (run through all the same Q's you did for a block design E)

- Can the same question be adequately addressed with a block design? (i.e., what is event-related design *adding* to your experiment?)

- What special confounds are there? (e.g., stimulus and baseline timing)

Caveats:

- Ensure appropriate conditions *within a run*. All the same issues of comparing activations across runs in a block design Exp apply here.

## Possible applications for event-related fMRI.

- Visual priming and object recognition – look for activation to only the primed object or look at activation *before* and *after* object recognition (i.e., very long events). (e.g., James, T. et al. (2000) *Current Biology*, 10, 1017-1024 and James, T. et al. (1999) *Neuroreport*, 10, 1019-1023).

- Exploring specific task components – e.g., preparatory set for pro vs. anti-saccades. (e.g., Connolly, J., et al. (in press) *Nature Neuroscience*)

- Exploring changes over time – e.g., effects of prism adaptation (Dancert, J. *in preparation*)

- Memory research – e.g., ideal for exploring remembering and forgetting – something that is impossible to do in blocked designs.

*and many, many more...*

## Blocked vs. Event-related

BLOCKED:



SPACED MIXED TRIAL:



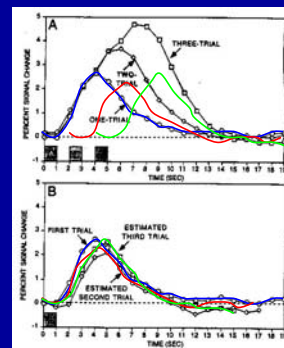
RAPID MIXED TRIAL:



Source: Buckner 1998

## Linearity of BOLD response

Dale & Buckner, 1997



Linearity:

"Do things really add up?"

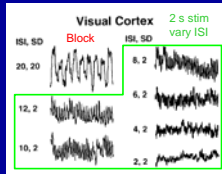
red = 2 - 1

green = 3 - 2

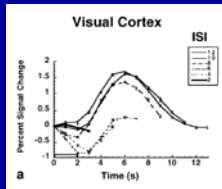
Sync each trial response to start of trial

Not quite linear but good enough (the noise in each trial is also non-linear – but this non-linearity is not large enough to cause huge problems).

## Spaced Mixed Trial: Constant ITI



Bandettini et al. (2000)  
 What is the **optimal trial spacing** (duration + inter-trial interval, ITI) for a **Spaced Mixed Trial** design with constant stimulus duration?

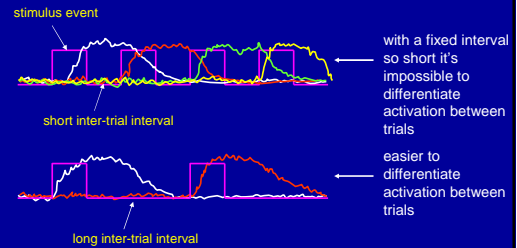


Sync with trial onset and average

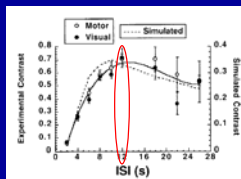
Source: Bandettini et al., 2000

## Spaced Mixed Trials Design Inter-trial intervals (ITIs).

• Stimulus duration and inter-trial-interval. The main idea is to let the HRF return to baseline before presenting your next trial.



## Optimal Constant ITI



Brief (< 2 sec) stimuli:  
 optimal trial spacing = 12 sec

For longer stimuli:  
 optimal trial spacing = 8 + 2\*stimulus duration

Effective loss in power of event related design:  
 = -35%  
 i.e., for 6 minutes of block design, run ~9 min ER design

Source: Bandettini et al., 2000

## Considerations and caveats.

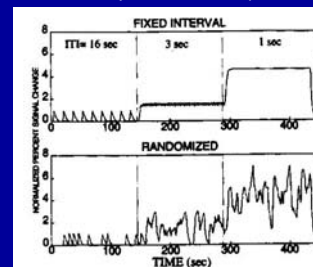
- Power – always a consideration! Whereas for block design you considered the duration and number of blocks for power issues, now you have to consider the **number of trials** per condition. (so overall duration of your experiment will increase)
- The timing of single events will always mean you have a lower SNR in event-related fMRI (for block design % signal change is in the range of 3 – 5 while for event-related fMRI you are often looking at changes of less than 1%)
- Block design is the sledgehammer (sometimes unavoidable and even ideal) while event-related designs have a little more finesse – but the trade off is in time (more trials needed often means longer runs) and power (lower SNR requires the greater number of trials)

## Rapid event-related fMRI.

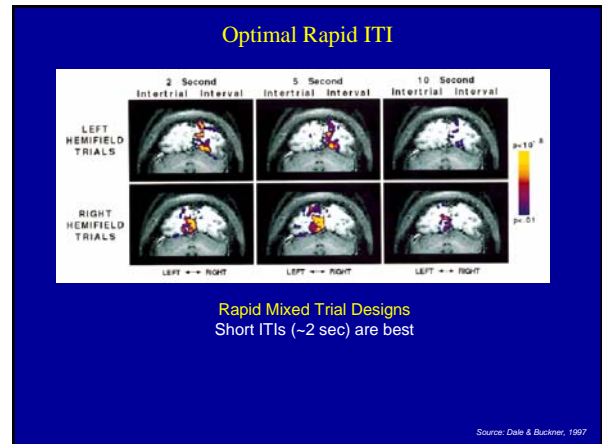
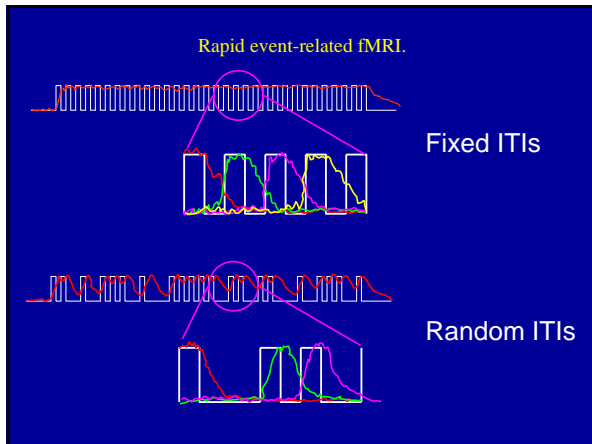
- In simple (I) event-related fMRI you allow the HRF to return to baseline after every trial.
- For rapid event-related fMRI, trials (or events in this case) are truly randomised as you would in a behavioural study and the HRF is *deconvolved* afterwards
- Power is an even bigger issue here – the differences in % signal change being smaller than in spaced event-related fMRI requiring some fancy stats.
- Two crucial components in your design:
  - make sure every possible *combination* of trial sequences is used (i.e., every trial type is preceded and followed by every other trial type an equal number of times)
  - jitter the ITI's – randomised ITI's are crucial for later deconvolution of the HRF (see fixed spaced example 3 slides back)

## Fixed vs. Random Intervals

If trials are jittered, ↓ ITI → ↑ power



Source: Burock et al., 1998

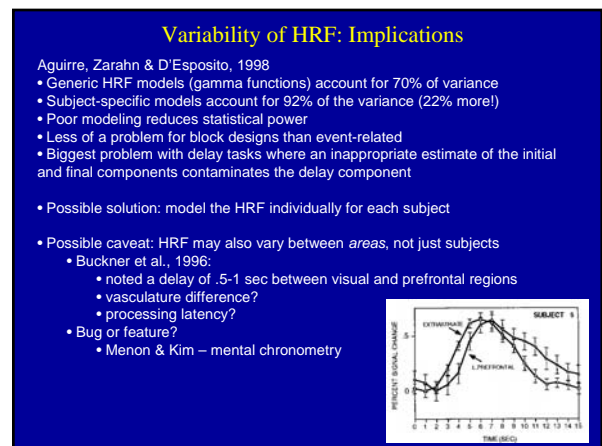
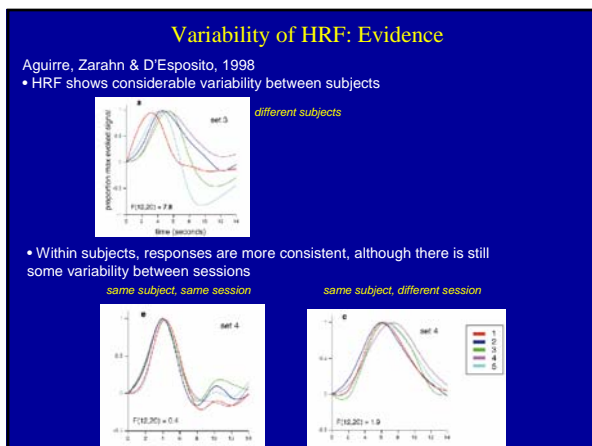
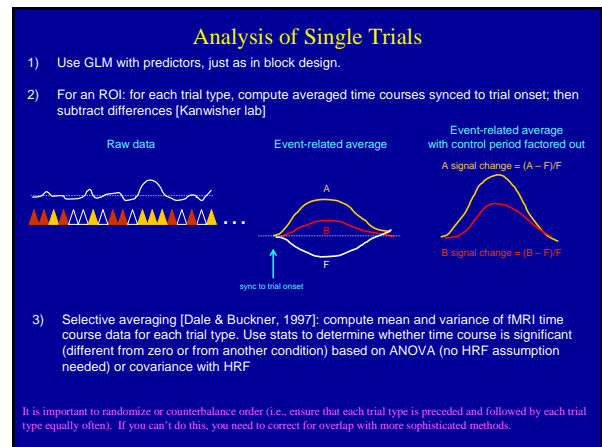


### ITI vs. Stimulus Duration

Bandettini and Cox 2000 make a brief mention of a 50/50 ratio of ITI to SD in rapid ER-fMRI

So the *mean* ITI should be roughly equal to the SD which is usually fixed.

Ollinger et al. 2001 discuss partial vs compound trials (partial trials are a sub-component of the compound trials) and suggest the best SNR is achieved with 40% of all trials being partial – not optimal from behavioural standpoint so 25% partial trials may be most practical.



### Advantages of Event-Related

- 1) Flexibility and randomization
  - eliminate predictability of block designs
  - avoid practice effects
- 2) Post hoc sorting
  - (e.g., correct vs. incorrect, aware vs. unaware, remembered vs. forgotten items, fast vs. slow RTs)
- 3) Can look at novelty and priming
- 4) Rare or unpredictable events can be measured
  - e.g., P300
- 5) Can look at temporal dynamics of response
  - Dissociation of motion artifacts from activation
  - Dissociate components of delay tasks
  - Mental chronometry

Source: Buckner & Braver, 1999