Advances in FDR for fMRI

Ruth Heller

Department of Statistics, University of Pennsylvania

Joint work with: Yoav Benjamini,
Nava Rubin, Damian Stanley, Daniel Yekutieli,
Yulia Golland, Rafael Malach

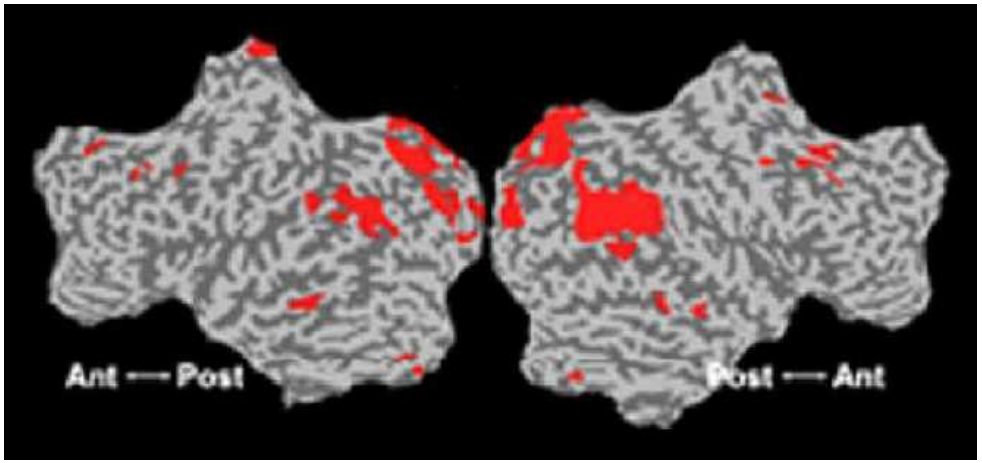
Overview

- Finding brain regions that respond to one cognitive task:
  1. Test clusters instead of voxels.
  2. Control the FDR on clusters.
- Finding brain regions that respond to several cognitive tasks:
  1. Test partial conjunction hypotheses.
  2. Control the overall FDR (OFDR) on a superimposed display.
- Group analysis: an alternative to fixed/random effects analysis.

Finding brain regions that respond to one task

Which regions in the brain respond to a cognitive task?

Figure 1: Results of a statistical analysis
Red brain regions of activity in the unfolded left and right hemispheres



Common approach: voxel-level analysis

- Which of my 100,000 voxels are active?
- If every voxel test is controlled for type I error at 0.05 level and there is no brain activity then we expect to declare 5000 voxels as active!
- How to control for false positives?
  - Control the probability of making at least one false positive (FWER, controlled by random field methods [Worsley et al., 2002]).
  - Control the expected proportion of false positives (FDR ,[Benjamini and Hochberg, 1995]).

    The more relevant error rate when a large number of voxels are tested, if we can tolerate few false discoveries as long as the proportion of false discoveries among discoveries is small.

Limitations of voxel-level analysis

- Does not exploit the signal's spatial structure.

  Common "remedy": smooth the data first. Limitations:

  - May reduce signal to noise ratio of highly localized activity.
  - May introduce signal into non-active voxels.

- Controlling false positives at voxel-level may not be relevant:

  - Findings are reported in terms of clusters of nearby voxels.
  - Isolated voxels detected are often not reported as findings.

Cluster-level analysis

Test clusters of contiguous voxels. Advantages:

- Clusters approximate better the units of interest (i.e. regions).
- Increased signal to noise ratio when pooling neighboring voxels together.
- Testing clusters reduces the multiple hypothesis problem.

How do we obtain the clusters?

- Approximate the cluster units from information **outside the data** to be tested, e.g. cluster nearby locations that are maximally correlated in a localizer experiment ([Heller et al., 2006]).
- Contiguous voxels above an a-priori fixed threshold ([Friston et al., 1994], [Nichols and Holmes, 2001]).

  The arbitrarily chosen threshold has a large impact on the analysis;
  The interpretation of discoveries from such an analysis is vague.

The cluster hypothesis test

- Let $C_1, C_2, \ldots, C_m$ be a partition of the brain cortex D into m clusters.
- For all $i = 1, \ldots, m$

$$H_i : \text{All voxels in } C_i \text{ are non-active}$$

- Derivation of test statistic for cluster i that was defined by outside data:

1. Spatially average the measured signal within the cluster.
2. Compute the test statistic of the spatially averaged time series.

FDR on Clusters

Definition in words:

Expected proportion of falsely rejected cluster hypotheses out of all cluster hypotheses rejected.

Formally:

Let I_0 be the subset of indices of clusters containing no signal,

$$\mathrm{FDR_c} = \mathrm{E}[\frac{\sum_{\mathrm{i}\in \mathrm{I}_0} \mathrm{R_i}}{\sum_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{R_i}} \mathrm{I}_{\{\sum_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{R_i} > 0\}}]$$

where $\mathrm{R_i}$ is the indicators of whether H_{0i} is rejected.

Extension: Weighted FDR on Clusters

Let w_i be the importance weight of cluster i,

$$\mathrm{WFDR} = \mathrm{E}[\frac{\sum_{\mathrm{i}\in\mathrm{I}_0} \mathrm{w_i R_i}}{\sum_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{w_i R_i}} \mathrm{I}_{\{\sum_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{R_i}>0\}}]$$

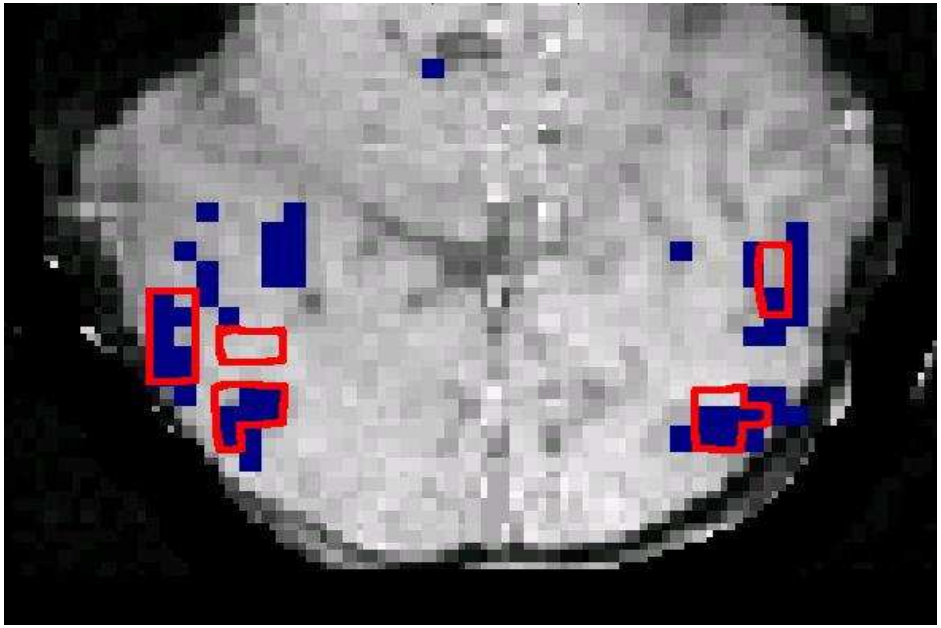
Example: If w_i corresponds on the cluster size,

WFDR is the expected proportion of area (volume) of false clusters out of all the area (volume) rejected.

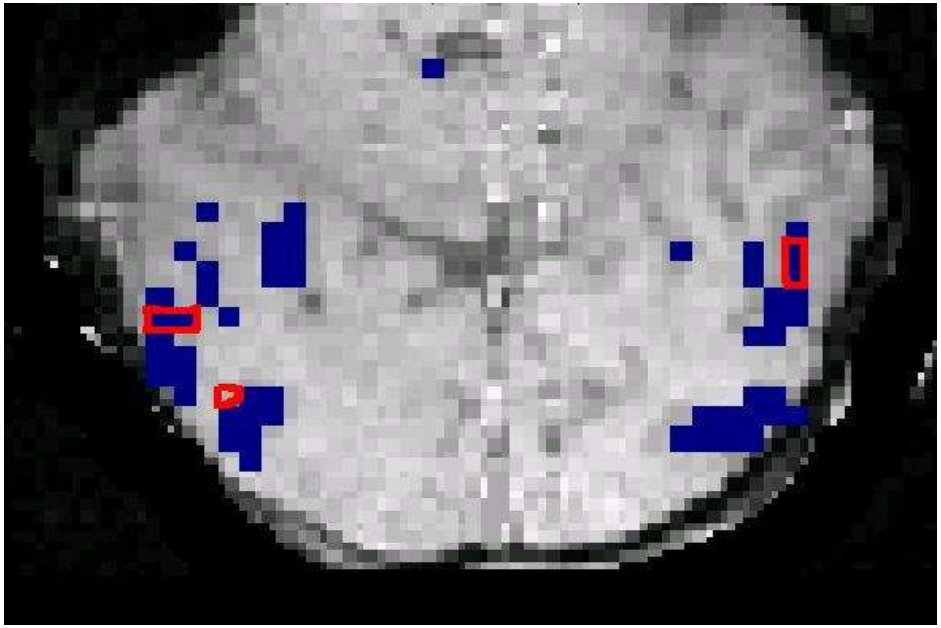
An illustration of the different FDR error measures

A (half) brain slice: blue is true signal, red is outline of cluster discoveries.

$\mathrm{FDRvoxel-level} = 13/43 = 30\%$
$\mathrm{FDRcluster-level} = 1/5 = 20\%$
$\mathrm{WFDRcluster-level} = 6/43 = 14\%$

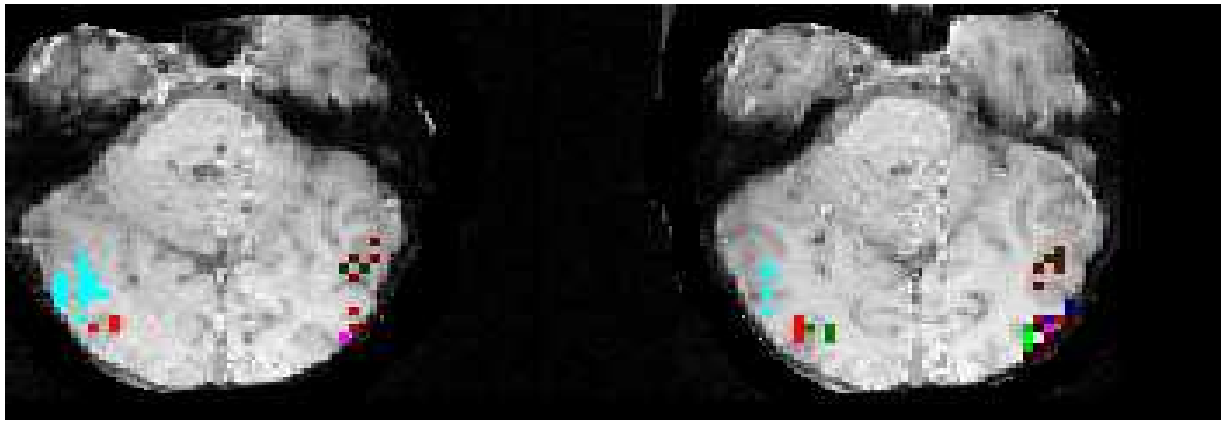


$\mathrm{FDRvoxel-level} = 1/5 = 20\%$
$\mathrm{FDRcluster-level} = 1/3 = 33\%$
$\mathrm{WFDRcluster-level} = 1/5 = 20\%$



fMRI Example: Control of FDR on Clusters

Sample slices 9-10: each detected cluster in a different color.



- Voxel-level testing with $q = 0.1$ discovered 55 voxels in 14 contiguous clusters.
- Cluster-level testing with $q = 0.1$ discovered 163 voxels in 13 clusters.

Control of FDR on clusters and on voxels

The Cluster Testing and Trimming (CTT) procedure
[Benjamini and Heller, 2007]:

- Cluster Testing stage: Apply a WFDR controlling procedure at level q.

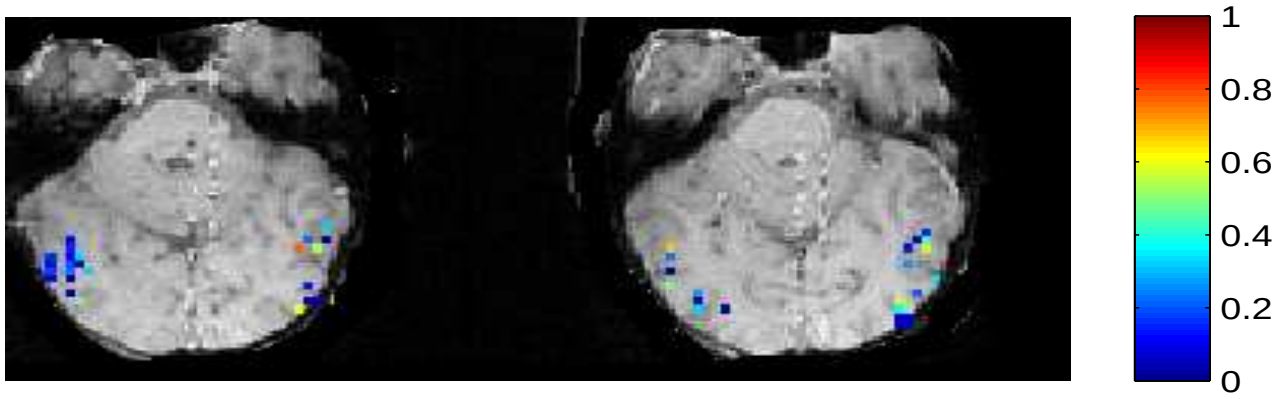
- Voxel Trimming stage: Within rejected clusters only

  1. Apply an FDR controlling procedure on the conditional p-values at level q_2.

  2. Keep all voxel discoveries in the clusters, trimming the others.

fMRI Example: Cluster Testing and Trimming

Sample slices 9-10: estimated voxel p-values within detected clusters.



- CTT with $q = 0.1$ and $q_2 = 0.25$ discovered 91 voxels in 13 clusters (FDR cutoff level 0.18).
- CTT with $q = 0.1$ and $q_2 = 0.1$ discovered 36 voxels in 13 clusters (FDR cutoff level 0.03).

Summary: How to best find activated brain regions

- We care about clusters, not voxels.
- Testing clusters can be much more powerful than testing voxels.
- We can weigh each cluster by importance and apply a WFDR controlling procedure.
- We conclude that one or more voxels within a rejected cluster contains signal.
- With the CTT procedure, we can conclude about voxels as well. We can choose a level of q_2 that is larger than the traditional level since the precision necessary is that of a general region, not exact spatial voxels.
- With the CTT procedure, we can focus attention to discovered clusters where a large proportion of voxels are discovered.

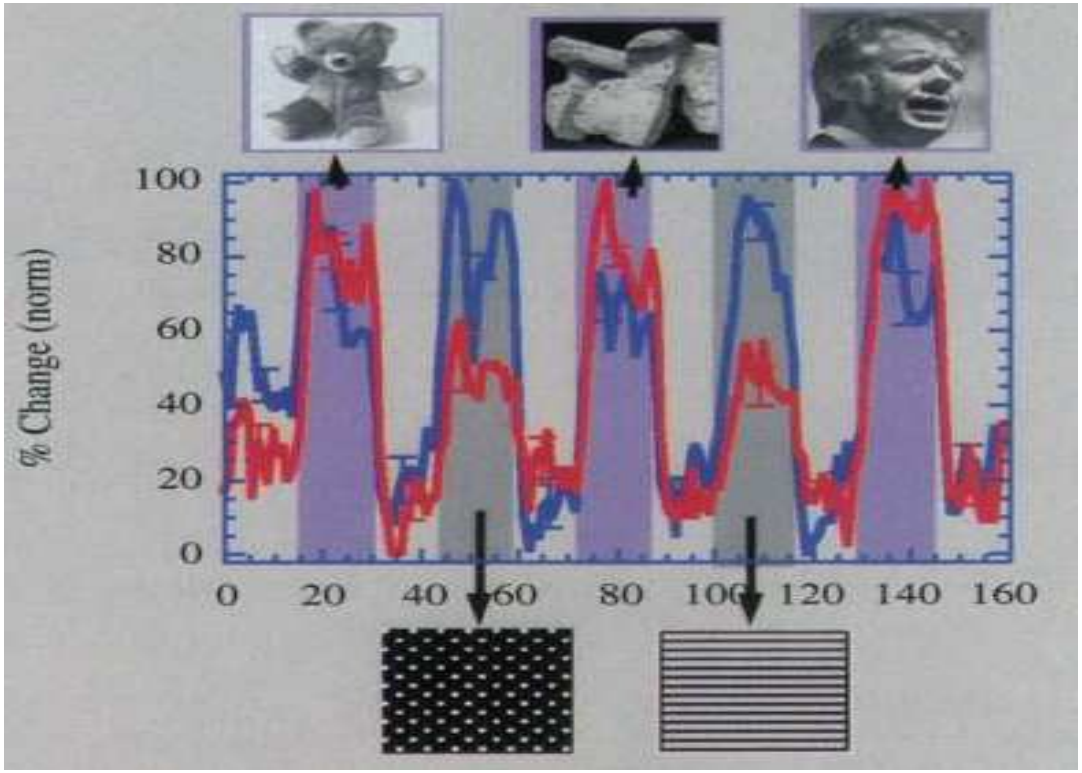
Overview

- Finding brain regions that respond to one cognitive task:
  1. Test clusters instead of voxels.
  2. Control the FDR on clusters.
- Finding brain regions that respond to several cognitive tasks:
  1. Test partial conjunction hypotheses.
  2. Control the overall FDR (OFDR) on a superimposed display.
- Group analysis: an alternative to fixed/random effects analysis.

An fMRI Vision Experiment

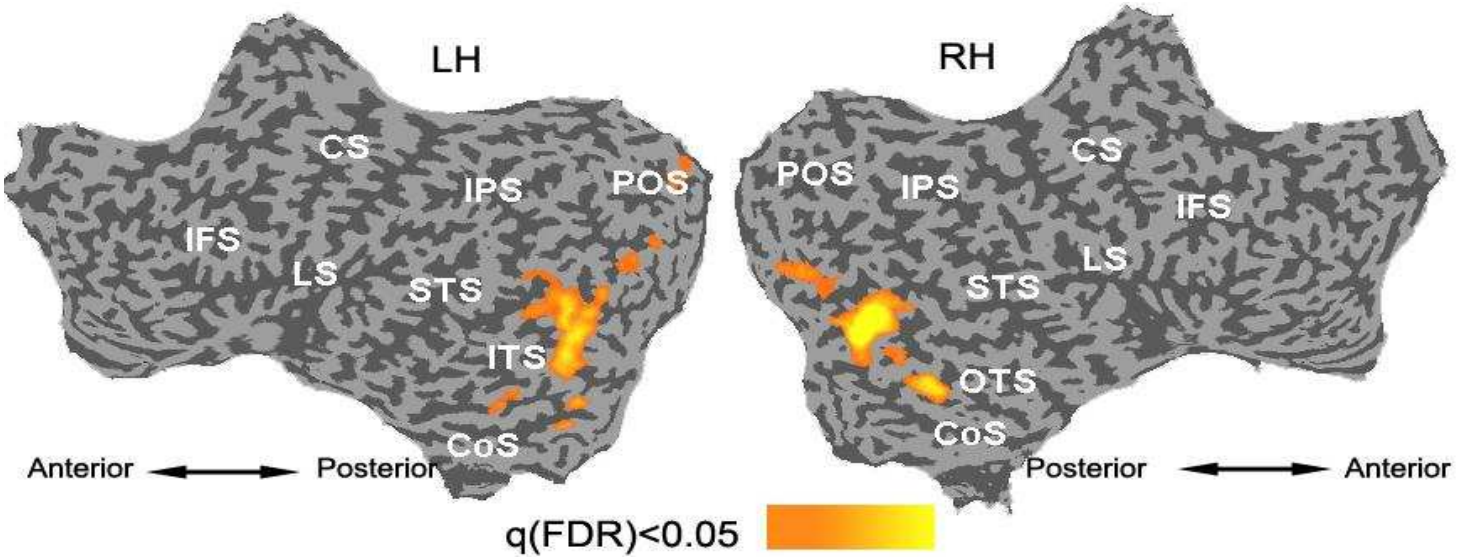
The brain activity is measured in tens of thousands of brain locations while the subject views 4 visual stimuli:
(i) faces (ii) houses (iii) common man-made objects and (iv) geometric patterns.

How to find the regions that were more active during (i)-(iii) than (iv)?



Which Regions are More Active During (i)-(iii) than (iv)?

Typical Analysis: Test in each voxel whether the average activation during (i)-(iii) is higher than during (iv), adjusting for the severe multiple comparisons problem using the BH procedure [Benjamini and Hochberg, 1995] at the 0.05 level.

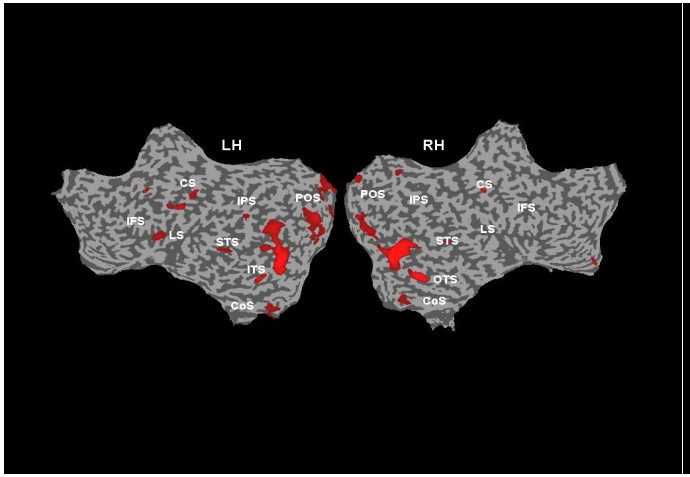


Are the discovered regions due to all or only some of the contrasts (i)-(iv), (ii)-(iv), (iii)-(iv)?

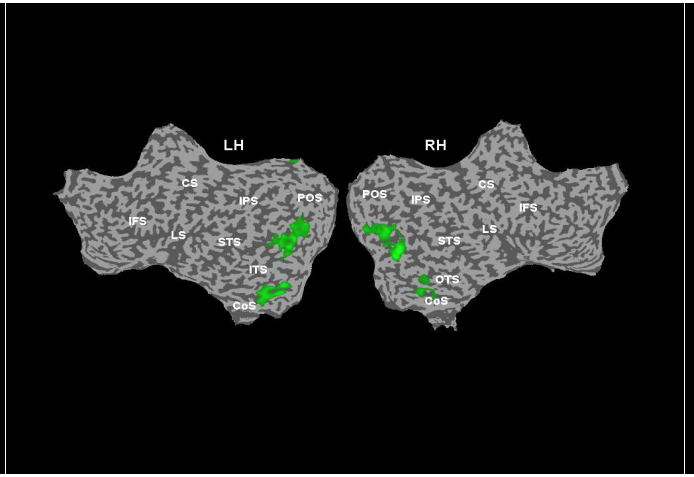
Separate Analysis for Each Contrast

For each contrast, used the BH procedure to control the FDR at the 0.05 level.

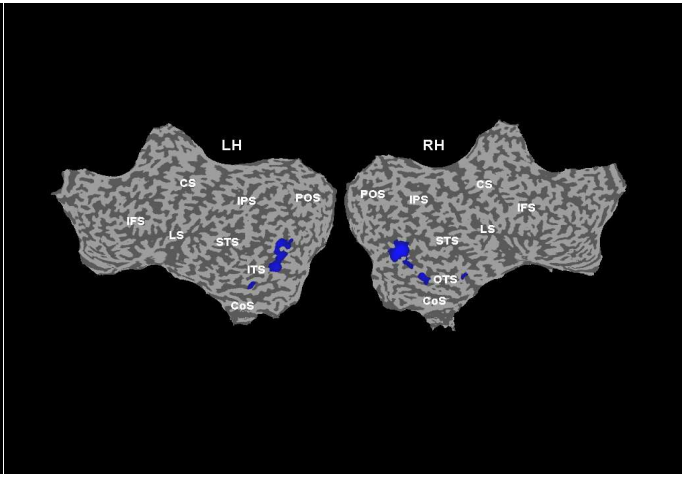
a) Faces vs. Patterns



b) Houses vs. Patterns



c) Objects vs. Patterns



The FDR is not necessarily controlled at the 0.05 level on combinations of maps:

- The intersection map may contain 0 locations that are more active during all of (i)-(iii) than during (iv);
- The union map may be expected to contain $0.05 \times 3 = 0.15$ of voxels where there is no increased activity during any of (i)-(iii) over (iv).

How to control the FDR on combinations of maps?

- Combine the test statistics in each voxel instead of the discoveries in each voxel.
- In each voxel, the combination of test statistics produces a p-value.
- This p-value is different for testing that:
  1. at least one contrast is positive;
  2. all contrasts are positive;
  3. most contrasts are positive.
- Apply an FDR controlling procedure on these p-values.

The Appropriate Test per Voxel

For n null hypotheses $H_1(s), \ldots, H_n(s)$ at voxel s

(In example, n=3 contrasts
$H_1(s) : \mu(i) - \mu(iv) = 0, H_2(s) : \mu(ii) - \mu(iv) = 0, H_3(s) : \mu(iii) - \mu(iv) = 0)$

1. The global null hypothesis $H^{1/n}(s)$ that all of $H_1(s), \ldots, H_n(s)$ are true.

   (In example, the null that all contrasts are zero.)
   Often too general to be scientifically meaningful.

2. The null $H^{n/n}(s)$ that at most $n-1$ of $H_1(s), \ldots, H_n(s)$ are false.

   (In example, the null that at most two contrasts are positive.)
   Difficult to reject in practice when screening a large # of such nulls.

3. The partial conjunction hypothesis $H^{u/n}(s)$ that at most $u-1$ of $H_1(s), \ldots, H_n(s)$ are false, $u = 2, \ldots, n-1$ ([Friston et al., 1999]).

   (In example with u=2, the null that at most one contrast is positive.)
   More informative than 1 and more powerful than 2.

For $H^{1/n}(s), H^{n/n}(s)$: Known *p*-Values

Let $p_{(1)}(s) \leq \ldots \leq p_{(n)}(s)$ be the sorted p-values for the n hypotheses.

- For testing $H^{n/n}(s)$: the maximum p-value $p^{n/n}(s) = p_{(n)}(s) \leq \alpha$.
- For testing $H^{1/n}$:
  1. For independent p-values, Fisher's combining method (among others):

$$p^{1/n}(s) = P(\chi^2_{2n} \geq -2 \sum_{j=1}^{n} \log p_j(s)) \leq \alpha$$

  2. For positive dependent p-values (e.g. stimuli (i)-(iii) compared to the same control stimulus (iv)), Simes test:

Reject $H^{1/n}(s)$ if $\exists \quad j \quad s.t. \quad p_{(j)}(s) \leq \frac{j}{n}\alpha \iff$

$$p^{1/n}(s) = \min_{j=1,\ldots,n} \{\frac{n}{j} p_{(j)}(s)\} \leq \alpha$$

  3. For dependent p-values, Bonferroni's test:

Reject $H^{1/n}$ if $\exists\, j \;\; p_{(j)}(s) \leq \frac{1}{n}\alpha \iff$

$$p^{1/n}(s) = n p_{(1)}(s) \leq \alpha$$

For $H^{u/n}(s), 1 < u < n$: New *p*-Values

For Independent p-values, the partial conjunction p-value motivated by the Fisher method is

$$p^{u/n}(s) = P(\chi^2_{2(n-u+1)} \geq -2\sum_{j=u}^{n} \log p_{(j)}(s))$$

If the p-values are 0.5, 0.02, 0.01 then

$$\begin{aligned} p^{1/3}(s) &= P(\chi^2_6 \geq -2(log(0.5) + log(0.02) + log(0.01))) = 0.005 \\ p^{2/3}(s) &= P(\chi^2_4 \geq -2(log(0.5) + log(0.02))) = 0.056 \\ p^{3/3}(s) &= 0.5 \end{aligned}$$

Theorem 1. If the set of null p-values at voxel s are independent , then $p^{u/n}$ is a p-value for testing $H^{u/n}(s)$:

$Pr(P^{u/n} \leq \alpha) \leq \alpha$ if $H^{u/n}(s)$ is true.

For $H^{u/n}(s), 1 < u < n$: More New *p*-Values

- The partial conjunction p-value motivated by the Simes method is

$$p^{u/n}(s) = \min_{j=1,\ldots n-u+1} \{ \frac{(n-u+1)}{j} p_{(u-1+j)}(s) \}$$

Theorem 2. If the set of null p-values at voxel s are independent or satisfy the PRDS property, then $p^{u/n}(s)$ is a p-value for testing $H^{u/n}(s)$.

- The partial conjunction p-value motivated by the Bonferroni method is

$$p^{u/n}(s) = (n-u+1) p_{(u)}(s)$$

Theorem 3. $p^{u/n}(s)$ is a p-value for testing $H^{u/n}(s)$.

Screening for partial conjunction hypotheses

1. For each voxel $s, s = 1, \ldots S$:

   (a) Specify the **screening hypothesis** $H^{u(s)/n(s)}(s)$.

   (b) Compute the partial conjunction p-value $p^{u(s)/n(s)}(s)$.

2. Apply the BH procedure at level q on $\{p^{u(s)/n(s)}(s) : s = 1, \ldots, S\}$.

Theorem 4. The above procedure controls the FDR at level q

1. if p-values across voxels are independent;

2. if p-values across voxels are positive dependent and within voxels are independent;

3. asymptotically, as $S \to \infty$, if the dependency across voxels is local.

Simulations suggest that the above procedure controls the FDR at level q for typical fMRI dependencies between and within voxels.

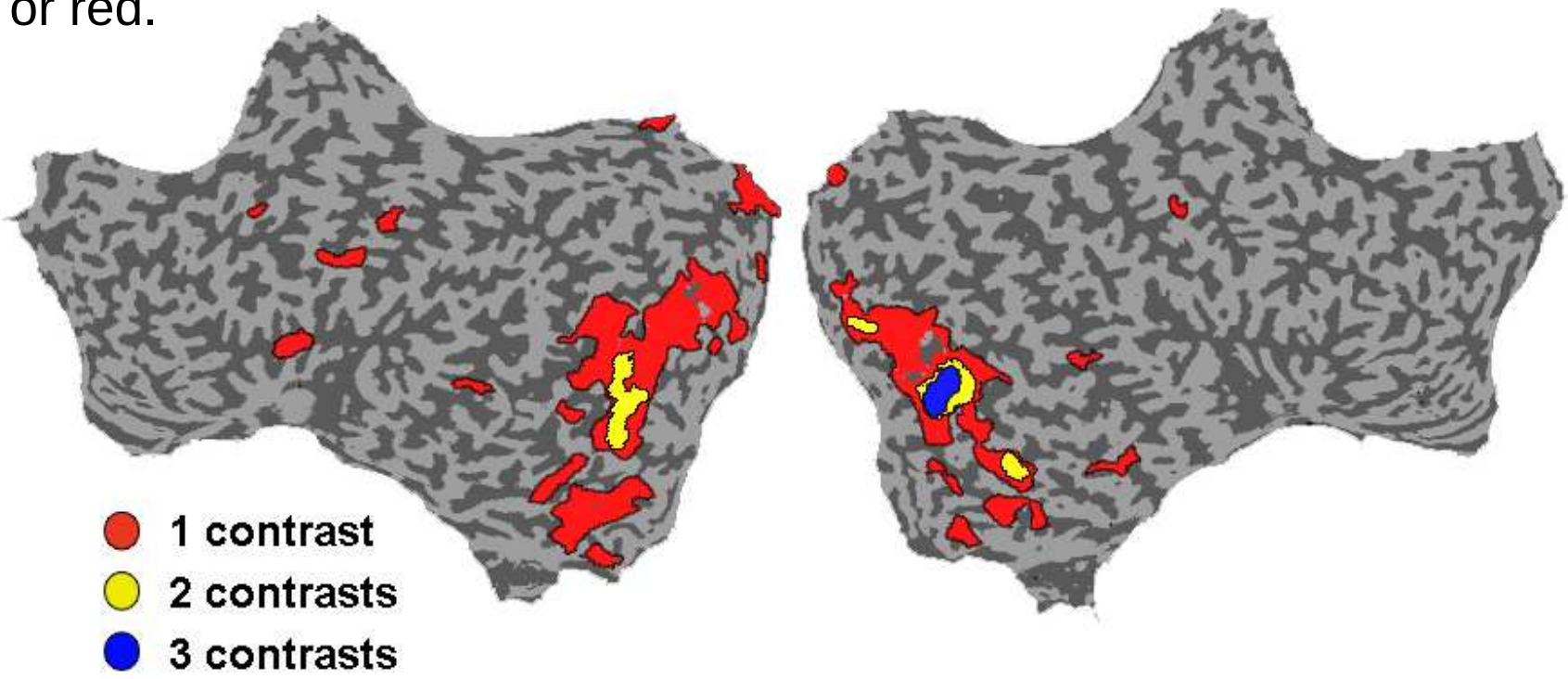
fMRI Example: Partial Conjunction of Contrasts

Experiment: Subject views 4 visual stimuli (i) faces (ii) houses (iii) common man-made objects and (iv) geometric patterns.

Goal: to find the regions that were more active during (i)-(iii) than (iv).

Method: Pooled p-values using Simes, BH procedure at level 0.05.

Results: Regions that were found to react to: all 3 contrasts are colored in blue; at least 2 contrasts are colored in blue or yellow; at least 1 contrast are colored in blue, yellow or red.



The Overall FDR (OFDR)

- A **discovered voxel** is a voxel for which at least one null hypothesis has been rejected.

  (in example, $H^{1/3}$)

- A **falsely discovered voxel** is a voxel for which at least one null hypothesis has been incorrectly rejected.

  (in example, $H^{1/3} \subset H^{2/3} \subset H^{3/3}$)

- The **overall FDR** (OFDR) is the expected proportion of falsely discovered voxels out of all discovered voxels.

  (in example, the expected proportion of discovered voxels where the largest u rejected exceeds the true number of positive contrasts)

OFDR Controlling Procedure for a Superimposed Display

1. For each voxel $s, s = 1, \ldots S$ Compute the global null p-value $p^{1/n(s)}(s)$.

2. Apply the BH at level q on $\{p^{1/n(s)}(s) : s = 1, \ldots, S\}$.

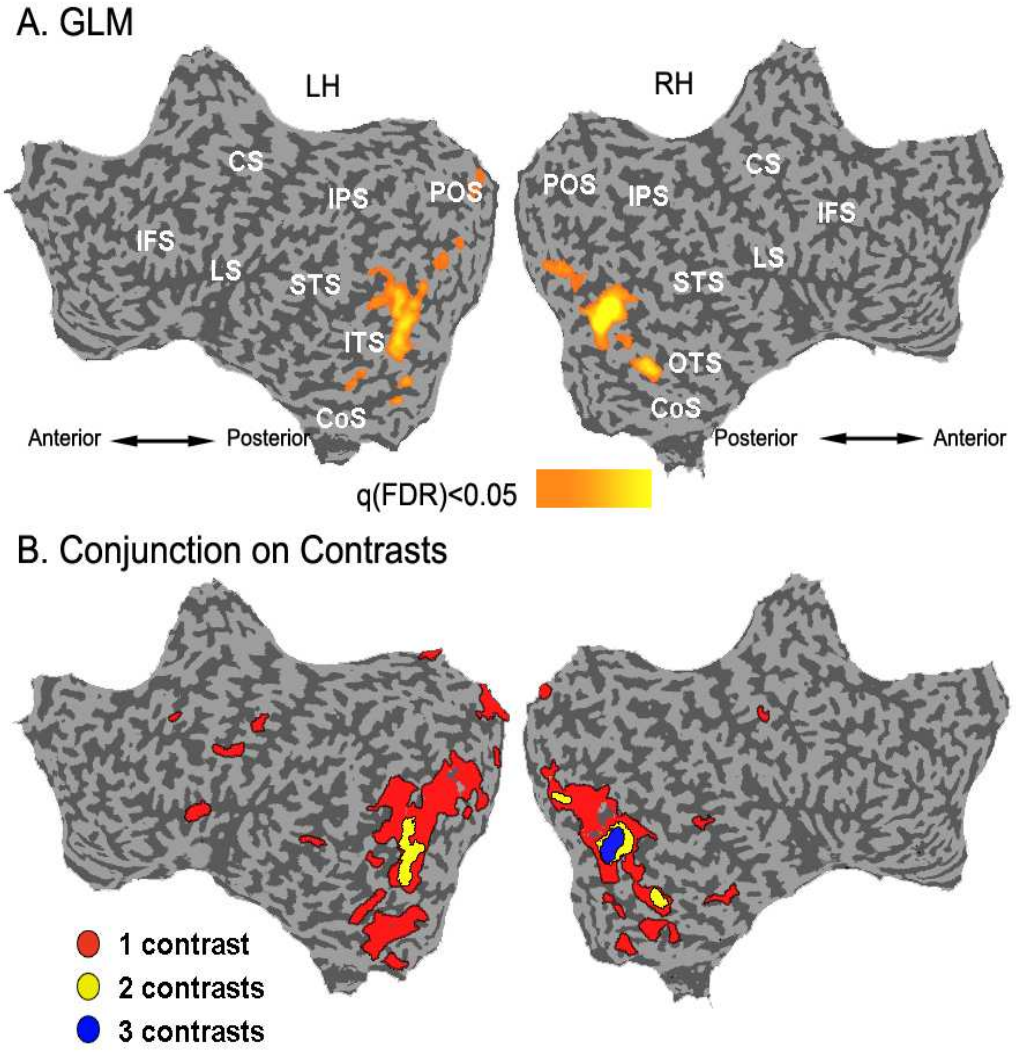
3. For each of the R voxels where the global null has been rejected, test sequentially the partial conjunction hypotheses $u = 2, 3, \ldots$ at level Rq/S :

$$u_{\max}(s) = \max_{u>1}\{p^{u/n(s)}(s) \leq Rq/S\}$$

For q=0.05, expect only 5% of $u_{\max}(s), s = 1, \ldots, S$ to exceed the true number of positive contrasts.

Single Subject

Faces, Houses, Objects versus Patterns



Group Analysis

- Fixed effects analysis is equivalent to screening whether at least $u = 1$ subject activates a location.
  Very weak scientific finding that can be driven entirely by one subject.

- Random effects analysis tests whether the population average task effect is positive under the assumptions:

  - The group is a random sample from the population.

  - The true subject task effects come from a normal distribution.
    Assumption is typically violated.

- Screening whether at least $u = 1, \ldots, n$ subjects activate a location:

  - Generalizes fixed effect analysis;

  - Is valid even when random effect analysis assumptions are violated.

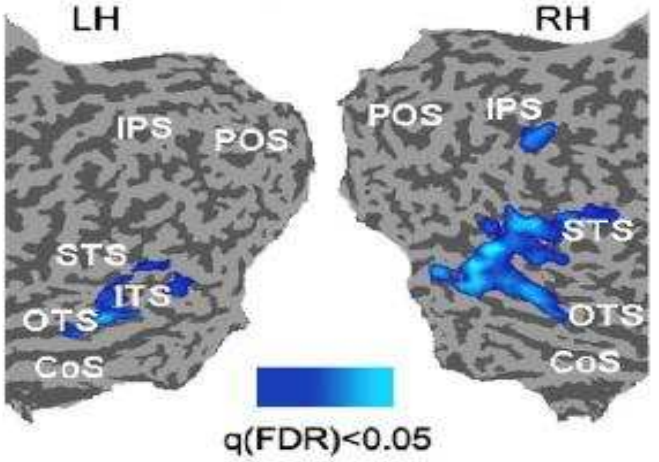
Group Analysis

FACE versus HOUSE

A. Random Effect, unsmoothed

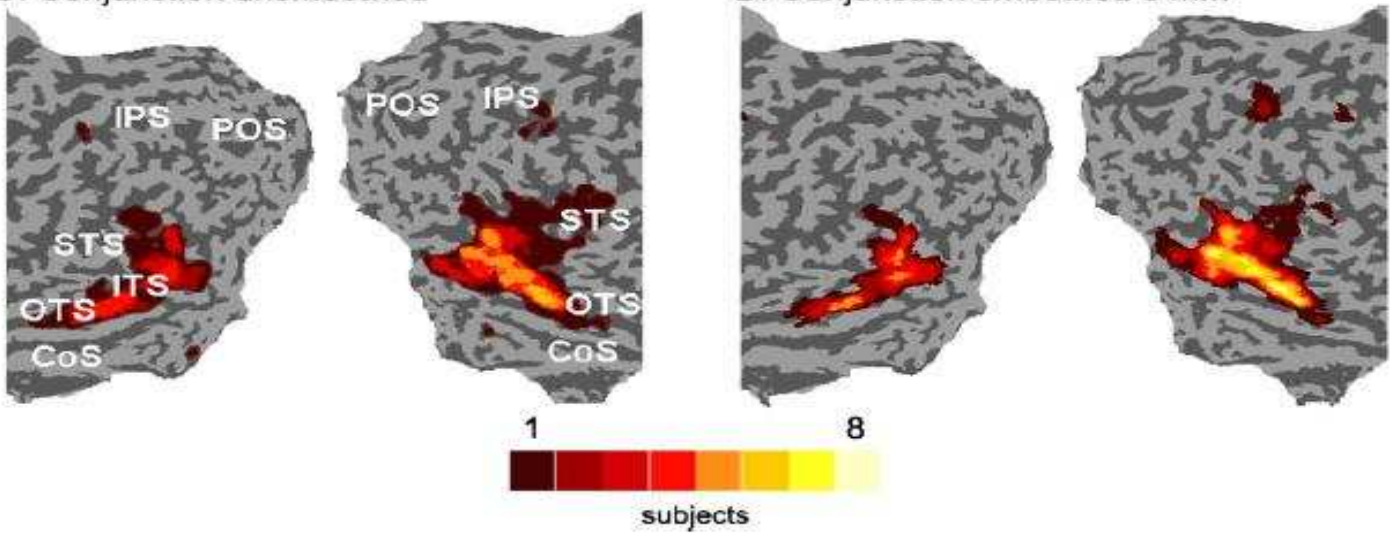
No Significant Voxels

B. Random Effect, smoothed 8mm



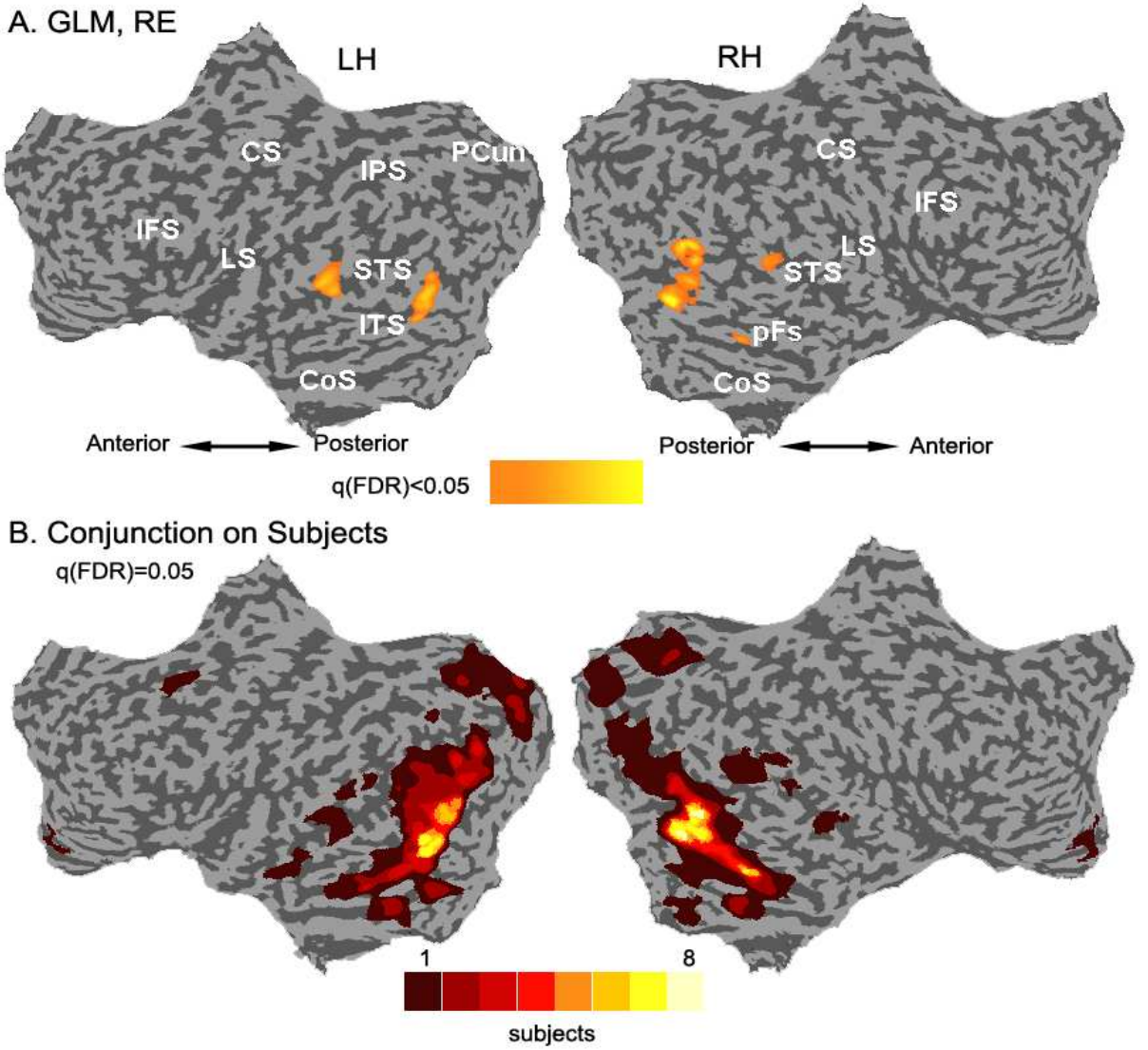
C. Conjunction unsmoothed

D. Conjunction smoothed 8 mm



Group Analysis

Faces, Houses, Objects > Patterns



Summary: Partial Conjunctions and Group Analysis

- The OFDR is an appropriate error measure for a superimposed display.
- Finding regions that react to more than one contrast:
  test the global null $u = 1$ in each voxel, then sequentially test for partial conjunction hypotheses $u = 2, 3, \ldots$.
- Group analysis (replacing contrast by subject):
  a lower bound on the proportion of subjects that activate each voxel.
- Meta-analysis of fMRI studies (replacing contrast by study):
  a lower bound on the proportion of **studies** that discover each voxel.
- Benefit from first clustering contiguous voxels, then combining cluster p-values rather than individual p-values per subject.
  How to find common clusters for group analysis?

[Benjamini and Heller, 2007] Benjamini, Y. and Heller, R. (2007). False discovery rates for spatial signals. *JASA - Theory and Methods*, 102(480):1272–1281.

[Benjamini and Heller, 2008] Benjamini, Y. and Heller, R. (2008). Screening for partial conjunction hypotheses. *Biometrics*, doi: 10.1111/j.1541-0420.2007.00983.x.

[Benjamini and Hochberg, 1995] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B Met.*, 57 (1):289–300.

[Friston et al., 1999] Friston, K., Holmes, A., Price, C., Buchel, C., and Worsley, K. (1999). Multisubject fmri studies and conjunction analyses. *NeuroImage*, 10:385–396.

[Friston et al., 1994] Friston, K., Worsley, K., Frackowiak, R., Mazziotta, J., and Evans, A. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, 1:214–220.

[Heller et al., 2007] Heller, R., Golland, Y., Malach, R., and Benjamini, Y. (2007). Conjunction group analysis: An alternative to mixed/random effect analysis. *NeuroImage*, 37(4):1178–85.

[Heller et al., 2006] Heller, R., Stanley, D., Yekutieli, D., Rubin, N., and Benjamini, Y. (2006). Cluster based analysis of fmri data. *NeuroImage*.