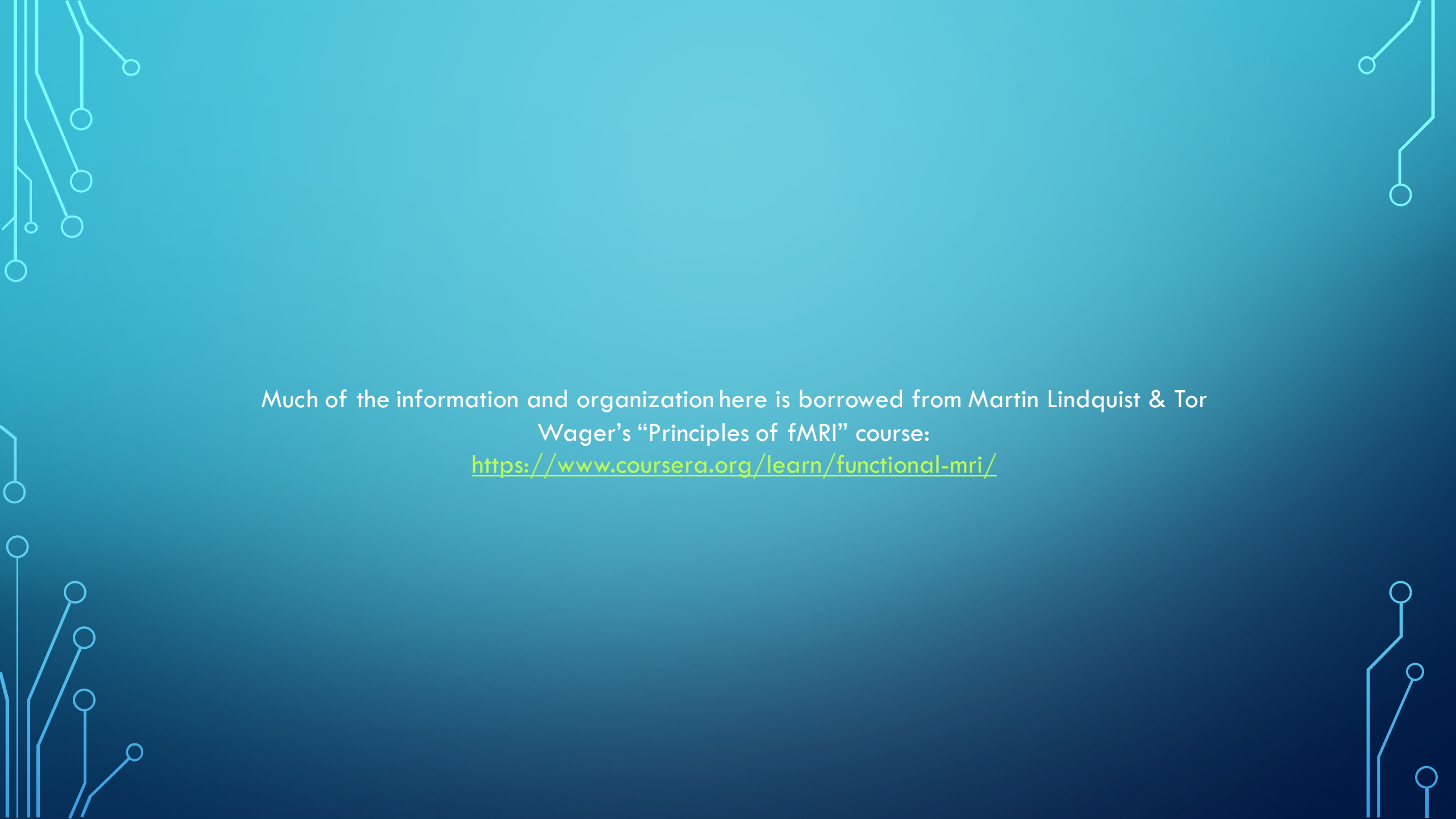




# CORRECTION METHODS IN FMRI

EMILY WASSERMAN

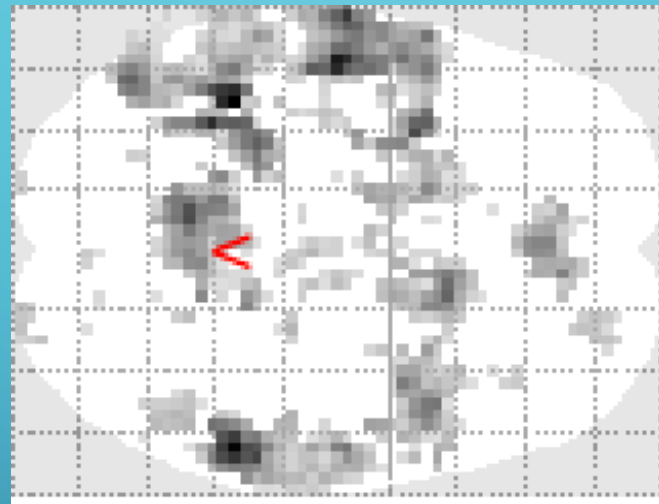
JULY 25, 2016

The background is a dark blue gradient. In the four corners, there are white, stylized circuit board traces. These traces consist of straight lines that turn at 90-degree angles, ending in small white circles. The traces are arranged in a way that suggests a network or data flow, with some lines extending from the edges towards the center.

Much of the information and organization here is borrowed from Martin Lindquist & Tor Wager's "Principles of fMRI" course:

<https://www.coursera.org/learn/functional-mri/>

# WHAT'S THE PROBLEM?

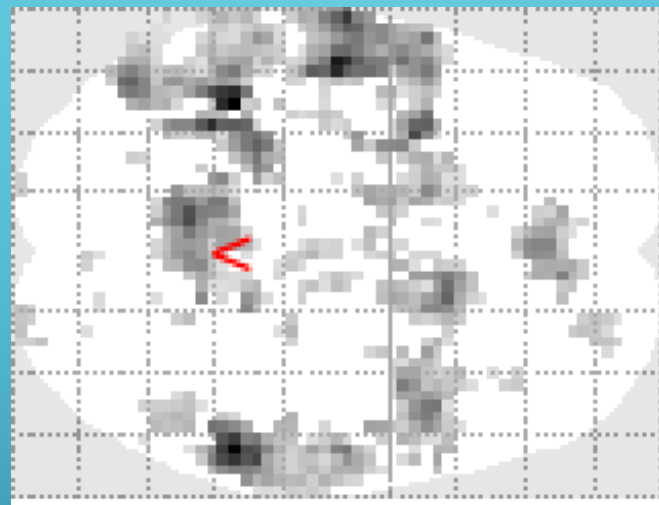


*unthresholded glass brain*

How many of these voxels are REALLY active?

# WHAT'S THE PROBLEM?

100,000 voxels



*unthresholded glass brain*

**X**

5% chance of Type I error



5,000 false positive voxels!

# FAMILIES OF TESTS

Formed by running the same statistical test  $N$  times

Correction methods use this value  $N$  to determine some new **threshold**, called  $u$



# FAMILYWISE ERROR CORRECTION

Let's control the **number of false positives out of ALL our voxels**



# FAMILYWISE ERROR CORRECTION

Let's control the **number of false positives out of ALL our voxels**

Family-wise  $H_0$ : There is **NO** activation in any voxel in the brain

$$H_0 = \bigcap_{i \in V} H_{0i}$$

$V$  = set of all voxels

# FAMILYWISE ERROR CORRECTION

If ANY voxel is truly activated, we can reject  $H_0$

But this means:

if that voxel is a false positive, we've committed a **family-wise Type I error**

$$H_0 = \bigcap_{i \in V} H_{0i}$$

$V$  = set of all voxels



# FAMILYWISE ERROR CORRECTION

If ANY voxel is truly activated, we can reject  $H_0$

But this means:

if that voxel is a false positive, we've committed a **family-wise Type I error**

$$P \left( \bigcup_{i \in V} \{T_i \geq u\} \mid H_0 \right) \leq \alpha$$

# FAMILYWISE ERROR CORRECTION

probability of family-wise  
Type I error

$T$ -statistic at voxel  $i$

threshold

$$P \left( \bigcup_{i \in V} \{T_i \geq u\} \mid H_0 \right) \leq \alpha$$

among all the voxels in the  
brain...

if the null is true...

# FAMILYWISE ERROR CORRECTION

Classic: **Bonferroni correction**

Simply divide your significance level by  $N$  – the number of tests in your family

$$P\left(\bigcup_{i \in V} \{T_i \geq u\} \mid H_0\right) \leq \frac{\alpha}{N}$$

$N = 100,000$  voxels

$\alpha = 0.05$

new significance level at each voxel:  $p < 0.0000005$

Kiss your results goodbye...

# THE TRADE-OFF: SENSITIVITY VS SPECIFICITY

*sensitive*

if there's an effect here,  
we'll definitely find it

ex.: no correction

*specific*

we can trust that anything  
we see isn't false

ex.: voxel-wise Bonferroni  
correction

# FAMILYWISE ERROR CORRECTION

Fancy: **Random Field Theory**

Account for the fact that fMRI data is highly **spatially correlated**:  
adjacent voxels are probably not independent

Control the probability of the **MAXIMUM T-statistic exceeding threshold**

$$P(\max_i \{T_i \geq u\} \mid H_0) \leq \alpha$$

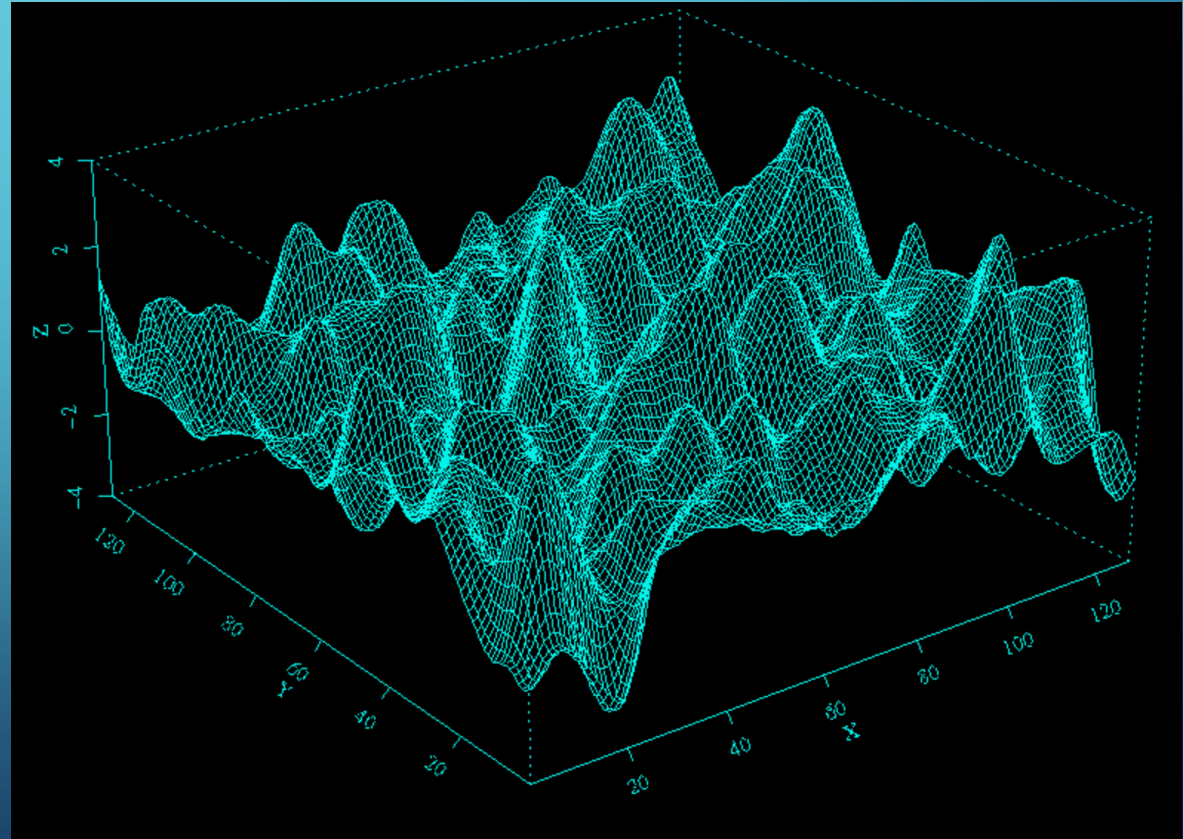
# WHAT'S A RANDOM FIELD?

"Set of random variables defined at every point in a D-dimensional space"

For us:

$D = 3$  (three-dimensional brain)

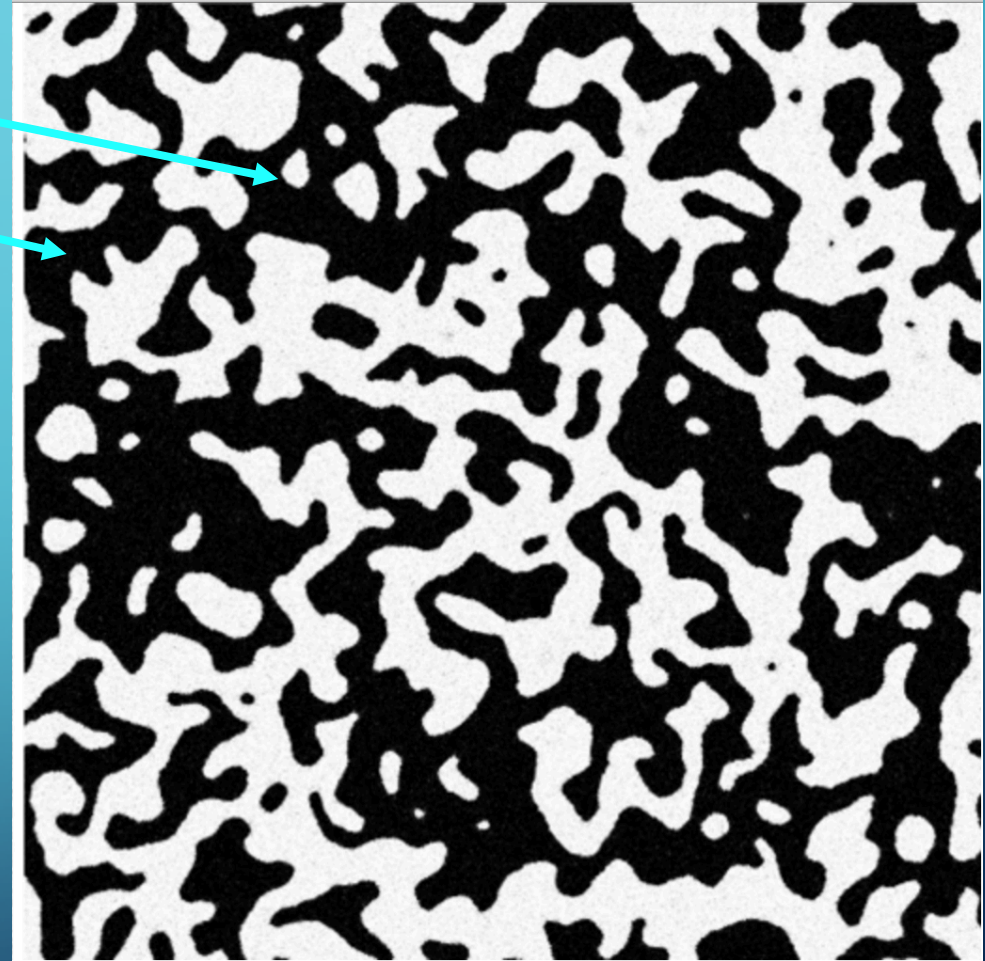
Gaussian, or **normal**, distribution



# BLOBS AND HOLES

An image of a random field has an **Euler characteristic:**

the number of blobs left in the image after thresholding, minus the number of holes



# BLOBS AND HOLES

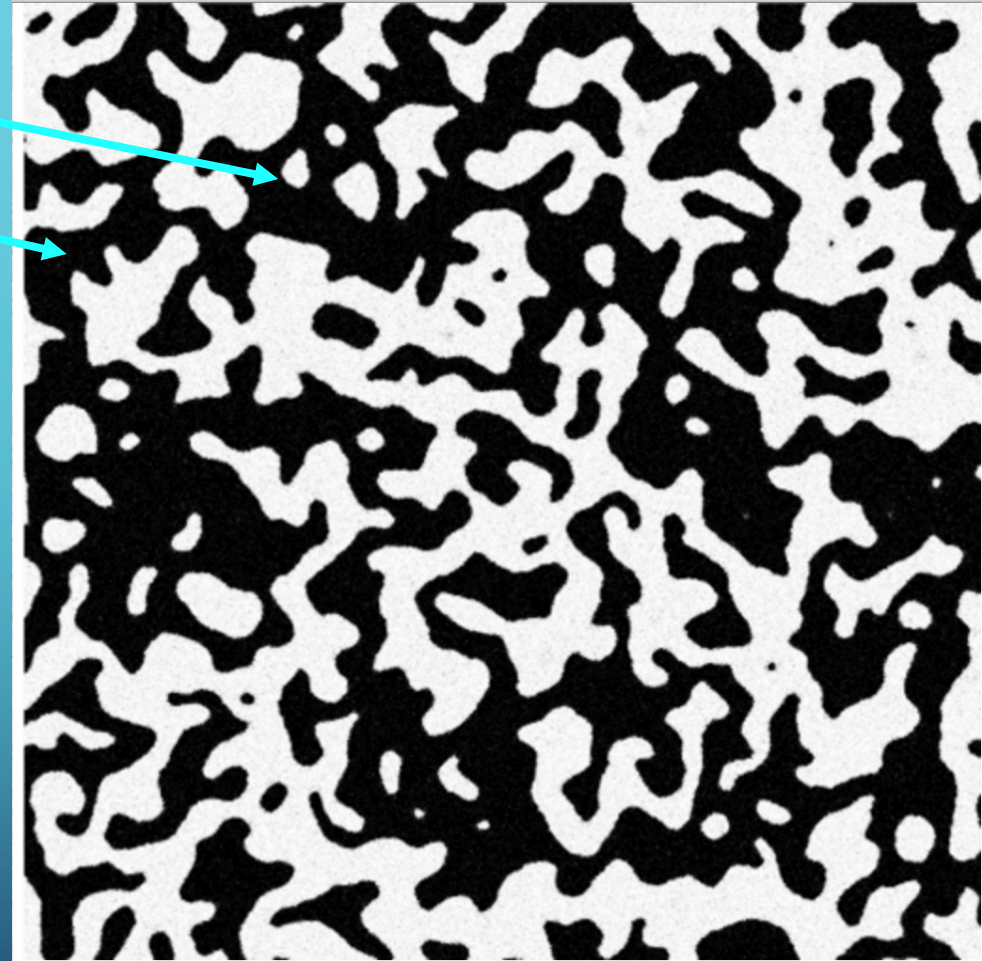
Our new claim:

If max T-statistic is above threshold  $u$ ,  
we expect the Euler characteristic to be **at least 1**.

But this seems harder to calculate...

Luckily, other people have already derived  
equations to solve for the expected Euler  
characteristic.

Thanks, math!





# FAMILYWISE ERROR CORRECTION

Controls the overall probability of making a Type I error:

**how many of my voxels are going to give me a false positive?**

But I might want to know instead:

**how many of my observed positive results are trustworthy?**

# FALSE DISCOVERY RATE CORRECTION

Let's control the proportion of positive results that are false positives

Obtains **greater sensitivity**, at the expense of specificity, relative to FWE

$$FWER = P(V \geq 1)$$

$$FDR = E\left(\frac{V}{R}\right)$$

note: EXPECTED, not guaranteed!

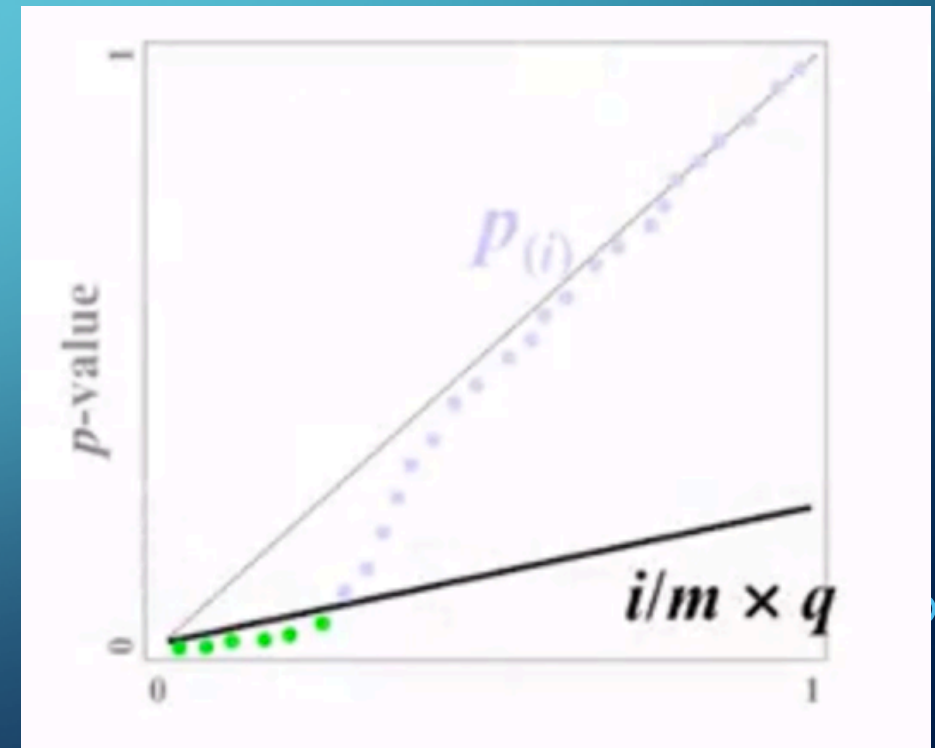
# FALSE DISCOVERY RATE CORRECTION

Select FDR threshold  $q$  – e.g., 0.05

Rank all the  $p$ -values over all the voxels in the brain

Find  $r$ : the largest  $i$  for which  $p_i$  is less than or equal to  $\frac{i}{m} \times q$

All  $p_i$  for which  $i \leq r$  are deemed significant  $p$ -values



# THE TRADE-OFF: SENSITIVITY VS SPECIFICITY

if you are controlling the rate of ALL false positives (FWER),  
you are *also* controlling the rate of false positives among all positives (FDR)

*sensitive*

if there's an effect here,  
we'll definitely find it

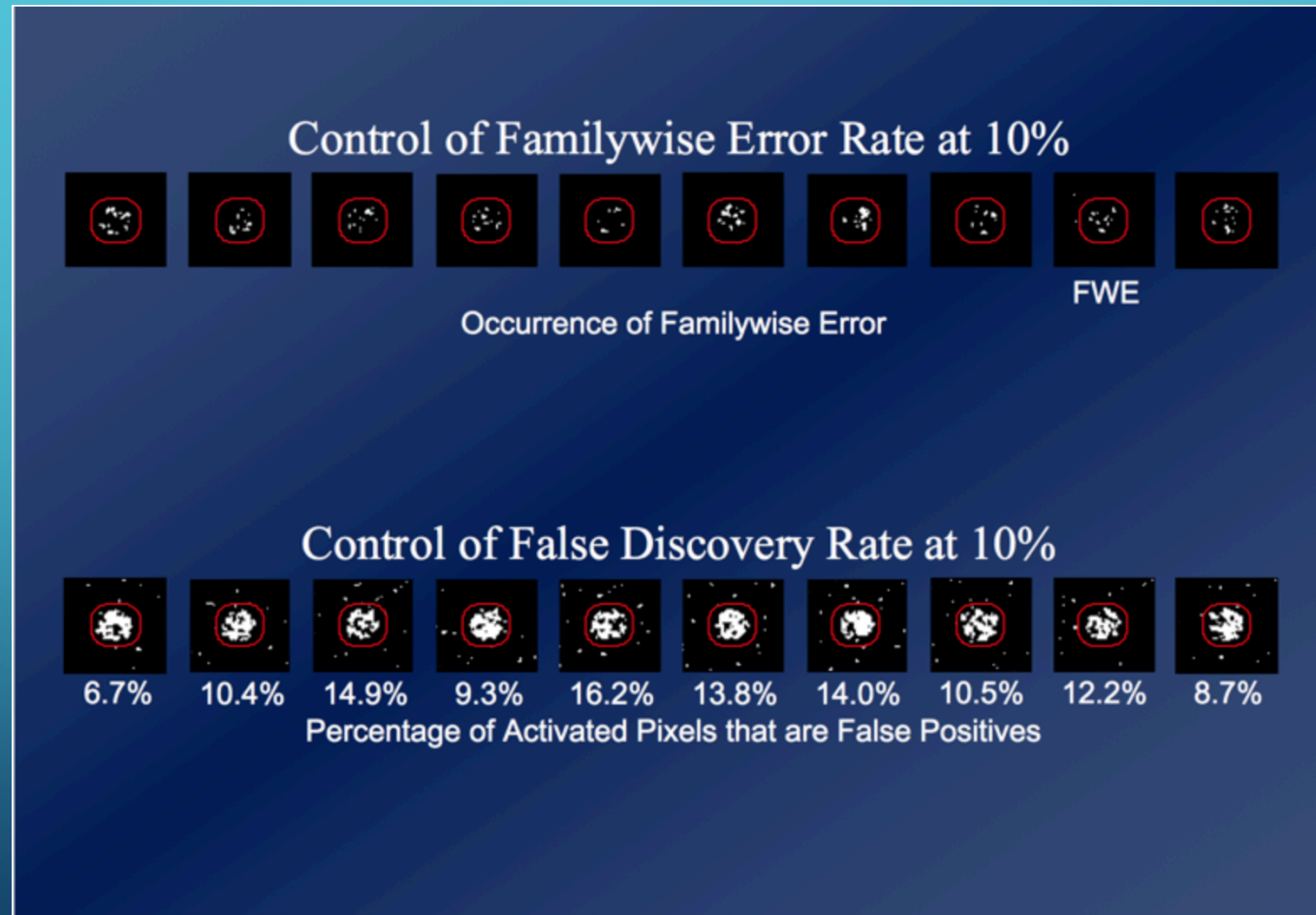
*specific*

we can trust that anything  
we see isn't false

FDR correction

FWER correction

# THE TRADE-OFF: SENSITIVITY VS SPECIFICITY



# AN ARGUMENT FOR SENSITIVITY OVER SPECIFICITY

doi:10.1093/scan/nsp052

SCAN (2009) 4, 423–428

## ***Tools of the Trade***

# **Type I and Type II error concerns in fMRI research: re-balancing the scale**

**Matthew D. Lieberman,<sup>1</sup> and William A. Cunningham<sup>2</sup>**

<sup>1</sup>Departments of Psychology, Psychiatry, & Biobehavioral Sciences, University of California, Los Angeles, and <sup>2</sup>Department of Psychology, Ohio State University, USA

# AN ARGUMENT FOR SENSITIVITY OVER SPECIFICITY

1. Stringent correction leads to inflated Type II errors
2. This leads to a bias toward large and simple effects (e.g., from visual or motor processes) rather than complex, subtle effects (e.g., from social cognition)
3. Correction procedures in fMRI claim to be in line with behavioral standards, but most behavioral papers don't correct
4. Type I errors will fail to replicate and thus will wash out in the literature anyway

# PERMUTATION TESTING

Why estimate a distribution when you can make your own?

The basic framework:

1. Permute something (labels, voxels, etc.) randomly on each of  $k$  iterations, and then run your statistical test
2. Compare your actual test statistic to the distribution of statistics you just generated

Some tools/methods: SnPM, AFNI's 3dClustSim, Scott Slotnick's permutation method



# SNPM

SPM toolbox (N stands for NONparametric)

Developed by Tom Nichols & co.

Free, open-source: <https://github.com/nicholst/SnPM-devel>

Downloads at: <http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/software/snpm>

Documentation at: <http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/software/snpm/man>

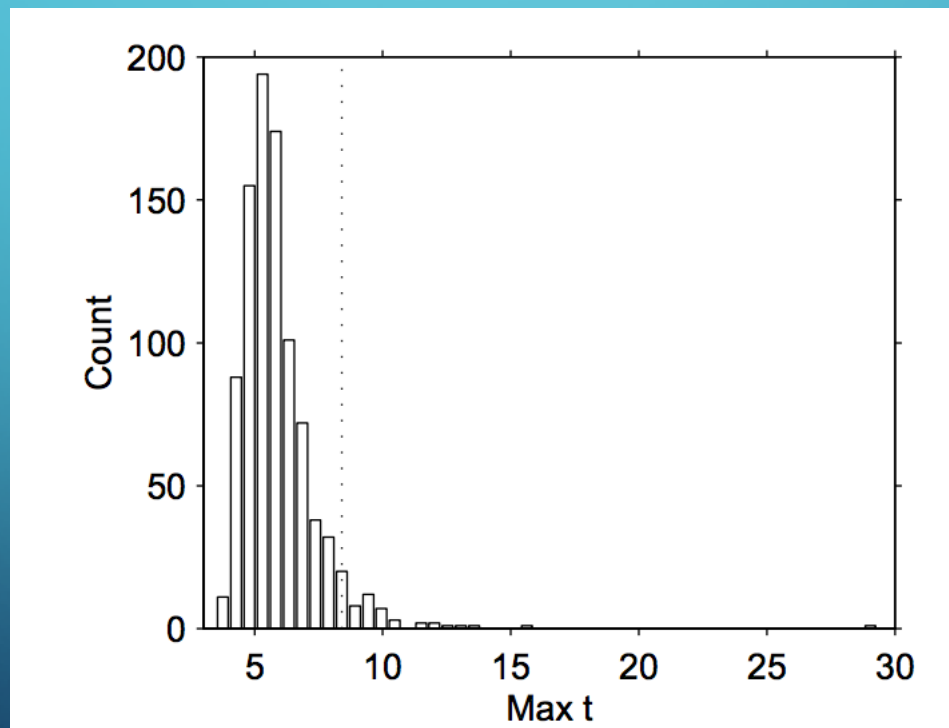
Installed on Pleiades at: `/usr/public/spm/spm12/toolbox/snpm`



# SNPM

Exchanges labels in the data for each iteration, then calculates maximum T (or maximum  $k$ , for cluster-wise correction)

Ex.: randomly shuffle the 'belief' and 'photo' labels assigned to items in a ToM localizer task



*Nichols & Holmes (2001)*

# SNPM

Can do both **voxel-level** and **cluster-level** correction

## Voxel-level:

Find a critical T-threshold based on permutation tests

Any voxel exceeding the threshold is considered significant

## Cluster-level:

Choose a T-threshold *a priori* and threshold all voxels

Find a critical cluster size  $k$  based on permutation tests

Any cluster larger than  $k$  is considered significant



# 3DCLUSTSIM

AFNI function; replaces older AlphaSim

AFNI is free and open-source, downloadable at: <https://afni.nimh.nih.gov/afni>

## CAVEAT:

AFNI is a command-line utility, so those who are less familiar with programming might find it a bit tougher than SPM initially

“Wait, but didn’t that paper say that 3dClustSim had a bug?”

Yes, and it’s been fixed.

A very fresh (7/20/2016) version of AFNI is installed on Pleiades at: `/usr/public/afni`



# 3DCLUSTSIM

Permutes voxels within a brain map, taking spatial smoothness into account, to generate a distribution of noise maps

# 3DCLUSTSIM

Sample command:

```
3dClustSim -nxyz 53 63 46 -dxyz 3 3 3 -fwhmxyz 11.01 10.55 11.01
```

# 3DCLUSTSIM

Sample output:

```
# 3dClustSim -nxyz 53 63 46 -dxyz 3 3 3 -fwhmxyz 11.01 10.55 11.01
# bi-sided thresholding
# Grid: 53x63x46 3.00x3.00x3.00 mm^3 (153594 voxels)
#
# CLUSTER SIZE THRESHOLD(pthr,alpha) in Voxels
# -NN 3 | alpha = Prob(Cluster >= given size)
# pthr | .10000 .05000 .02000 .01000
# -----|-----
0.050000 414.1 465.7 536.0 588.3
0.020000 193.5 216.7 250.0 273.0
0.010000 126.0 143.0 164.7 178.0
0.005000 87.4 99.6 116.0 127.9
0.002000 56.9 65.7 77.5 85.0
0.001000 42.2 49.3 58.0 65.5
0.000500 31.6 37.5 45.1 51.0
0.000200 21.5 26.3 32.5 36.8
0.000100 16.0 19.8 25.2 29.4
++ Clock time now = 54.3 s
```

# A FINAL WORD FROM TOM NICHOLS

“ •My take homes from this exercise have been:

•No matter what method you're using, if you go to town on a P-value on the razor's edge of  $P=0.05000$ , you *lean heavily on the assumptions of your method*, and any perturbation of the data (or slight failure of the assumptions) would likely give you a result on the other side of the boundary.

”



# A FINAL WORD FROM TOM NICHOLS

“ • In an ideal world, all studies, good and bad, would be published with full data sharing...each *clearly marked* with their varying strengths of evidence (***no correction*** < ***FDR voxel-wise*** < ***FDR cluster-wise*** < ***FWE cluster-wise*** < ***FWE voxel-wise***).

”

# A FINAL WORD FROM TOM NICHOLS

- **Complete reporting of results**, i.e. filing of statistical maps in public repositories, must happen!
- Data sharing **must also come**.

”

# TOWARDS DATA SHARING



A public repository of unthresholded statistical maps, parcellations, and atlases of the human brain



## Open Science Framework

A scholarly commons to connect the entire research cycle



# REFERENCES

Holmes, A. (1996). Non-parametric analysis of statistic images from functional brain mapping. *Journal of Cerebral Blood Flow and Metabolism*.

Lieberman, M., & Cunningham, W. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *SCAN*.

Nichols, T., & Holmes, A. (2001). Nonparametric analysis of PET functional neuroimaging experiments: a primer. *Human Brain Mapping*.

Penny, Will (2005). Statistical Inference, Multiple Comparisons, Random Field Theory. <http://www.fil.ion.ucl.ac.uk/spm/course/slides05/ppt/infer.ppt>

Winkler, A., Ridgway, G., Webster, M., Smith, S., & Nichols, T. (2014). Permutation inference for the general linear model. *NeuroImage*.