Significance Testing and Neuroimagery

Edouard Machery
University of Pittsburgh

fMRI

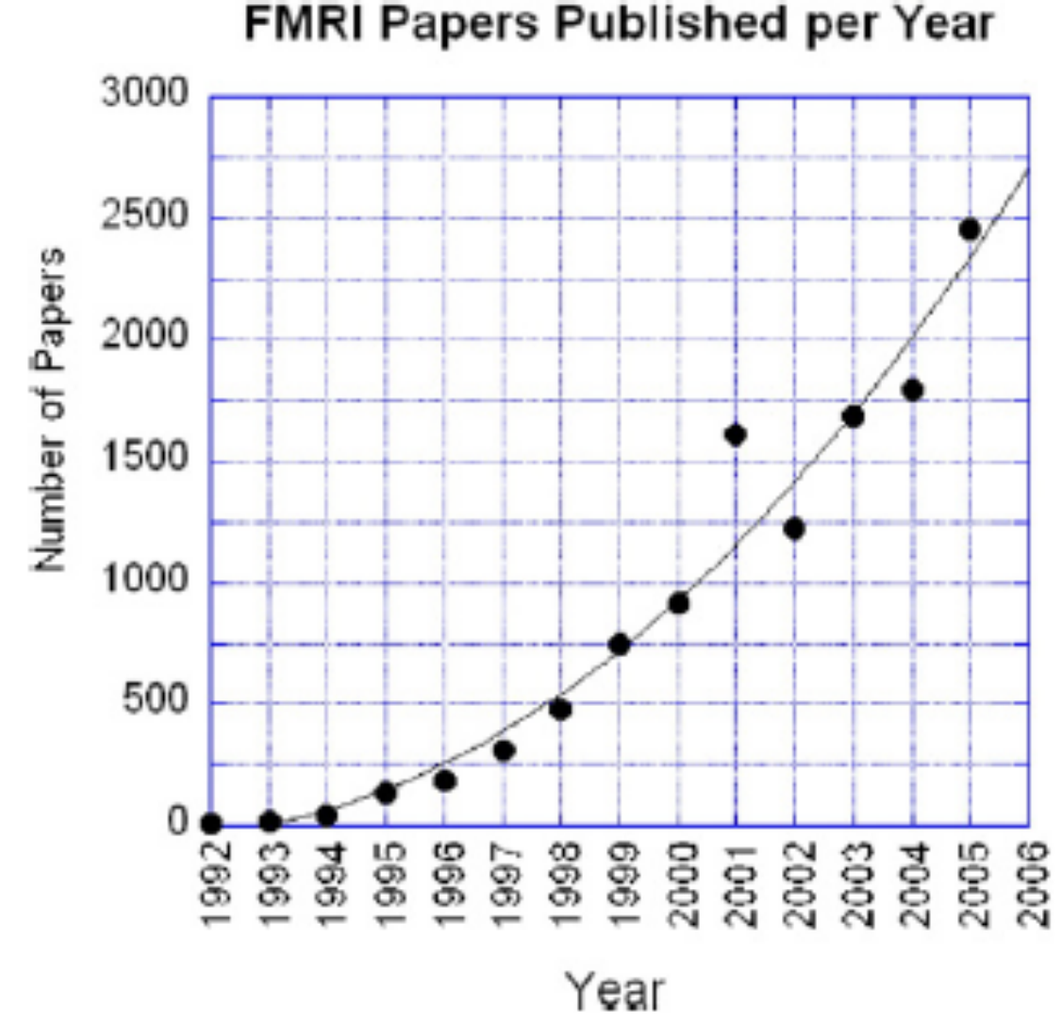
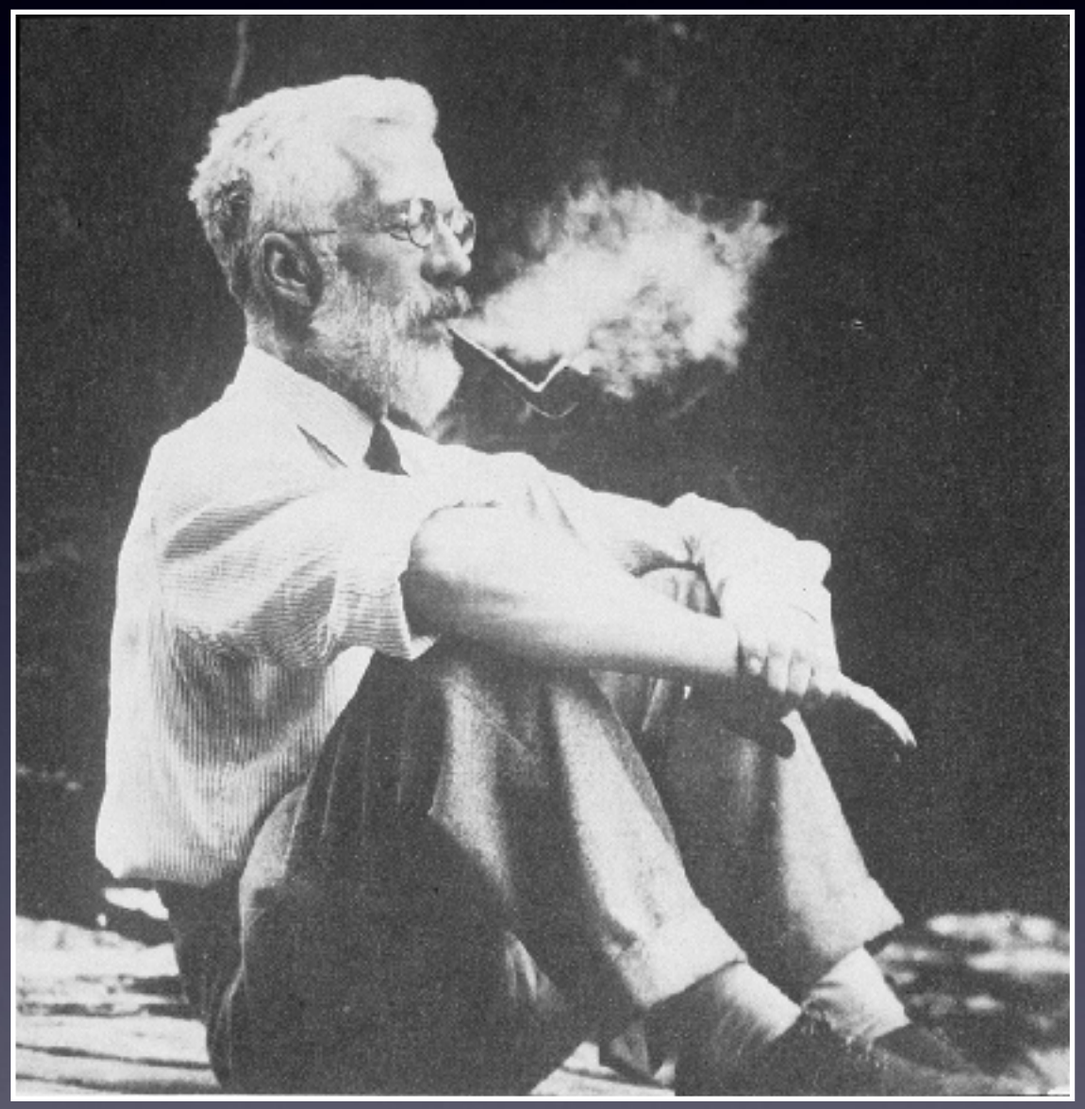


Fig. 1. Number of papers with keyword, title, or abstract containing "fMRI" or "functional MRI" published per year.

Bandettini, 2007

Significance Testing



The Issue

Does neuroimagery's reliance on significance tests prevent imagery data to be evidence for or against functional hypotheses (as Colin Klein [2010, *BJPS*] has argued)?

Take-Home Message

no

Plan

1. Significance tests in neuroimagery
2. Colin Klein's argument
3. Significance tests when the null is false
4. Application to neuroimagery

Plan

1. Significance tests in neuroimagery
2. Colin Klein's argument
3. Significance tests when the null is false
4. Application to neuroimagery

Believing in free will causes people to behave in a morally appropriate manner (Vohs & Schooler, 2008)

<u>Substantive hypothesis</u>
Causal relation or correlation between constructs

People who read an essay
free will are <u>more likely</u> to
a given task than people w
an irrelevant essay

If the probability of obtaining a standardized difference between the means of a given size conditional on the hypothesis that people who read an essay denying free will are as likely to cheat as people who read a control essay (p) is low ($< \alpha$), then the null hypothesis is rejected and it is inferred that people who read an essay denying free will are more likely to cheat than people who read a control essay.

People who read an e
denying free will are <u>equa</u>
to cheat in a specific
people who read an

Given this hypothe
derive the sampli
standardized difference between the means in the anti-free-will and control conditions.

Because H_0 is specific, it is possible to derive the sampling distribution of a given statistic given H_0.

Null Hypothesis Significance Testing (NHST)

1. To test theory *T*, assuming a statistical model (typically, $X \sim N(\theta,\sigma)$), a statistical hypothesis (H_A: $\theta \neq 0$) is derived .
2. H_A cannot be directly assessed: Since it is a range hypothesis, it does not allow for the computation of a sampling distribution for the statistic of interest (e.g., *F*).
3. One formulates a null hypothesis (H_0), which allows for the computation of this sampling distribution and contradicts H_A (e.g., H_0: $\theta = 0$).
4. If the *p*-value of the statistic computed from the data is below the significance level (α), then H_0 is rejected.
5. If H_0 is rejected, then H_A is accepted, which provides some support for *T*.

NHST in Cognitive Neuroscience

The statistical analyses involved in testing cognitive-neuroscientific hypotheses often rely on Null Hypothesis Significance Testing.

General Linear Model in Whole-Brain Scans

Build a model made of regressors of interest and nuisance regressors (e.g., head motions) obtained by convolving variables with a hemodynamic response function and, depending on the contrast, test voxel by voxel whether the regression coefficients differ from 0 or from one another.

Toy Example

t- and F- tests in ROI Analyses

For each subject, identify a Region of Interest (e.g., rTPJ) functionally by means of a functional localizer task. In the main task, test whether the percent signal change (averaged across each voxel of the RoI and across subjects) differ from 0 for the experimental conditions or differ across experimental conditions by means of a t- or F-test.

Example

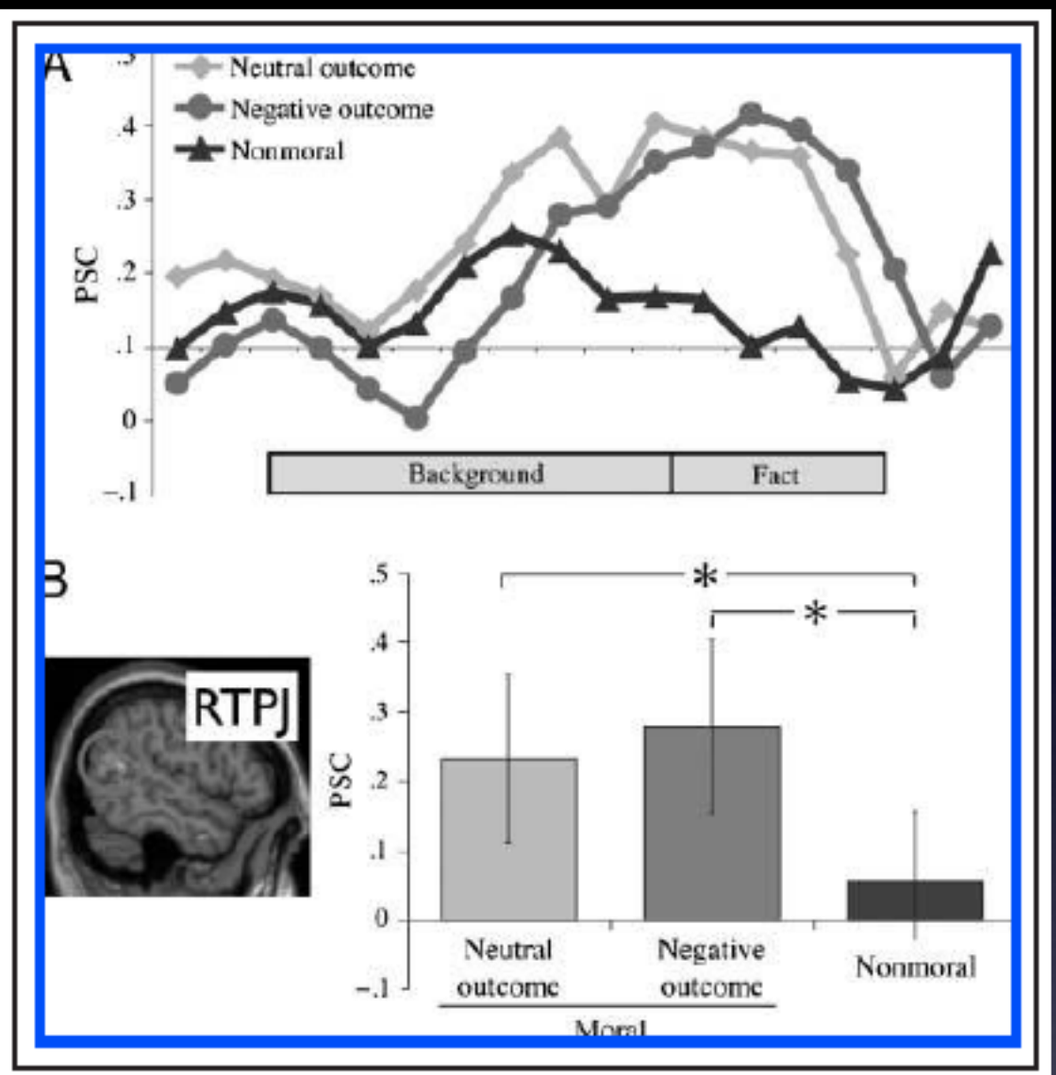


Figure 2. (A) PSC from rest in the RTPJ over time in Experiment 1. Background information was presented for the first 12 sec. Fact was presented for the second 6 sec. Labels have been shifted forward 4.5 sec relative to the stimulus timing to account for hemodynamic lag. (B) PSC at Time 2 (fact) in the RTPJ in Experiment 1. (*Left*) Brain regions where the BOLD signal was higher for (nonmoral) stories about beliefs than (nonmoral) stories about physical representations ($n = 14$, random effects analysis, $p < .0001$, uncorrected). These data were used to define ROIs, that is, RTPJ. (*Right*) The PSC was significantly greater for moral facts (neutral and negative outcomes, lighter bars) than nonmoral facts (darker bar). Error bars represent standard error.

Young & Saxe, 2008
Exp. 1

Plan

1. Significance tests in neuroimagery
2. Colin Klein's argument
3. Significance tests when the null is false
4. Application to neuroimagery

Klein (2010, *BJPS*)

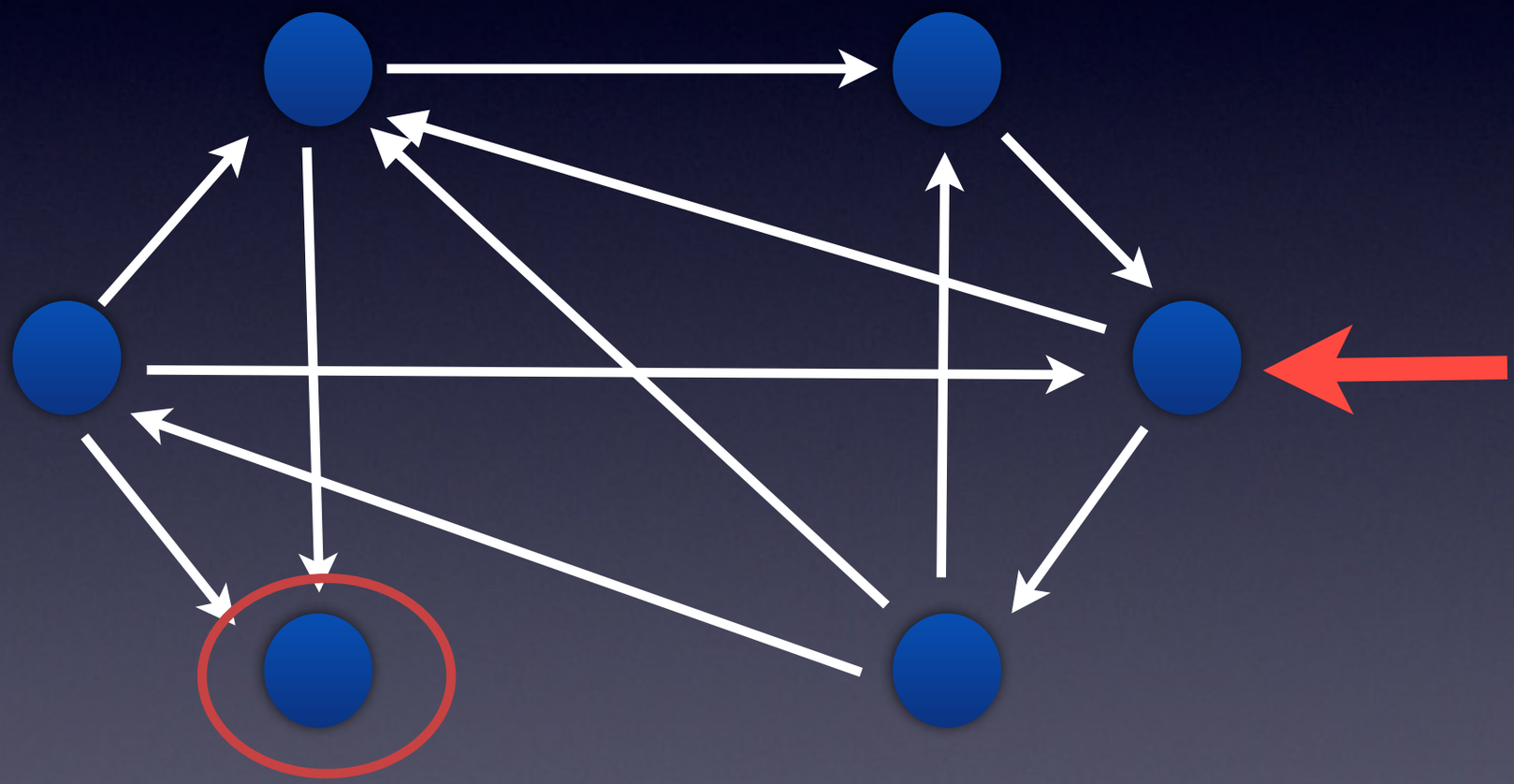
pictures of "brain activity" associated with fMRI experiments are poor evidence for
neuroimages present the results of null hypothesis significance
tests performed on fMRI data. Significance tests alone cannot provide evidence about the
functional structure of causally dense systems, including the brain.
should be seen as indicating regions where further data analysis is warranted. This

Inspired by Klein (2010)

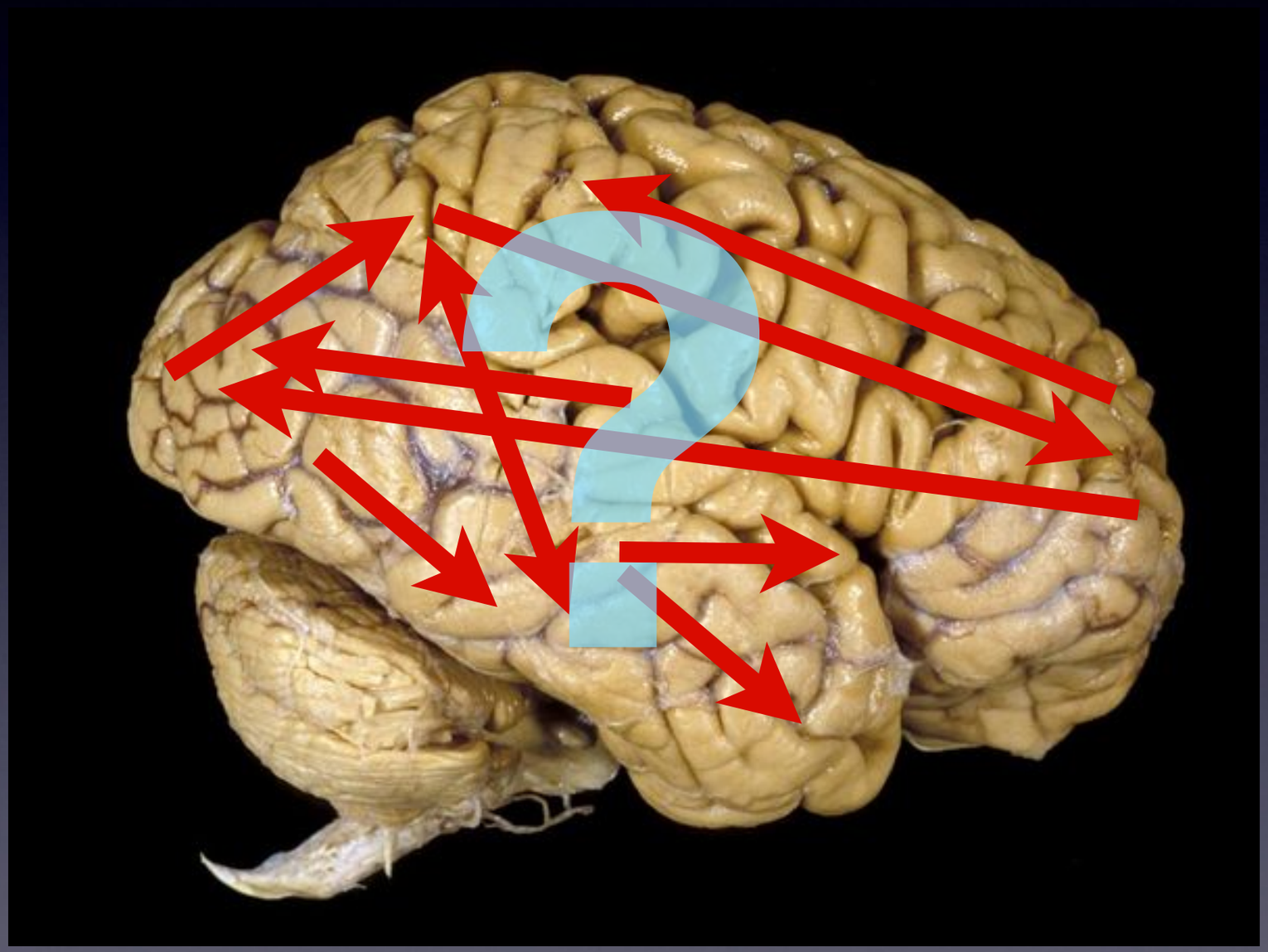
Step 1

1. Any change induced in a variable of a causally dense system causes a change in all the other variables.
2. The brain is a causally dense system.
3. An experimental task induces a change in the BOLD signal in the brain areas and voxels functionally involved in completing this task.
4. Hence, whether or not area or set of voxels A is functionally involved in completing task T, completing T induces a change in the BOLD signal in A.
5. Hence, changes in the BOLD signal cannot support functional hypotheses.

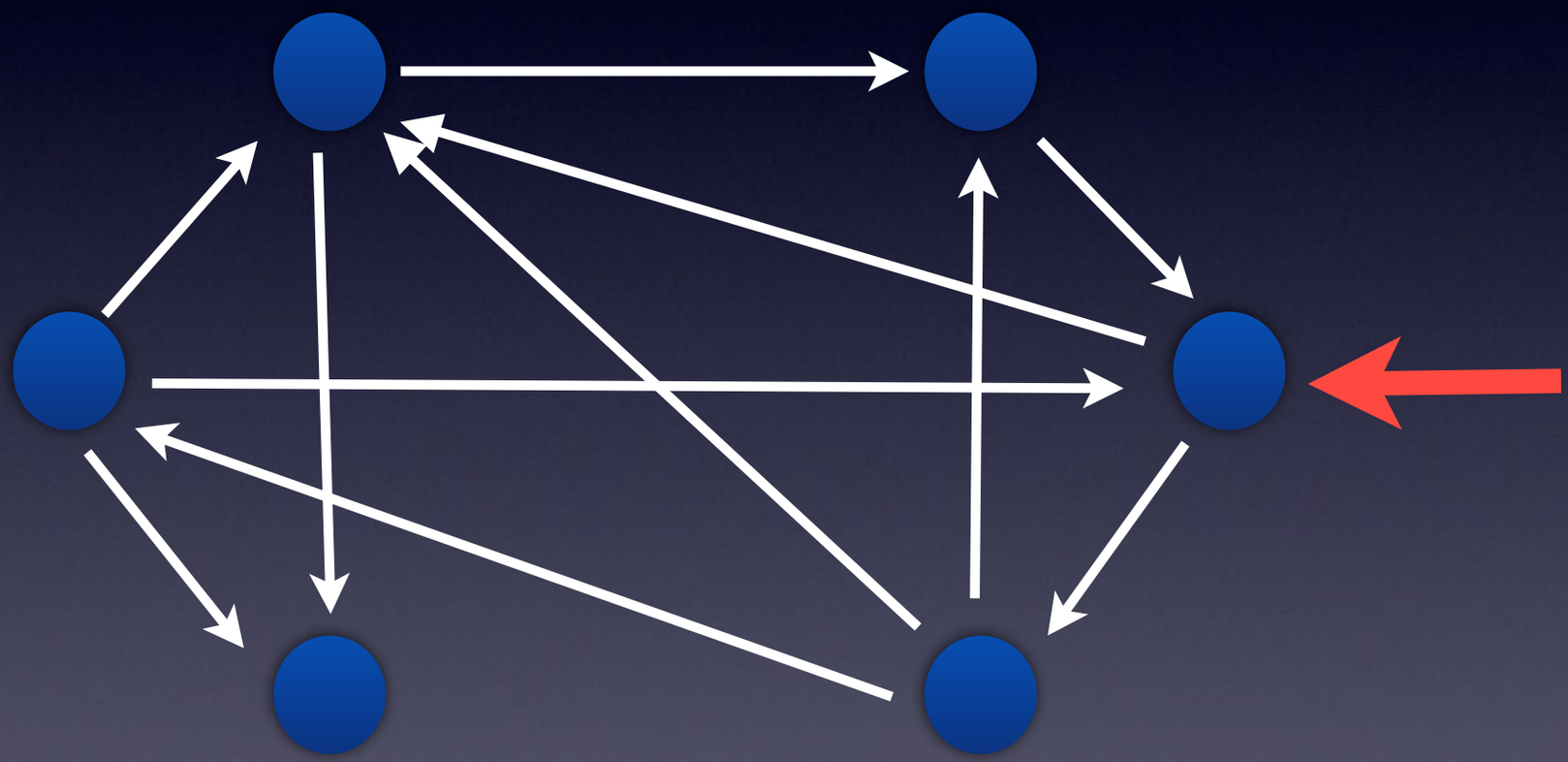
Causally Dense Systems (Pr. 1)



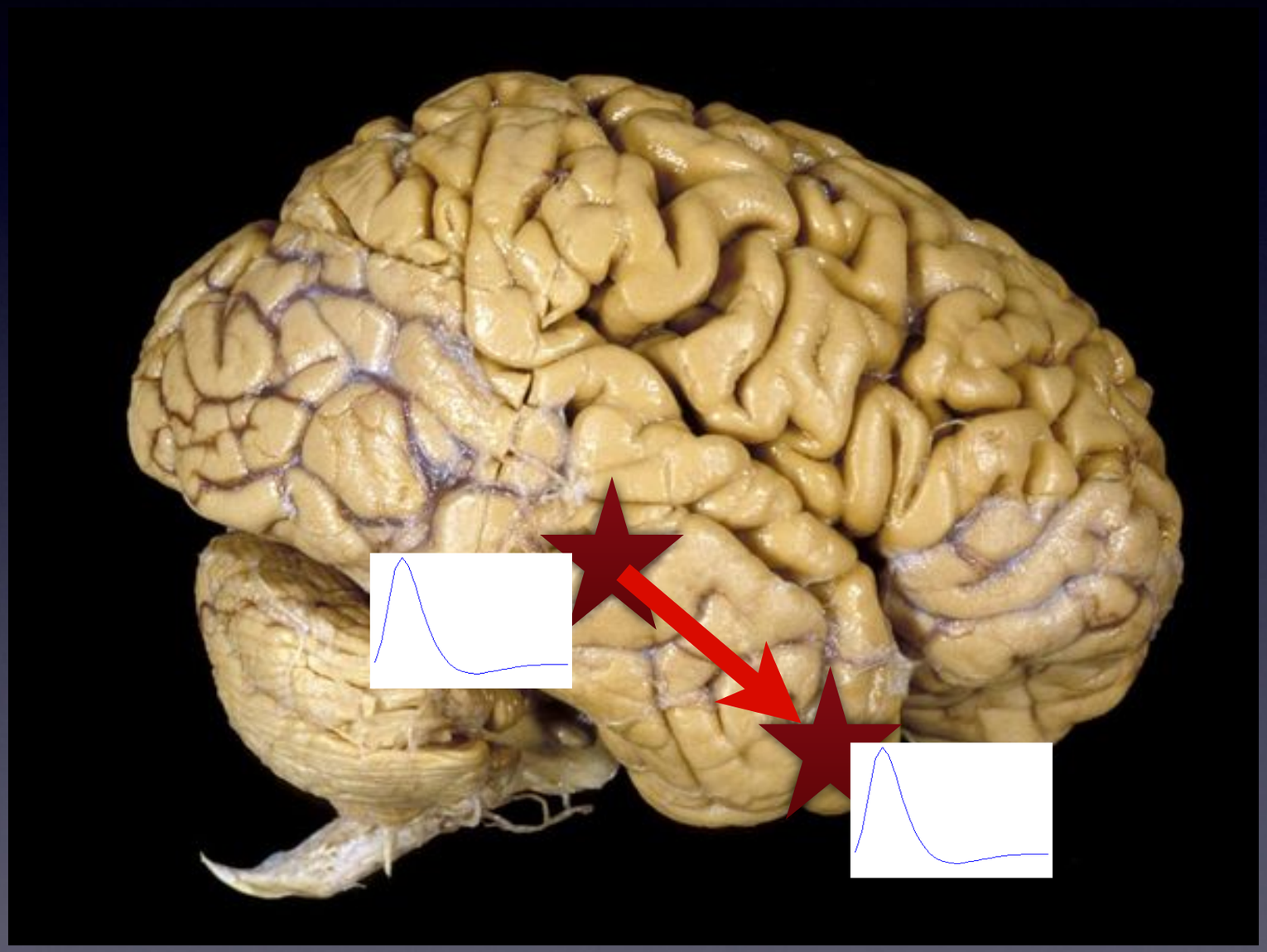
Is the Brain Causally Dense (Pr. 2)?



Experiments as Inducing Changes (Pr. 3)

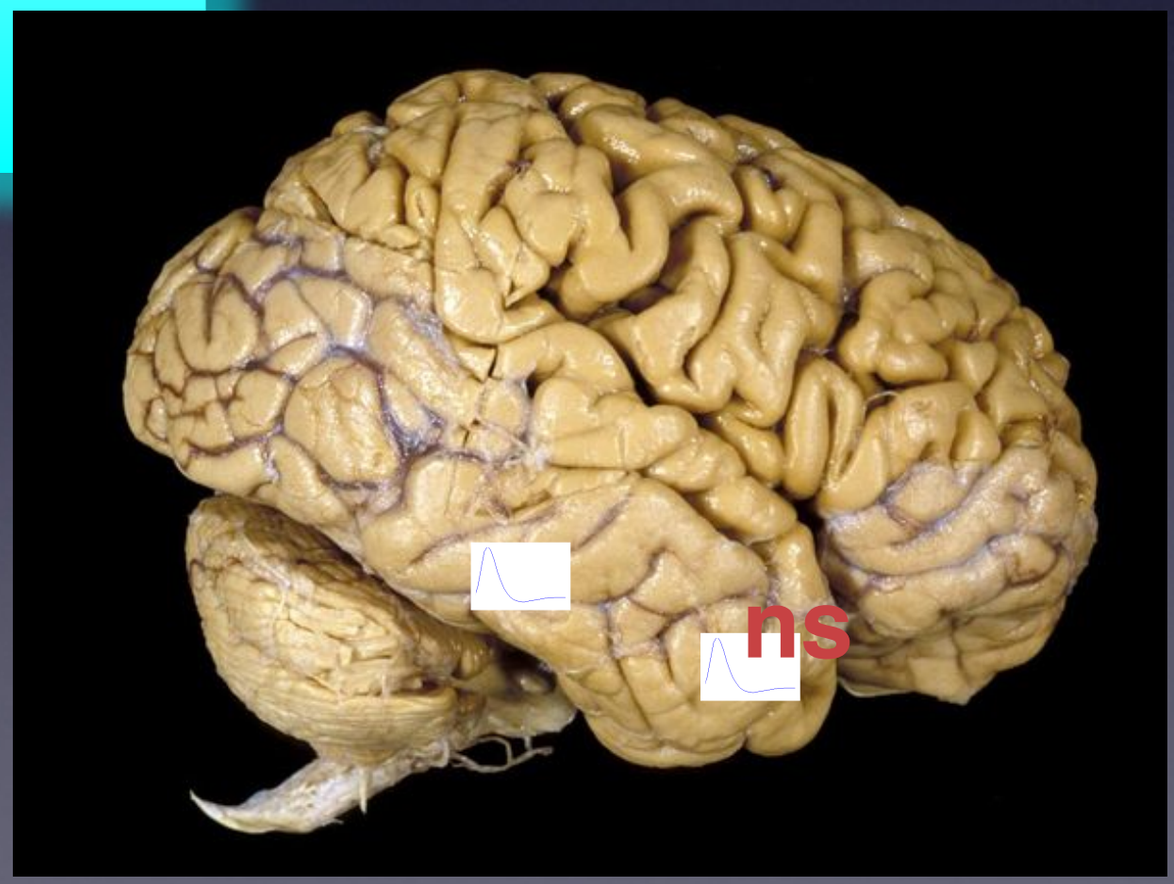


Functional Hypotheses and Brain Activation (Clsions 4 & 5)



Response

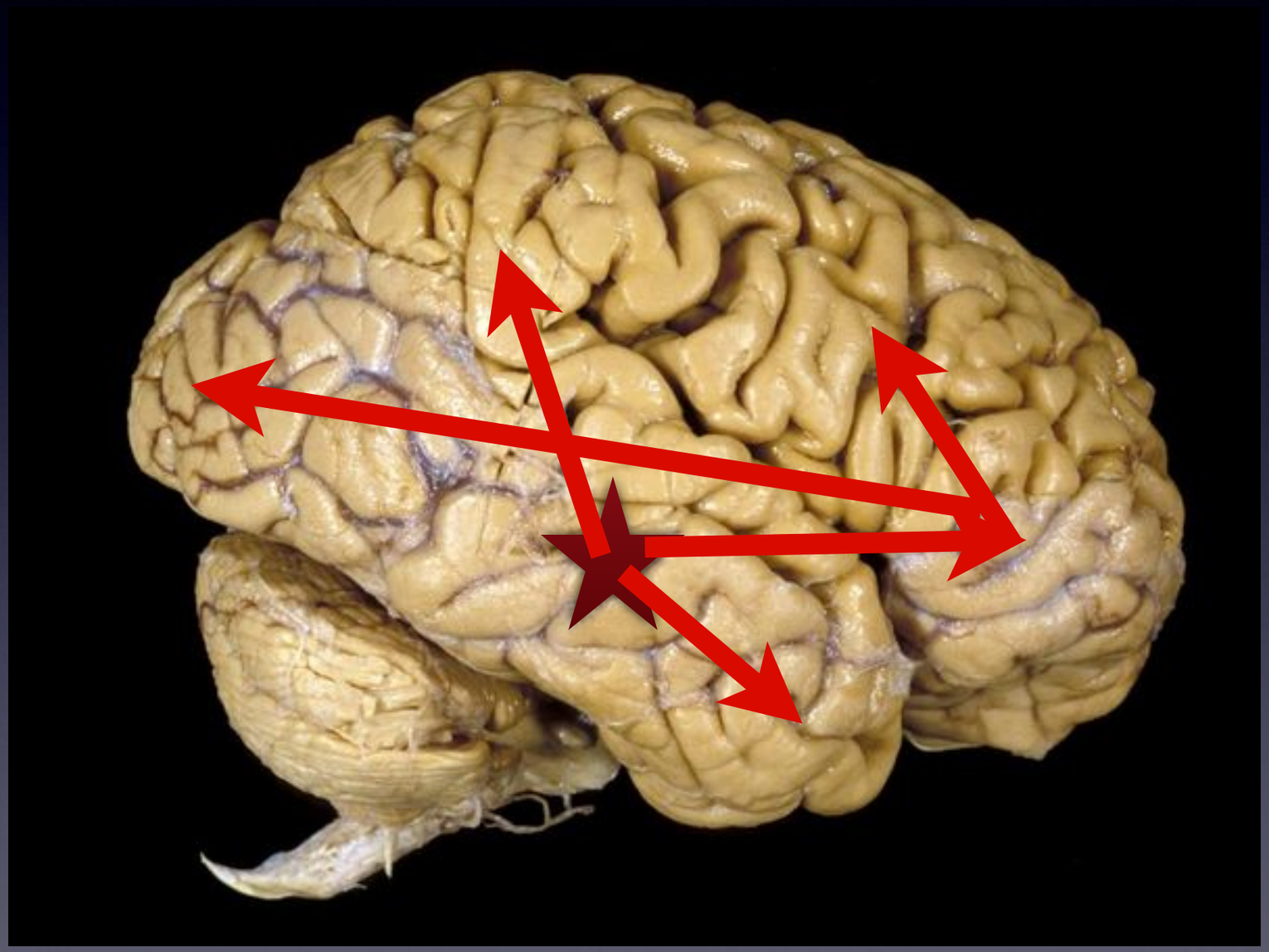
But surely not every change in the BOLD signal will be significant, and only *significant* changes matter for confirming functional hypotheses.



Step 2

1. An empirical result is statistically significant if and only if the relevant *p*-value is below the significance level.
2. The significance level is set at a particular value to control the probability of rejecting the null hypothesis when it is true.
3. Any experimental task causes a change in the BOLD signal of any brain area or voxel.
4. Hence, in significance tests in fMRI-based studies, the significance level should be set at 0.
5. Hence, all changes in the BOLD signal should be treated as being significant.

No Risk of False Positives



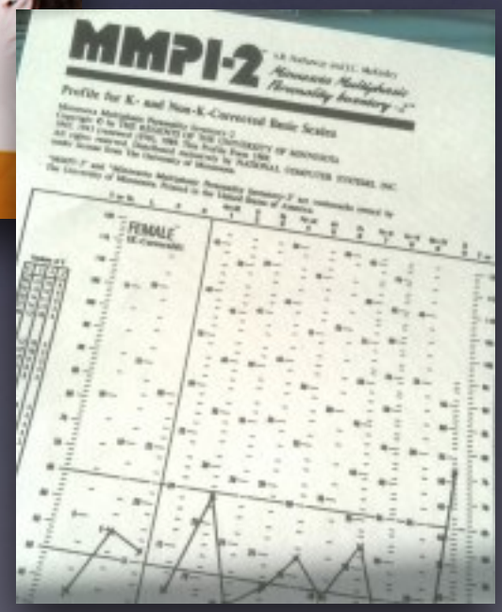
Plan

1. Significance tests in neuroimagery
2. Colin Klein's argument
3. Significance tests when the null is false
4. Application to neuroimagery

The Problem

What is the point of doing a significance test if the null hypothesis is bound to be false?

A General Problem



Significance Tests When the Null is false

When the null is bound to be false, NHST is used to test whether the effect size is null *or nearly so*.

Significance Tests When the Null is false

When it is very unlikely to obtain a statistic of a given size or a larger one if the *point* null hypothesis (H_0) is true, scientists are *in fact* rejecting a *range* null hypothesis, H_0', e.g.,:

H_0': $\theta = 0 +/- \delta$ (where δ is a trivial value).

and they accept the following type of alternative hypothesis H_A':

H_A': $| \theta | > \delta$.

Type-I Error Probability

The probability of rejecting H_0' if H_0' is true is at most the power of the test computed assuming a trivial effect size (δ).

Plan

1. Significance tests in neuroimagery
2. Colin Klein's argument
3. Significance tests when the null is false
4. Application to neuroimagery

NHST and the Study of Causally Dense Systems

Statistical hypothesis
A non-trivial effect ($|\theta|>\delta$)

~~**Statistical hypothesis**~~
~~A trivially small effect ($\theta=0+/-\delta$)~~

NHST

NHST and the Study of Causally Dense Systems

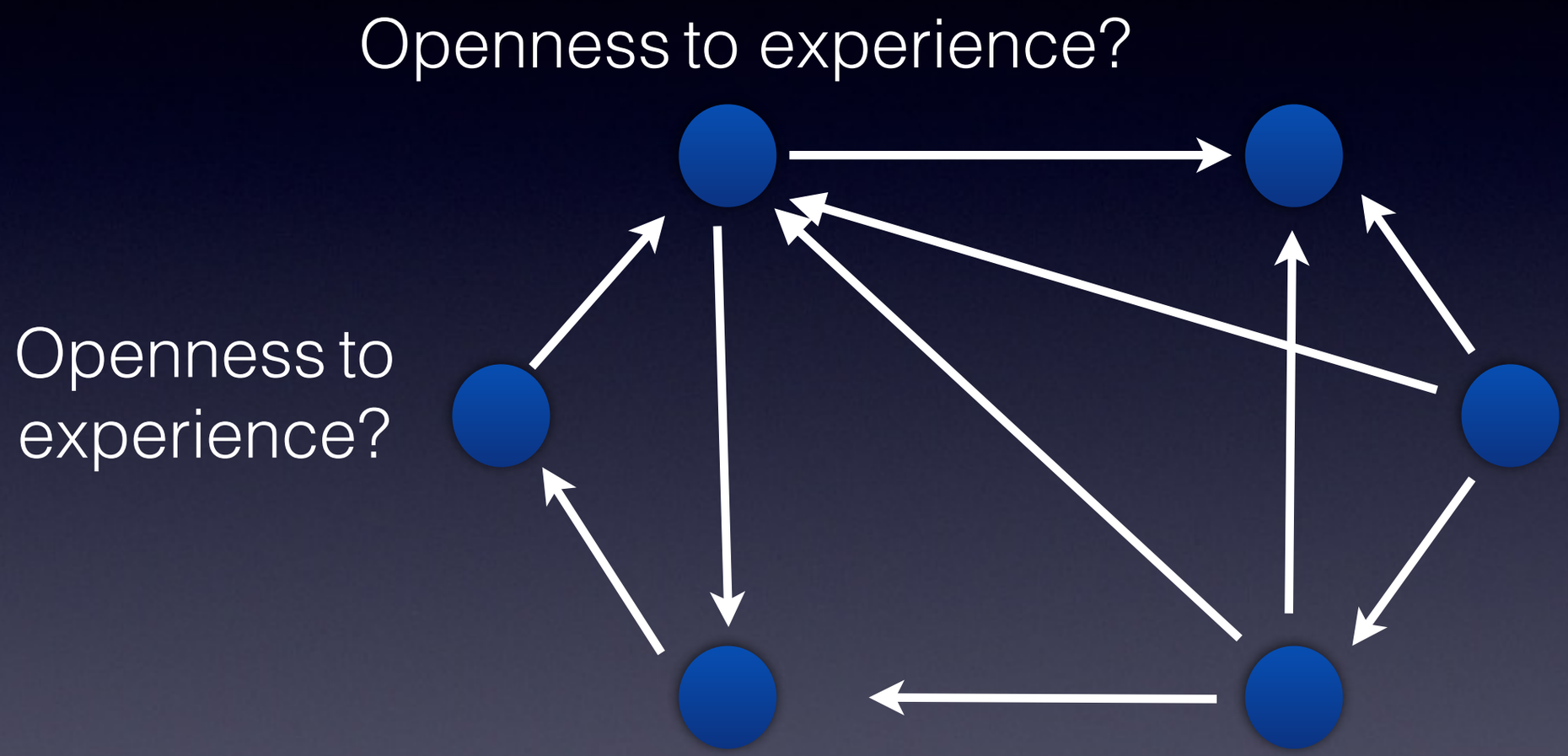
Substantive hypothesis is true
(e.g., A is an important cause of B)

Substantive hypothesis is false
(A is not an important cause of B)

Statistical hypothesis
A non-trivial effect ($|\theta| > \delta$)

Statistical hypothesis
A trivially small effect ($\theta = 0 +/- \delta$)

NHST and the Study of Causally Dense Systems



Does birth order have an *important* causal influence on openness to experience?

NHST and the Study of Causally Dense Systems

<u>Substantive hypothesis is true</u>
Birth order is an important cause of openness to experience

<u>Substantive hypothesis is false</u>
Birth order is not an important cause of openness to experience

<u>Statistical hypothesis</u>
A non-trivial difference between the openness to experience of first and second born children

<u>Statistical hypothesis</u>
A trivial difference between the openness to experience of first and second born children)

NHST

Application to Neuroimagery

Functional hypothesis is true
Area A plays a role in φ-ing

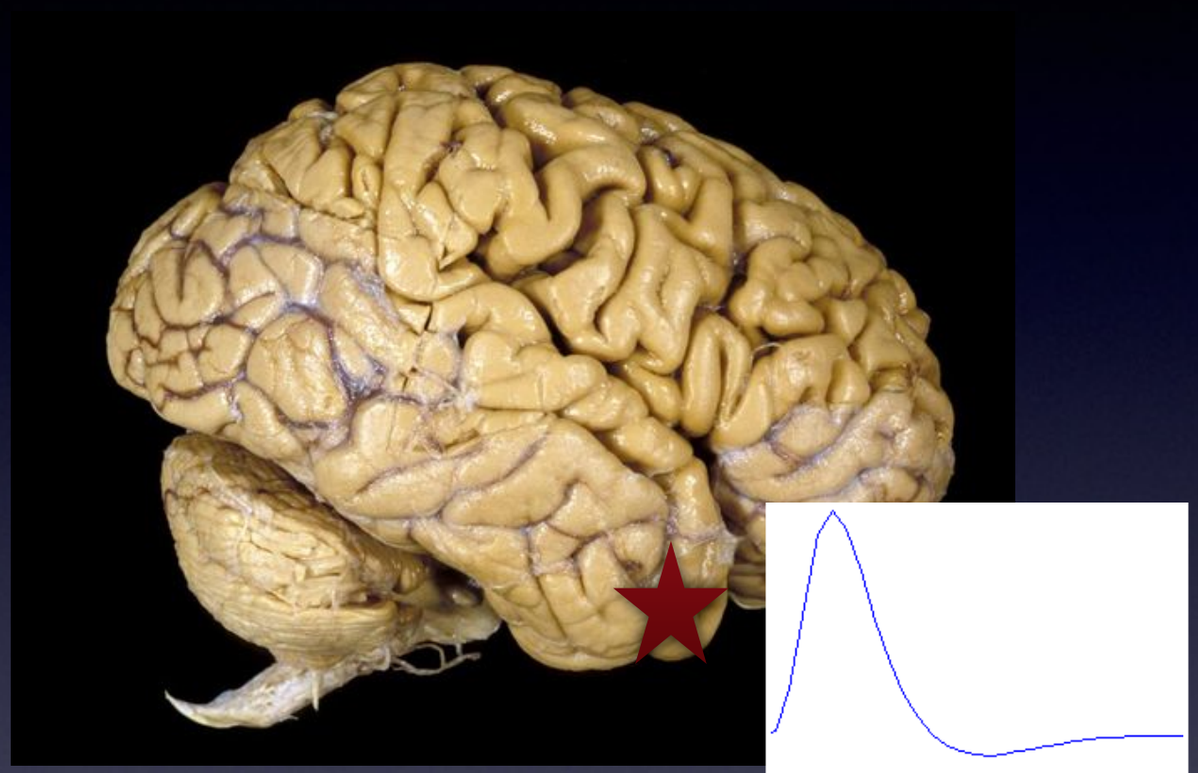
Functional hypothesis is false
Area A plays no role in φ-ing

Statistical hypothesis
A non-trivial contrast ($|\theta| > \delta$)

Statistical hypothesis
A trivially small contrast ($\theta = 0 +/- \delta$)

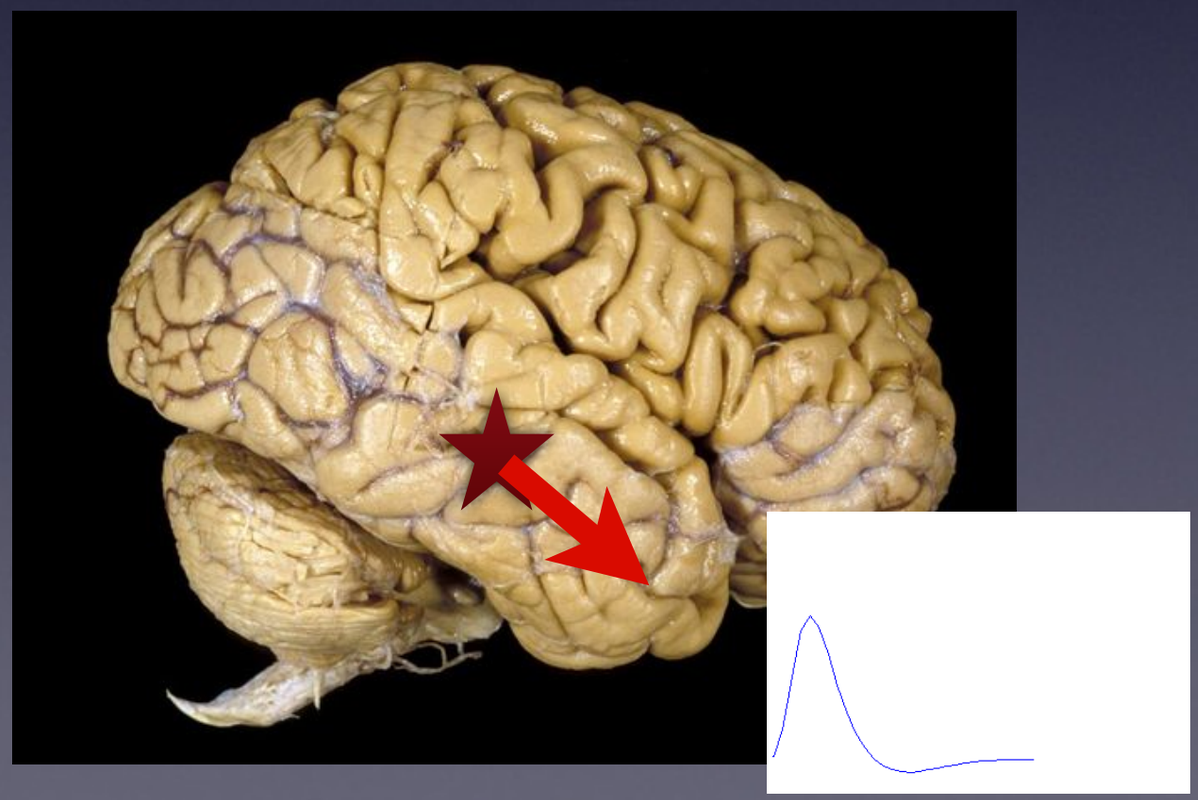
NHST

Application to fMRI



significant

n.s.



Application to Neuroimagery

Functional hypothesis is true
Area A plays a role in φ-ing

Statistical hypothesis
A non-trivial contrast ($|\theta| > \delta$)

Functional hypothesis is false
Area A plays no role in φ-ing

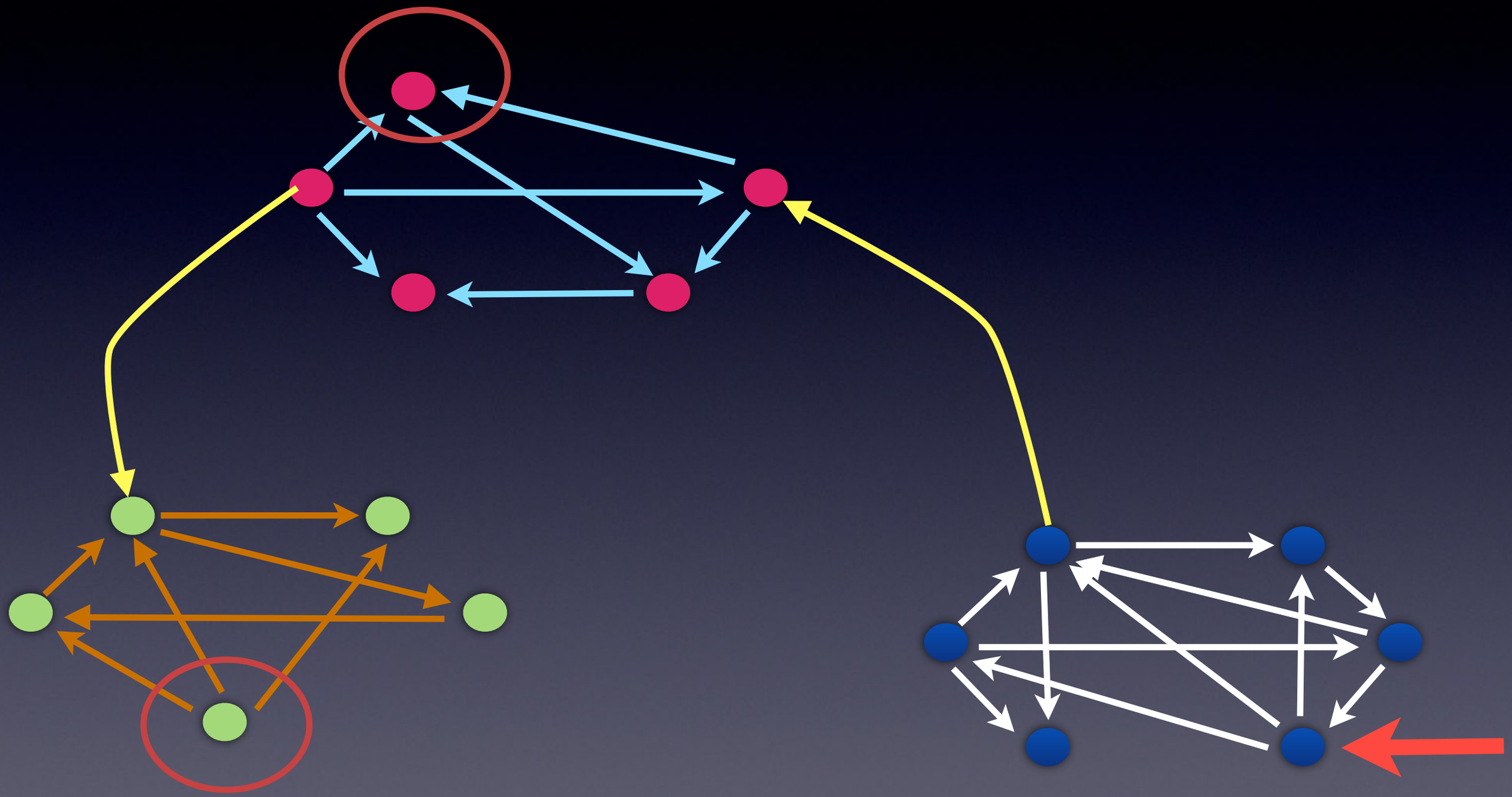
Statistical hypothesis
A trivially small contrast ($\theta = 0 +/- \delta$)

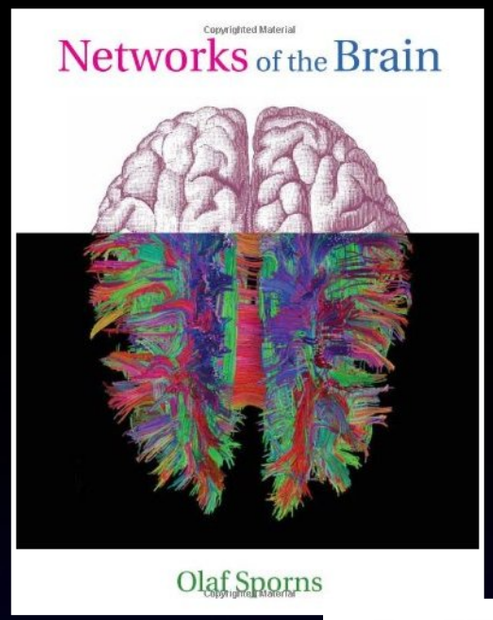
Response 1

1. The more modular a system is, the more likely it is that changes within a module don't have a large influence on the values of nodes in other modules.

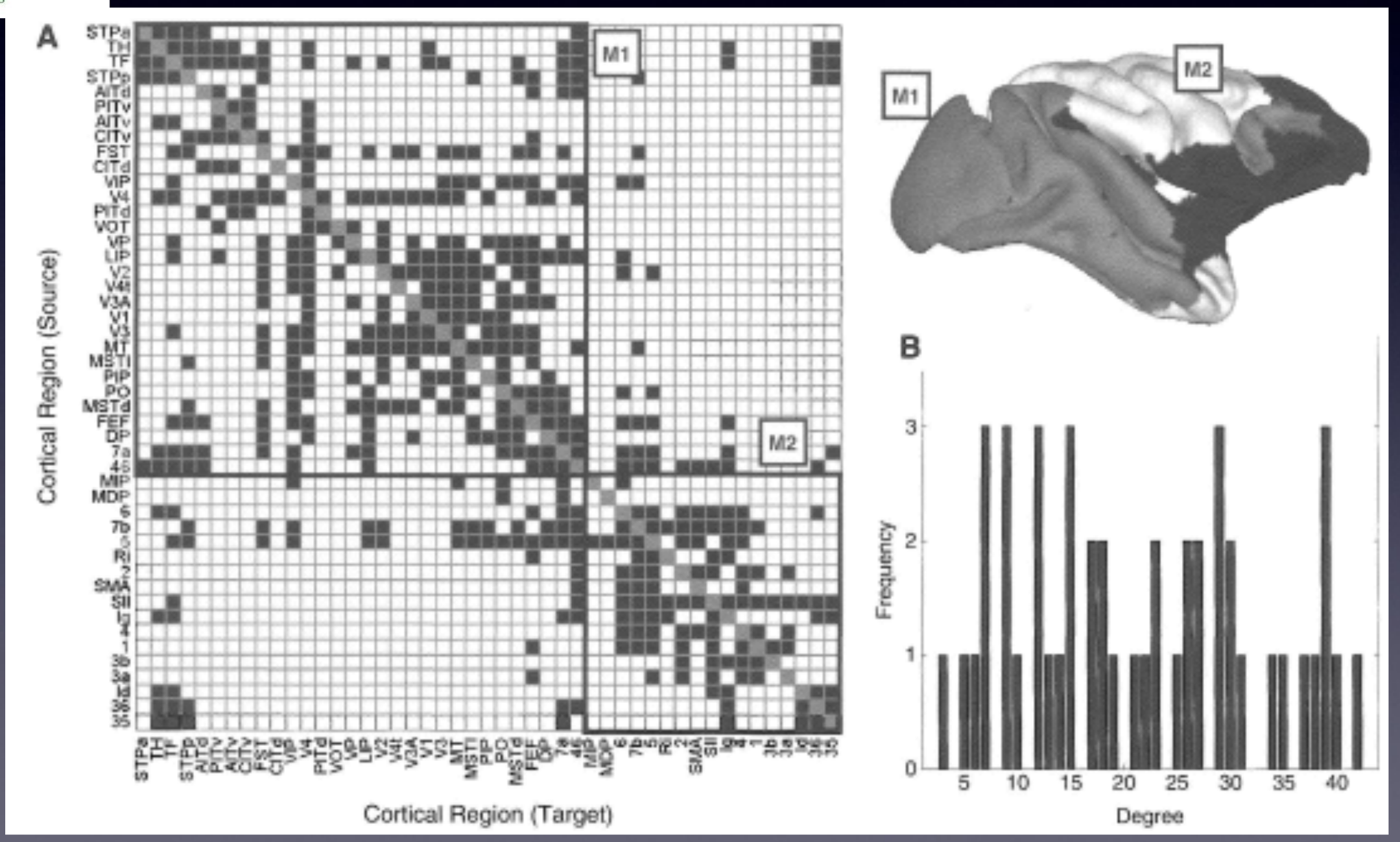
2. At the level of analysis of brain areas (the relevant level for fMRI), the brain is to a substantial extent modular.

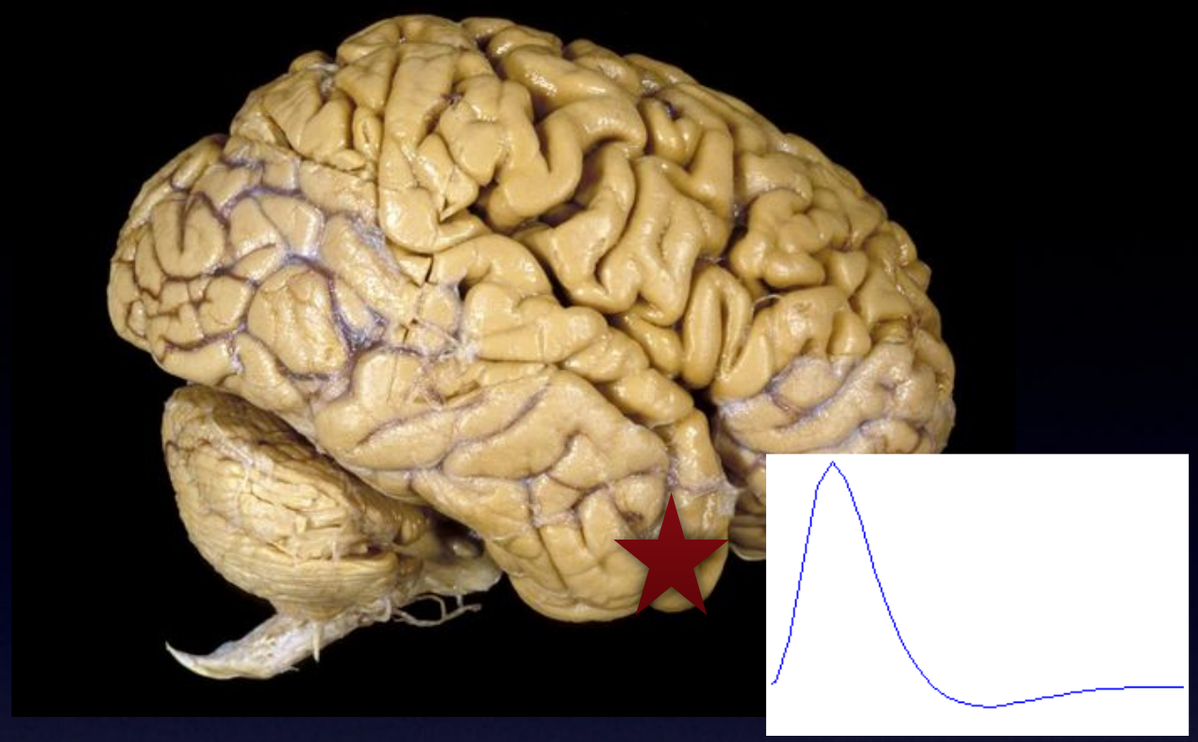
Claim 1

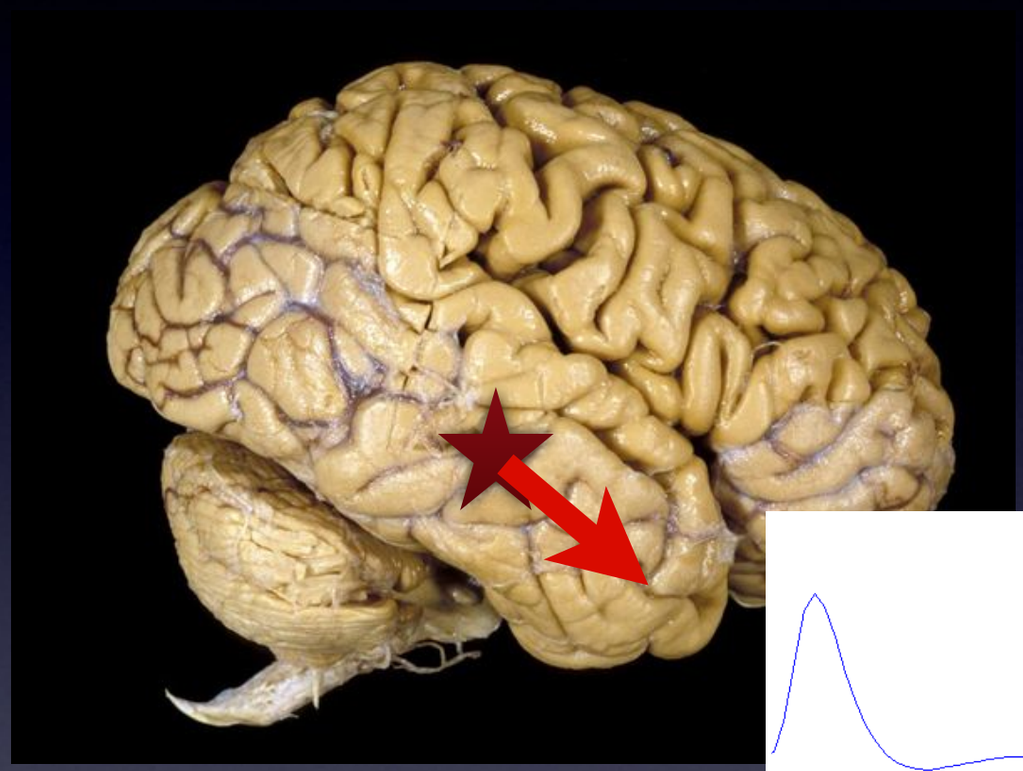


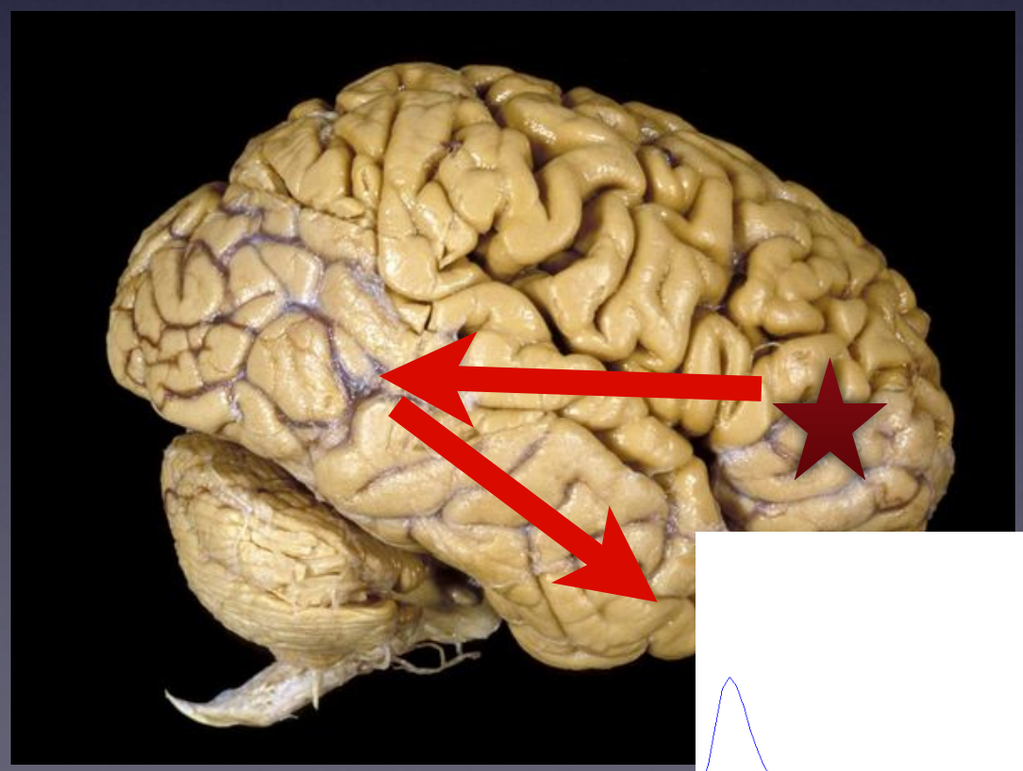

Copyrighted Material
Networks of the Brain
Olaf Sporns

Claim 2

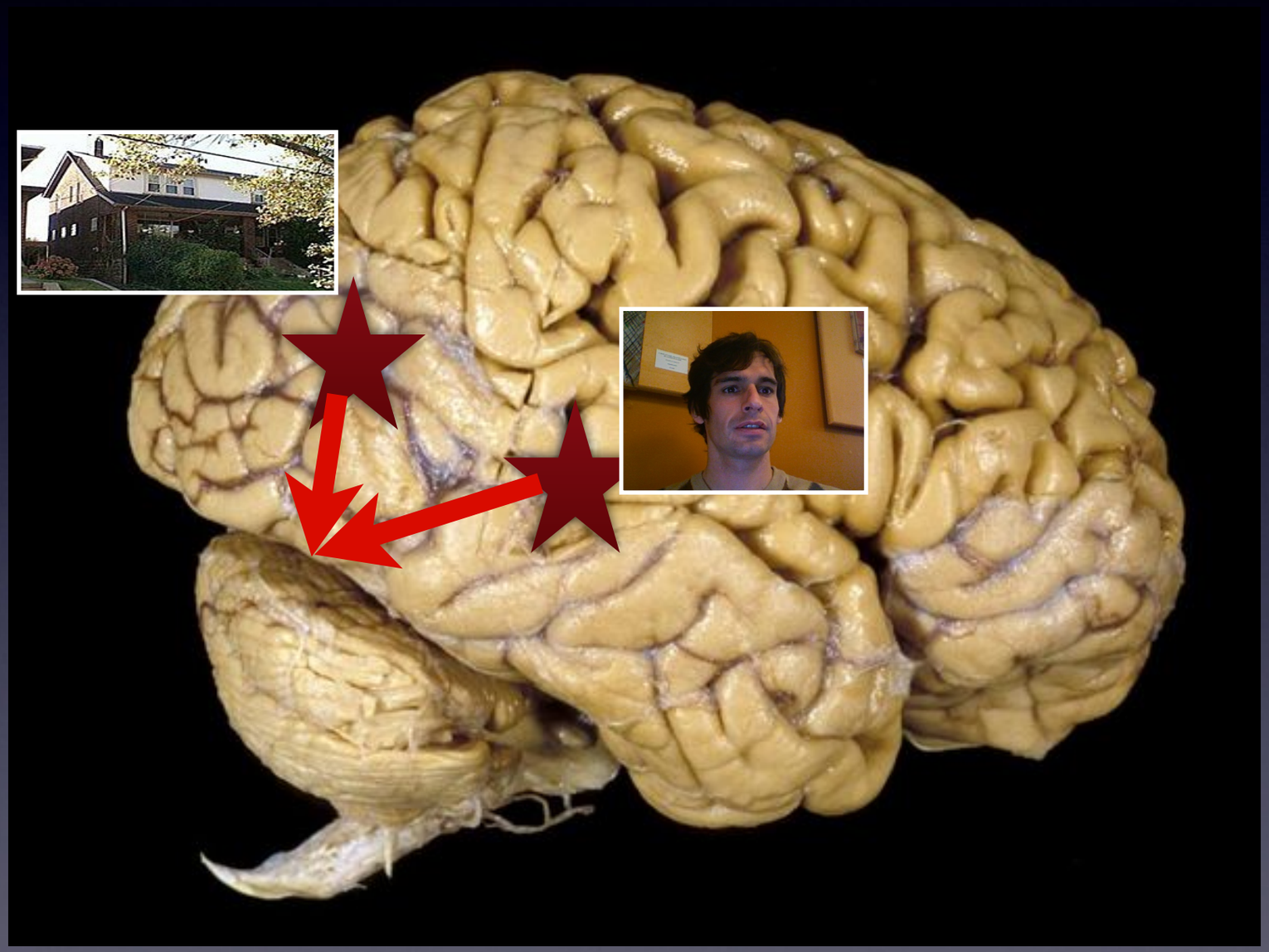

A
Cortical Region (Source)
Cortical Region (Target)
M1
M2
B
Frequency
Degree
0
1
2
3
5
10
15
20
25
30
35
40







Response II



Upshot

NHST is used to test functional hypotheses about brain areas or networks because (1) cognitive neuroscientists are really testing approximate nulls and (2) because the change in the BOLD signal in a voxel is not trivial only when the functional hypothesis is true.

Take-Home Message

Despite the brain's causal density, the role played by significance tests in neuroimagery does not prevent neuroscientists from using brain imagery data to test functional hypotheses about brain areas or networks.

If Interested

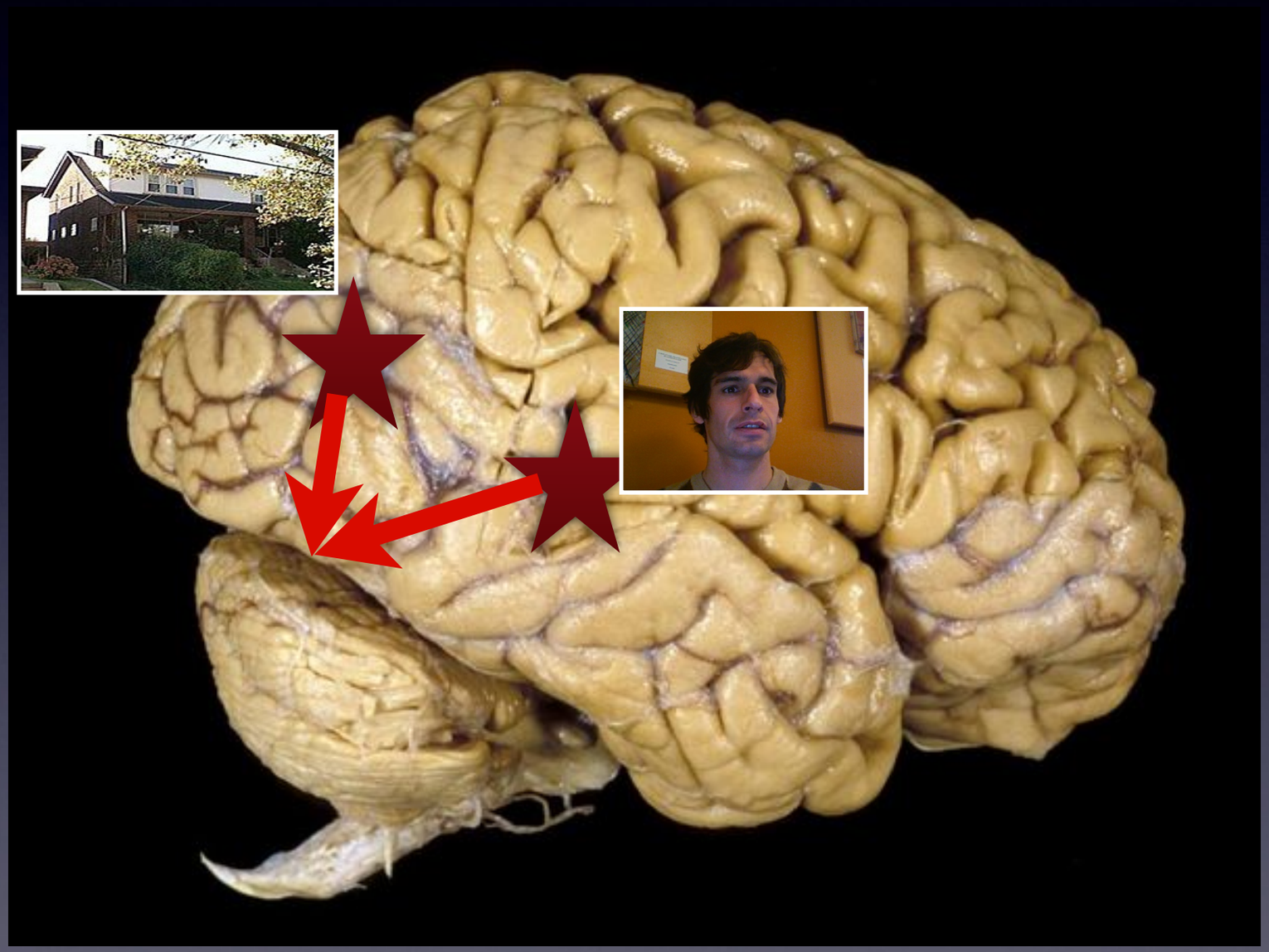
- Machery, E. Forthcoming. Significance testing in neuroimagery. In M. Sprevak and J. Kallerstrup (Eds.), *New Waves in Philosophy of Mind*.


APPLAUSE

A Tempting Response

Is it really true that null hypotheses are always false given that *contrasts* (viz. differences in changes in the BOLD signal) are tested?

Toy Example



Rejoinder to Reply I

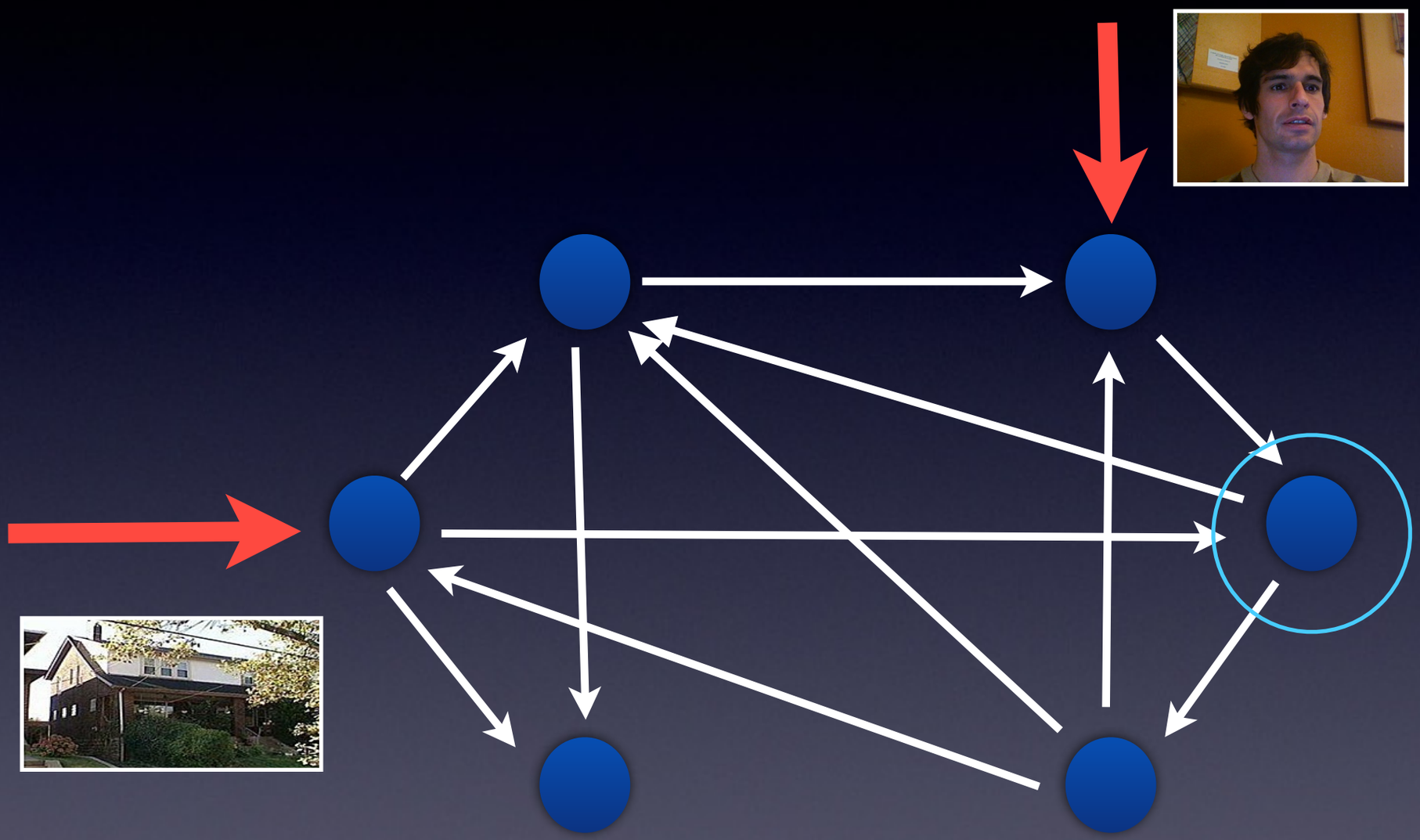


Figure 2.6

An example of an adjacency matrix and a degree distribution. (A) The adjacency matrix records the presence (black square) and absence (white square) of corticocortical connections between regions of the macaque cortex (Honey et al., 2007; Sporns et al., 2007). Many of the connections are symmetrical, and two main modules, corresponding to mostly visual (M1) and mostly somatomotor regions (M2), are indicated in the anatomical surface plot at the upper right. (B) The degree distribution (indegree plus outdegree for each node) is broad, with degrees ranging from 3 to 42. Area abbreviations (after Felleman and Van Essen, 1991): VP, ventral posterior; V1, visual area 1; MT, middle temporal; V3, visual area 3; V2, visual area 2; MSTd, medial superior temporal (dorsal); MSTl, medial superior temporal (lateral); V4, visual area 4; DP, dorsal prelunate; LIP, lateral intraparietal; VIP, ventral intraparietal; FEF, frontal eye field; FST, floor of superior temporal; PO, parieto-occipital; PIP, posterior intraparietal; V3A, visual area V3A; V4t, V4 transitional; AITv, anterior inferotemporal (ventral); PITv, posterior inferotemporal (ventral); CITv, central inferotemporal (ventral); CITd, central inferotemporal (dorsal); PITd, posterior inferotemporal (dorsal); VOT, ventral occipitotemporal; MDP, medial dorsal parietal; MIP, medial intraparietal; 46, area 46; 7a, area 7a; 5, area 5; 7b, area 7b; 6, area 6; AITd, anterior inferotemporal (dorsal); STPa, superior temporal polysensory (anterior); Ig, insular cortex (granular); STPp, superior temporal polysensory (posterior); TF, TF; TH, TH; 2, area 2; 4, area 4; 1, area 1; SII, secondary somatosensory area; SMA, supplemental motor area; 3a, area 3a; 3b, area 3b; Ri, retroinsular cortex; 35, area 35; 36, area 36; Id, insular cortex.