

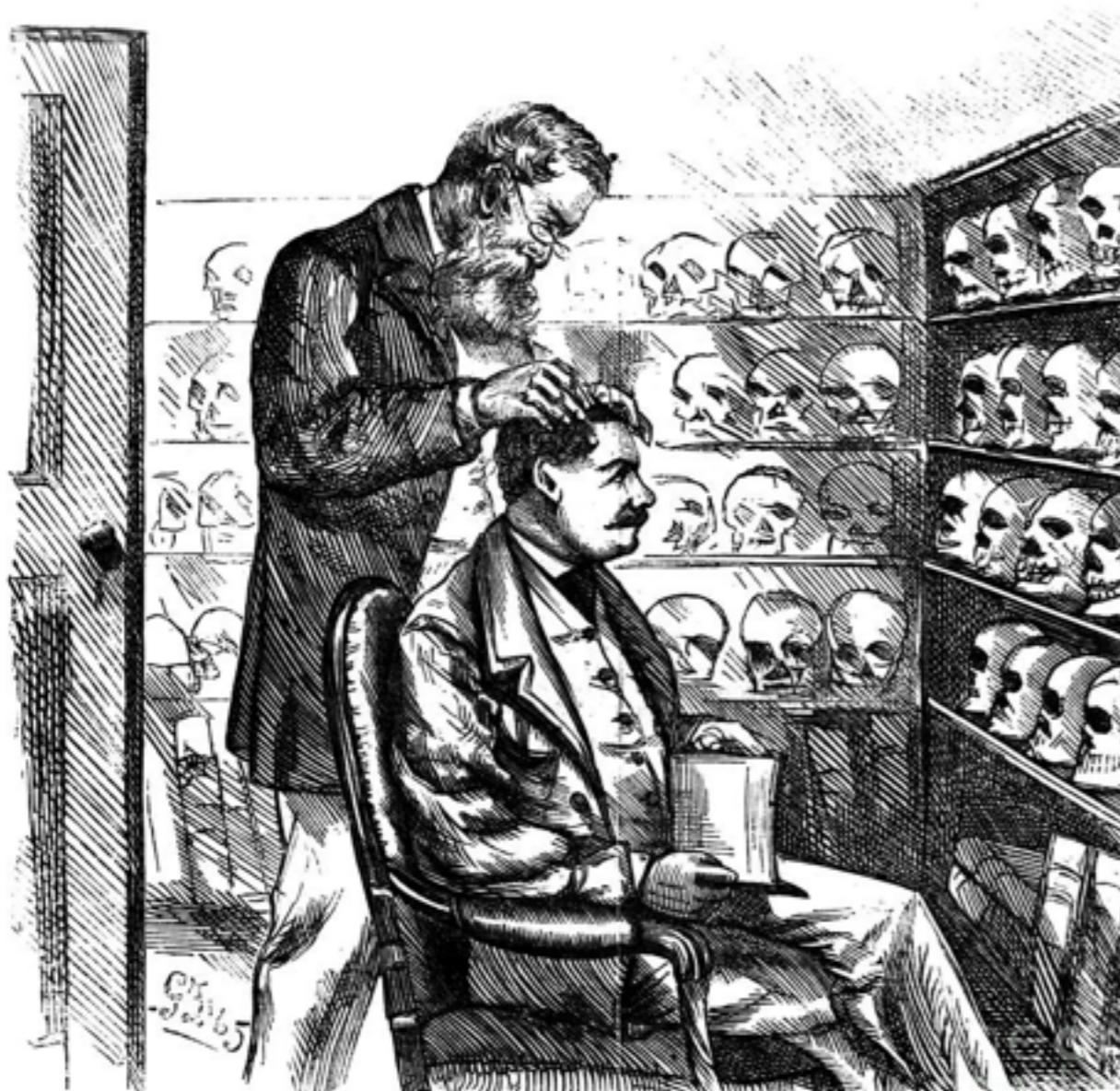
Improving the Reproducibility of Neuroimaging Research

Russell Poldrack

Department of Psychology
Stanford University



How confident are we in our scientific approach?



Science in crisis (?)

Open access, freely available online

298 | NATURE | VOL 485 | 17 MAY 2012

BAD COPY

IN THE WAKE OF HIGH-PROFILE CONTROVERSIES, PSYCHOLOGISTS ARE FACING UP TO PROBLEMS WITH REPLICATION.

BY ED YONG

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

PLoS Medicine | www.plosmedicine.org

0696

August 2005 | Volume 2 | Issue 8 | e124

Raise standards for preclinical cancer research

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

29 MARCH 2012 | VOL 483 | NATURE | 531

The Economist
World politics Business & finance Economics Science & technology Culture

Problems with scientific research

How science goes wrong

Scientific research has changed the world. Now it needs to change itself

Oct 19th 2013 | From the print edition

Like 119 Tweet 1,365



Why does reproducibility matter?

Growth in a Time of Debt

By CARMEN M. REINHART AND KENNETH S. ROGOFF*

American Economic Review: Papers & Proceedings 100 (May 2010): 573–578
<http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.2.573>

Reinhart & Rogoff have clearly exerted a major influence in recent years on public policy debates over the management of government debt and fiscal policy more broadly. Their findings have provided significant support for the austerity agenda that has been ascendant in Europe and the United States since 2010. - Herndon et al., 2013

Is the evidence for austerity based on an Excel spreadsheet error?

By Brad Plumer April 16, 2013

“Reinhart and Rogoff appear to have made an error with one of their Excel spreadsheet formulas. By typing AVERAGE(L30:L44) at one point instead of AVERAGE(L30:L49), they left out Belgium, a key counterexample [to their claim]”

Debt, Growth and the Austerity Debate

By CARMEN M. REINHART and KENNETH S. ROGOFF APRIL 25, 2013

Last week, three economists at the University of Massachusetts, Amherst, released a [paper](#) criticizing our findings. They correctly identified a spreadsheet coding error that led us to miscalculate the growth rates of highly indebted countries since World War II.

Neuroimaging: a perfect storm for irreproducibility



Why Most Published Research Findings Are False

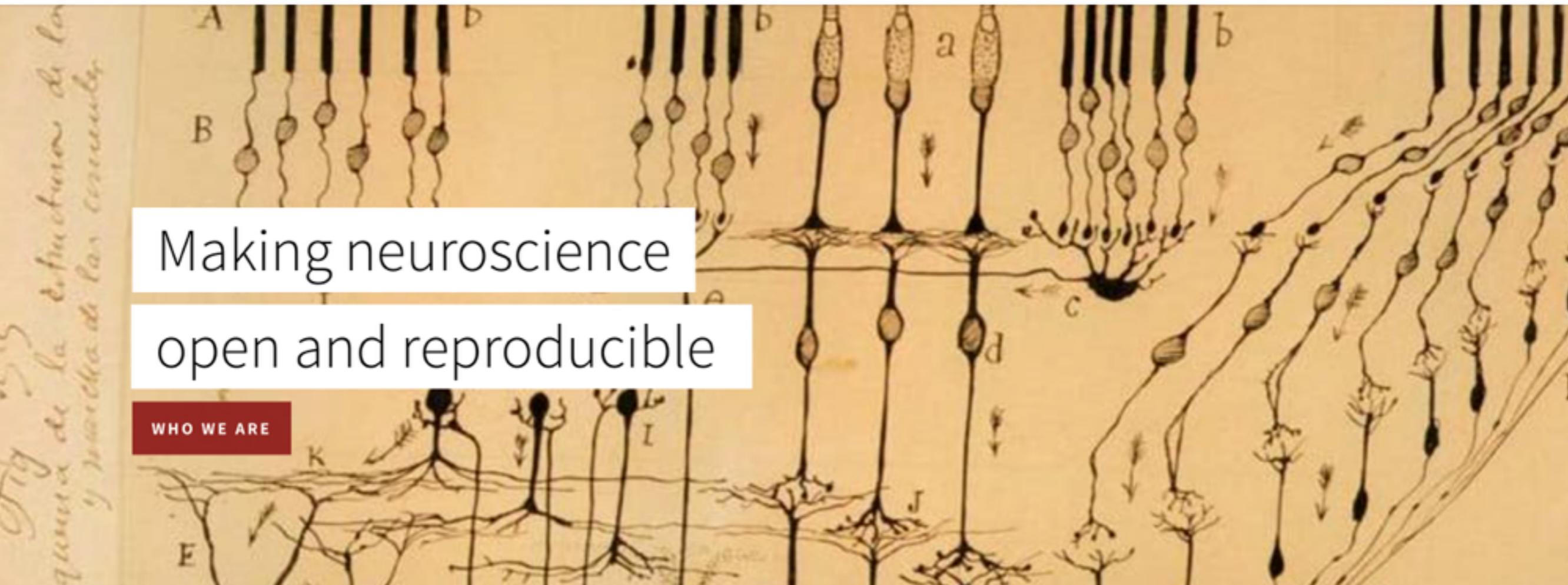
John P. A. Ioannidis

 PLoS Medicine | www.plosmedicine.org

0696

August 2005 | Volume 2 | Issue 8 | e124

- Research findings are less likely to be true:
 - The smaller the studies conducted in a scientific field
 - The smaller the effect sizes in a scientific field
 - The greater the number and the lesser the selection of tested relationships in a scientific field
 - The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field
 - The greater the financial and other interests and prejudices in a scientific field
 - The hotter a scientific field (with more scientific teams involved)



Making neuroscience open and reproducible

WHO WE ARE

Reproducibility matters

Neuroscience research is the basis for critical decisions about health and society. Our first goal as researchers is to ensure that the results of our research will stand the test of time.

Enabling better research

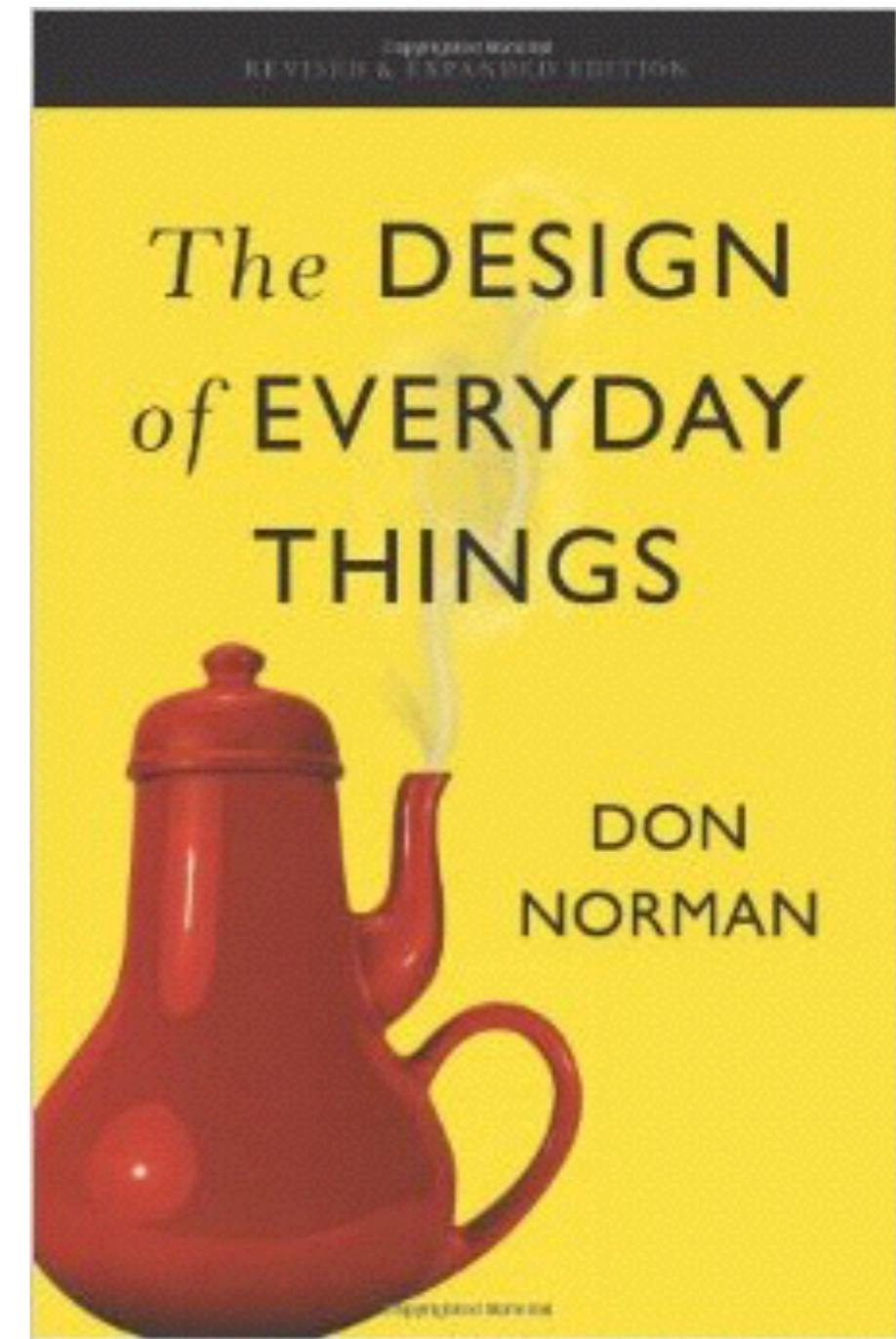
We are expanding the OpenfMRI project into a free and open platform that will enable the analysis and sharing of neuroimaging data, harnessing the power of high-performance computing to improve the quality of research.

From data to discovery

Our platform will provide neuroimaging researchers with leading-edge tools to analyze and share large datasets, with a focus on quantifying the reproducibility of the results.

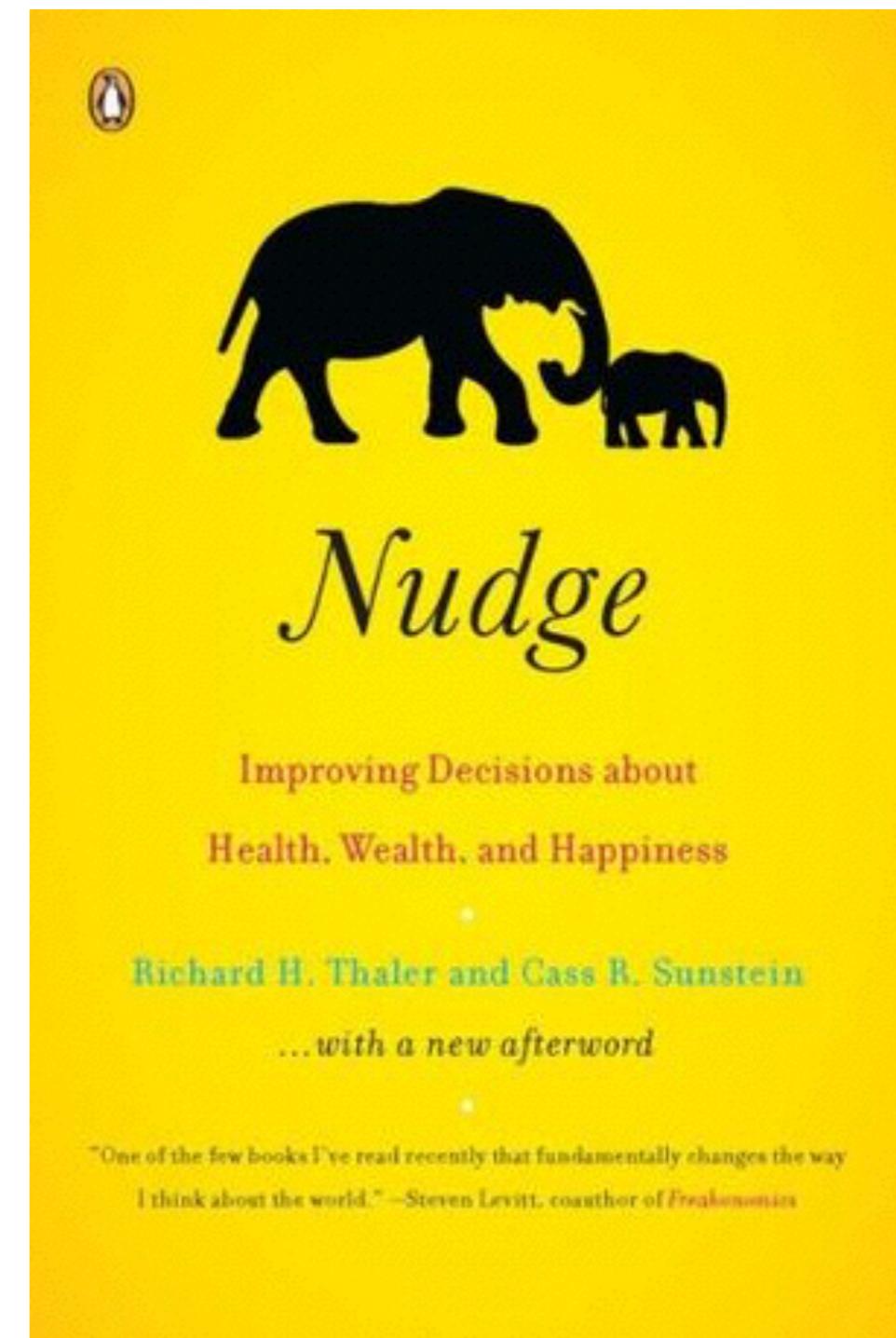
<http://reproducibility.stanford.edu>

Designing a more reproducible scientific enterprise



Improving the choice architecture of science

- Choice architecture
 - particular set of features that drive people toward or away from particular choices
- Nudges
 - Improving incentives
 - Using the power of defaults
 - Providing feedback
 - Expecting and prevent errors



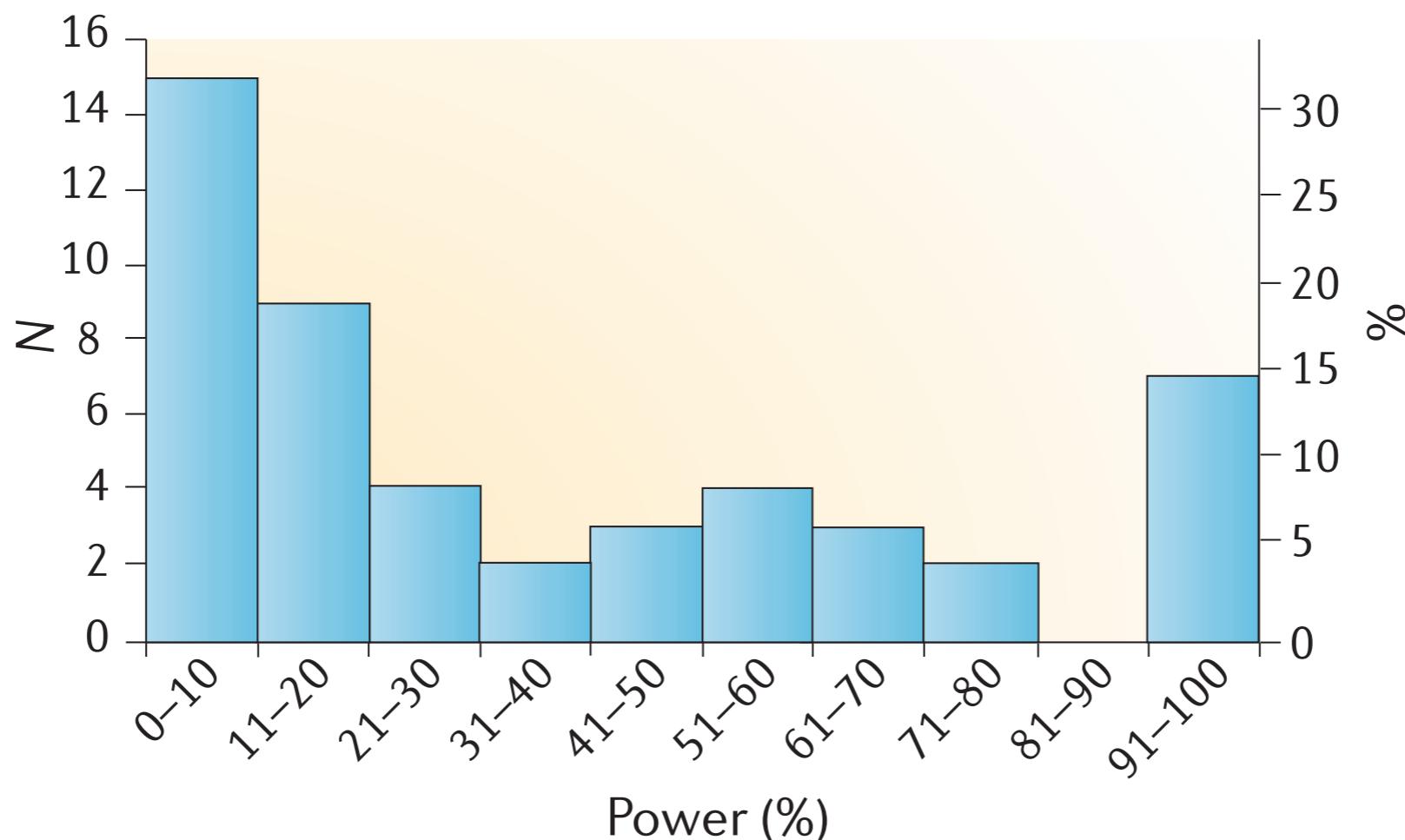
Threats to reproducibility: Low power

Power failure: why small sample size undermines the reliability of neuroscience

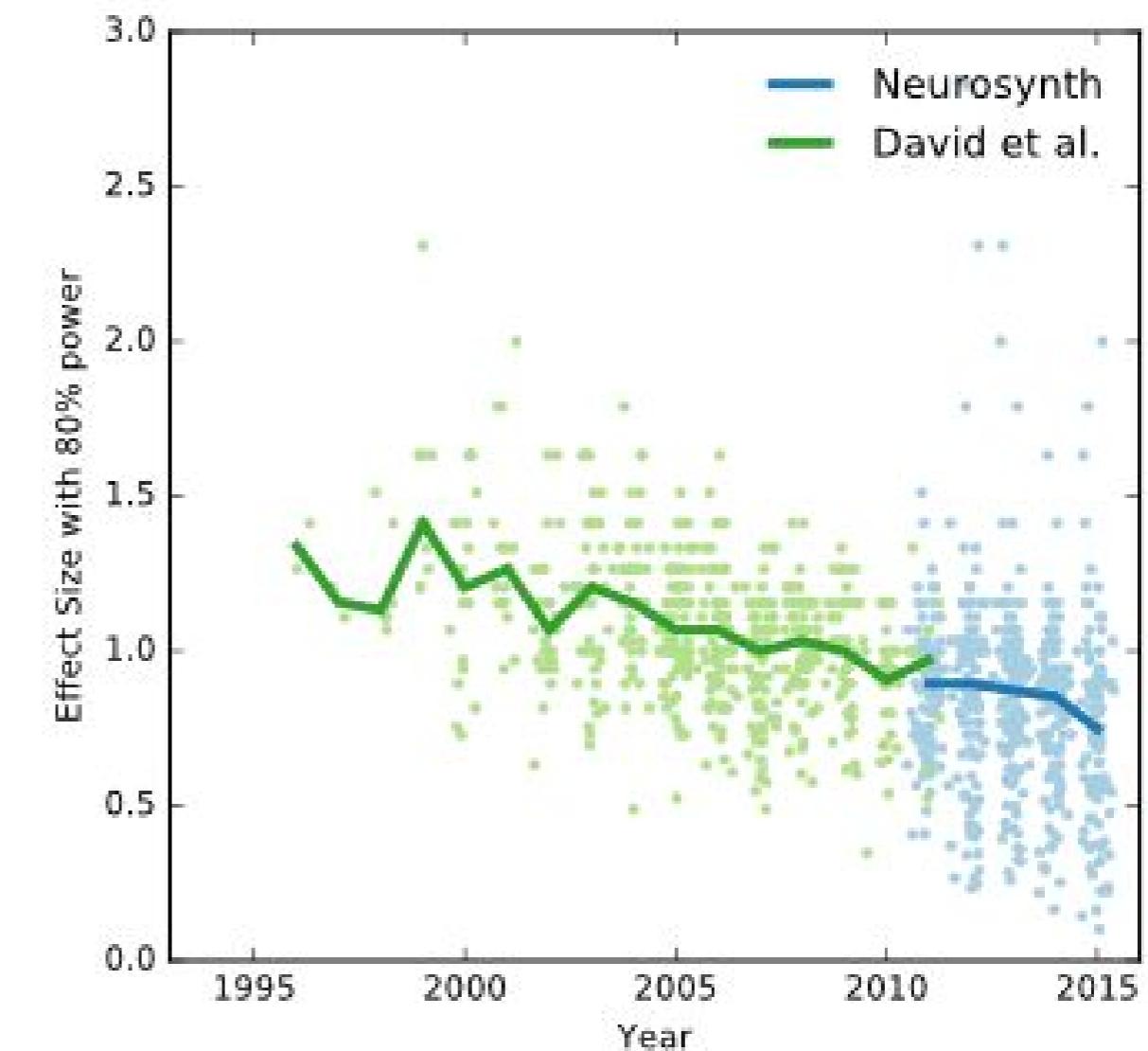
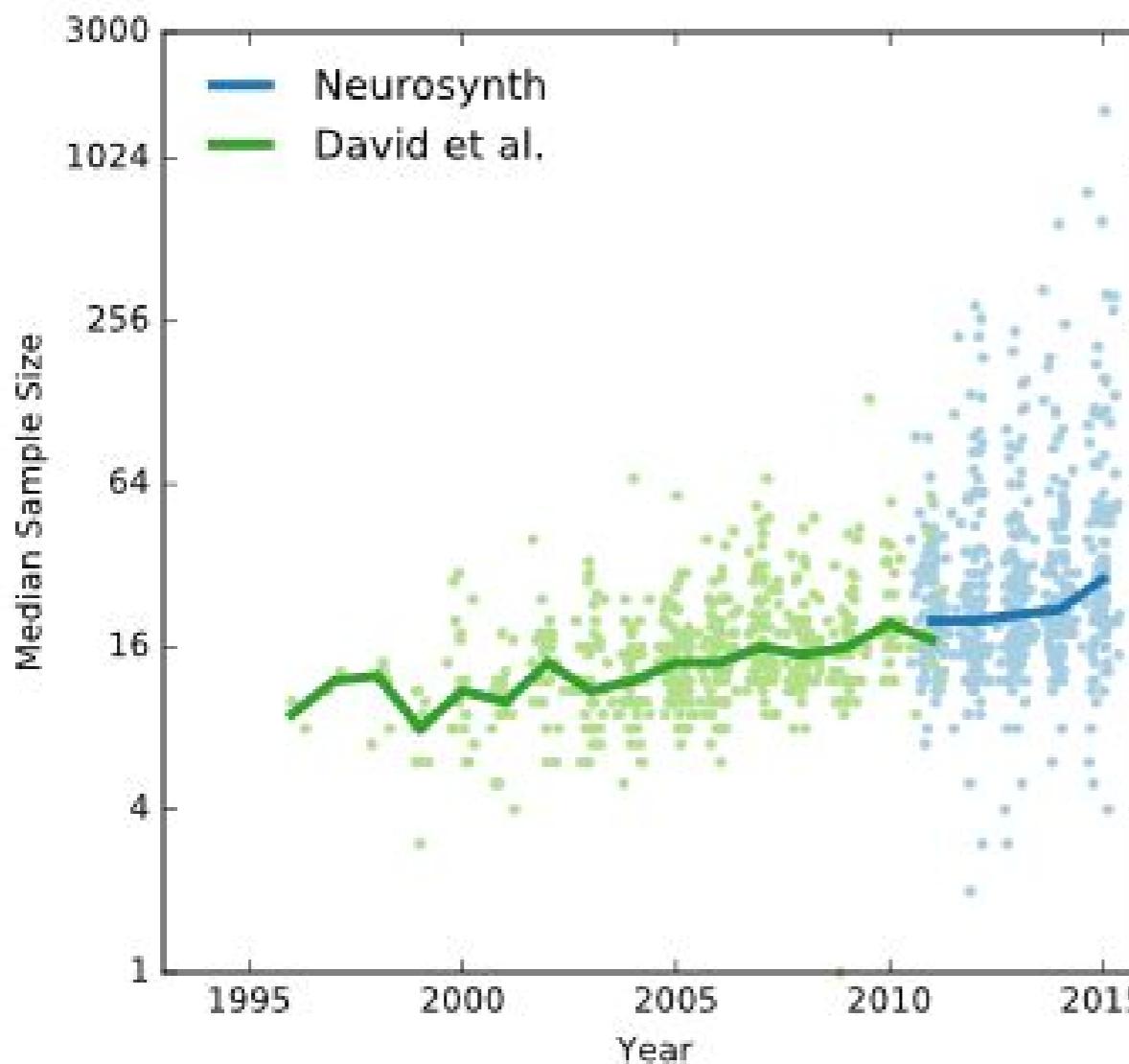
Katherine S. Button^{1,2}, John P. A. Ioannidis³, Claire Mokrysz¹, Brian A. Nosek⁴, Jonathan Flint⁵, Emma S. J. Robinson⁶ and Marcus R. Munafò¹

NATURE REVIEWS | NEUROSCIENCE

VOLUME 14 | MAY 2013 | 365



Sample size and power in fMRI studies

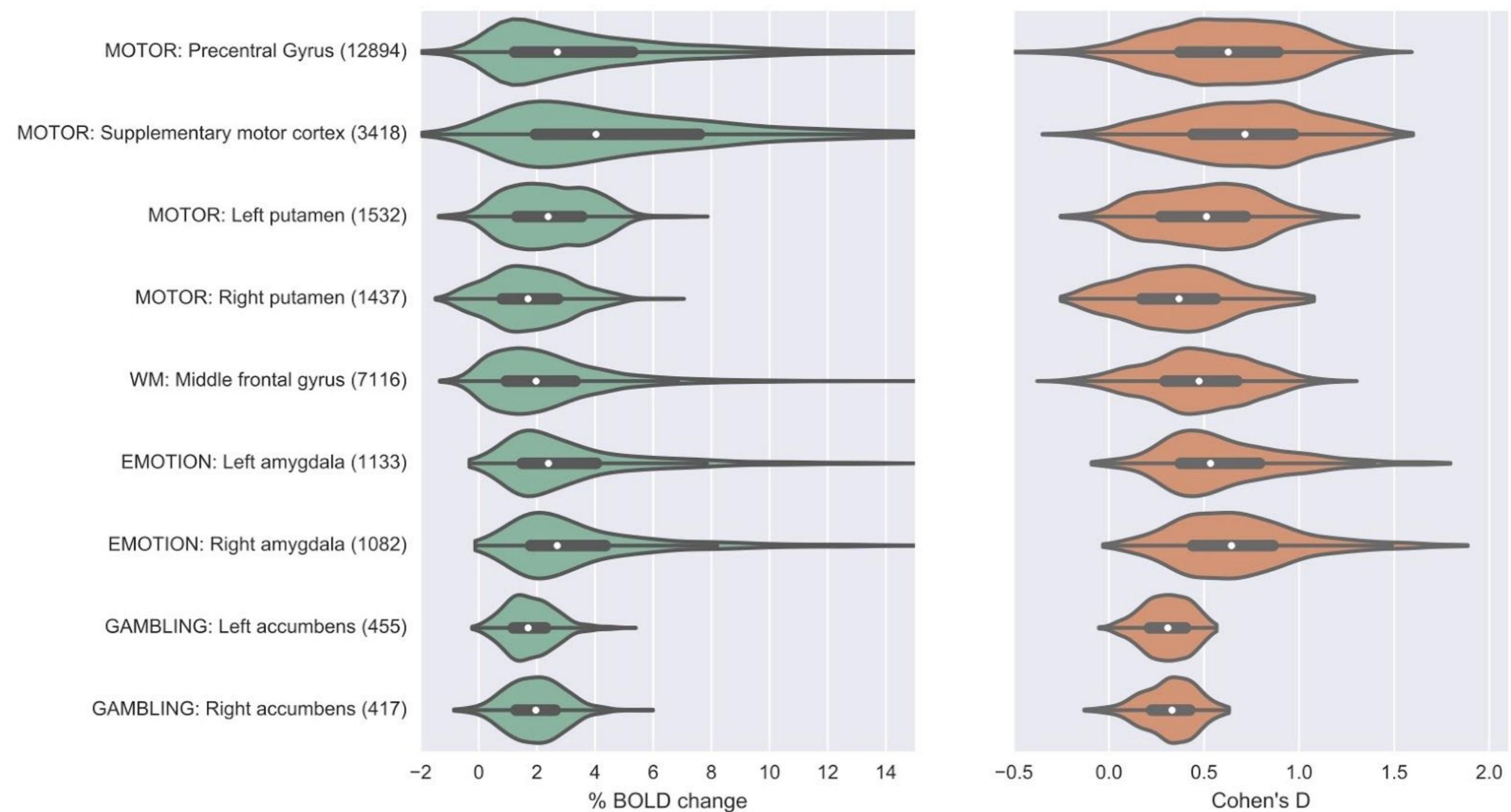


Median estimated sample size in 2015 = 28.5
Median effect size with 80% power = 0.75

Thanks to Sean David and Tal Yarkoni
for sample size data

[https://github.com/poldracklab/
ScanningTheHorizon](https://github.com/poldracklab/ScanningTheHorizon)

What are realistic effect sizes for fMRI?



Estimated from HCP task data
using combined anatomical + neurosynth ROIs

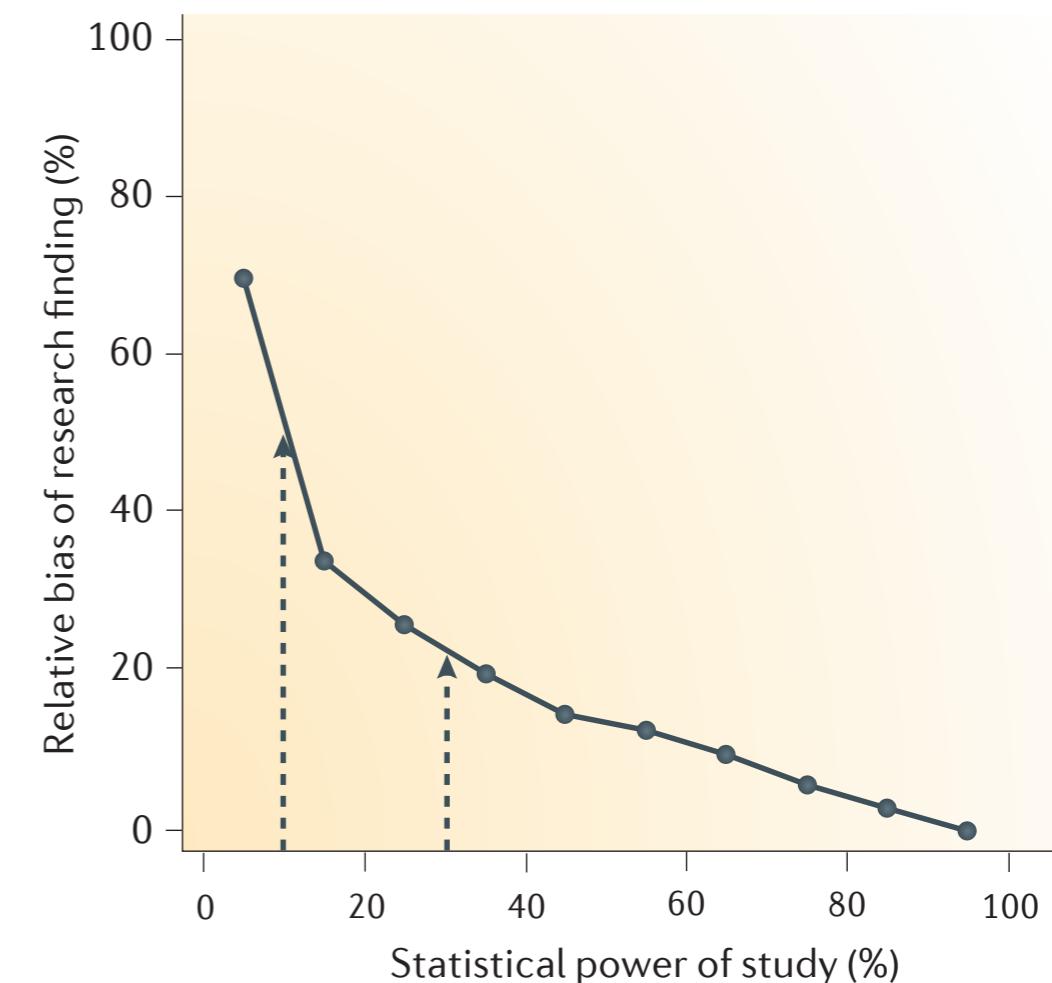
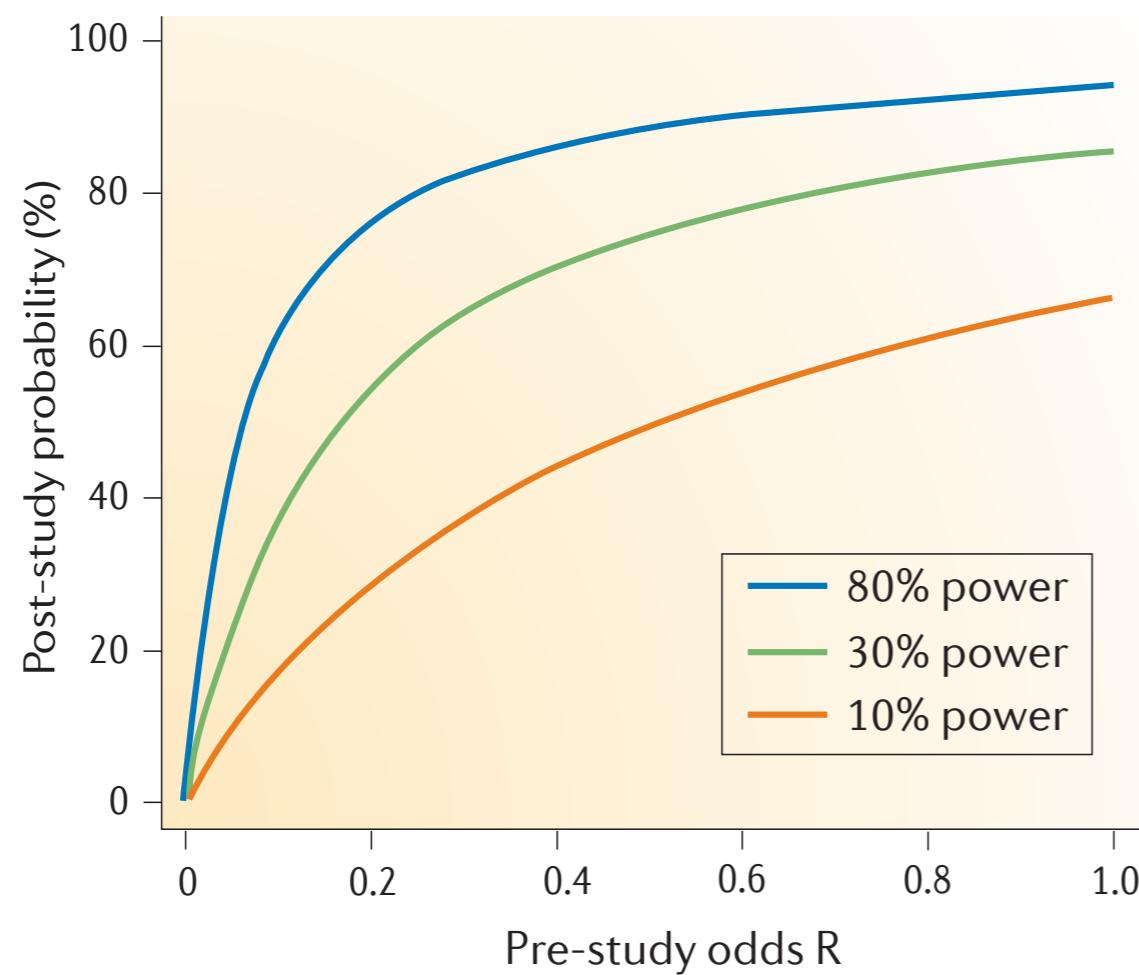
[https://github.com/poldracklab/
ScanningTheHorizon](https://github.com/poldracklab/ScanningTheHorizon)

Low power -> unreliable science

Positive Predictive Value (PPV): The probability that a positive result is true

Winner's Curse: overestimation of effect sizes for significant results

$$PPV = ([1 - \beta] \times R) / ([1 - \beta] \times R + \alpha)$$



Button et al., 2013

- “My result isn’t significant, so I need to add more subjects...”

Sample size flexibility

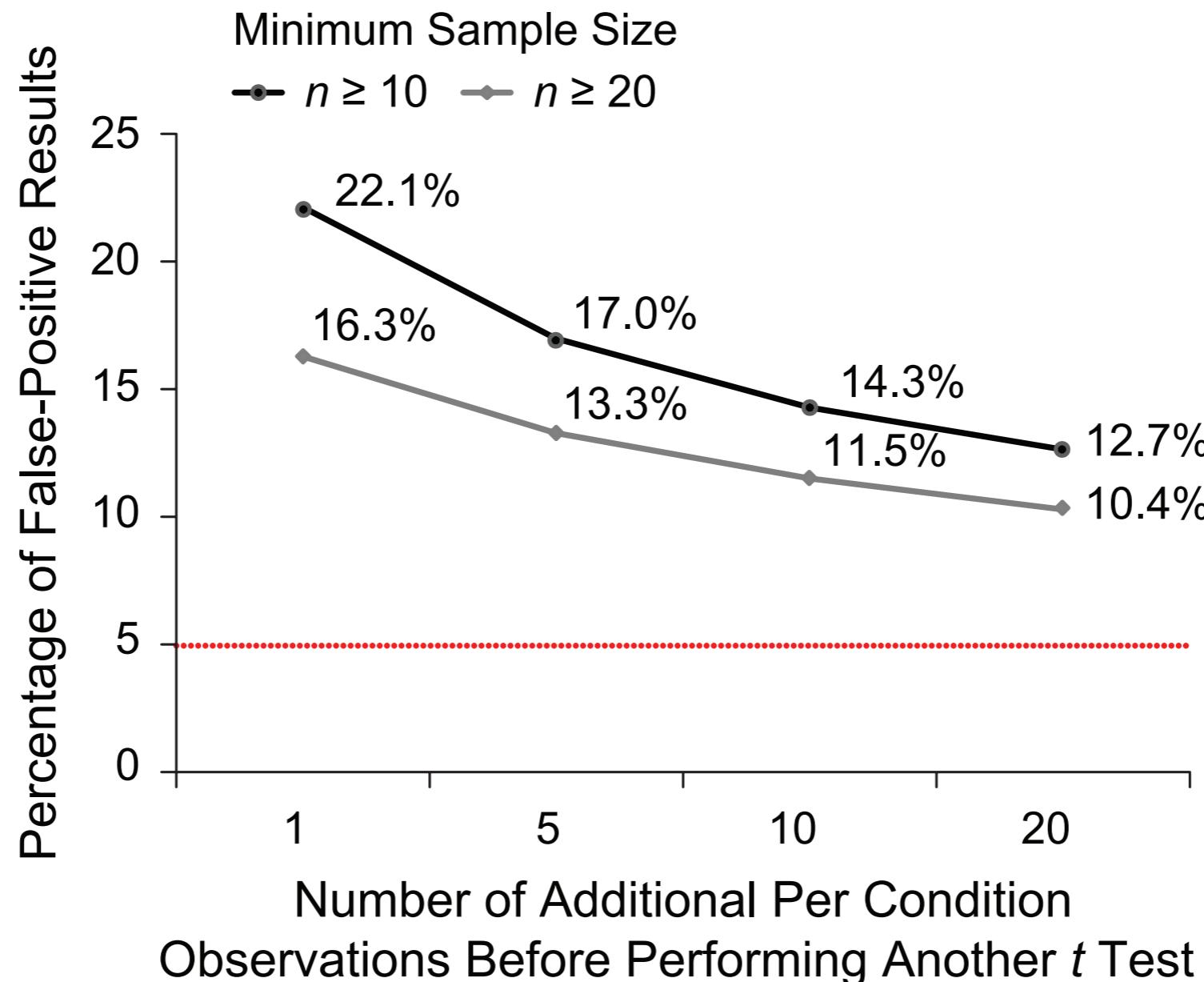


Fig. I. Likelihood of obtaining a false-positive result when data collection ends upon obtaining significance ($p \leq .05$, highlighted by the dotted line). The figure depicts likelihoods for two minimum sample sizes, as a function of the frequency with which significance tests are performed.

-Simmons et al., 2011, Psychological Science

Improvement: always predetermine sample size

neuropowertools.org

NeuroPower

1. Load data 2. Estimate model 3. Power

Select your statistical parametric map for a certain contrast (T or Z) in nifti format (.nii, NOT .nii.gz).

Bladeren... spmT_0001.nii
Upload complete

Are the values Z- or T-values?

T

What is your peakforming threshold?

units = p-value

0.01

How many subjects?

18

Is the study a one- or two-sample test?

One-sample

How do you want the smoothness to be defined?

Estimate from the data
 Manual input

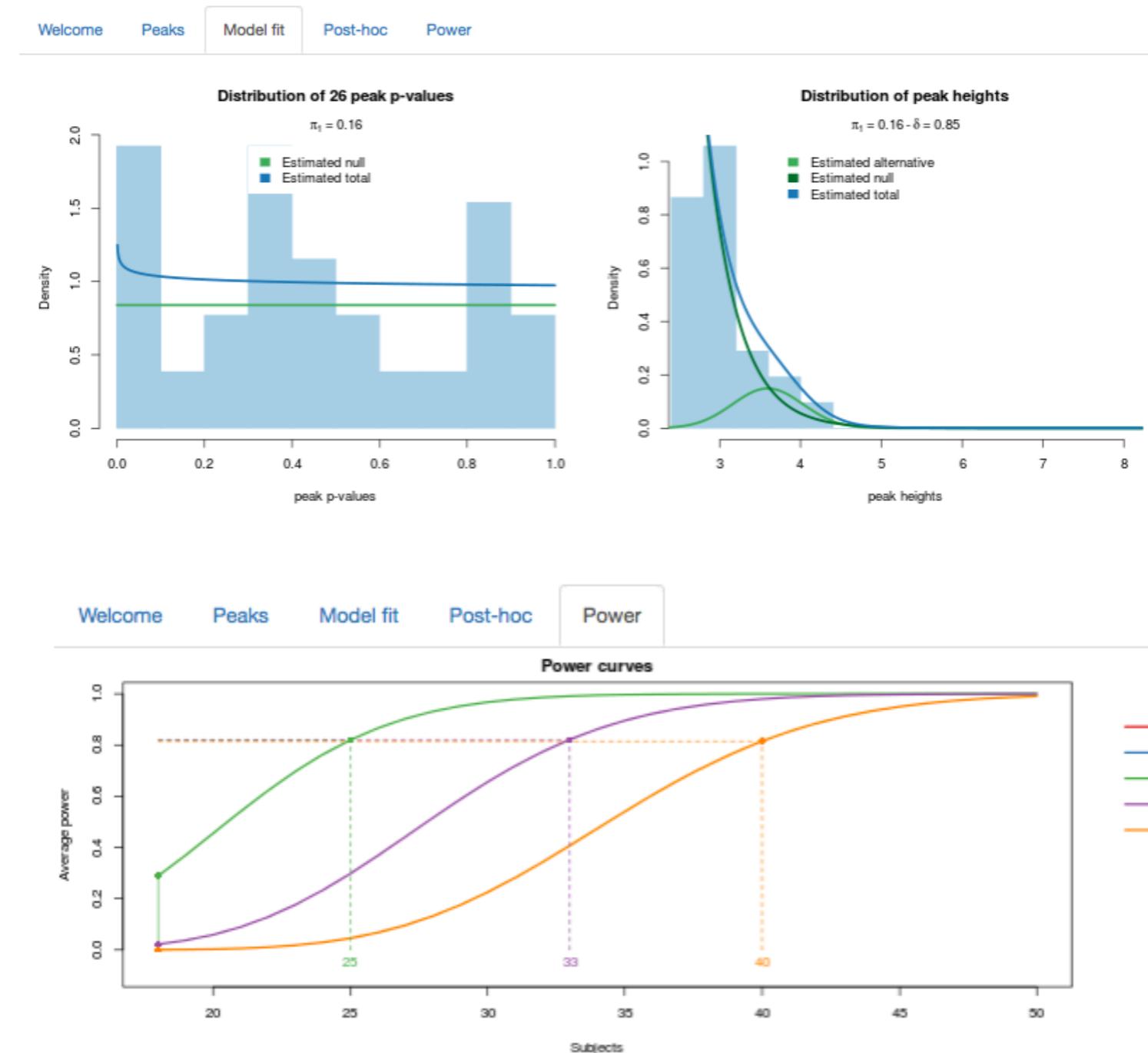
If manually: what is the FWHM in mm? (eg. '[8,8,8]')

[8,8,8]

If manually: What is the voxelsize? (eg. '[2,2,2.3]')

[3.9,3.9,4]

Extract peaks

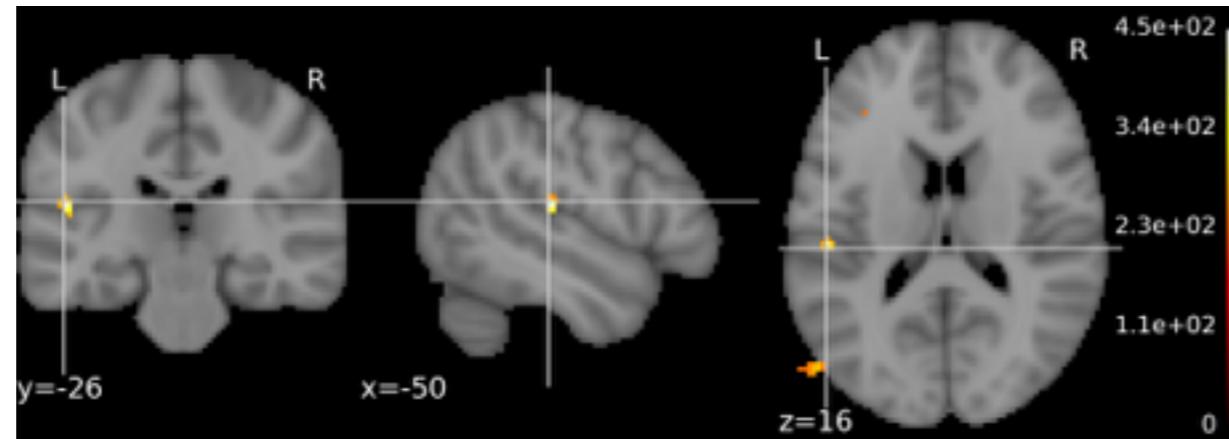


Joke Durnez

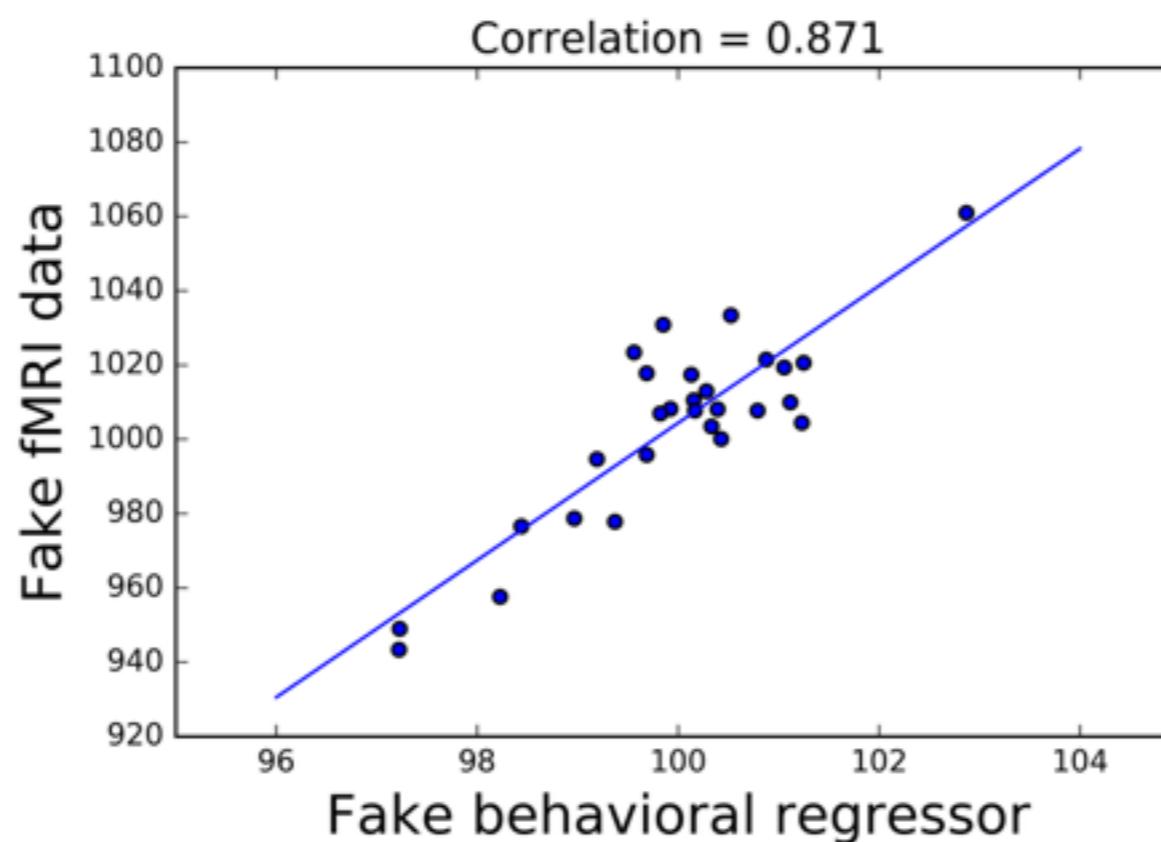
poldracklab.org

Threats to reproducibility: high dimensionality

Correlation between random simulated behavioral variable
and activation across 28 subjects



~220,000 voxels
 $p < 0.001$
10 voxels cluster threshold



<https://github.com/poldracklab/ScanningTheHorizon>

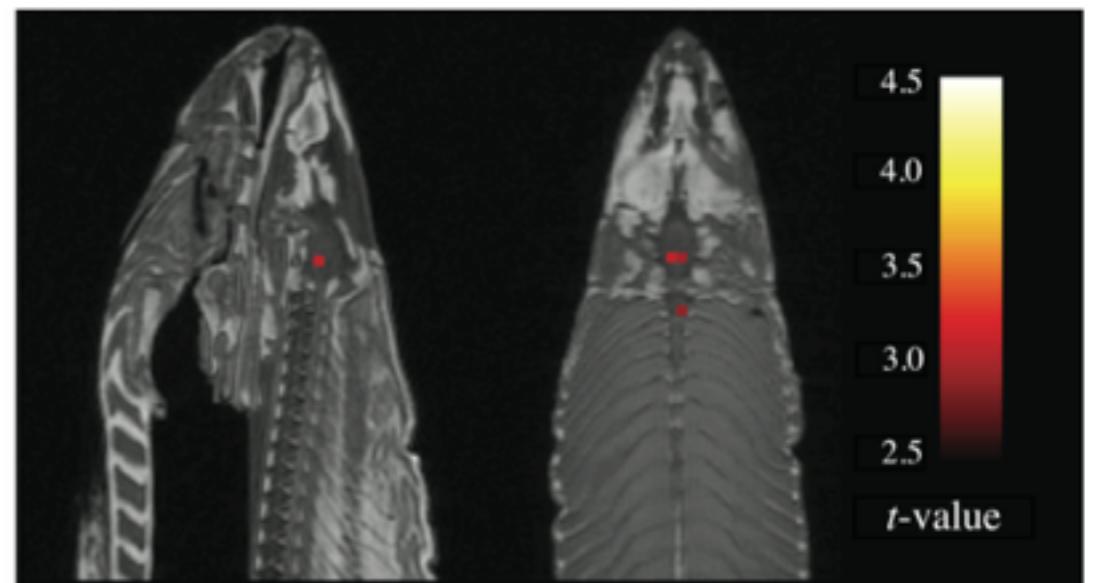
Need for statistical corrections

Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction

Craig M. Bennett¹, Abigail A. Baird², Michael B. Miller¹, and George L. Wolford³

Subject. One mature Atlantic Salmon (*Salmo salar*) participated in the fMRI study. The salmon was approximately 18 inches long, weighed 3.8 lbs, and was not alive at the time of scanning.

Task. The task administered to the salmon involved completing an open-ended mentalizing task. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.



A t -contrast was used to test for regions with significant BOLD signal change during the photo condition compared to rest. The parameters for this comparison were $t(131) > 3.15$, $p(\text{uncorrected}) < 0.001$, 3 voxel extent threshold.

Several active voxels were discovered in a cluster located within the salmon's brain cavity (Figure 1, see above). The size of this cluster was 81 mm^3 with a

Identical t -contrasts controlling the false discovery rate (FDR) and familywise error rate (FWER) were completed. These contrasts indicated no active voxels, even at relaxed statistical thresholds ($p = 0.25$).

Improvement: Insist on proper corrections

- Nonparametric approaches are generally best

Threats to reproducibility: Methodological flexibility

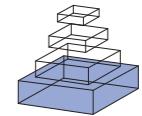
- Using standard FSL analysis options
- 69,120 possible analysis workflows

Processing step	Reason	Options	Number of plausible options
Motion correction	Correct for head motion during scanning	Interpolation [linear vs. sinc] Reference volume [single vs. mean]	4
Slice timing correction	Correct for differences in acquisition timing of different slices	No/before motion correction/after motion correction	3
Field map correction	Correct for distortion due to magnetic susceptibility	Yes/No	2
Spatial smoothing	Increase SNR for larger activations and ensure assumptions of Gaussian random field theory	FWHM [4/6/8 mm]	3
Spatial normalization	Warp individual brain to match a group template	Method [linear/nonlinear]	2
High pass filter	Remove low-frequency nuisance signals from data	Frequency cutoff [100, 120]	2
Head motion regressors	Remove remaining signals due to head motion via statistical model	Yes/No If Yes: 6/12/24 parameters or single timepoint “scrubbing” regressors	5
Hemodynamic response	Account for delayed nature of hemodynamic response to neuronal activity	Basis function [single-gamma, double-gamma] Derivatives [none/shift/dispersion]	6
Temporal autocorrelation model	Model for the temporal autocorrelation inherent in fMRI signals.	Yes/no	2
Multiple comparison correction	Correct for large number of comparisons across the brain	Voxel-based GRF, Cluster-based GRF, FDR, nonparametric	4
Total possible workflows			69,120

Threats to reproducibility: Methodological flexibility

frontiers in
NEUROSCIENCE

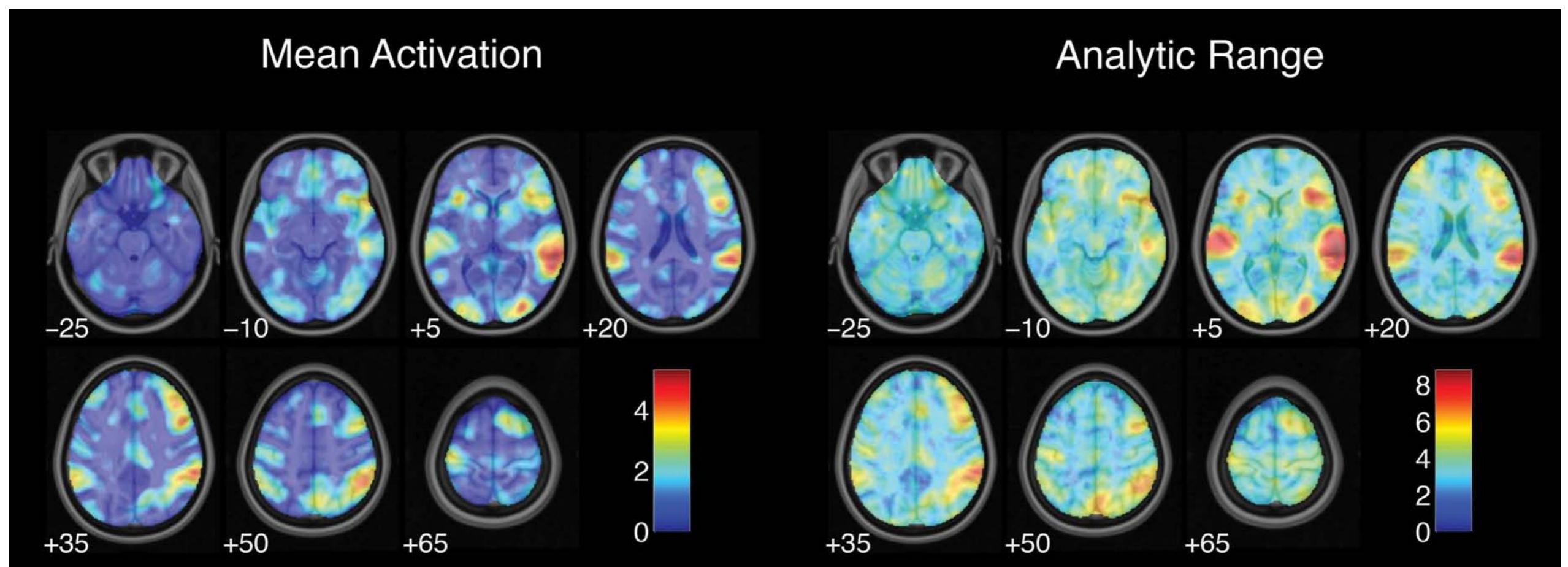
ORIGINAL RESEARCH ARTICLE
published: 11 October 2012
doi: 10.3389/fnins.2012.00149



On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments

Joshua Carp*

6,912 pipelines



Multiple comparison correction

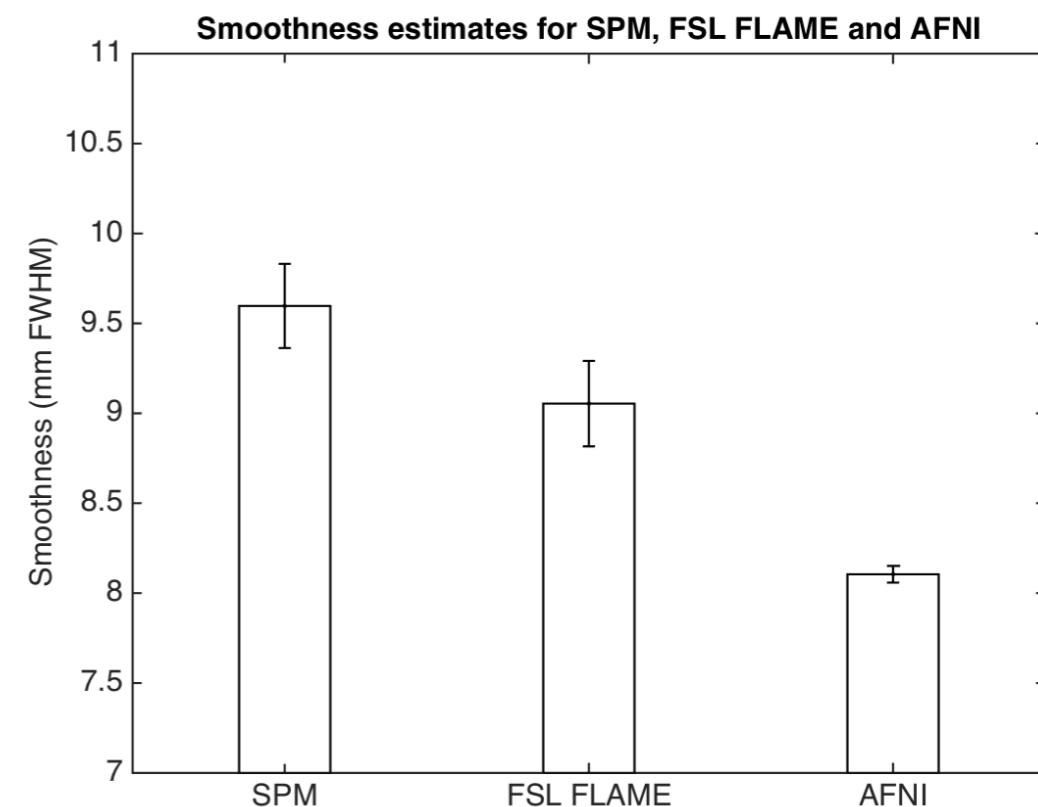
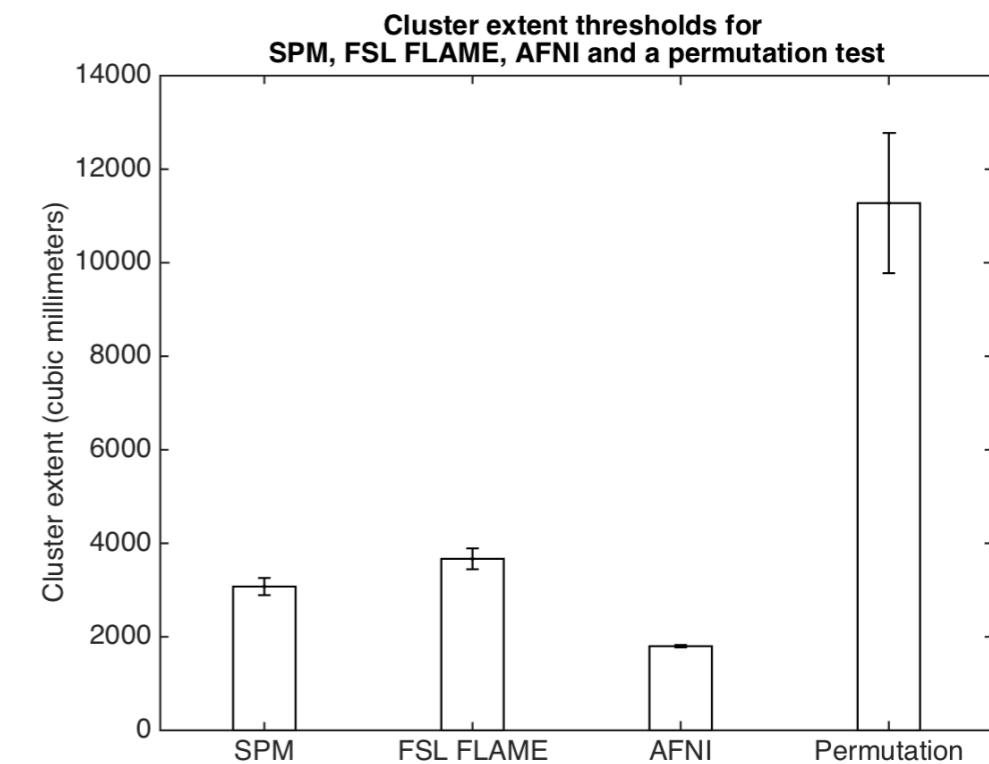
- Assessed latest 100 papers matching query for fMRI activation studies
 - 65 reported whole-brain activation data
 - Good news
 - only 3 papers reported uncorrected results
 - Bad news
 - 11% of papers analyzed data using SPM/FSL but then corrected for multiple comparisons using AFNI's alphasim/3dclustsim
 - Why is this a problem?

Poldrack et al., submitted

poldracklab.org

Communal p-hacking?

- Eklund et al. (2016, PNAS)
 - “a 15 year old bug was found in 3dClustSim while testing the three software packages (the bug was fixed by the AFNI group as of May 2015, during preparation of this manuscript). The effect of the bug was an underestimation of how likely it is to find a cluster of a certain size (in other words, the p-values reported by 3dClustSim were too low).”
 - AFNI also underestimated smoothness due to long tails

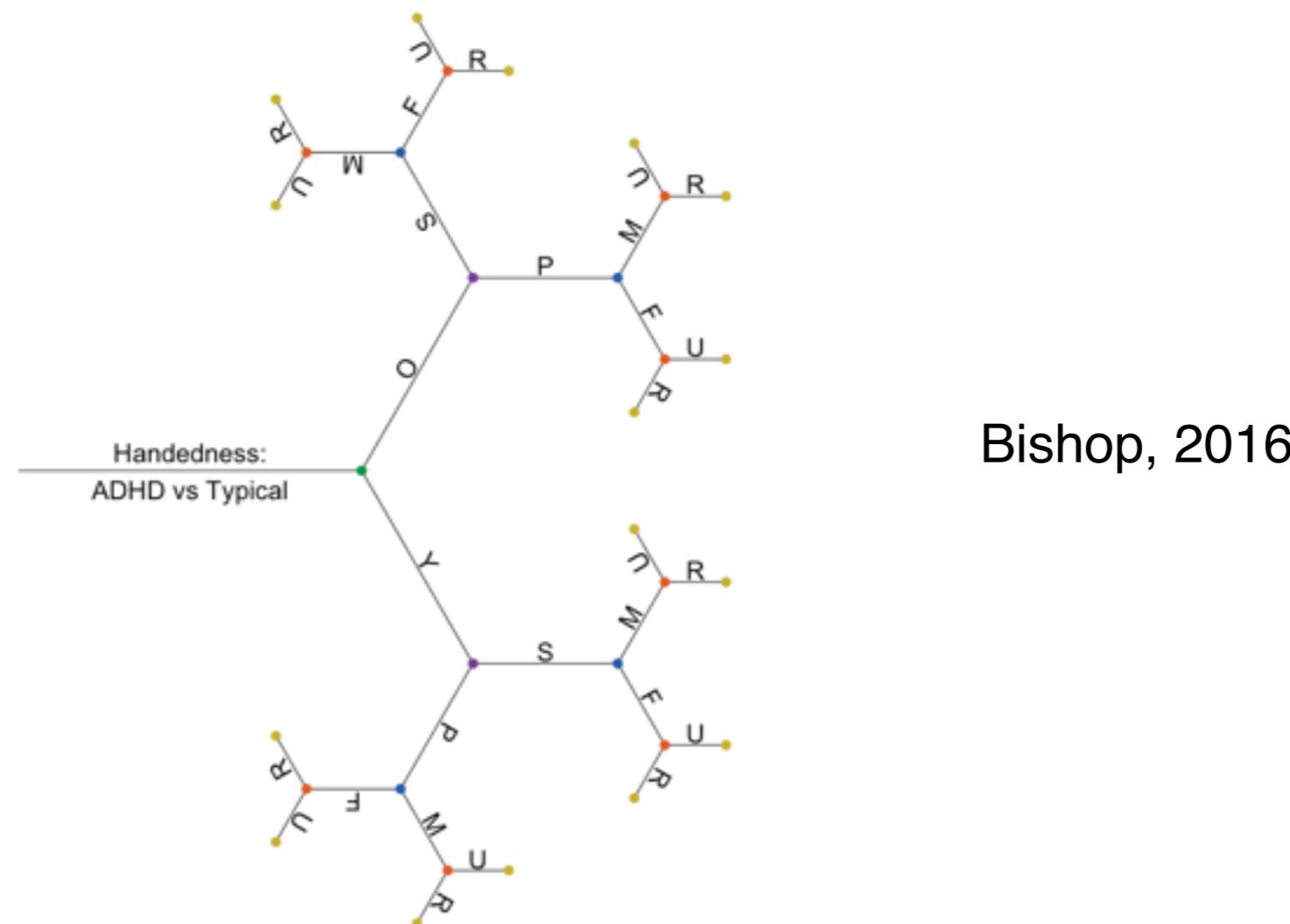


The garden of forking paths

The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time*

Andrew Gelman[†] and Eric Loken[‡]

14 Nov 2013



Improvement: Study pre-registration

- Register sample size and analysis plan up front
- This does not prevent exploratory analysis
 - But planned and exploratory analyses must be clearly delineated in the paper

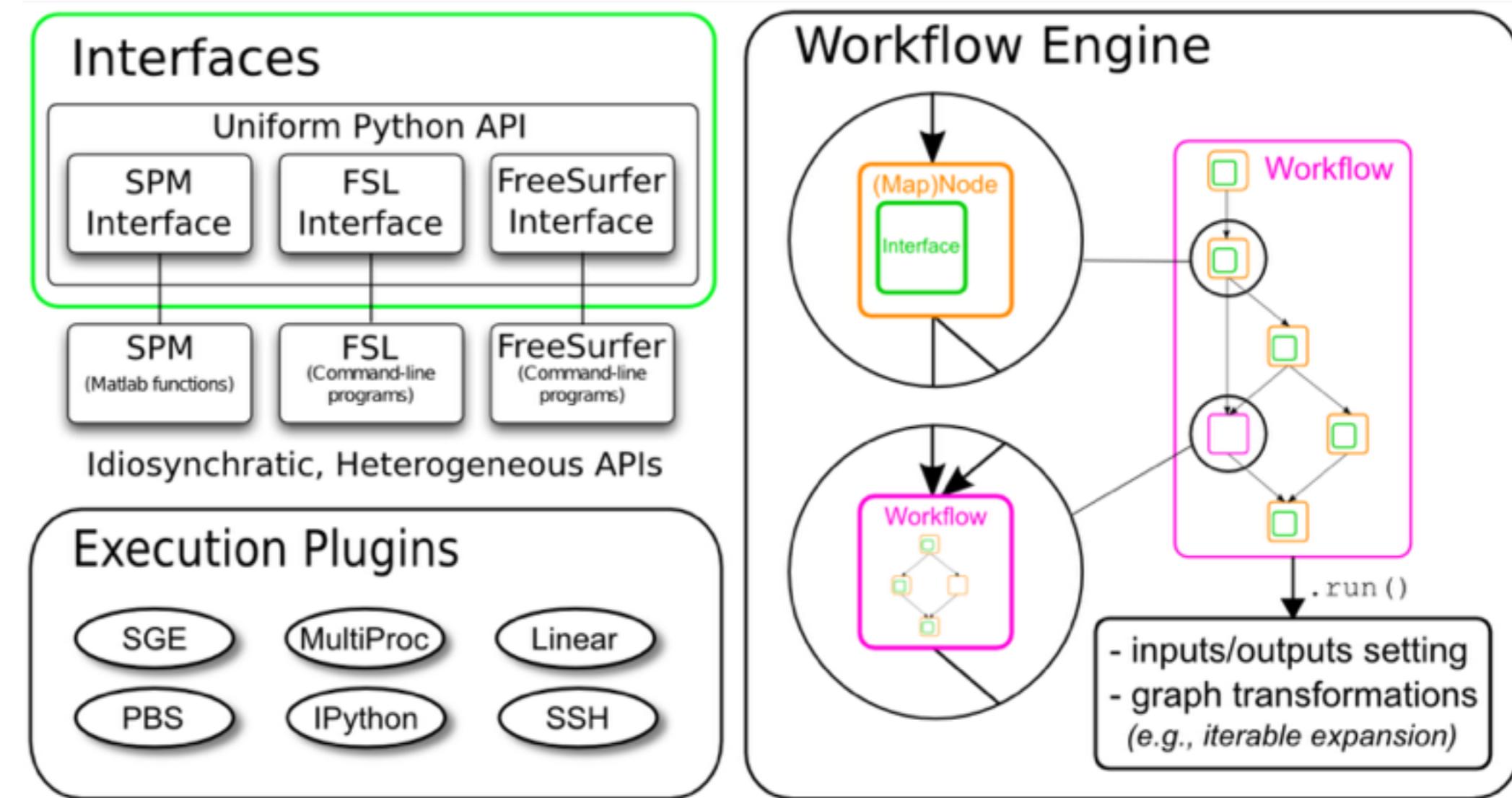
The screenshot shows the Open Science Framework interface. At the top, there's a dark header with the "Open Science Framework" logo and a "Browse" dropdown. Below the header, a blue navigation bar contains the project title "MRIQC: Predicting quality in manual MR..." and links for "Files", "Wiki", "Analytics", and "Forks". A light blue banner at the bottom of the navigation bar states: "This project is a registration of [this project](#); the content of the project has been frozen and cannot be edited." The main content area displays the project summary: "MRIQC: Predicting quality in manual MRI assessment protocols using no-reference image quality measures." Below the summary, contributor information is listed: "Contributors: Oscar Esteban, Chris Gorgolewski, Sanmi Koyejo, Russell Poldrack". Other details include: "Registration Supplement: Prereg Challenge", "Date registered: 2016-05-03 01:17 PM", "Date created: 2016-04-29 10:08 AM", and "Category: Project".

<http://www.russpoldrack.org/2016/09/why-preregistration-no-longer-makes-me.html>

Improvement: Mapping the garden

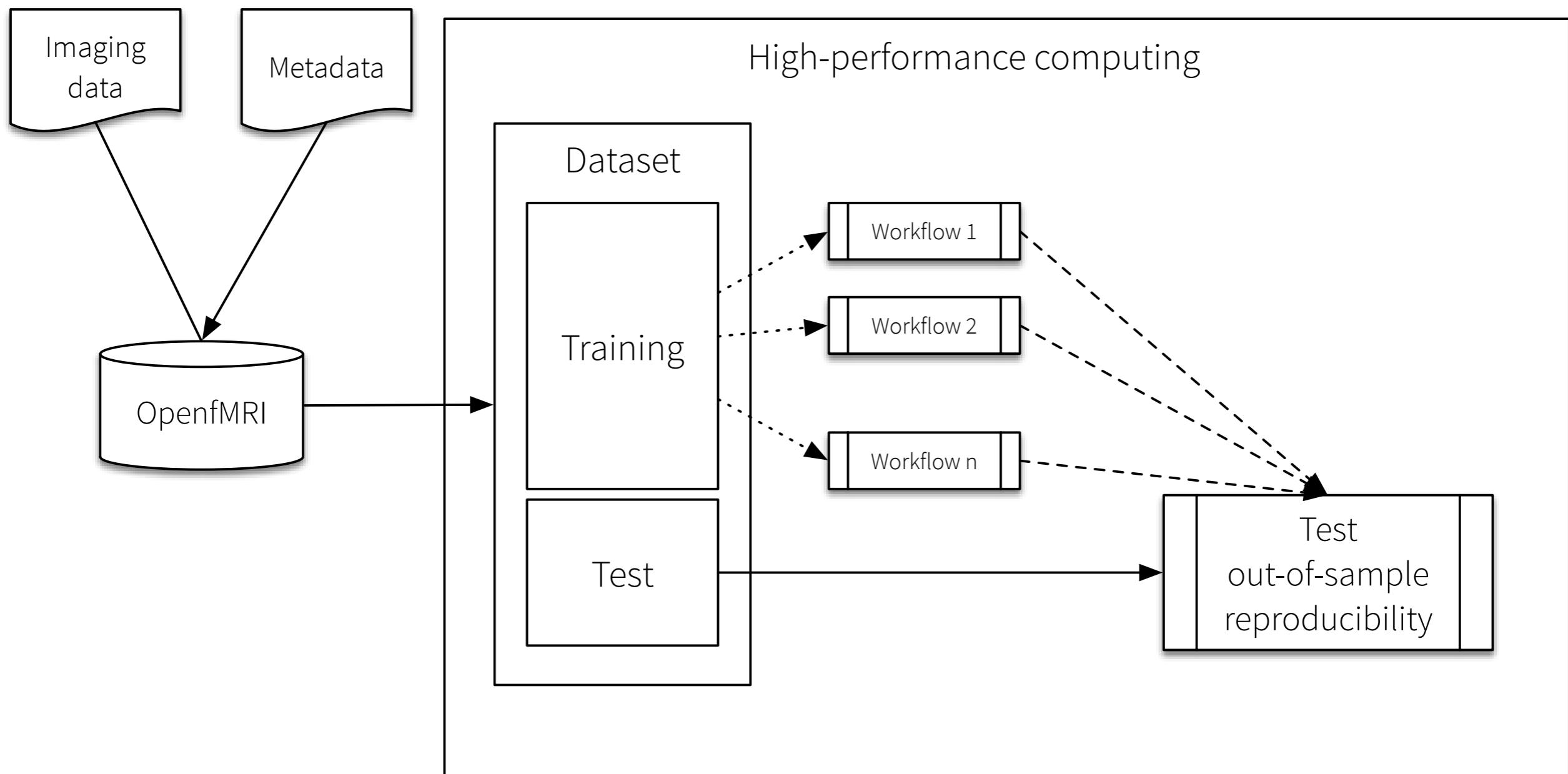


nipy.org



Tracking every analysis
allows a full characterization of exploration

Improvement: Quantifying “vibration of effects”



Focusing on finding generalizable results,
rather than hitting the $p < 0.05$ jackpot

Threats to reproducibility: software errors



Geoffrey Chang

Structure of MsbA from *E. coli*: A Homolog of the Multidrug Resistance ATP Binding Cassette (ABC) Transporters

Geoffrey Chang* and Christopher B. Roth

Multidrug resistance (MDR) is a serious medical problem and presents a major challenge to the treatment of disease and the development of novel therapeutics. ABC transporters that are associated with multidrug resistance (MDR-ABC transporters) translocate hydrophobic drugs and lipids from the inner to the outer leaflet of the cell membrane. To better elucidate the structural basis for the "flip-flop" mechanism of substrate movement across the lipid bilayer, we have determined the structure of the lipid flippase MsbA from *Escherichia coli* by x-ray crystallography to a resolution of 4.5 angstroms. MsbA is organized as a homodimer with each subunit containing six transmembrane α -helices and a nucleotide-binding domain. The asymmetric distribution of charged residues lining a central chamber suggests a general mechanism for the translocation of substrate by MsbA and other MDR-ABC transporters. The structure of MsbA can serve as a model for the MDR-ABC transporters that confer multidrug resistance to cancer cells and infectious microorganisms.

www.sciencemag.org SCIENCE VOL 293 7 SEPTEMBER 2001

Structure of the ABC Transporter MsbA in Complex with ADP·Vanadate and Lipopolysaccharide

Christopher L. Reyes and Geoffrey Chang*

Select members of the adenosine triphosphate (ATP)-binding cassette (ABC) transporter family couple ATP binding and hydrolysis to substrate efflux and confer multidrug resistance. We have determined the x-ray structure of MsbA in complex with magnesium, adenosine diphosphate, and inorganic vanadate ($Mg\text{-ADP-V}_i$) and the rough-chemotype lipopolysaccharide, Ra LPS. The structure supports a model involving a rigid-body torque of the two transmembrane domains during ATP hydrolysis and suggests a mechanism by which the nucleotide-binding domain communicates with the transmembrane domain. We propose a lipid "flip-flop" mechanism in which the sugar groups are sequestered in the chamber while the hydrophobic tails are dragged through the lipid bilayer.

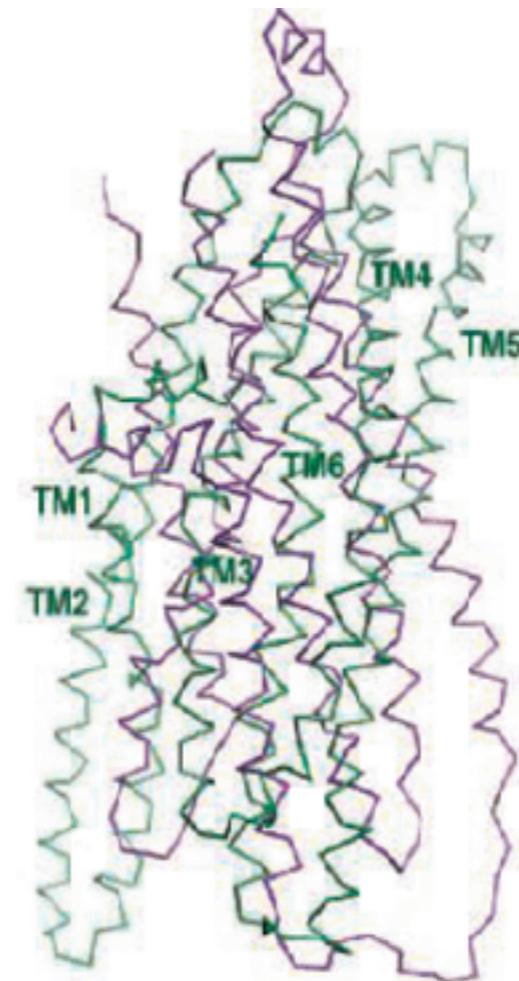
13 MAY 2005 VOL 308 SCIENCE www.sciencemag.org

X-ray Structure of the EmrE Multidrug Transporter in Complex with a Substrate

Owen Purnillos, Yen-Ju Chen, Andy P. Chen, Geoffrey Chang*

EmrE is a prototype of the Small Multidrug Resistance family of efflux transporters and actively expels positively charged hydrophobic drugs across the inner membrane of *Escherichia coli*. Here, we report the x-ray crystal structure, at 3.7 angstrom resolution, of one conformational state of the EmrE transporter in complex with a translocation substrate, tetraphenylphosphonium. Two EmrE polypeptides form a homodimeric transporter that binds substrate at the dimerization interface. The two subunits have opposite orientations in the membrane and adopt slightly different folds, forming an asymmetric antiparallel dimer. This unusual architecture likely confers unidirectionality to transport by creating an asymmetric substrate translocation pathway. On the basis of available structural data, we propose a model for the proton-dependent drug efflux mechanism of EmrE.

23 DECEMBER 2005 VOL 310 SCIENCE www.sciencemag.org



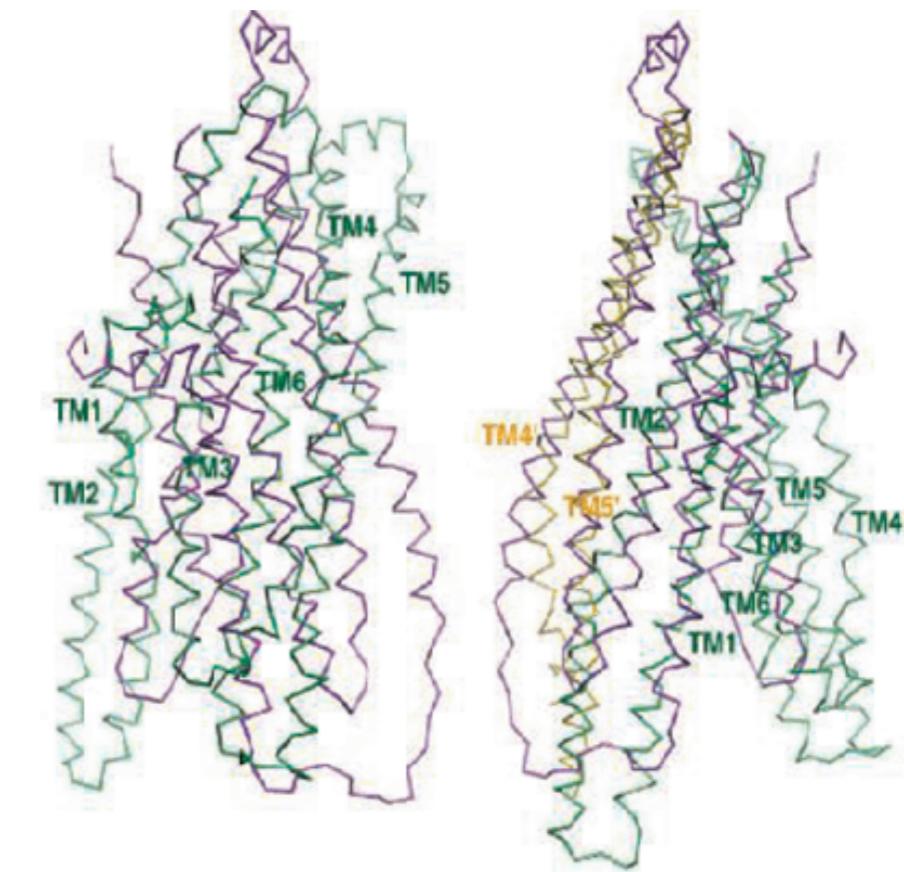
Threats to reproducibility: software errors

Retraction

WE WISH TO RETRACT OUR RESEARCH ARTICLE “STRUCTURE OF MsbA from *E. coli*: A homolog of the multidrug resistance ATP binding cassette (ABC) transporters” and both of our Reports “Structure of the ABC transporter MsbA in complex with ADP•vanadate and lipopolysaccharide” and “X-ray structure of the EmrE multidrug transporter in complex with a substrate” (1–3).

The recently reported structure of Sav1866 (4) indicated that our MsbA structures (1, 2, 5) were incorrect in both the hand of the structure and the topology. Thus, our biological interpretations based on these inverted models for MsbA are invalid.

An in-house data reduction program introduced a change in sign for anomalous differences. This program, which was not part of a conventional data processing package, converted the anomalous pairs (I^+ and I^-) to (F^- and F^+), thereby introducing a sign change. As the diffraction data collected for each set of MsbA crystals and for the EmrE crystals were processed with the same program, the structures reported in (1–3, 5, 6) had the wrong hand.



Small errors can have big effects

```
# 23-class classification problem

skf=StratifiedKFold(labels,8)

if trainsvm:
    pred=N.zeros(len(labels))
    for train,test in skf:
        clf=LinearSVC()
        clf.fit(data[train],labels[train])           data[:,train]
        pred[test]=clf.predict(data[test])           data[:,test]
```

Results:
93% accuracy

Results:
53% accuracy

CORRECTION

Correction: The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science

Stephan Lewandowsky, Gilles E. Gignac, Klaus Oberauer

The dataset included two notable age outliers (reported ages 5 and 32757).

Specifically, the statement on page 9 “age turned out not to correlate with any of the indicator variables” is incorrect. It should read instead “age correlated significantly with 3 latent indicator variables (Vaccinations: .219, $p < .0001$; Conservatism: .169, $p < .001$; Conspiracist ideation: -.140, maximum likelihood $p < .0001$, bootstrapped $p = .004$), and straddled significance for a fourth (Free Market: .08, $p = .05$).”

```
In [1]: age=32757
```

```
In [2]: assert age>12 and age<120
```

```
-----  
AssertionError
```

```
Traceback (most recent call last)
```

```
<ipython-input-2-37de876b5fda> in <module>()
```

```
----> 1 assert age>12 and age<120
```

```
AssertionError:
```

Improvement: The principle of assumed error

- Whenever you find a seemingly good result (e.g. one that fits your predictions), assume that it occurred due to an error in your code
- “The first principle is that you must not fool yourself and you are the easiest person to fool” - R. Feynman

Improvement: Software testing and validation

- Smoke tests and unit tests may be useful but are not sufficient
- For complex analyses:
 - Parameter recovery: Generate data for which the true answer is known, and assess ability of code to recover the correct answer
 - Randomization: Generate data for which the null hypothesis of no relationship should be true on average, and ensure that the observed false positive rate is accurate (cf. Eklund et al., 2016, PNAS)

<http://www.russpoldrack.org/2016/08/the-principle-of-assumed-error.html>

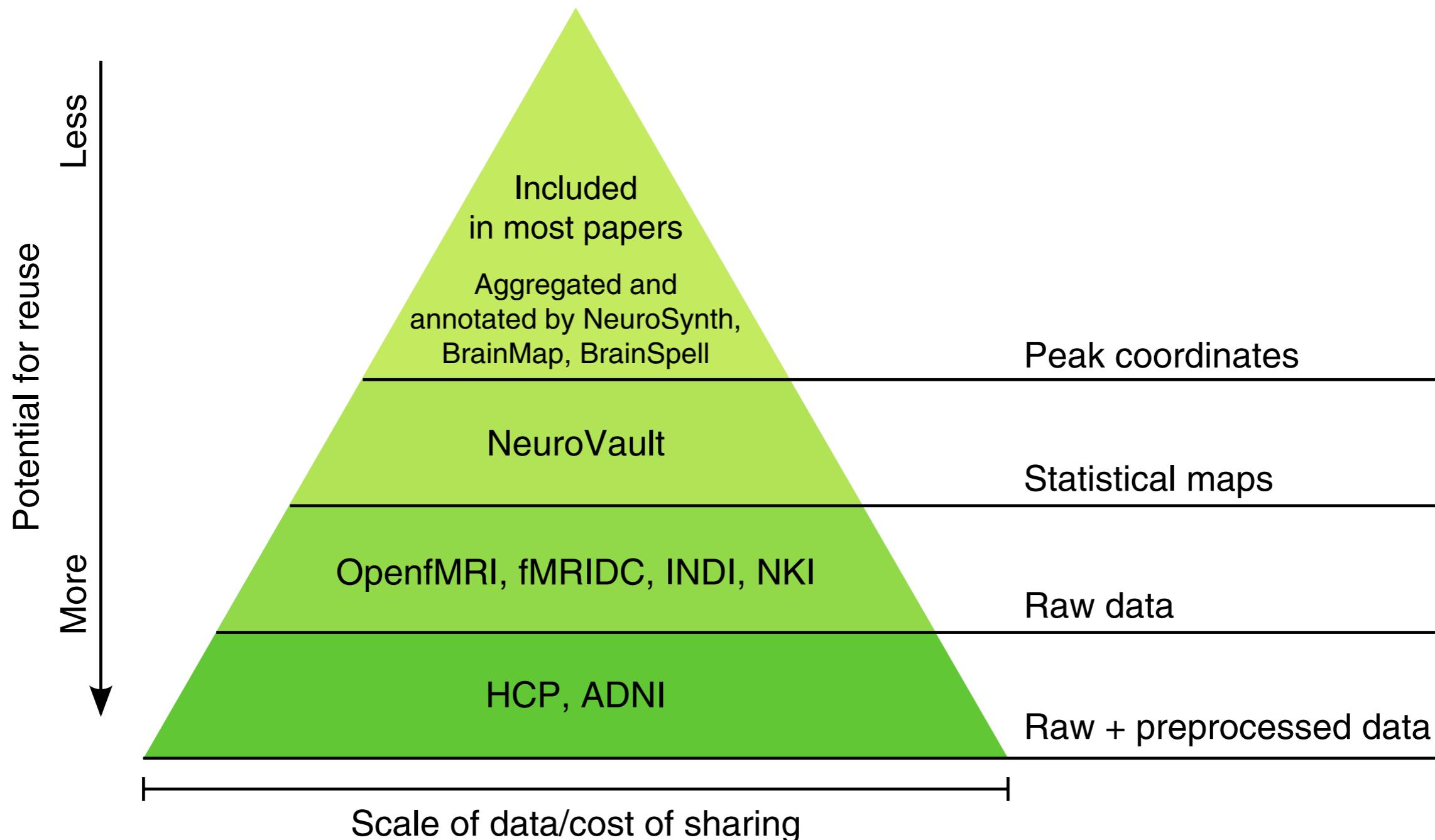
Threats to reproducibility: Study reporting and transparency

- In 22 of the 65 papers we analyzed for multiple comparison procedures, it was impossible to identify precisely which correction technique was used
 - beyond generic terms such as “cluster based correction”

Improvement: Better study description

- OHBM Committee on Best Practices in Data Analysis and Sharing (COBIDAS) report
 - www.humanbrainmapping.org/cobidas/
- In the future, tools may be able to automatically generate standards-compliant methods text from a nipype workflow

Improvement: Data Sharing



Poldrack & Gorgolewski, 2014

Improvement: Brain Imaging Data Structure (BIDS)

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

- » Data publication and archiving
- » Research data

Received: 18 December 2015

Accepted: 19 May 2016

Published: 21 June 2016

The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments

Krzysztof J. Gorgolewski¹, Tibor Auer², Vince D. Calhoun^{3,4}, R. Cameron Craddock^{5,6}, Samir Das⁷, Eugene P. Duff⁸, Guillaume Flandin⁹, Satrajit S. Ghosh^{10,11}, Tristan Glatard^{7,12}, Yaroslav O. Halchenko¹³, Daniel A. Handwerker¹⁴, Michael Hanke^{15,16}, David Keator¹⁷, Xiangrui Li¹⁸, Zachary Michael¹⁹, Camille Maumet²⁰, B. Nolan Nichols^{21,22}, Thomas E. Nichols^{20,23}, John Pellman⁶, Jean-Baptiste Poline²⁴, Ariel Rokem²⁵, Gunnar Schaefer^{1,26}, Vanessa Sochat²⁷, William Triplett¹, Jessica A. Turner^{3,28}, Gaël Varoquaux²⁹ & Russell A. Poldrack¹

● sub-control01

○ anat

- sub-control01_T1w.nii.gz
- sub-control01_T1w.json
- sub-control01_T2w.nii.gz
- sub-control01_T2w.json

○ func

- sub-control01_task-nback_bold.nii.gz
- sub-control01_task-nback_bold.json
- sub-control01_task-nback_events.tsv
- sub-control01_task-nback_physio.tsv.gz
- sub-control01_task-nback_physio.json
- sub-control01_task-nback_sbref.nii.gz

○ dwi

- sub-control01_dwi.nii.gz

<http://bids.neuroimaging.io>

Open sharing is associated with better science

OPEN  ACCESS Freely available online



Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results

Jelte M. Wicherts*, Marjan Bakker, Dylan Molenaar

Psychology Department, Faculty of Social and Behavioral Sciences, University of Amsterdam, Amsterdam, The Netherlands

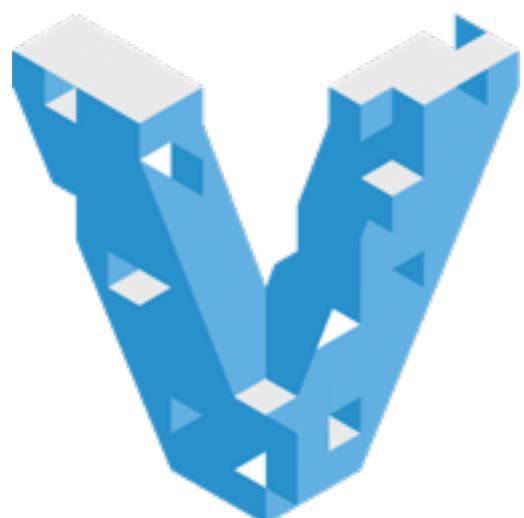
Won't I get scooped?

- Maybe...
 - But online publication documents priority
- Howard Aiken
 - “Don't worry about people stealing an idea. If it's original, you will have to ram it down their throats.”
- Gary King
 - “The thing that matters the least is being scooped. The thing that matters the most is being ignored.”

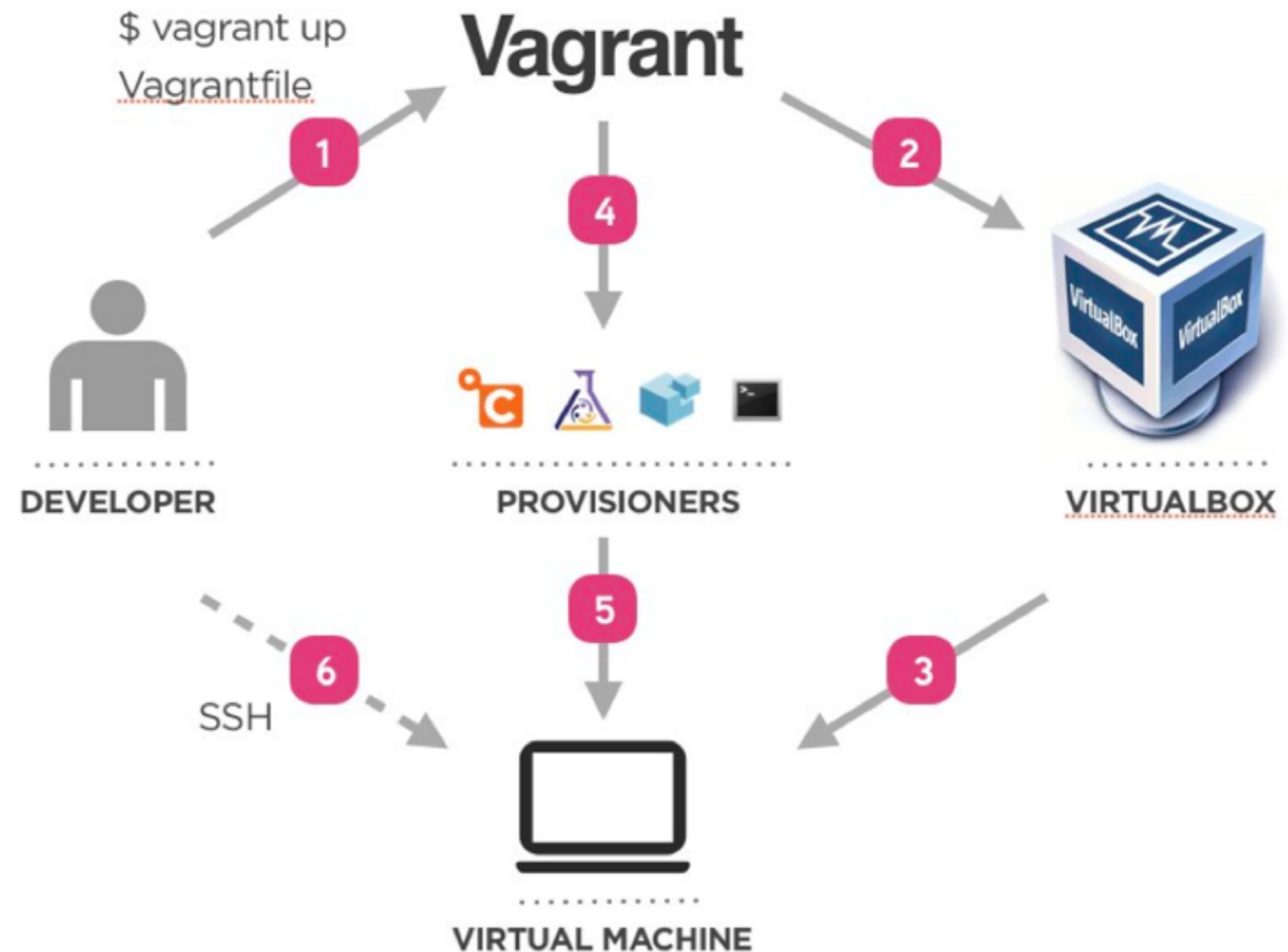
Improvement: Sharing of analysis platforms

- “an article about a computational result is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result.” - Buckheit & Donoho, 1995
- The tale of myconnectome

Virtual machines as tools for reproducible science



VAGRANT



27 Google Calc × <https://www.google.com/calendar> × <https://twitter.com> × <https://app.gutenberg.org> × [MTS Nature](#) × [Shared with me](#) × [Handling M](#) × [poldrack/my...](#) Russ

GitHub, Inc. [US] <https://github.com/poldrack/myconnectome-vm>

Apps 27 Google Calendar - W SERA Axess PubMed sklearn nipype

This repository Search Pull requests Issues Gist

 poldrack / myconnectome-vm   

Virtual machine setup for MyConnectome data analysis — Edit

86 commits 1 branch 0 releases 3 contributors

Branch: **master** myconnectome-vm / +

Merge pull request #14 from vsoch/master ...
poldrack authored 23 days ago latest commit 3b45da4ddb 

LICENSE Initial commit 2 months ago

README.md Update README.md 24 days ago

Vagrantfile removing supervisor controller from application - will be run with st... 23 days ago

 README.md

MyConnectome-VM: A virtual machine to implement MyConnectome analyses.

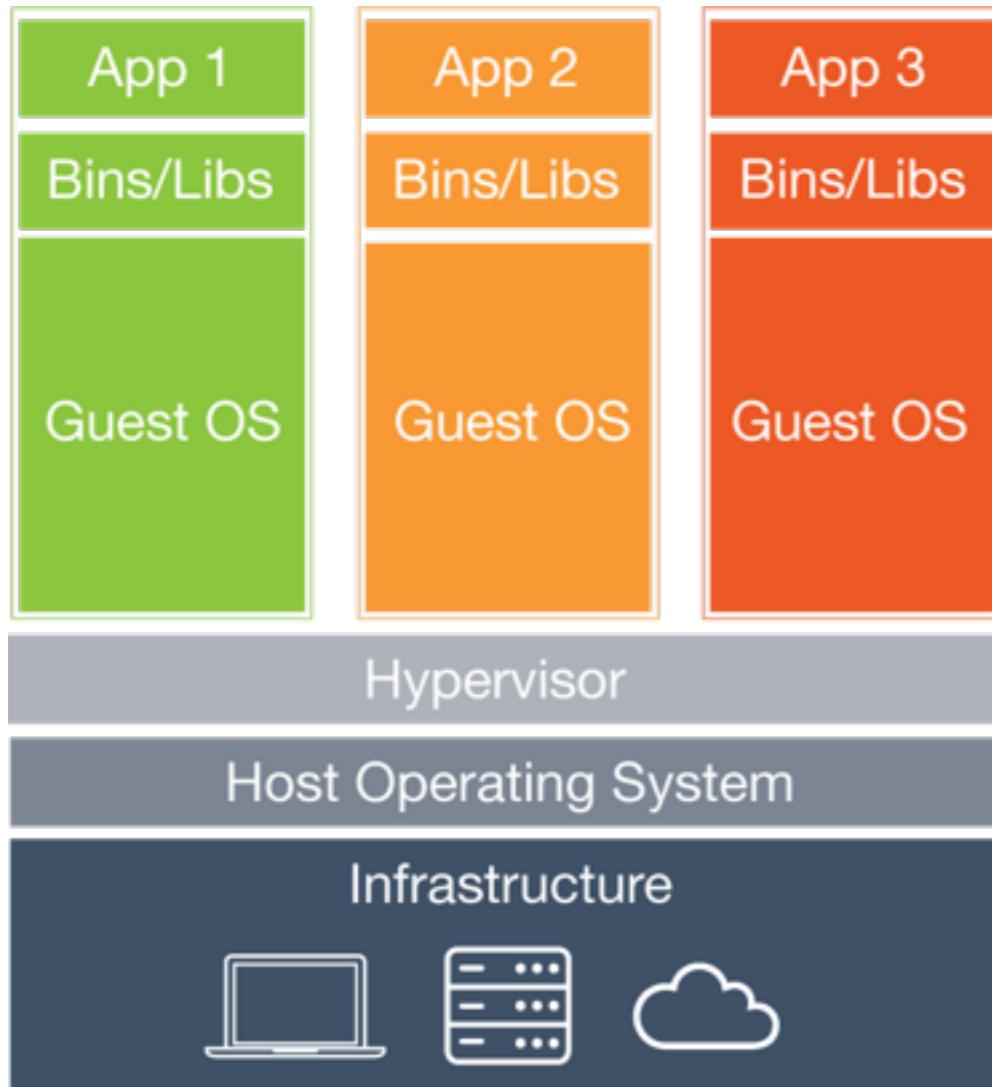
The MyConnectome project is a project meant to investigate the relations between mind, brain, and body across an extended period of time in a single individual. One of the major goals of the project is to serve as a testbed for reproducible analysis practices. For this reason, we have released the data and as much code as possible for the processing and analyses.

 SSH clone URL
git@github.com:poldrack/myconnectome-vm 
You can clone with HTTPS, SSH, or Subversion. 

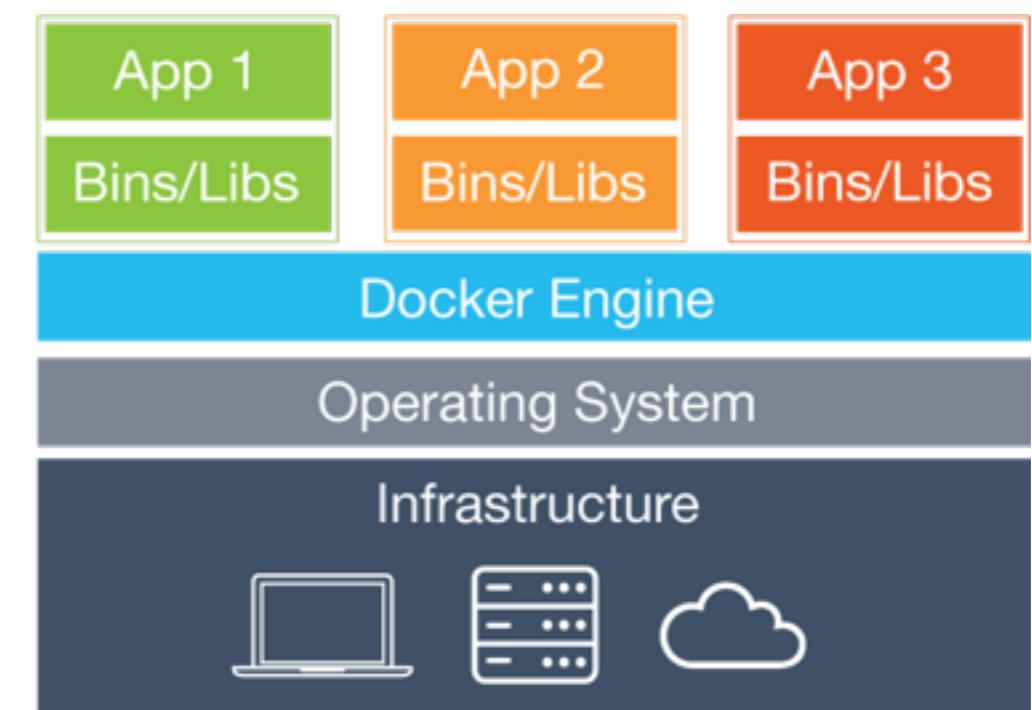
 Clone in Desktop
 Download ZIP

nihms-470697.pdf yoi140096.pdf Show All

Container computing: Docker



Virtual machines



Docker containers



BIDS Apps

A collection of containerized neuroimaging workflows and pipelines that accept datasets organized according to the Brain Imaging Data Structure (BIDS).

<http://bids.neuroimaging.io>

Repositories

People 25

Teams 1

Settings

Filters ▾

Find a repository...

New repository

ndmg

Updated 2 hours ago

TeX ★ 0 ⚡ 0

niak

Updated 3 days ago

Python ★ 0 ⚡ 0

fmriprep

Updated 3 days ago

★ 0 ⚡ 0

People

25 >



[Invite someone](#)

Conclusions

- We need to redesign the choice architecture of fMRI methods so that it prevents rather than affords fooling ourselves
- We will almost certainly have fewer successes and more failures
- But the resulting discoveries will have a greatly likelihood of being true

Acknowledgments

Tal Yarkoni
Tom Nichols



The Poldrack Lab @ Stanford

<http://reproducibility.stanford.edu>

Data sets and code will be made available at www.openfmri.org



James S. McDonnell
Foundation

