

# **Reproducibility in the Life Sciences**

Jean-Baptiste Poline

MNI, McConnell Brain Imaging Centre,  
McGill, Montreal

**Part I: Reproducibility: background**

**Part II : Etiology of Irreproducibility**

**Part III :Some therapeutic proposals**

# Part I: Reproducibility: background

Part II : Etiology of Irreproducibility

Part III :Some therapeutic proposals

- 53 papers examined at Amgen in preclinical cancer research
- Papers were selected that described something completely new and in very high impact factor journals
- **Scientific findings were confirmed in only 6 (11%)**

Begley and Ellis, Nature, 2012



FEATURE ARTICLE



REPRODUCIBILITY IN CANCER BIOLOGY

## Challenges for assessing replicability in preclinical cancer biology

- \* Not enough information in papers
- \* Analytical flexibility is particularly pervasive in neuroimaging analysis (cf workflow tools Nipype, Niflow, Pydra, ...)

# Altered Brain Activity in Unipolar Depression Revisited Meta analyses of Neuroimaging Studies

Veronika I. Müller, PhD, Edna C. Cieslik, PhD, Ilinca Serbanescu, MSc, Angela R. Laird, PhD, Peter T. Fox, MD, and Simon B. Eickhoff, MD

**RESULTS**—In total, 57 studies with 99 individual neuroimaging experiments comprising in total 1058 patients were included; 34 of them tested cognitive and 65 emotional processing. Overall analyses across cognitive processing experiments ( $P > .29$ ) and across emotional processing experiments ( $P > .47$ ) revealed no significant results. Similarly, no convergence was found in analyses investigating positive (all  $P > .15$ ), negative (all  $P > .76$ ), or memory (all  $P > .48$ ) processes. Analyses that restricted inclusion of confounds (eg, medication, comorbidity, age) did not change the results.

Stein et al., 2012, Nature Genetics, study of the hippocampal volume in more than 10k+7k subjects

Previously identified candidate polymorphisms associated with hippocampal volume in general showed little association within our meta-analysis :(

*Stein et al, Nat. Gen. 2013*

## Essay

## Why Most Published Research Findings Are False

John P. A. Ioannidis

2005. *PLoS Medicine*, 2(8), e124. doi:  
10.1371/journal.pmed.0020124

“There is increasing concern about the reliability of biomedical research, with recent articles suggesting that up to 85% of research funding is wasted.”

Bustin, S. A. (2015). The reproducibility of biomedical research: Sleepers awake!  
*Biomolecular Detection and Quantification*

## THE LANCET

[Online First](#) [Current Issue](#) [All Issues](#) [Special Issues](#) [Multimedia](#) [Information for Authors](#)

All Content

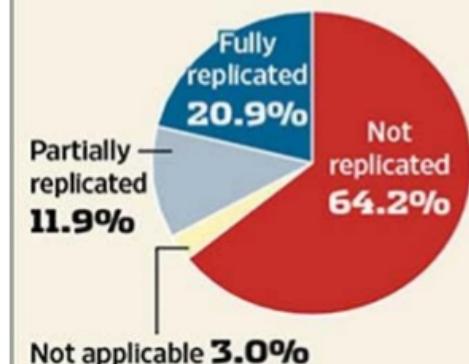
Search

Advanced Search

Research: increasing value, reducing waste

## No Cure

When Bayer tried to replicate results of 67 studies published in academic journals, nearly two-thirds failed.



Source: Nature Reviews Drug Discovery

**nature** International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archives

News & Comment > News > 2015 > May > Article

NATURE | NEWS

First results from psychology's largest reproducibility test

# Credibility Crisis

**Los Angeles Times | BUSINESS**

LOCAL U.S. WORLD BUSINESS SPORTS ENTERTAINMENT HEALTH STYLE TRAVEL

## Science has lost its way, at a big cost to humanity

Researchers are rewarded for splashy findings, not for double-checking accuracy. So many scientists looking for cures to diseases have been building on ideas that aren't even true.

**Science** AAAS.ORG FEEDBACK HELP LIBRARIANS All Science Journals Enter Search Text GUEST ALERTS ACCESS RICH NEWS SCIENCE JOURNALS CAREERS MULTIMEDIA COLLECTIONS

**Science** The World's Leading Journal of Original Scientific Research, Global News, and Commentary. Science Home Current Issue Previous Issues Science Express Science Products My Science About the Journal

Home > Science Magazine > 17 January 2014 > McNutt, 343 (6168): 229

Article Views Science 17 January 2014: Vol. 343 no. 6168 p. 229 DOI: 10.1126/science.1250475

Summary Full Text Full Text (PDF)

**Reproducibility** Marcia McNutt

» Marcia McNutt is Editor-in-Chief of Science. Science advances on a foundation of trusted data, an approach that scientists use to gain confidence. The community was shaken by reports that a troublingly large number of results were not reproducible. Because confidence in results is crucial to the scientific method, we are announcing new initiatives to address this problem. In this week's *Science*, for preclinical studies (one of the targets of the National Institutes of Health), we are increasing transparency. Authors will indicate how they handled (such as how to deal with outliers), which ensure a sufficient signal-to-noise ratio, whether the experimenter was blind to the conduct of the experiments, and whether the experiments were conducted in accordance with guidelines.

**Article Tools**

- Save to My Folders
- Download Citation
- Alert Me When Article is Cited
- Post to CiteULike
- E-mail This Page
- Rights & Permissions
- Commercial Reprints and E-Prints
- View PubMed Citation

Related Content

**TheScientist** EXPLORING LIFE. INSPIRING INNOVATION

## NIH Tackles Irreproducibility

The federal agency speaks out about how to improve the quality of scientific research.

By Jef Akst | January 28, 2014

nature.com | Sitemap Login Register

**nature** International weekly journal of science

Home News & Comment Research Careers & Jobs Current Issue Archive Audio & Video For Authors

Archive > Volume 496 > Issue 7446 > Editorial > Article

NATURE | EDITORIAL

Announcement: Reducing our irreproducibility

24 April 2013

PDF Rights & Permissions

Over the past year, *Nature* has published a string of articles that highlight the need to improve the reliability and reproducibility of published research (collected and discussed in this editorial).

**nature** International weekly journal of science

Menu Advanced search Search Go

archive > volume 483 > issue 7391 > editorials > article

NATURE | EDITORIAL

## Must try harder

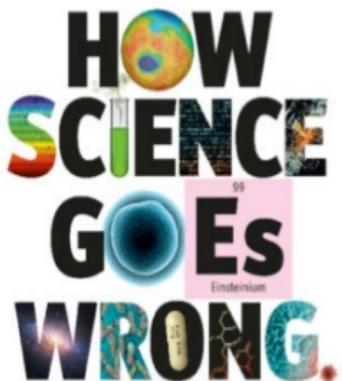
*Nature* 483, 509 (29 March 2012) | doi:10.1038/483509a  
Published online 28 March 2012

PDF Citation Reprints Rights & permissions Article metrics

Too many sloppy mistakes are creeping into scientific papers. Lab heads must look more rigorously at the data — and at themselves.

**The Economist**

Washington's lawyer surplus  
How to do a nuclear deal with Iran  
Investment tips from Nobel economists  
Junk bonds are back  
The meaning of Sachin Tendulkar

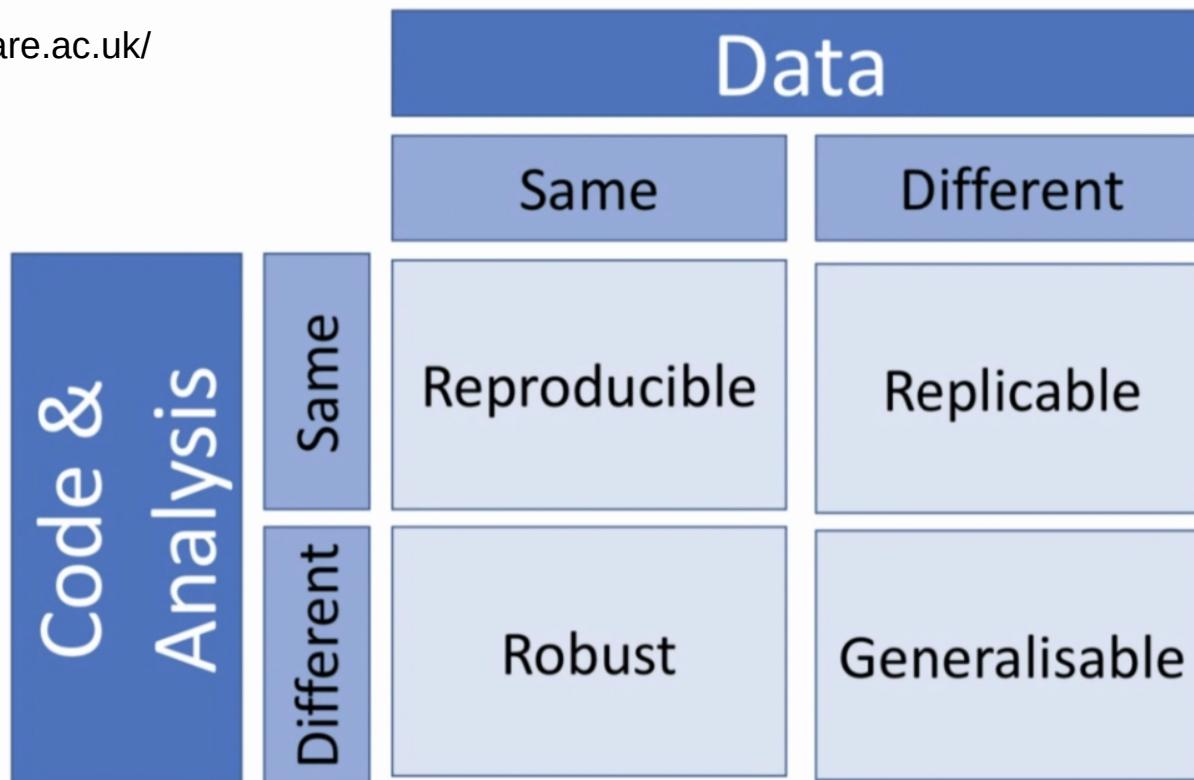


# NIH plans to enhance reproducibility

**Francis S. Collins** and **Lawrence A. Tabak** discuss initiatives that the US National Institutes of Health is exploring to restore the self-correcting nature of preclinical research.

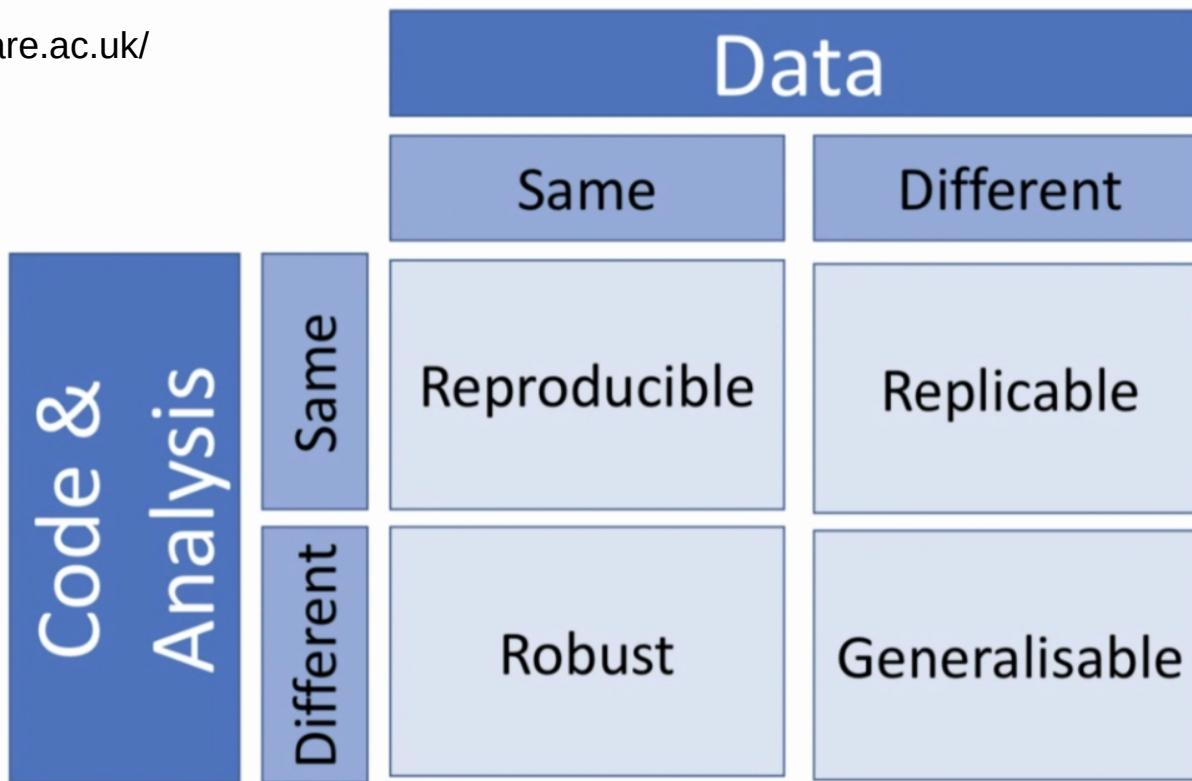
Collins and Tabak. 2014. Nature 505: 612–13.

<https://www.software.ac.uk/>



- My (preferred) way of thinking about reproducibility: **only talk about generalizability** across ... (Data, Software, Time, Scanner, Stimuli, ... etc)

<https://www.software.ac.uk/>



- My (preferred) way of thinking about reproducibility: **only talk about generalizability** across ... (Data, Software, Time, Scanner, Stimuli, ... etc)

**Part I: Reproducibility: background**

**Part II : Etiology of Irreproducibility**

**Part III :Some therapeutic proposals**

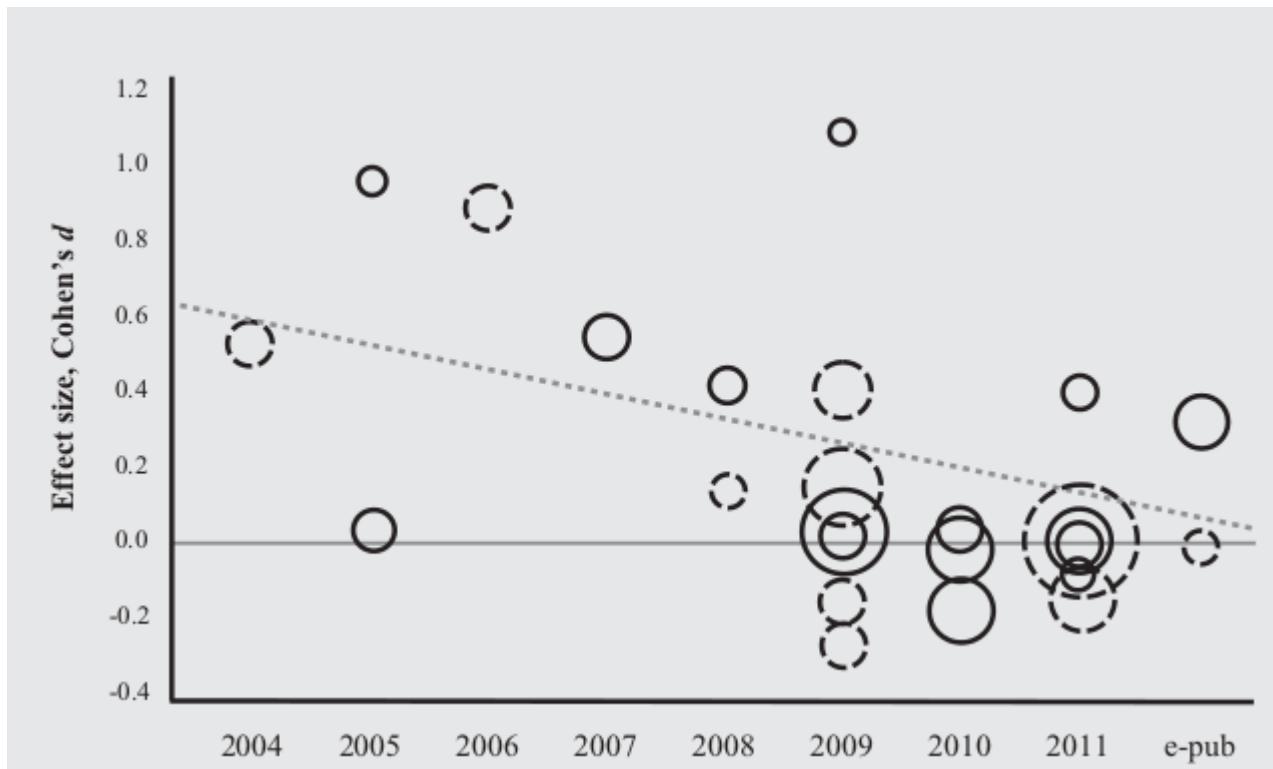
1. Poor statistical procedures
2. Issues in data and software
3. A cultural issue: Publication practices and research incentives

## 1. Poor statistical procedures

- Small Ns and effect sizes
- Power issues
- P-hacking - harking

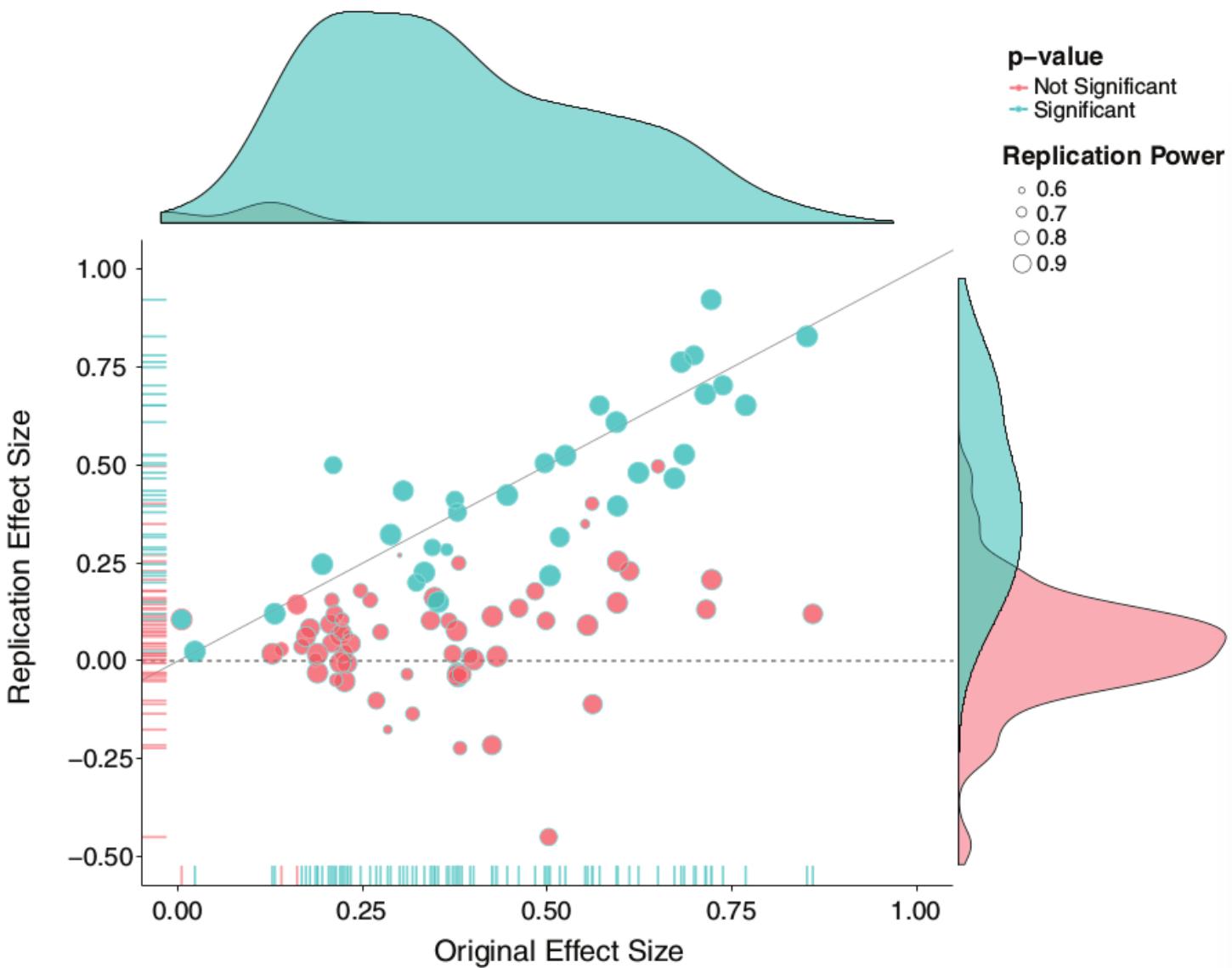
## 2. Issues in data and software

## 3. A cultural issue: Publication practices and research incentives



Molendijk, 2012: BDNF and hippocampal volume

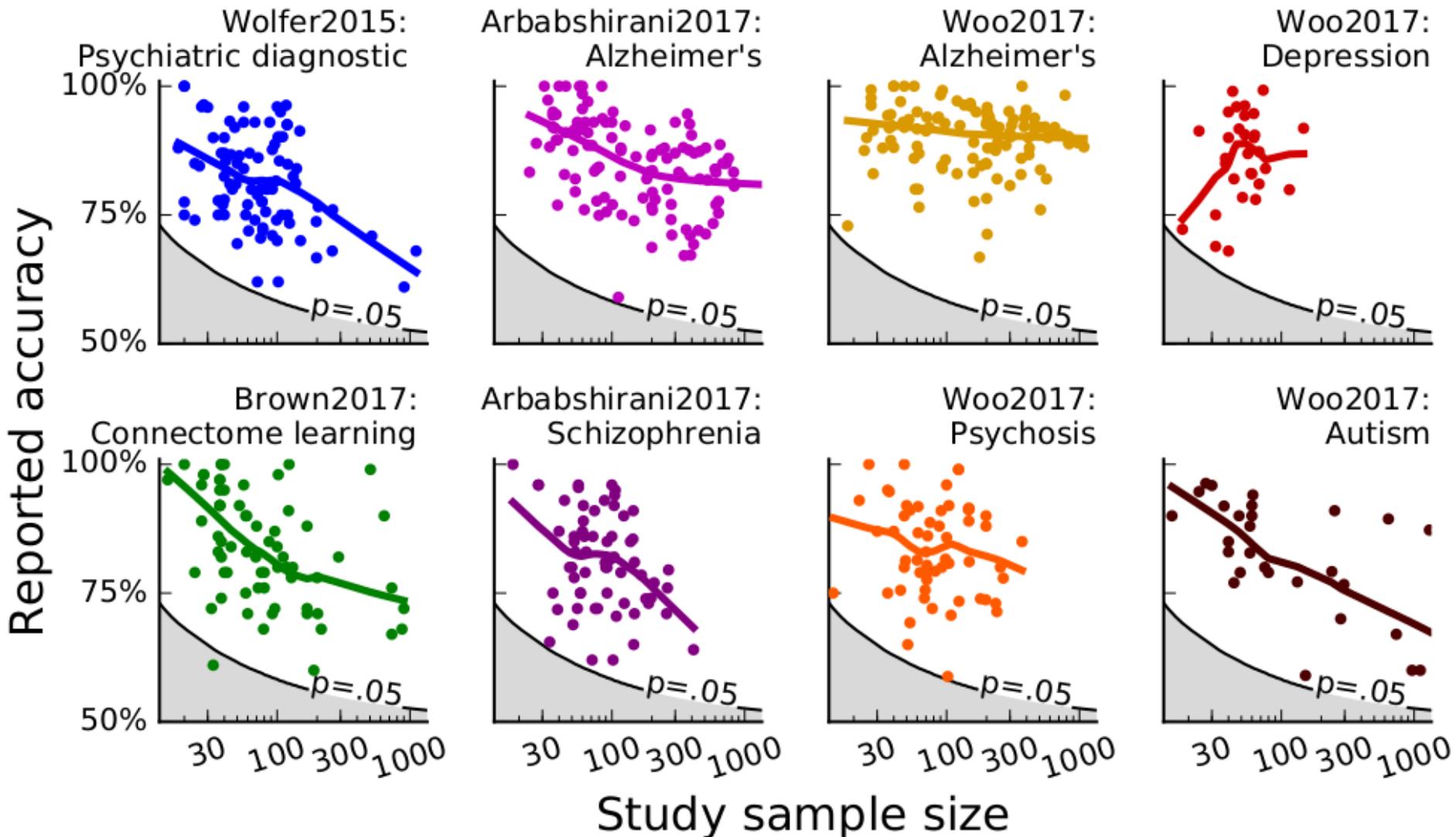
See also : Mier, 2009: COMT and DLPFC



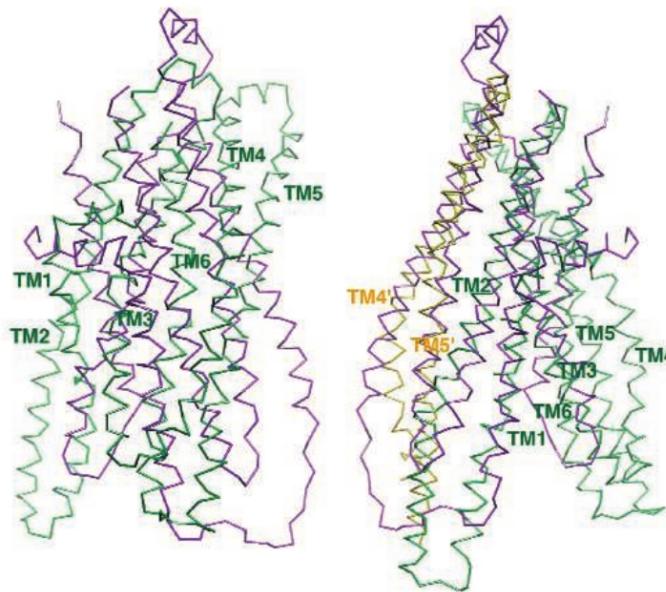
\* The mean **effect size** ( $r$ ) of the replication effects ( $M r = 0.197$ ,  $SD = 0.257$ ) was **half the magnitude** of the mean effect size of the original effects ( $M r = 0.403$ ,  $SD = 0.188$ )

\* **39%** of effects were rated to have replicated the original effect

# Sample size issue in ML



1. Poor statistical procedures
2. Issues in data and software
  - Research software issues: motivation and examples
  - Across OS / implementation
  - Across software and parametrization
  - Across pipeline: “analytical flexibility”
  - Issues in data
3. A cultural issue: Publication practices and research incentives



**Flipping fiasco.** The structures of MsbA (purple) and Sav1866 (green) overlap little (left) until MsbA is inverted (right).

- G. Chang: 3 Science, 1 PNAS, 1 J Mol Biol retracted
- “... a homemade data-analysis program had flipped two columns of data...”,
- “... inherited from another lab...”
- The code was distributed and used by others

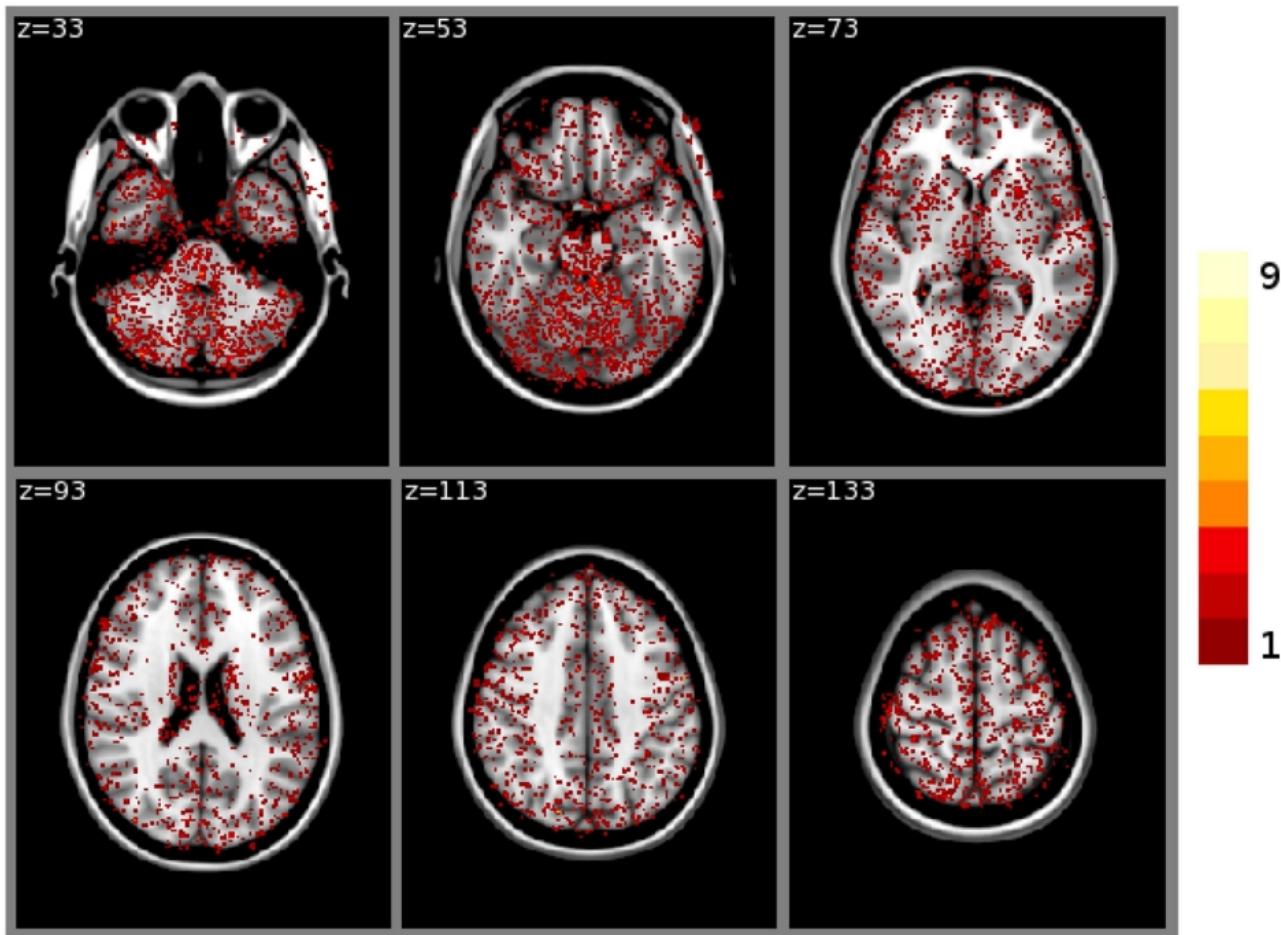
- Potti et al., Nat. Med. 2006, 2008 vs Baggerly and Coombes, “Forensic analysis”, Annals of applied Stat., 2009
- Choose cell lines that are most sensitive / resistant to a drug, use expression profiles to build a model that predicts patient response

Baggerly and Coombes Forensic:

***“with poor documentation and irreproducibility even well meaning investigator may argue for drug that are contraindicated to some patients”***

**“the most common errors are simple (e.g., row or column offsets); conversely, the most simple errors are common.”**

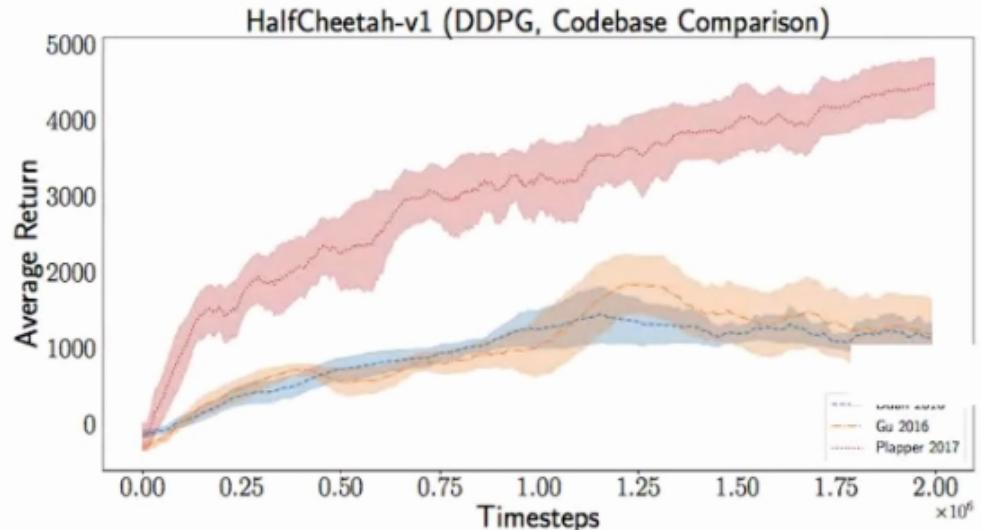
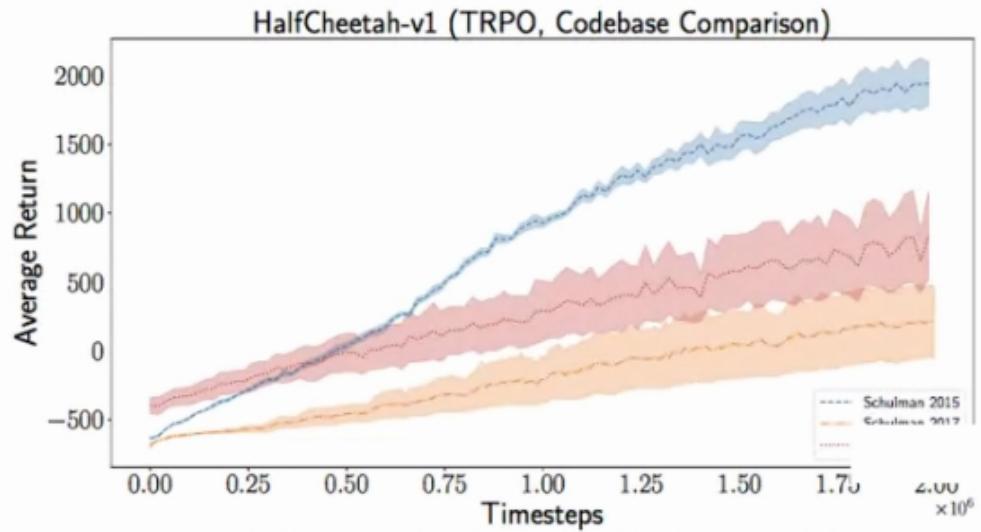
# Across OS Across implementations



**FIGURE 2 |** Sum of binarized differences between cortical tissue classifications obtained on cluster A and cluster B (FSL FAST, build 1,  $n = 150$  subjects). All binarized differences were resampled to the default MNI152 volume template.

Glatard et al, 2015, F. in Neuroinformatics

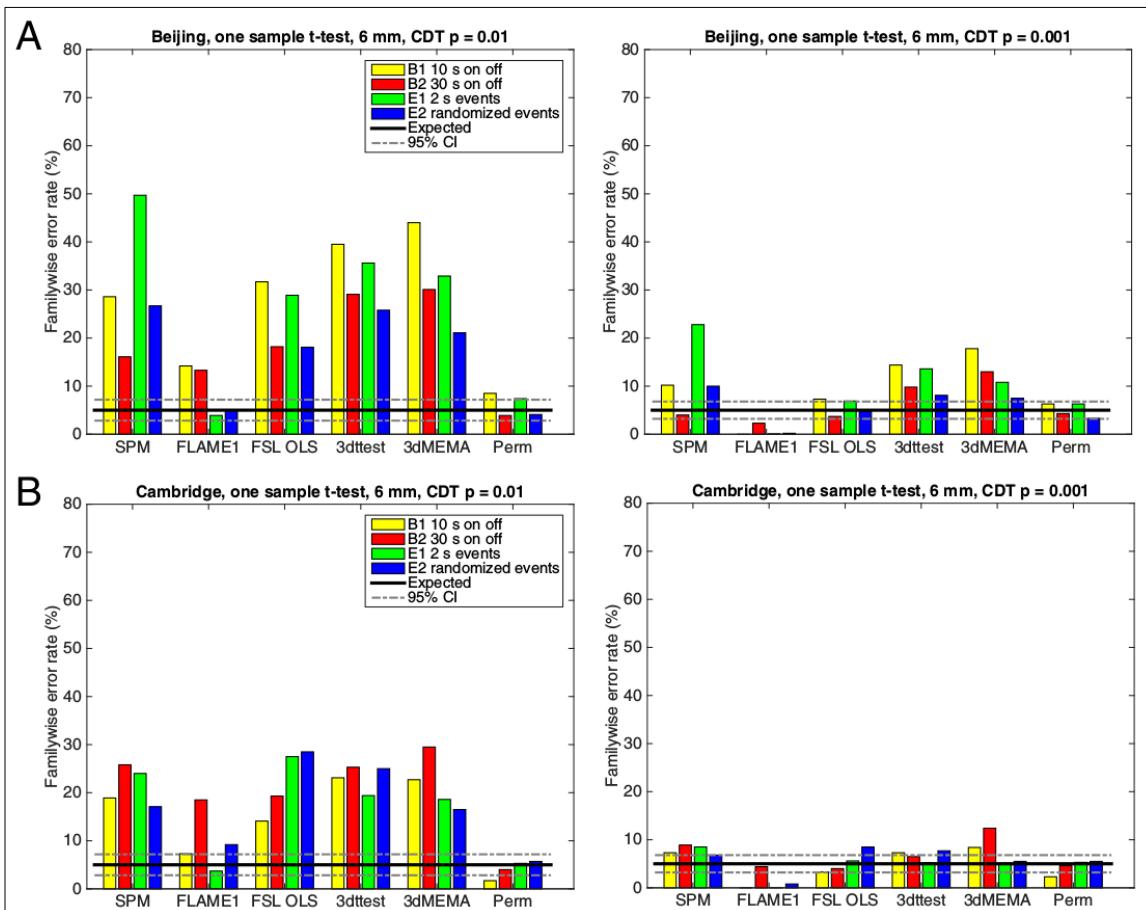
- Same algorithm
  - Top graph: TRPO
  - Bottom graph: DDPG
- Same domain
- Simulation environment
- Different implementations



Credit : Joelle Pineau

# **Across parametrizations of software**

## **and across software**

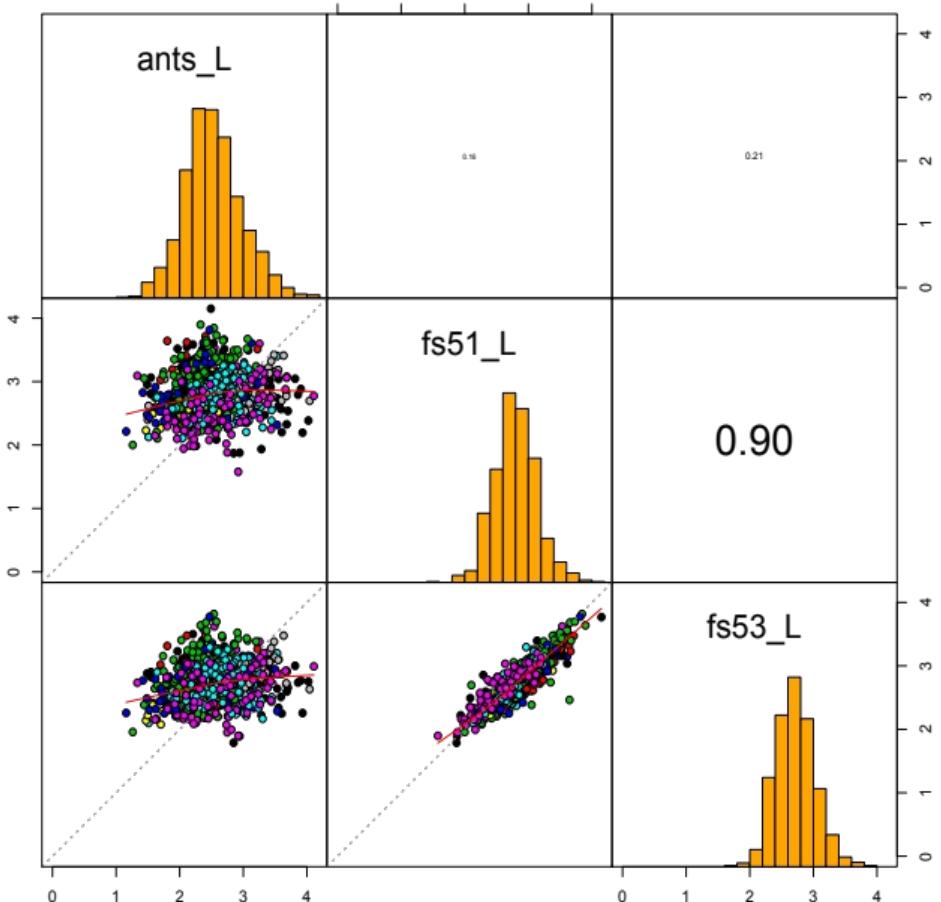


Eklund et al., PNAS, 2016 :

- Low threshold issue
- High threshold issue with Paradigm E1 ?
- Ad hoc procedure leads to around 70% FPR

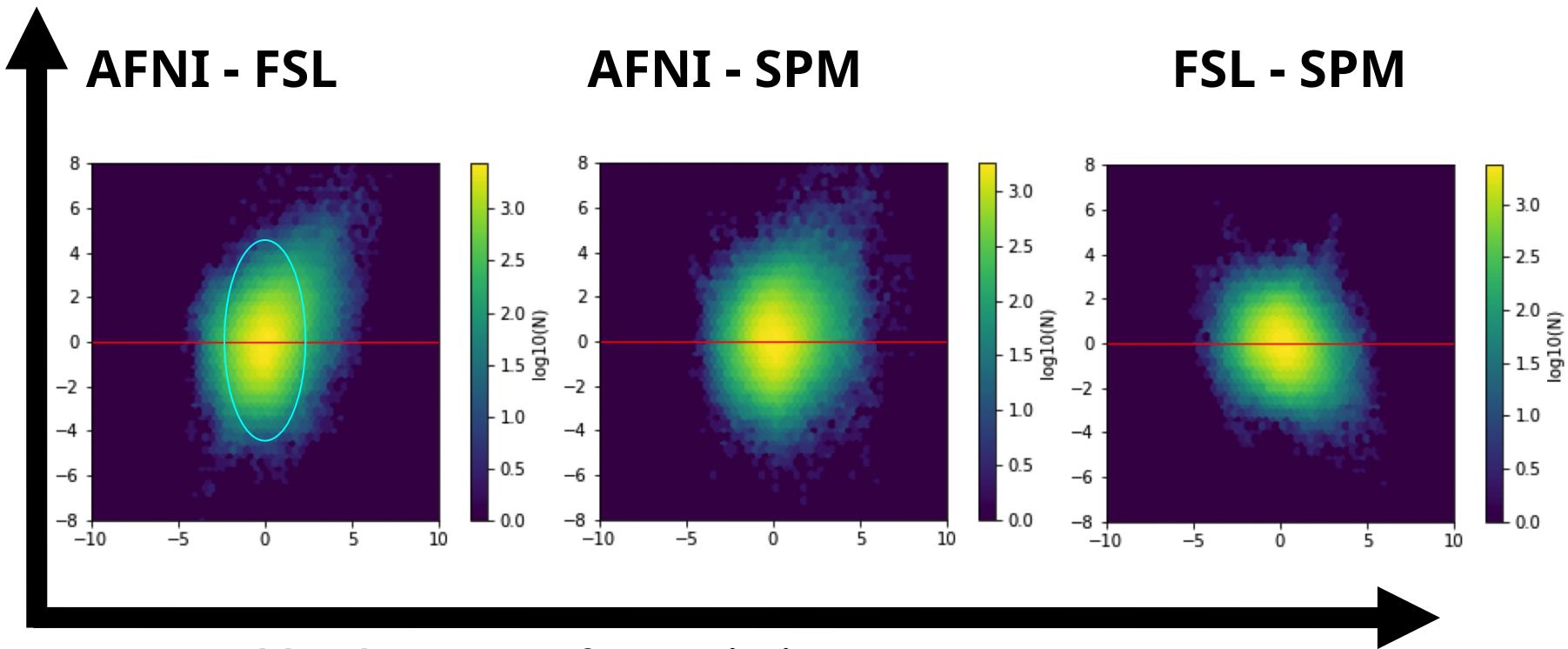
- Estimated 3,500 papers affected by low threshold ?
- But 13000 w/o multiple comparisons ?

## Size of the left caudal anterior Cingulate



Dickie E et al., 2017

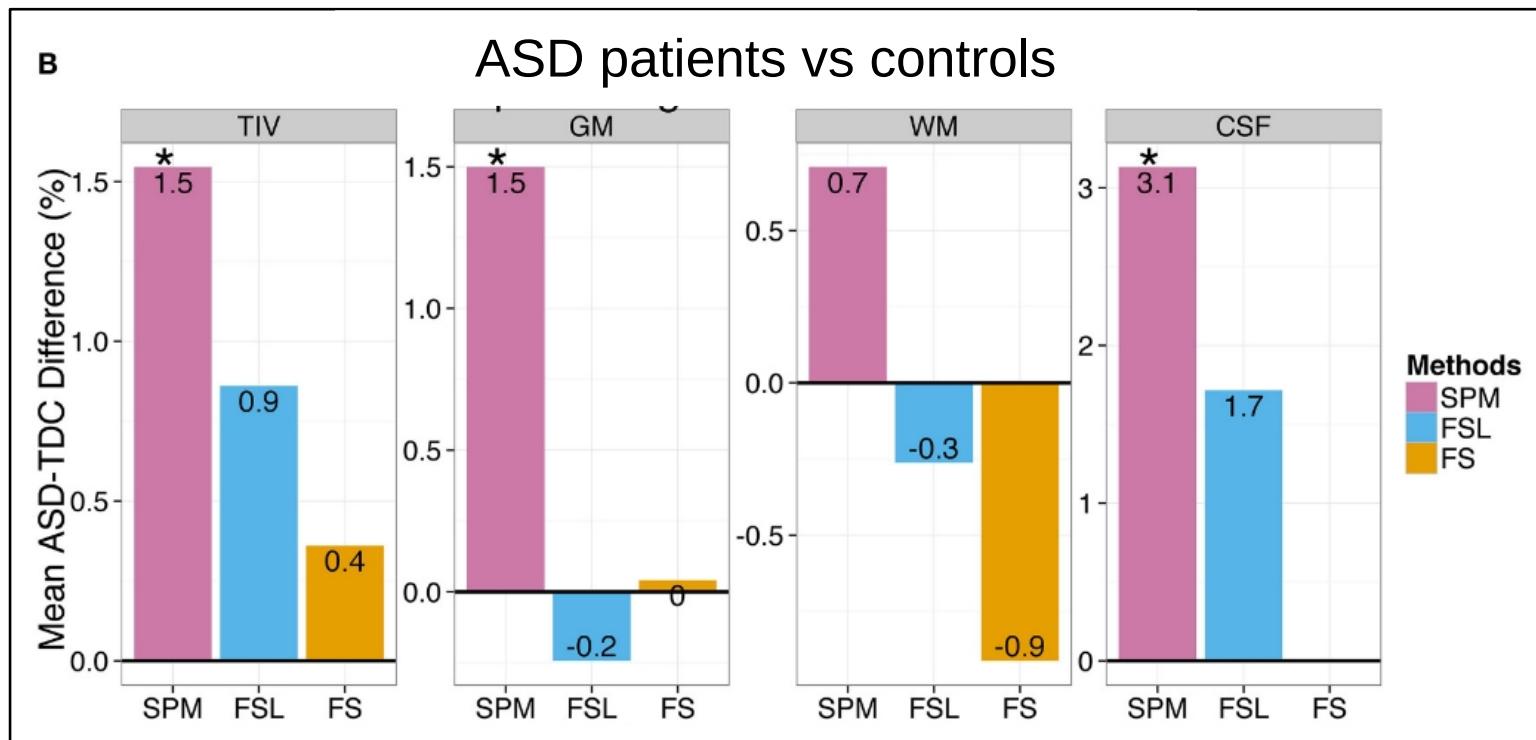
Y = Diff. of t-statistics



- ▷ Plots similar to expected variation if **independent** was fed into each package

Alex Bowring, Camille Maumet, Thomas Nichols

G. Katuwal, f. in Brain Imaging Methods, 2016

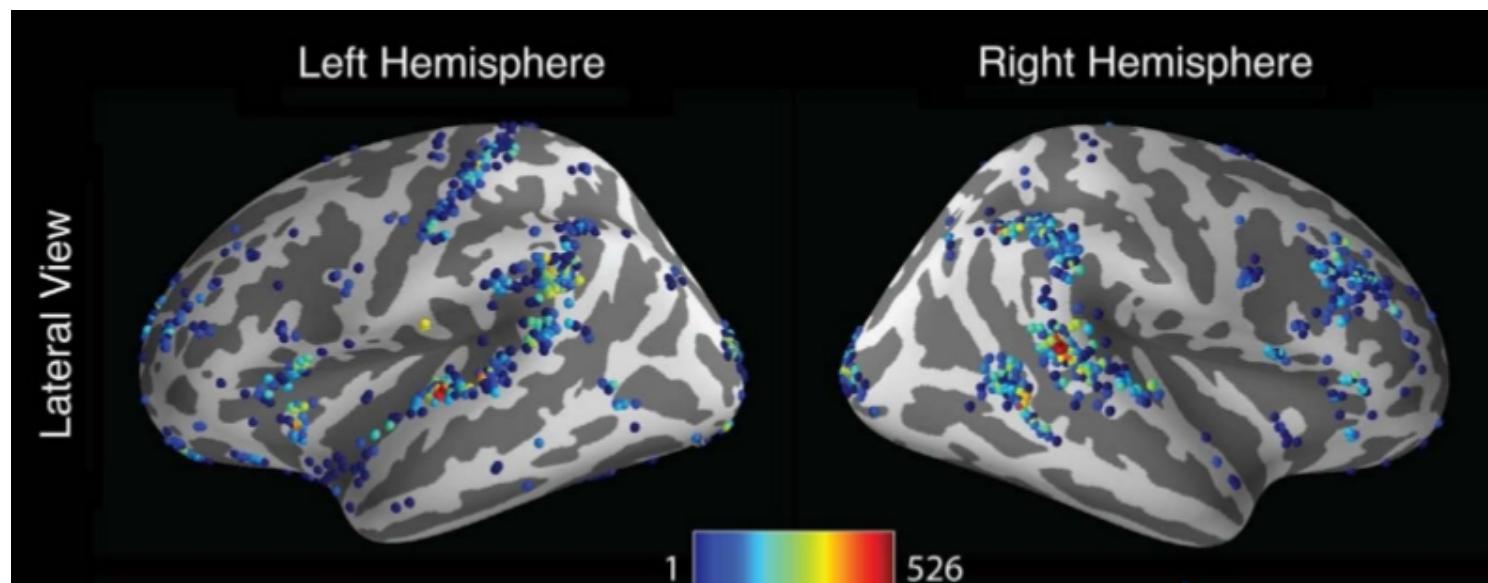


- Change from FSL to SPM?
- Change from v.1.12 to v.2.1 ?
- Change from cluster A to cluster B? Glatard et. al., finsc, 2015

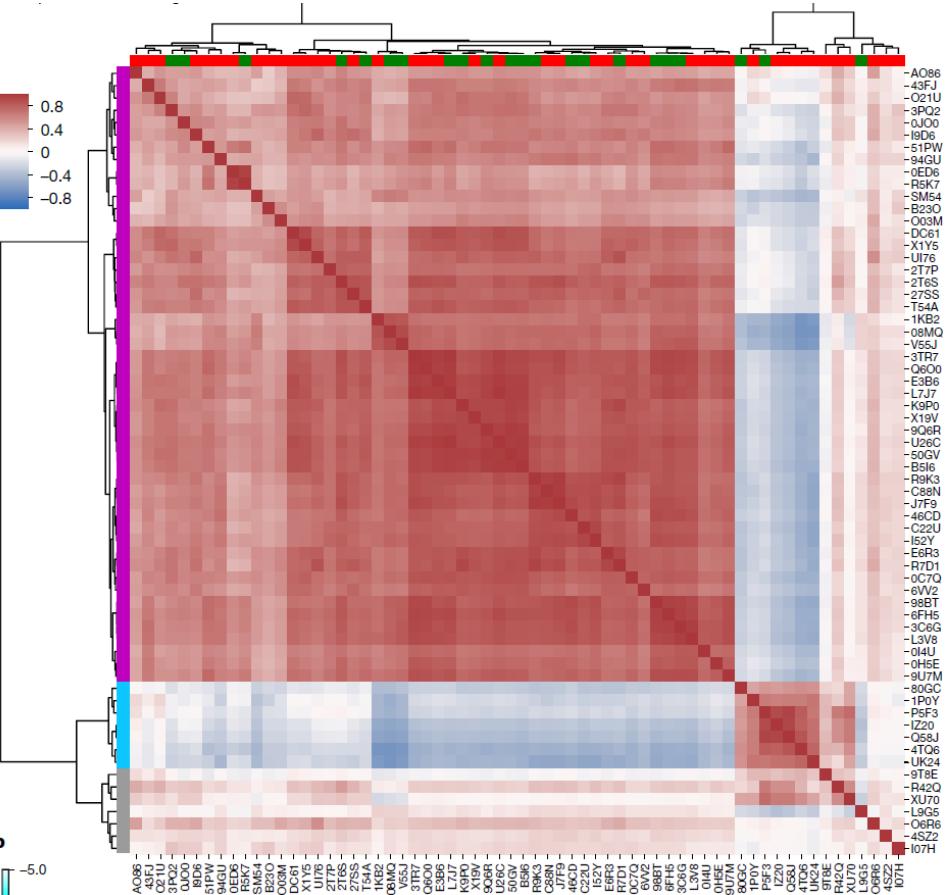
# Across possible pipelines

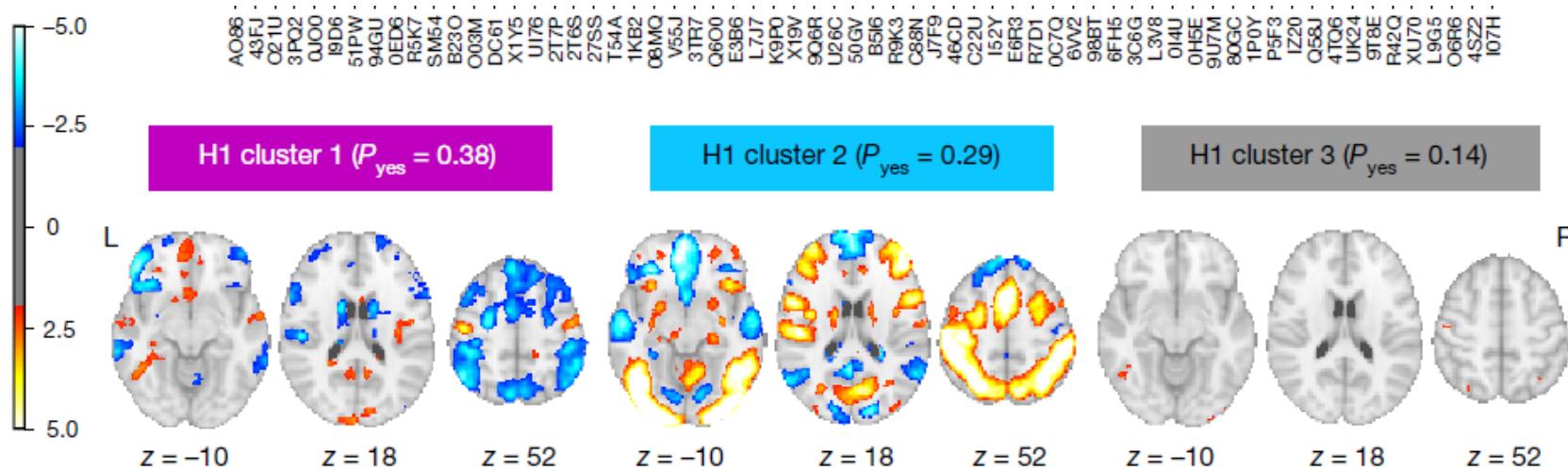
J. Carp, *f. Neuroscience*, 2012

- A **single** event-related fMRI experiment to a large number of unique analysis procedures
- Ten analysis steps for which multiple strategies appear in the literature : **6,912 pipelines**
- Plotting the maximum peak

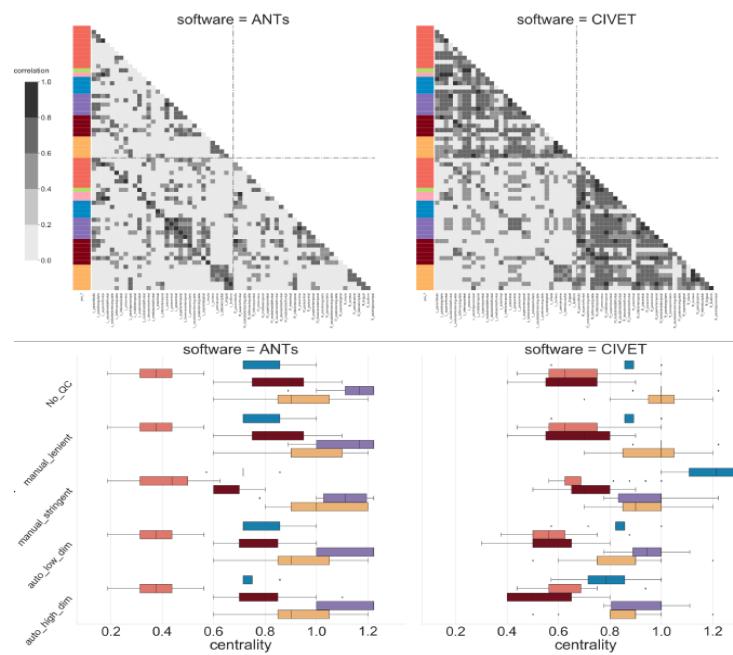


- 70 independent teams analyzing the same fMRI dataset
- No team had the same pipeline
- Results show three “clusters”
- Even within clusters decision to reject H0 varies





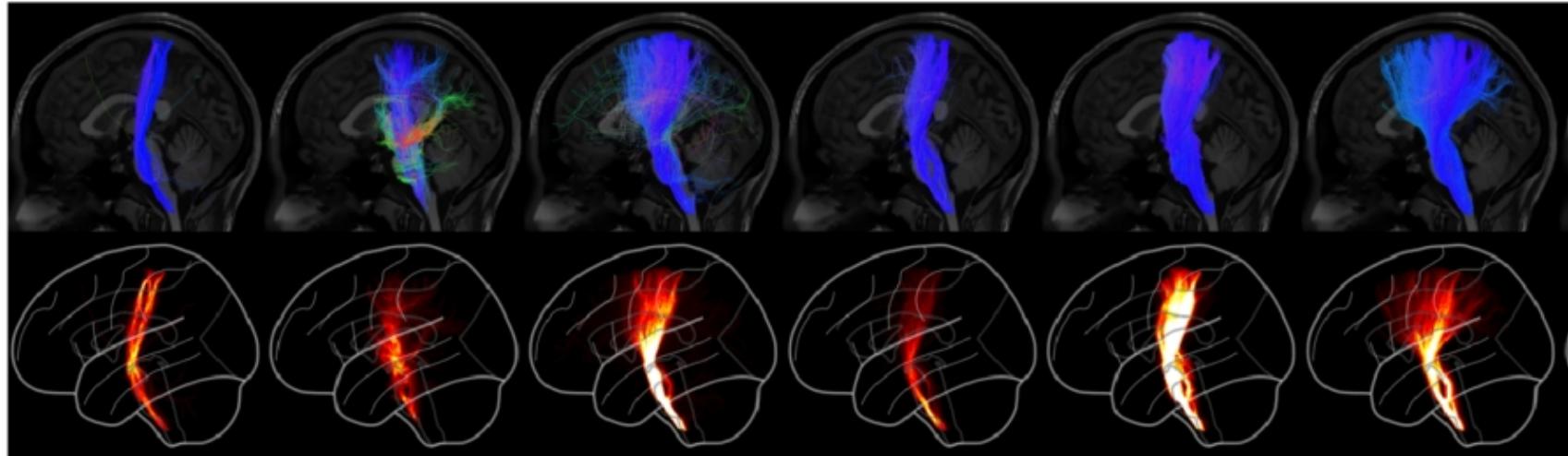
## Analytical flexibility: N. Bhagwat, 2020



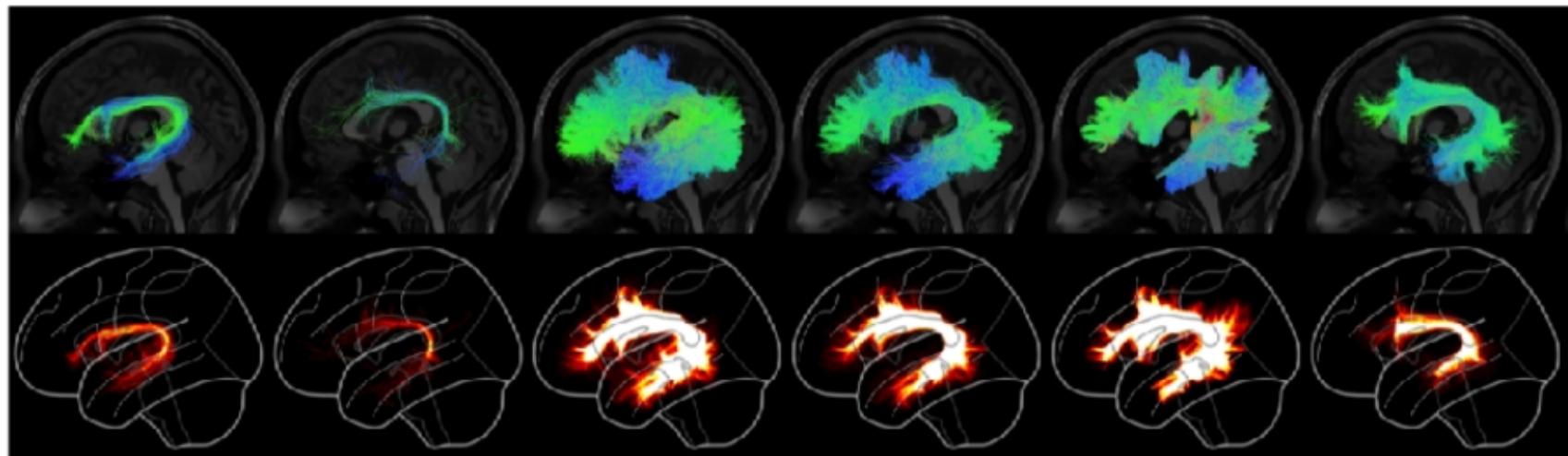
What happens when 42 groups dissect 14 white matter bundles on the same dataset?

*KG Schilling ... M Descoteaux 2020 (~150 authors)*

CST



AF



- A less rare case than usually thought !
- No license
- Database not containing what it describes
- Wrong QC – QC unreliable
- Headers of files are not correct (cf the Left/Right issue)
- Provenance of data is lost
- **SAM1 SAM2 SAM3:**  
<https://www.youtube.com/watch?v=N2zK3sAtr-4>

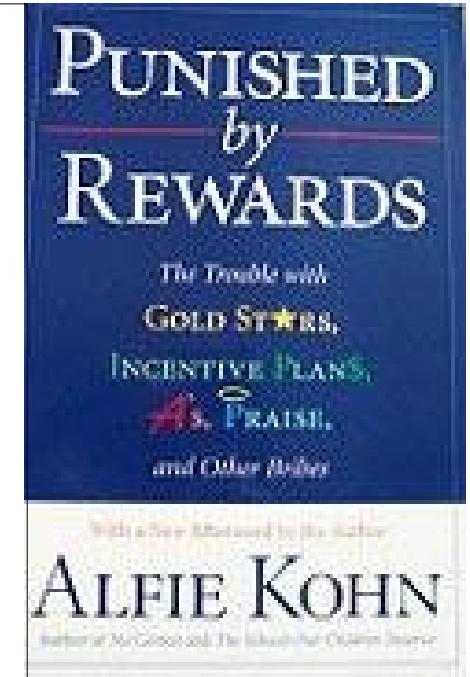
- From HCP:

"With the releases of FreeSurfer 7.X, there have been some regressions in surface placement performance when running FreeSurfer inside the HCP Pipelines. At this time, *I would recommend sticking with FreeSurfer 6.0* while we get these issues sorted out."

# Three causes

1. Poor statistical procedures
2. Issues in data and software
3. A cultural issue: Publication practices and research incentives

- Publication = the only “currency” for researchers, universities
- The high competition incites researchers to keep data and code as “assets” and to get as many authorships as possible
- The current incentive system promotes poorly reproducible research



ROYAL SOCIETY  
OPEN SCIENCE

[rsos.royalsocietypublishing.org](http://rsos.royalsocietypublishing.org)

Research



The natural selection  
of bad science

Paul E. Smaldino<sup>1</sup> and Richard McElreath<sup>2</sup>

<sup>1</sup>Cognitive and Information Sciences, University of California, Merced, CA 95343, USA

<sup>2</sup>Department of Human Behavior, Ecology, and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

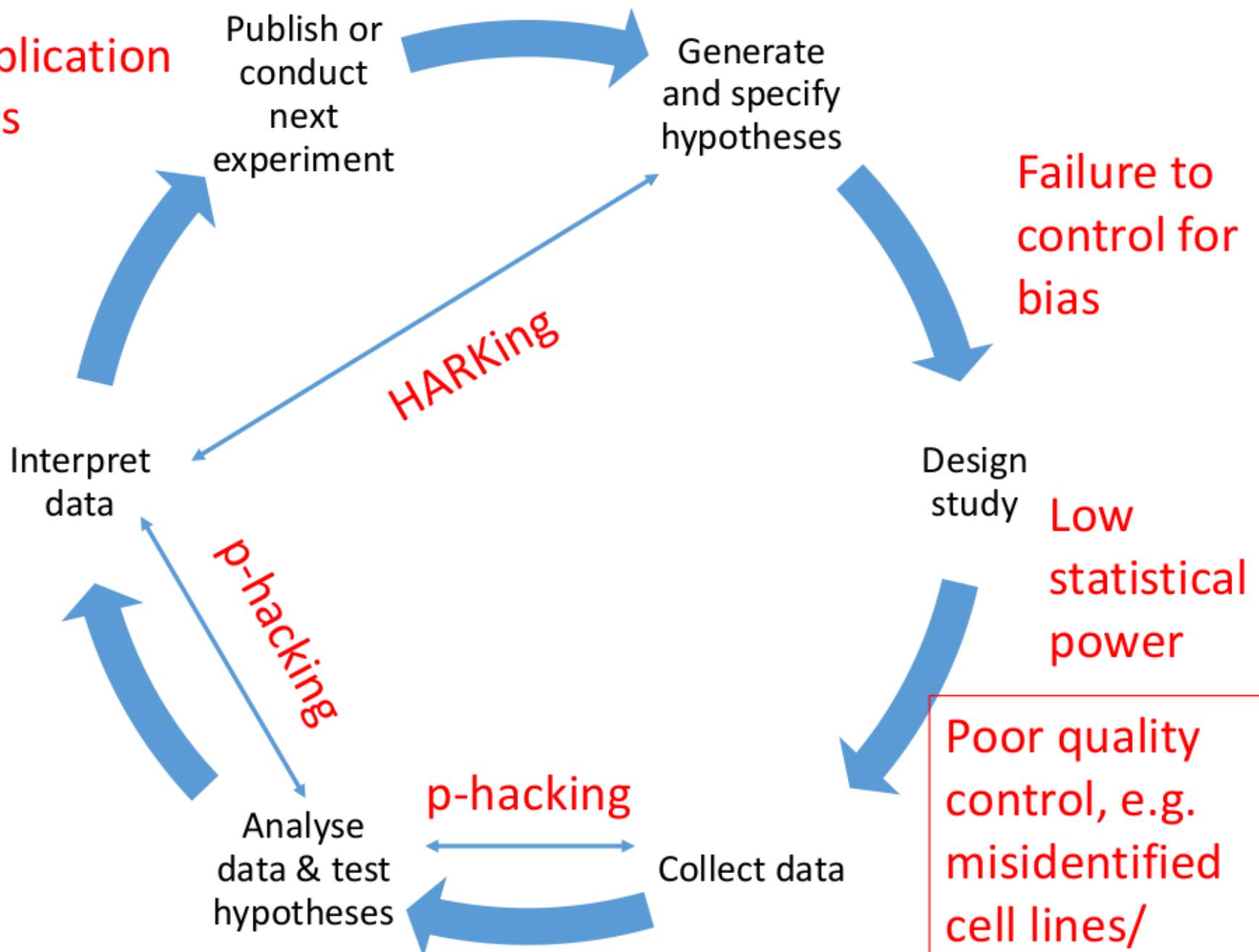
**Objectives** US national pilot study to

- (1) test the feasibility of online administration of the Bioethical Issues in Biostatistical Consulting (BIBC) Questionnaire
- (2) determine the prevalence and relative severity of a broad array of bioethical violations requests that are presented to biostatisticians by investigators seeking biostatistical consultations; and
- (3) establish the sample size needed for a full-size phase II study.

**Conclusion:** **clear evidence** that researchers make requests of their biostatistical consultants that are rated as **severe violations** and occur **frequently**

Wang et al. 2017. BMJ Open 7 (11): 2017.

# Publication bias



**Part I: Reproducibility: background**

**Part II : Etiology of Irreproducibility**

**Part III : Some therapeutic proposals**

- Make/use the tools that change the culture ex:  
used standard to harmonize data
- Improve training and open practices
- Develop community, training & standards
- Change the incentives

# schema.org

What if we had google power for datasets ?

# BIDS to RDF (NIDM)

dicomdir/

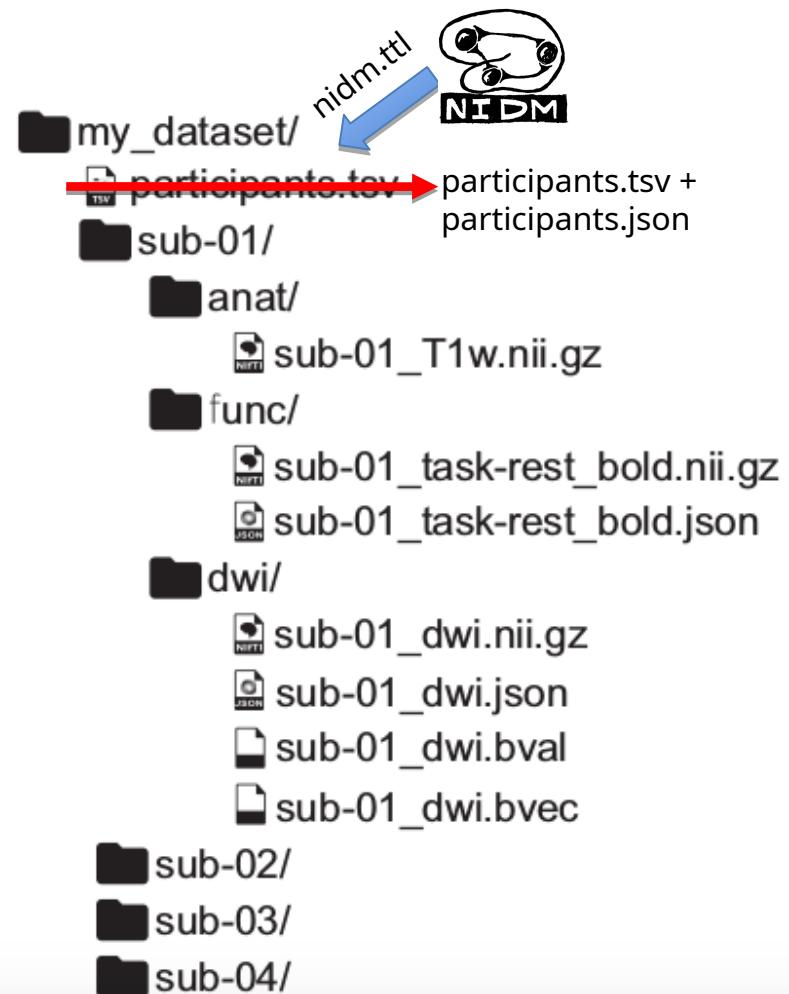
1208200617178\_22/

- 1208200617178\_22\_8973.dcm
- 1208200617178\_22\_8943.dcm
- 1208200617178\_22\_2973.dcm
- 1208200617178\_22\_8923.dcm
- 1208200617178\_22\_4473.dcm
- 1208200617178\_22\_8783.dcm
- 1208200617178\_22\_7328.dcm
- 1208200617178\_22\_9264.dcm
- 1208200617178\_22\_9967.dcm
- 1208200617178\_22\_3894.dcm
- 1208200617178\_22\_3899.dcm

1208200617178\_23/

1208200617178\_24/

1208200617178\_25/



# Participants.tsv file:

ADHD200 - Brown

sub-0026001
sub-0026002
sub-0026004
sub-0026005

## Participants.tsv file:

participant_id	gender	age	handedness	verbal_iq	performance_iq	full4_iq	qc_rest_1	qc_anatomical_1
26001	Male	16.92	Right	133	104	120	Pass	Pass
26002	Male	15.68	Right	106	106	107	Pass	Pass
26004	Female	14.99	Right	119	123	125	Pass	Pass
26005	Female	15.16	Right	116	131	126	Pass	Pass
26009	Male	16.91	Left	113	81	97	Pass	Pass
26014	Female	16.21	Right	101	102	102	Pass	Pass
26015	Female	15.2	Right	127	98	113	Pass	Pass
26016	Male	16.07	Right	120	96	109	Pass	Pass
26017	Female	14.56	Right	95	87	89	Pass	Pass
26022	Male	17.83	Right	105	111	109	Pass	Pass
26024	Female	17.77	Right	89	83	85	Pass	Pass
26027	Female	11.28	Right	108	103	106	Pass	Pass
26030	Female	14.51	Right	121	119	123	Pass	Pass
26039	Female	14.19	Right	125	117	124	Pass	Pass
26040	Female	13.67	Right	111	93	102	Pass	Pass
26041	Female	13.68	Right	129	120	128	Pass	Questionable
26042	Female	13.82	Right	106	110	109	Pass	Pass

# Describe data with standard terminology

## Link data

### RDF triples form (linked) graphs

<Bob> <is a> <person>.

<Bob> <is a friend of> <Alice>.

<Bob> <is born on> <the 4th of July 1990>.

<Bob> <is interested in> <the Mona Lisa>.

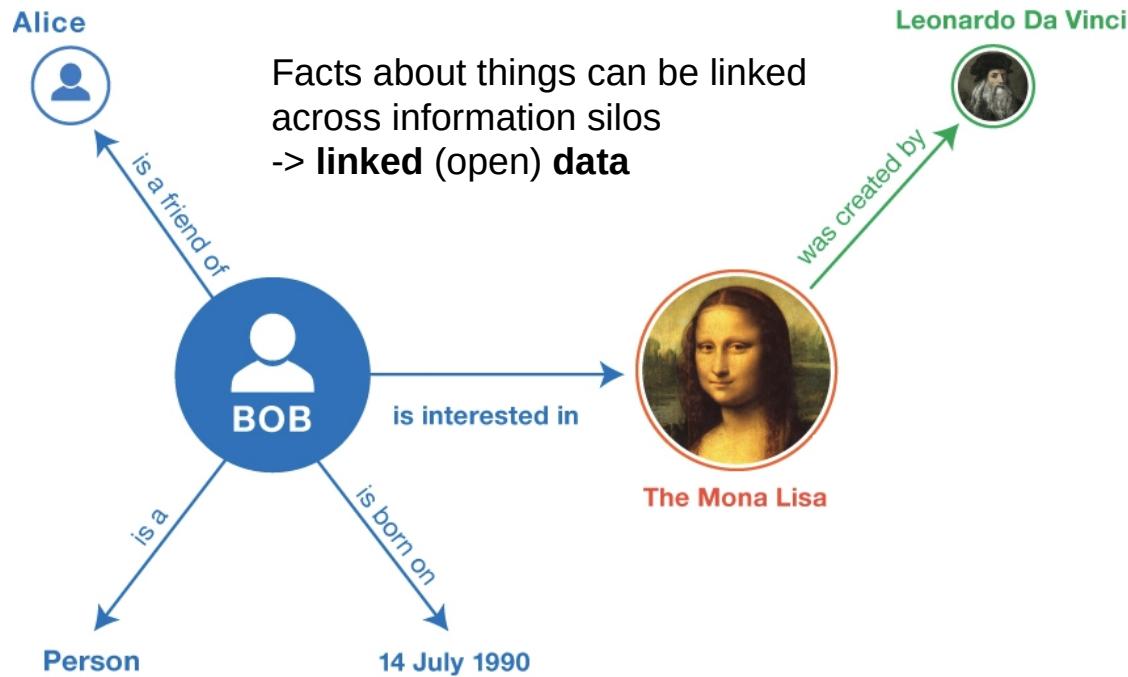
<the Mona Lisa> <was created by> <Leonardo da Vinci>.

vocabularies? -> many!

Theory



Practice



<https://www.w3.org/TR/rdf11-primer/>

# Tools / infrastructure

- How much should I invest in learning the tools?
  - Which tools ? Can you take them with you?
  - What are your personal and technical constraints / priorities?
- Be part of a community with similar mindsets
  - You may be seating in an environment that is not conducive of reproducible research
- Re-think reproducibility in terms of efficiency for your PI next student if not for yourself

# Tools / infrastructure

- How much should I invest in learning the tools?
  - Which tools ? Can you take them with you?
  - What are your personal and technical constraints / priorities?
- Be part of a community with similar mindsets
  - You may be seating in an environment that is not conducive of reproducible research
- Re-think reproducibility in terms of efficiency for your PI next student if not for yourself

# Tools / infrastructure

- How much should I invest in learning the tools?
  - Which tools ? Can you take them with you?
  - What are your personal and technical constraints / priorities?
- Be part of a community with similar mindsets
  - You may be seating in an environment that is not conducive of reproducible research
- Re-think reproducibility in terms of efficiency for your PI next student if not for yourself

# What can we do ?

- Make the tools that change the culture
- Improve open & reproducible practices
- Develop community, training & standards
- Change the incentives

# Pre-registration -registered report

- Promoted by many (eg Chris Chambers)
- Explain in details your hypothesis methods and software
- The pre-registered report is peer reviewed by a journal
- If accepted, results (even null results) will be published
- Does not preclude exploratory analyses
- A good idea even just for yourself

**Is there any reason not to do it ?**

# Solution: Preregistration with OSF

[Search](#)[Support](#)[Donate](#)[Sign Up](#)[Sign In](#)

## There's a better way to manage your research

OSF is a free, open platform to support your research and enable collaboration.

COMMENT · 10 SEPTEMBER 2019

## What's next for Registered Reports?

Reviewing and accepting study plans before results are known can counter perverse incentives. Chris Chambers sets out three ways to improve the approach.

Chris Chambers

- The rise of Git (GitHub, GitLab, ...) repositories
  - Many are open, collaborative
  - From lab-based “one software” to collaborative open source
- The rise of “FAIR” data format and containers
  - BIDS
  - BIDSapps
- The (slower) rise of standardized provenance information
  - NIDM - W3C prov



- Sustainability is in all the minds – governance?
- Critically linked to sustainability
  - 8. Be transparent in governance and operations
  - 9. Involve community in governance and decision making
  - 10. Be transparent on sustainability - financial and technical

**OPEN**

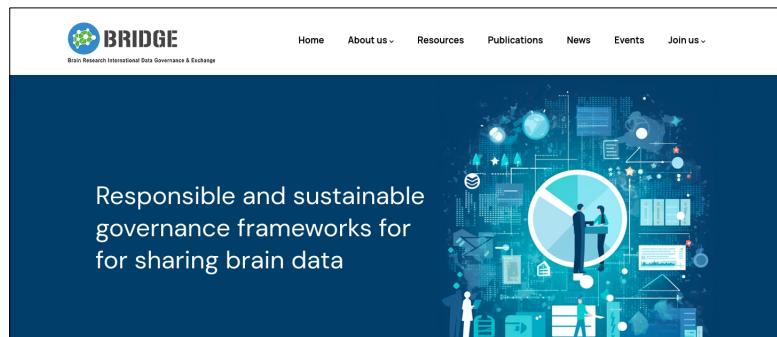
**COMMENT**

## Recommendations for repositories and scientific gateways from a neuroscience perspective

**It's the governance, stupid!: TAPIC: a governance framework to strengthen decision making and implementation [Internet]**

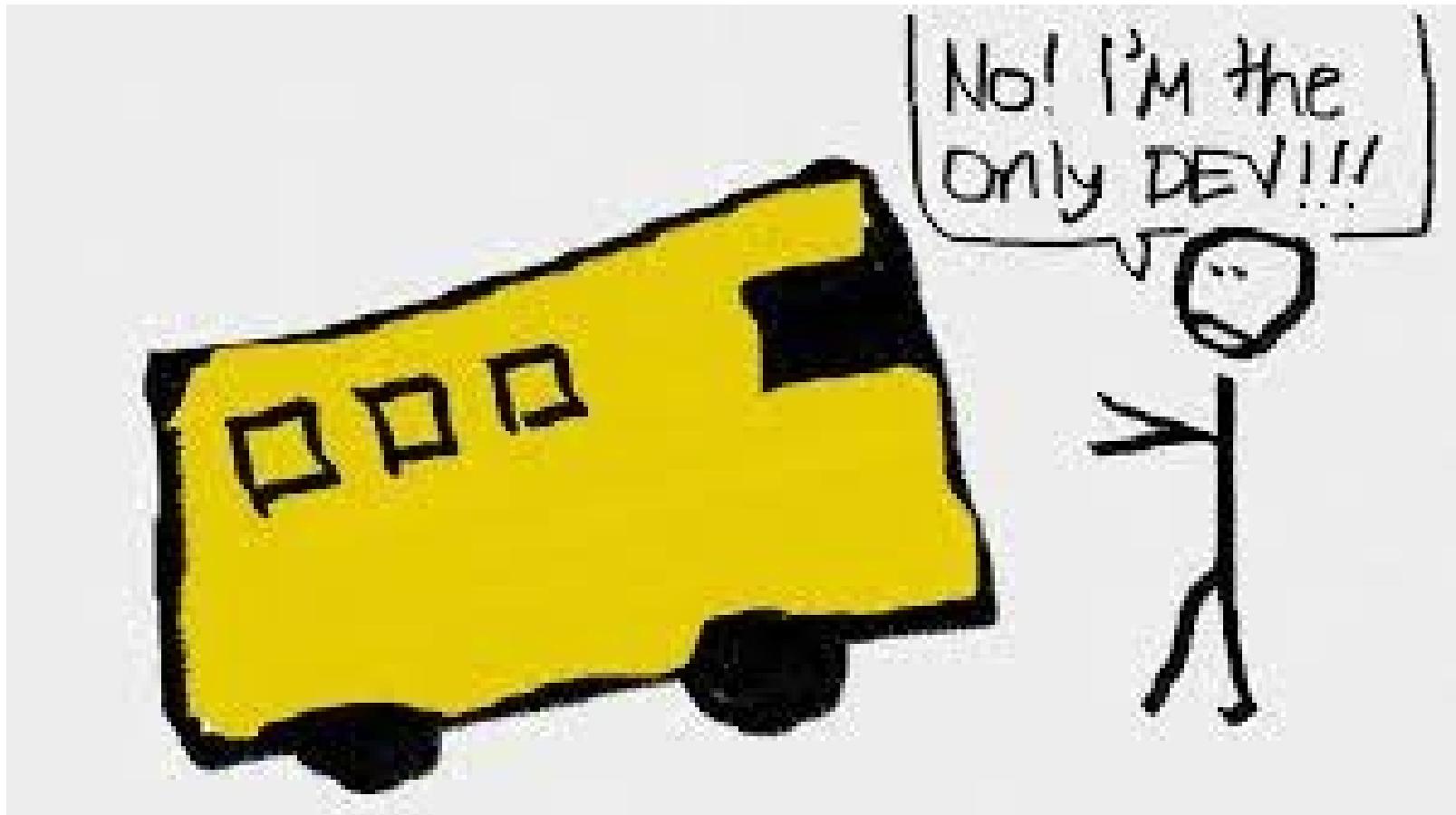
Scott L. Greer <sup>1</sup>, Nikolai Vasev <sup>2</sup>, Holly Jarman <sup>1</sup>, Matthias Wismar <sup>3</sup>, Josep Figueras <sup>3</sup>

Copenhagen (Denmark): European Observatory on Health Systems and Policies; 2019.  
European Observatory Policy Briefs.



The screenshot shows the BRIDGE (Brain Research International Data Governance & Exchange) website. The header includes the logo, navigation links for Home, About us, Resources, Publications, News, Events, and Join us, and a search bar. The main content area features a dark blue background with a circular graphic showing two figures in a digital environment, and the text "Responsible and sustainable governance frameworks for sharing brain data".

# Distributed development and governance: Nilearn



- Software / Code should receive equivalent level of scrutiny than narrative
- Software / Code should have an equivalent level of recognition
- Software need tests
  - Unit testing for refactoring
  - Simulated / known results for integration
- Infrastructures for reviewing and publishing notebooks

# Building Research Platforms

## SCIENTIFIC DATA

110110  
0111101  
11011110  
011101101

OPEN

### Comment: High-quality science requires high-quality open data infrastructure

Susanna-Assunta Sansone<sup>1</sup>, Patricia Cruse<sup>2</sup> & Mark Thorley<sup>3</sup>

Received: 26 January 2018

Accepted: 29 January 2018

Published: 27 February 2018



OPEN

COMMENT

### Recommendations for repositories and scientific gateways from a neuroscience perspective

Malin Sandström<sup>1</sup>✉, Mathew Abrams<sup>1</sup>, Jan G. Bjaalie<sup>2</sup>, Mona Hicks<sup>3</sup>, David N. Kennedy<sup>4</sup>, Arvind Kumar<sup>5</sup>, Jean-Baptiste Poline<sup>6</sup>, Prasun K. Roy<sup>7</sup>, Paul Tiesinga<sup>8</sup>, Thomas Wachtler<sup>9</sup> & Wojtek J. Goscinski<sup>10,11</sup>

- Make the tools that change the culture
- Improve training and open practices
- Develop community, training & standards
- Change the incentives

- The rise of Git (GitHub, GitLab, ...) repositories
  - Many are open, collaborative
  - From lab-based “one software” to collaborative open source
- The rise of “FAIR” containers
  - BIDS
- The (slow) rise of standardized provenance information
  - NIDM - W3C prov

# Training: NIH P41 ReproNim

*ReproNim*

HOME

PROJECTS

TRAINING

CONTACT

*ReproNim: A Center for Reproducible  
Neuroimaging Computation  
(Discover, Replicate, Innovate)Repeat*

TELL ME MORE

**REPRONIM INTRO**

**DATA PROCESSING**

**REPRODUCIBILITY BASICS**

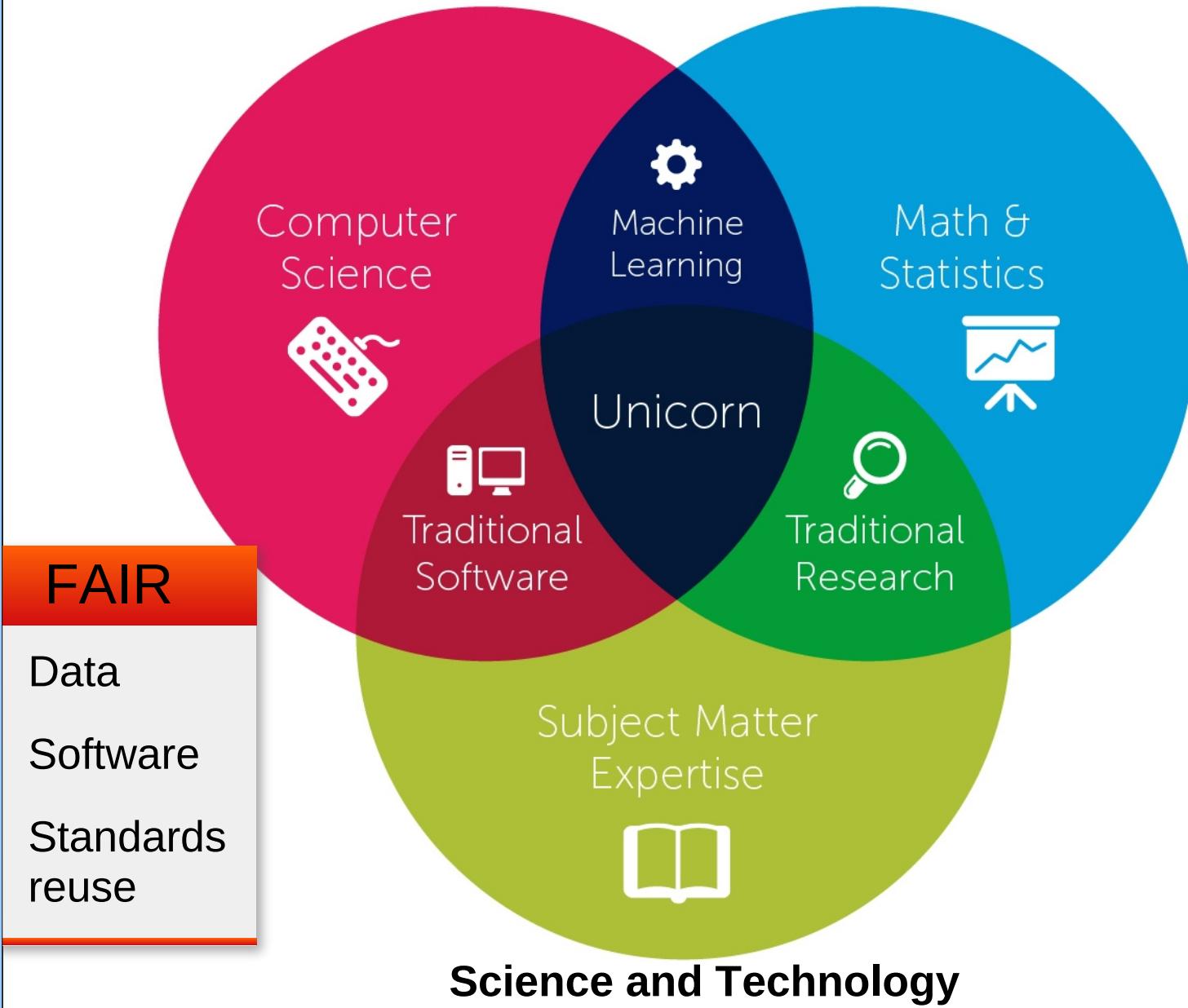
**STATISTICS**

**FAIR DATA**

# Data Science

Ethical Scholarly communications  
Epistemology / lessons from the past  
How to collaborate and teach

Research



- Make the tools that change the culture
- Improve training and open practices
- Develop community standards
- Change the incentives

# Changing the publication model



Reproducibility: A tragedy of errors, Allison et al, 2016, Nature

# Data & Software research objects

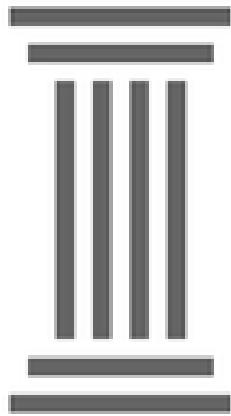
DATA	SOFTWARE
DataCite	CiteSoftware
FORCE11 JDDCP	software.ADORE
Scientific Data / ...	JOSS/ ...
Data packages	Containers
Governance (GDPR vs OpenNeuro?)	Governance (Scipy vs Numpy ?)

# What is needed ?

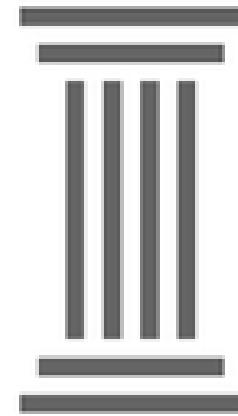


An open publishing platform for  
the neuroscience community

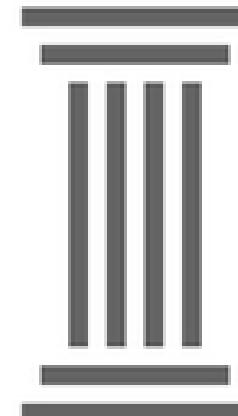
APERTURE



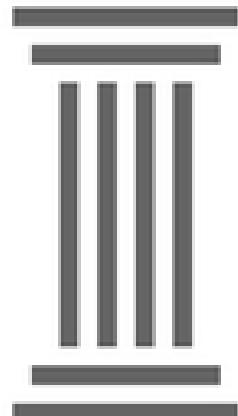
High-Quality  
Publications



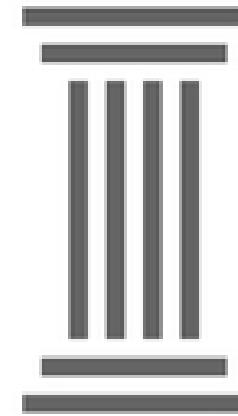
Open-Access



Low Cost



Diversity of  
published  
Research Objects



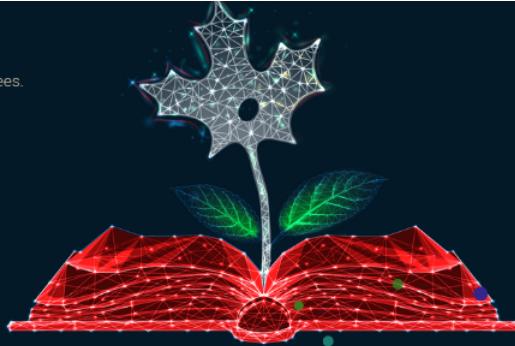
Community  
Driven

# NeuroLibre

NeuroLibre is a preprint server for **reproducible** data analyses.

Committed to publishing curated Jupyter notebooks with zero article processing charges or subscription fees.

 Submit a preprint to NeuroLibre →



 Explore Reproducible Preprints

 Documentation

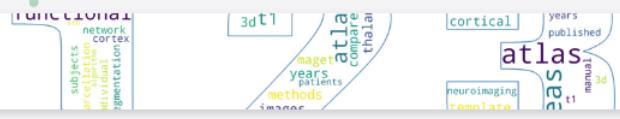
 Learn More

 Parcellating the parcellation issue - a proof of concept for reproducible analyses using Neurolibre

 NiMARE: Neuroimaging Meta-Analysis Research Environment

[Jupyter Notebook](#) [JavaScript](#) [Python](#)

Taylor Salo, Tal Yarkoni, Thomas E. Nichols, Jean-Baptiste Poline, Murat Bilgel, Katherine L. Bottenhorn, Dorota Jarecka, James D. Kent, Adam Kimbler, Dylan M. Nielson, Kendra M. Oudyk, Julio A. Peraza, Alexandre Pérez, Puck C. Reenders, Julio A. Yanes, Angela R. Laird



 An interactive meta-analysis of MRI biomarkers of myelin

[Jupyter Notebook](#)

Matteo Mancini, Agah Karakuza, Julien Cohen-Adad, Mara Cercignani, Thomas E. Nichols, Nikola Stikov



# Thank you

- Lab@McGill: <https://neurodatascience.github.io/>
- **McGill-Concordia-UdM** colleagues: T. Glatard, P. Bellec, C. Greenwood, A. DeGuise and many others
- **ReproNim** colleagues: D. Kennedy, D. Keator, S. Ghosh, M. Martone, J. Grethe, M. Hanke, Y. Halchenko
- **Berkeley** colleagues: S. Van der Walt, M. Brett, J. Millman, Dan Lurie, M. D'Esposito, et al
- **Pasteur** colleagues: G. Dumas, R. Toro, T. Bourgeron, and others
- **Paris** colleagues: B. Thirion, G. Varoquaux, V. Frouin, et al
- **Funders**: HBHL, HBHL NeuroHub, NIH, NIMH, NSERC, Compute Canada