Reproducibility: what you need to know from epistemology to statistics

Jean-Baptiste Poline

MNI, Brain Imaging Centre, McGill, Montreal
HWNI, UC Berkeley

Part I: Reproducibility: background

Part II : Etiology of Irreproducibility

Part III : Epistemology and statistics

Part IV :Some therapeutic proposals

Part I: Reproducibility: background

Part II : Etiology of Irreproducibility

Part III : Epistemology and statistics

Part IV :Some therapeutic proposals

Amgen replication

- 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

JAMA Psychiatry 2017

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.

Imaging Genetics GWAS

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

Funding agencies reaction

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.

The problem is widespread

Essay

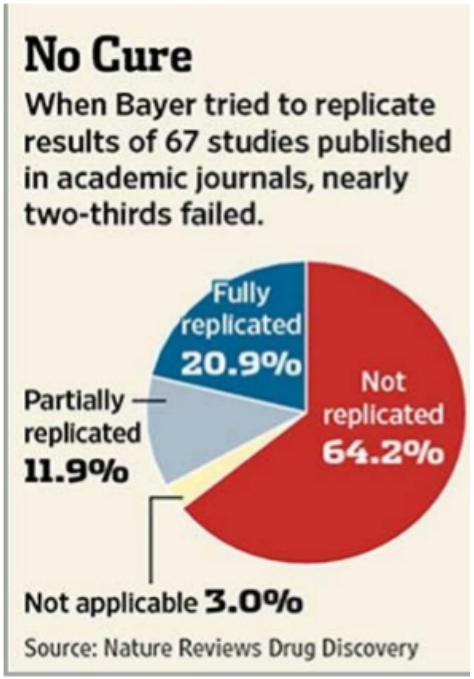
Why Most Published Research Findings Are False

John P. A. Ioannidis

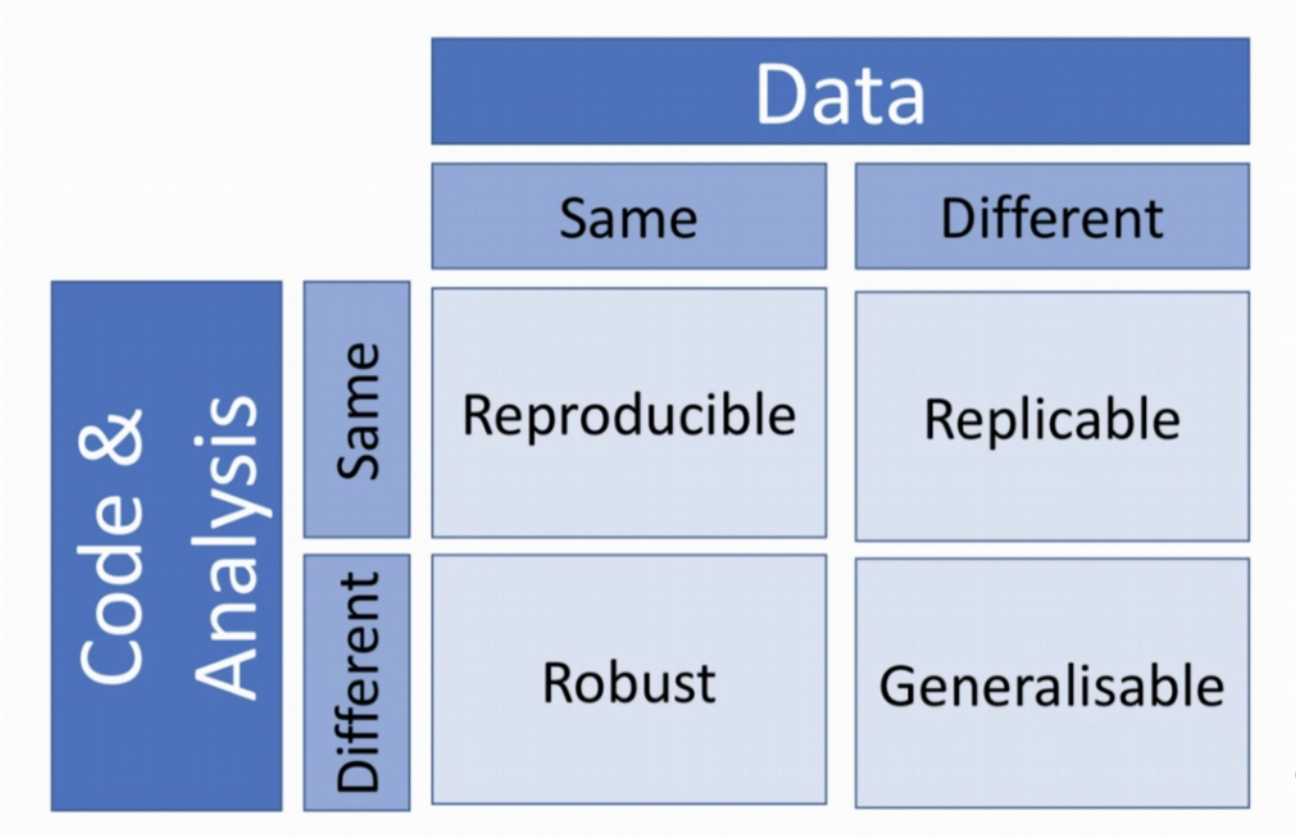
2005. PLoS Medicine, 2(8), e124. doi: 10.1371/journal.pmed.0020124

"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*



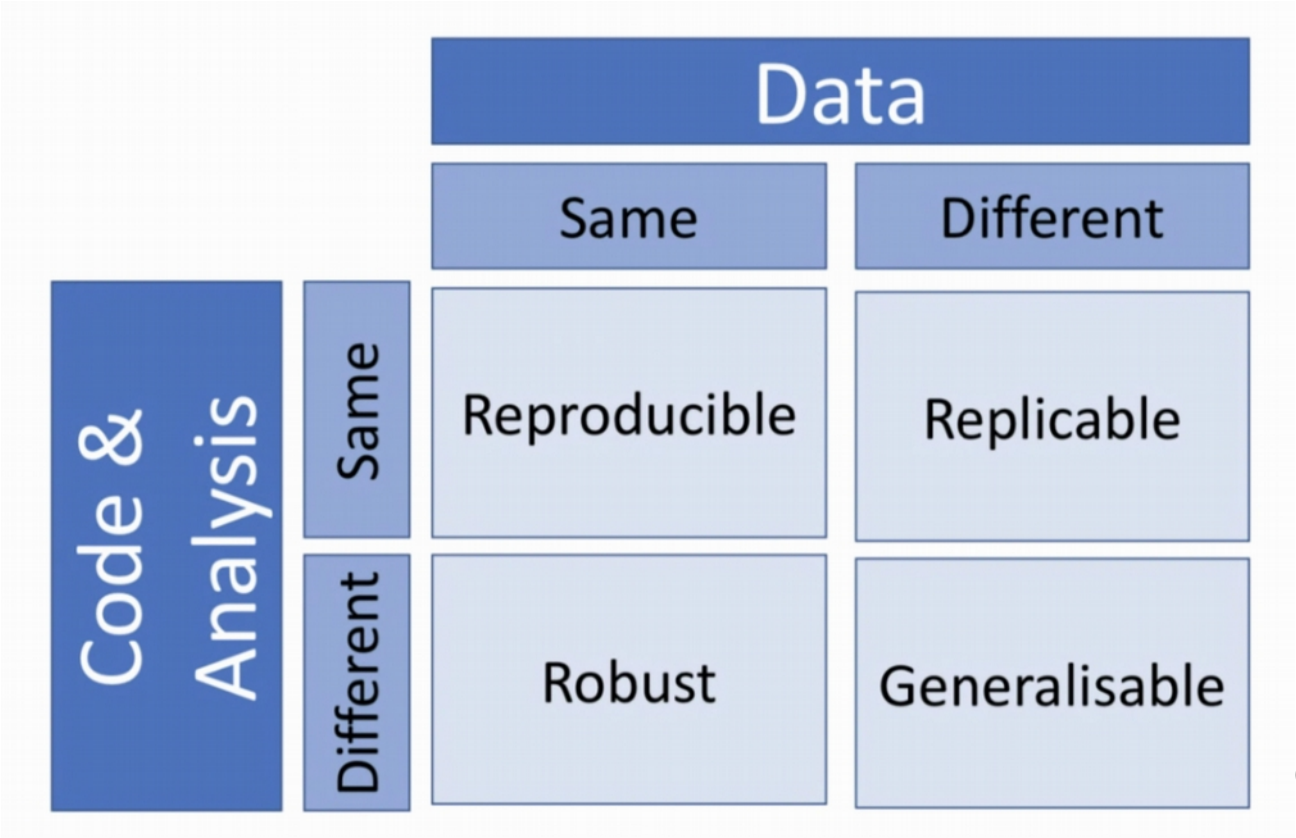
Some Definitions



Credit: J.Pineau

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

Some Definitions



Credit: J.Pineau

- My (prefered) 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 : Epistemology and statistics

Part IV :Some therapeutic proposals

Three causes

1. Poor statistical procedures

2. Issues in data and software

3. A cultural issue: Publication practices and research incentives

Three causes

1. Poor statistical procedures

2. Issues in data and software

3. A cultural issue: Publication practices and research incentives

Understanding statistics

Consider a typical medical research study, for example designed to test the efficacy of a drug, in which a null hypothesis H_0 ('no effect') is tested against an alternative hypothesis H_1 ('some effect'). Suppose that the study results pass a test of statistical significance (that is P-value <0.05) in favor of H_1. What has been shown?

1. H_0 is false.
2. H_1 is true.
3. H_0 is probably false.
4. H_1 is probably true.
5. Both (1) and (2).
6. Both (3) and (4).
7. None of the above.

Results

Table 1 Quiz answer profile

Answer	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Number	8	0	58	37	6	69	12
Percent	4.2	0	30.5	19.5	3.2	36.3	6.3

Westover, M.B., Westover, K., Bianchi, M., 2011. Significance testing as perverse probabilistic reasoning. BMC medicine 9, 20.

P-values ...

- P-values can indicate how incompatible the data are with a specified statistical model.
- P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
- Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.

P-values ...

- P-values can indicate how incompatible the data are with a specified statistical model.
- P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
- Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.

P-values ...

- P-values can indicate how incompatible the data are with a specified statistical model.
- P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
- Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.

- Proper inference requires full reporting and transparency. P-values and related analyses should not be reported selectively.
- A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
- By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

- Proper inference requires full reporting and transparency. P-values and related analyses should not be reported selectively.
- A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
- By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

- Proper inference requires full reporting and transparency. P-values and related analyses should not be reported selectively.
- A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
- By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

P-Hacking

- A long history - Simmons and Simonsohn 2011
- Most often not intentional - and can be difficult to detect
  - As soon as some summary are seen ?
  - Necessity to “visualize data”
- P-hacking test
  - Based on a well known fact (?) : p-values are uniformly distributed !
  - P-curves : Simonsohn et al, 2014
- Evil P-value

  http://www.repronim.org/module-stats/03-p-values/

  github.com/repronim/module-stats/notebooks/evil-p.ipynb
- P-hacking exercise

  github.com/repronim/module-stats/notebooks/P-value-exercise.ipynb

P-hacking - solutions ?

- Pre-registration
- Ban p-values
- Change \alpha
- Complement with other statistics !

Significance

The lack of reproducibility of scientific research undermines public confidence in science and leads to the misuse of resources when researchers attempt to replicate and extend fallacious research findings. Using recent developments in Bayesian hypothesis testing, a root cause of nonreproducibility is traced to the conduct of significance tests at inappropriately high levels of significance. Modifications of common standards of evidence are proposed to reduce the rate of nonreproducibility of scientific research by a factor of 5 or greater.

Johnson, V.E. (2013). Revised standards for statistical evidence. PNAS 110, 19313–19317.

Effect size - definition

What is the non standardized effect ?

Imagine 2 groups (1 and 2):

$$\mu = \bar{x_1} - \bar{x_2}$$

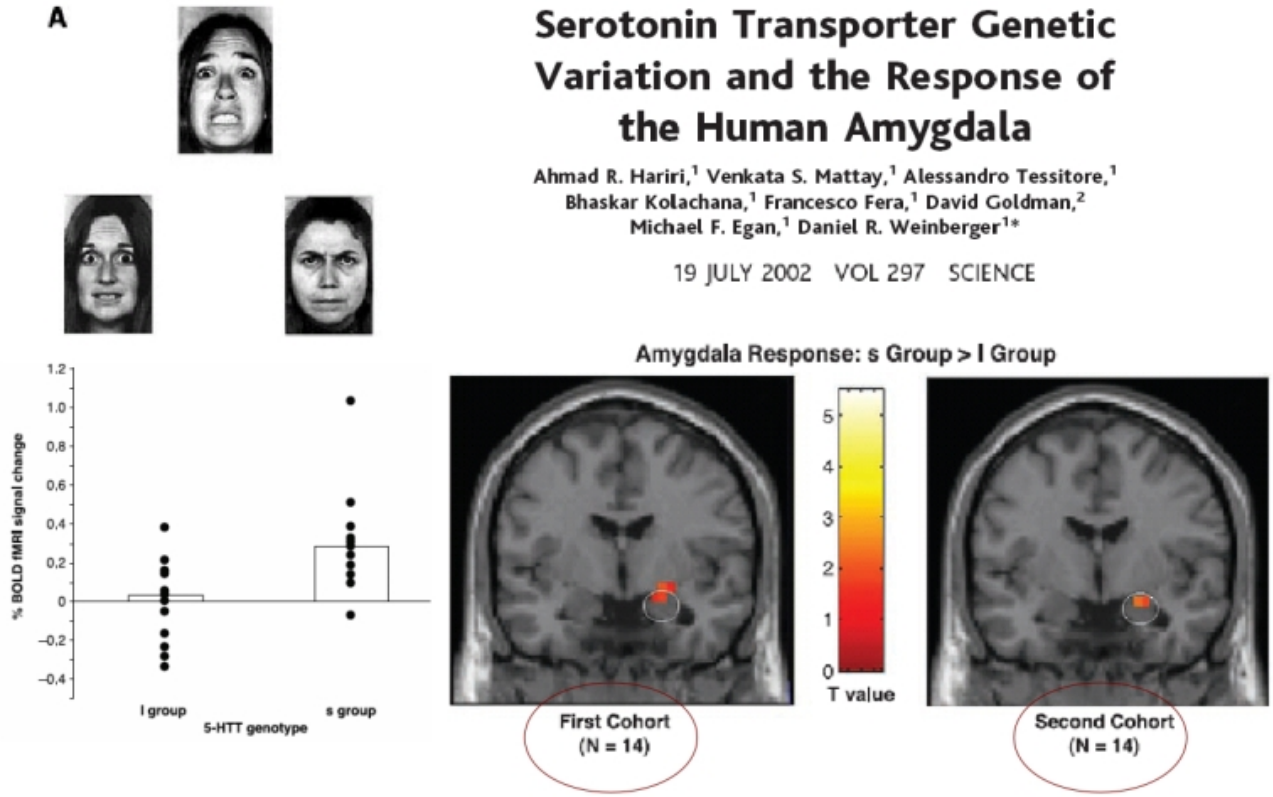
What is the standardized effect ? (eg Cohen's d)

$$d = \frac{\bar{x_1} - \bar{x_2}}{\sigma} = \frac{\mu}{\sigma}$$

"Z" : Effect accounting for the sample size

$$Z = \frac{\mu}{\sigma/\sqrt{n}}$$

Small N studies



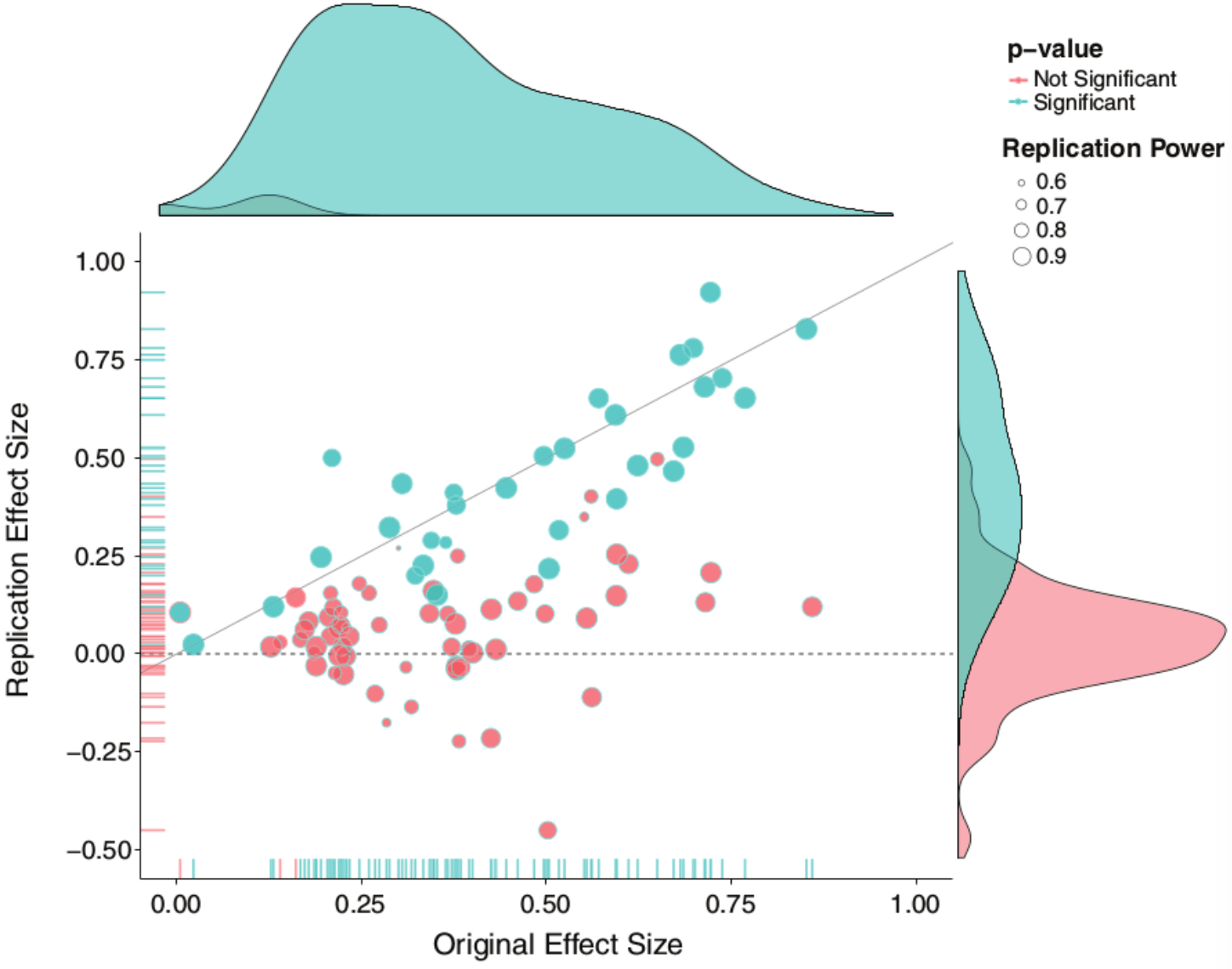
- Authors report
  $m_1 = .28, m_2 = .03, \mathrm{SDM}_1 = 0.08, \mathrm{SDM}_2 = 0.05, N_1 = N_2 = 14$
- How do we compute the effect size ?

Effect size: in practice ?

First, compute the standard deviation of the data from the SDM

- get σ from SDM : $\sigma = \sqrt{14-1} \times \mathrm{SDM}$
- Combine the σ to have one estimation across the groups
  - formula easy to recompute or find
- $\sigma = \sqrt{14-1} \times \mathrm{SDM}$, $d = \frac{m_1 - m_2}{\sigma} = 1.05$
- What is the percentage of variance explained ?
- Write the estimated model: $Y = [1 \ldots 1]^t [m_1 - m_2] + \mathrm{residual}$
- Compute the total sum of square $Y^t Y$, then the proportion:
- $V_e = \frac{(n_1+n_2)(m_1-m_2)^2}{n_1 s_1^2 + n_2 s_2^2 + (n_1+n_2)(m_1-m_2)^2} > 40\%$

B. Nosek, Estimating the reproducibility of psychological science, Science 2015

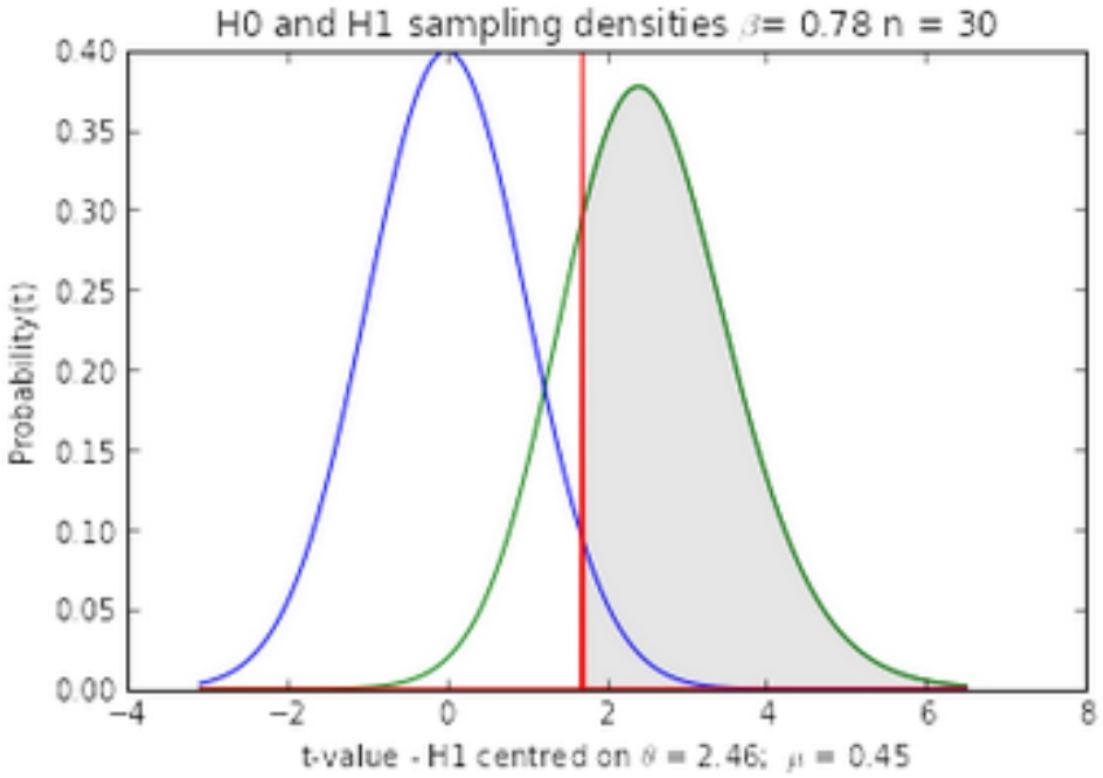


* The mean **effect size** (r) of the replication effects (M r = 0.197, SD = 0.257) **was half the magnitude** of the mean effect size of the original effects (M r = 0.403, SD = 0.188)

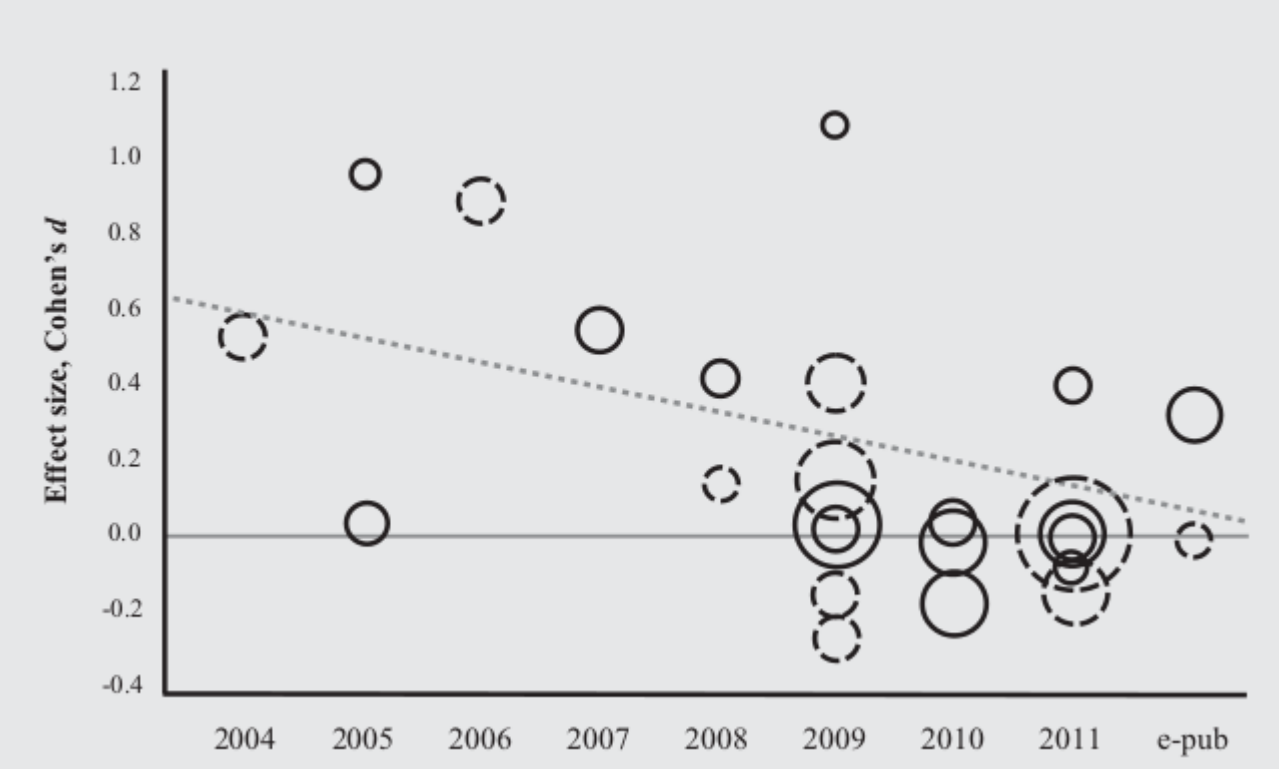
* **39%** of effects were rated to have replicated the original effect

Power - definition

Decision/H	H0 True	H1 True
reject	α (type I)	$1 - \beta$ (Power)
not reject	1 - α	β (type II)



Statistics: ~~The~~ One problem



Molendijk, 2012: BDNF and hippocampal volume

See also : Mier, 2009: COMT and DLPFC

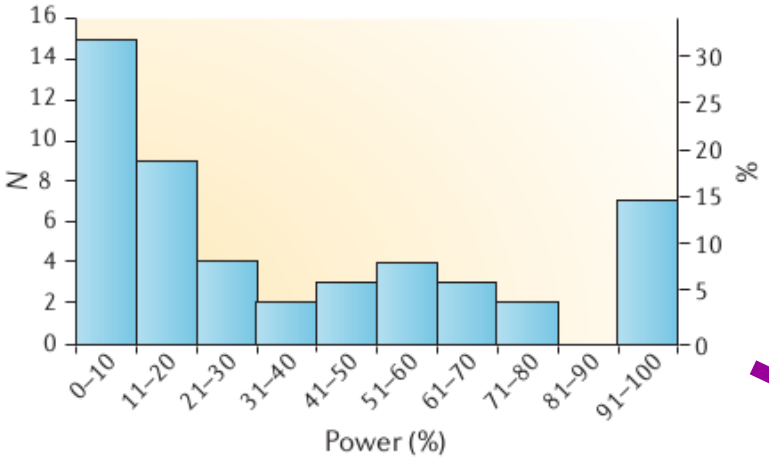
Power issues

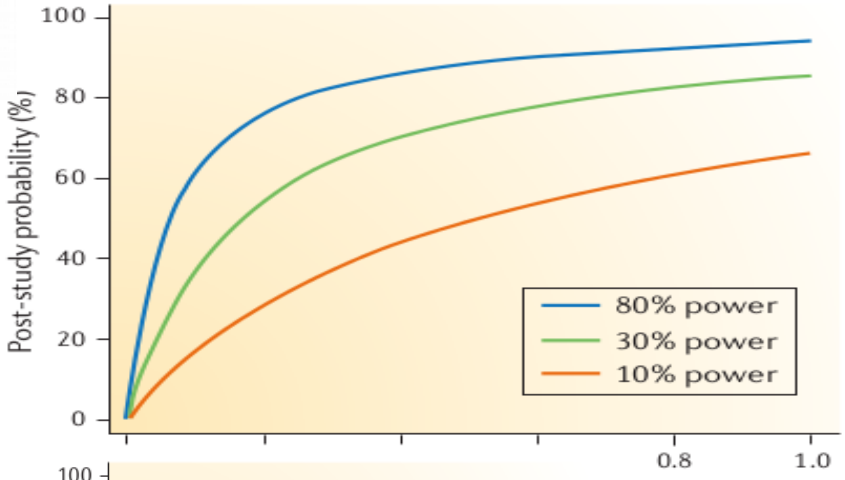
Open access, freely available online

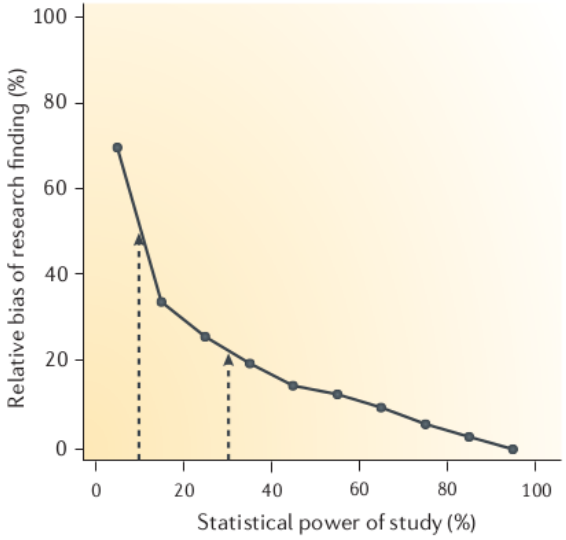
Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

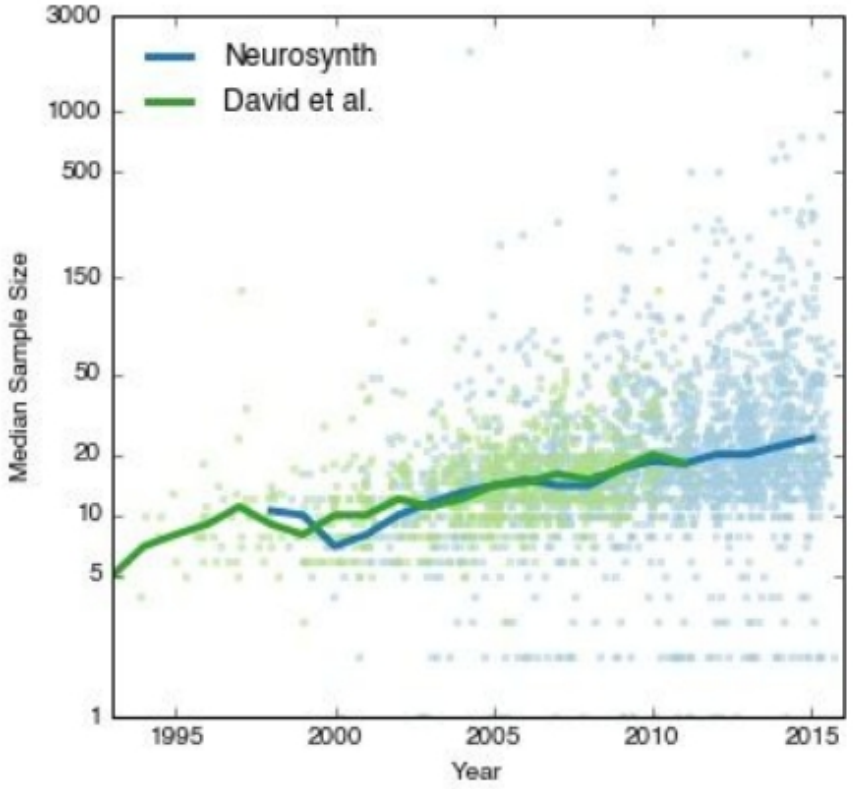


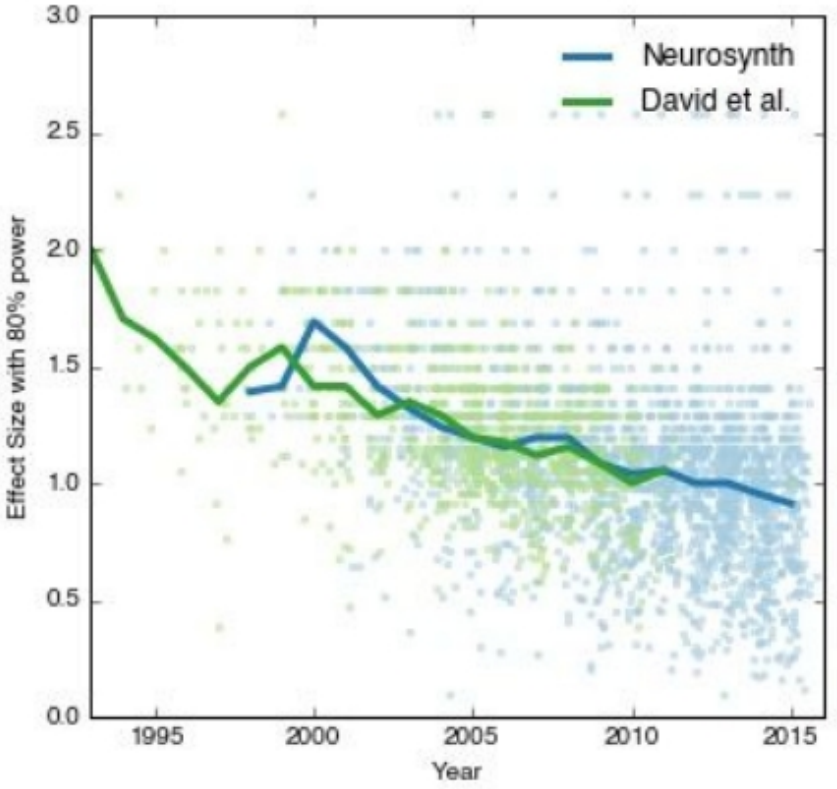




Button et al., NNR, 2013

Feeling the Future





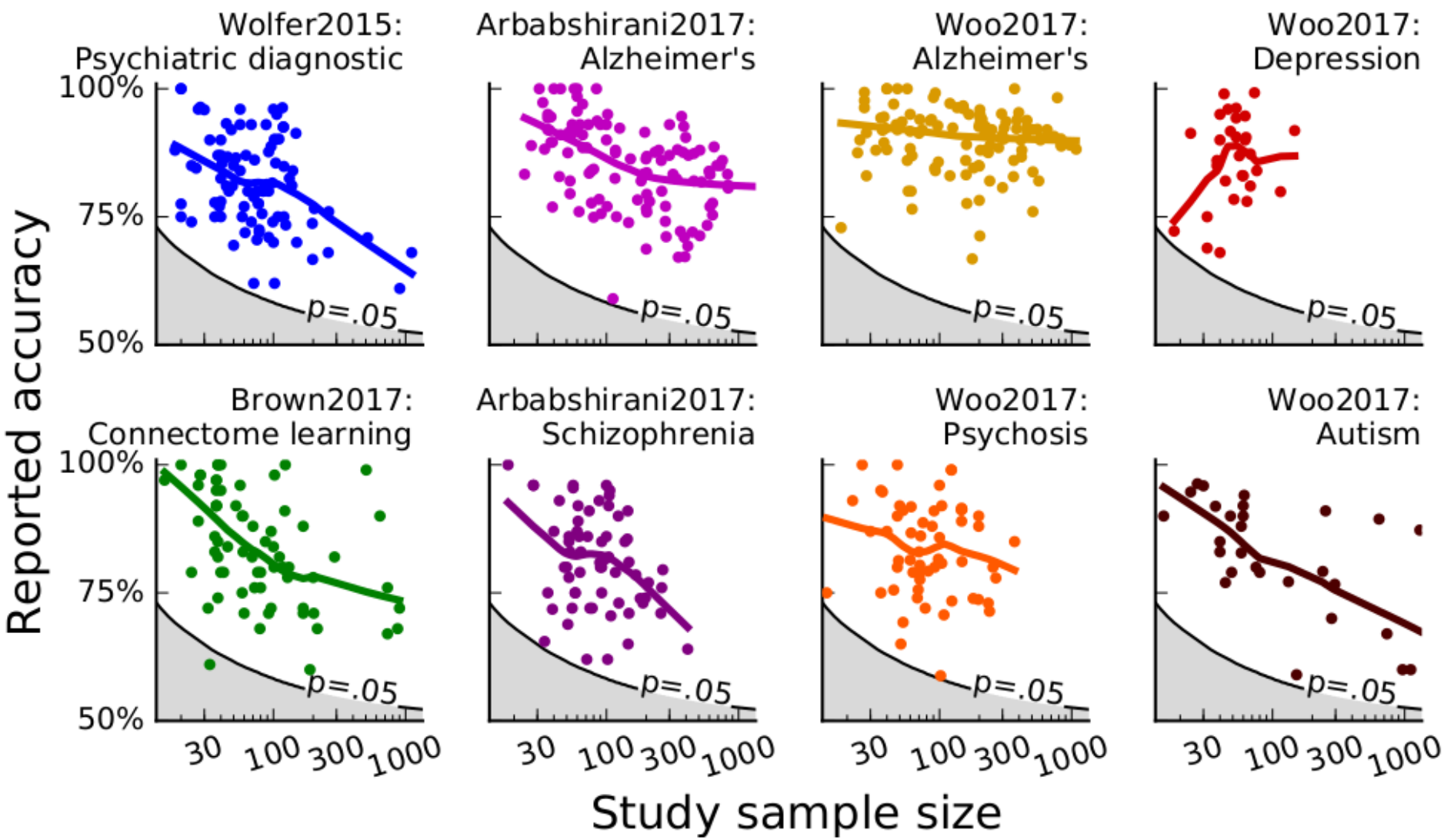
Poldrack et al., PNAS, 2016

Feeling the Future

Paradigm	Intersection mask	mask size (vox)	Cohen D			BOLD		
			P10	median	P90	P10	median	P90
MOTOR	Bilateral Precentral Gyrus	12894	0.158	0.628	1.070	0.505	2.707	8.582
	Bilateral Supplementary motor cortex	3418	0.211	0.716	1.197	0.911	4.033	12.510
	Left putamen	1532	0.114	0.513	0.864	0.586	2.388	4.318
	Right putamen	1437	-0.008	0.369	0.749	-0.045	1.696	3.609
WM	Bilateral Middle frontal gyrus	7116	0.101	0.474	0.837	0.130	0.986	2.504
EMOTION	Left amygdala	1133	0.265	0.534	1.065	0.516	1.198	3.379
	Right amygdala	1082	0.308	0.645	1.140	0.581	1.350	3.557
GAMBLING	Left accumbens	455	0.138	0.310	0.461	0.369	0.849	1.440
	Right accumbens	417	0.141	0.332	0.488	0.373	0.981	1.618

With effect size = 0.5 => Power ~ 30%

Sample size issue in ML



Varoquaux, 2017, PsyArXiv, Neuroimage, Cross validation failure

Positive Predictive Value

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

- Positive Predictive Value : The probability that the alternative hypothesis is true knowing that the test is significant

Deriving the PPV … probability?

- Objective / Physical : property of the nature or system
  - Associated with a **collective**
- Subjective: a degree of belief (“Evidential probability”)
- Frequentist: limit of frequency across random trials
- Bayesian: as reasonable expectation representing a state of knowledge

  Often part of the definition:

  P(A and B) = P(A) P(B given A) = P(A) P(B|A)

Recall Bayes

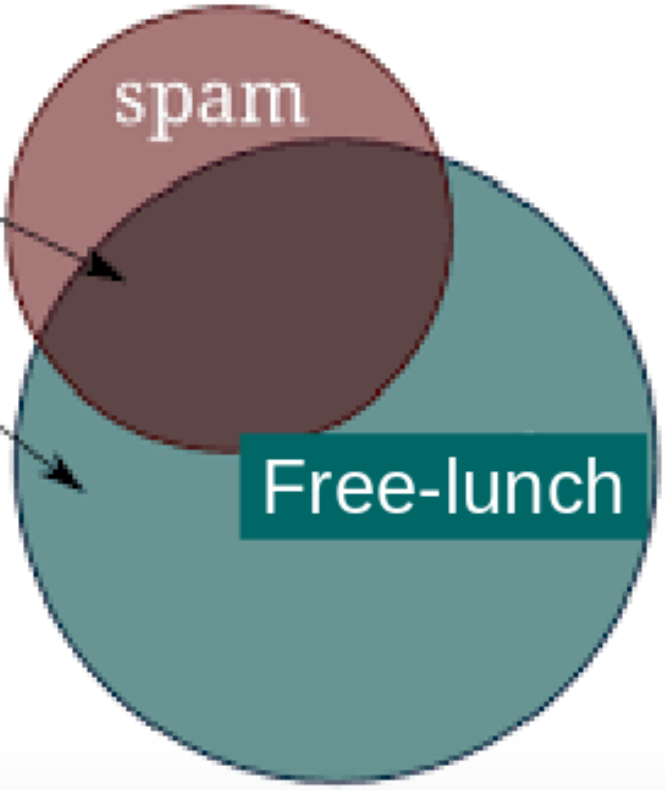
$$P(H, D) = P(H|D)P(D)$$

$$P(H, D) = P(D|H)P(H)$$

$$P(H|D) = \frac{P(D, H)}{P(D)} = \frac{P(D|H)P(H)}{P(D)}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{P(B|A)P(A)}{P(B)}$$



Deriving PPV

http://www.repronim.org/module-stats/05-PPV/

Bayes and likelihood functions

Can we use these to choose between theories ?

- Likelihood: P(D | H) : Careful: not a frequentist probability !

$$P(H_1|D) = \frac{P(D,H_1)}{P(D)} = \frac{P(D|H_1)P(H_1)}{P(D)}$$

$$P(H_0|D) = \frac{P(D,H_0)}{P(D)} = \frac{P(D|H_0)P(H_0)}{P(D)}$$

- Ratio: posterior odds = Bayes-factor $\times$ prior-odds

Bayes and likelihood functions

Can we use these to choose between theories ?

- Likelihood: P(D | H) : Careful: not a frequentist probability !

$$P(H_1|D) = \frac{P(D,H_1)}{P(D)} = \frac{P(D|H_1)P(H_1)}{P(D)}$$

$$P(H_0|D) = \frac{P(D,H_0)}{P(D)} = \frac{P(D|H_0)P(H_0)}{P(D)}$$

- Ratio: posterior odds = Bayes-factor $\times$ prior-odds

- Define : $P_1 = P(H_1)$ and $P_0 = P(H_0)$

$$PPV = \frac{(1-\beta)P_1}{(1-\beta)P_1 + \alpha P_0}$$

- With $R = P_1/P_0$ and $W = 1-\beta$:

$$PPV = \frac{WR}{WR + \alpha}$$

- Wikipedia (ML): PPV = number-of-true-positive / (number-of-true-positive + number-of-false-positive)

Ioannidis 2005

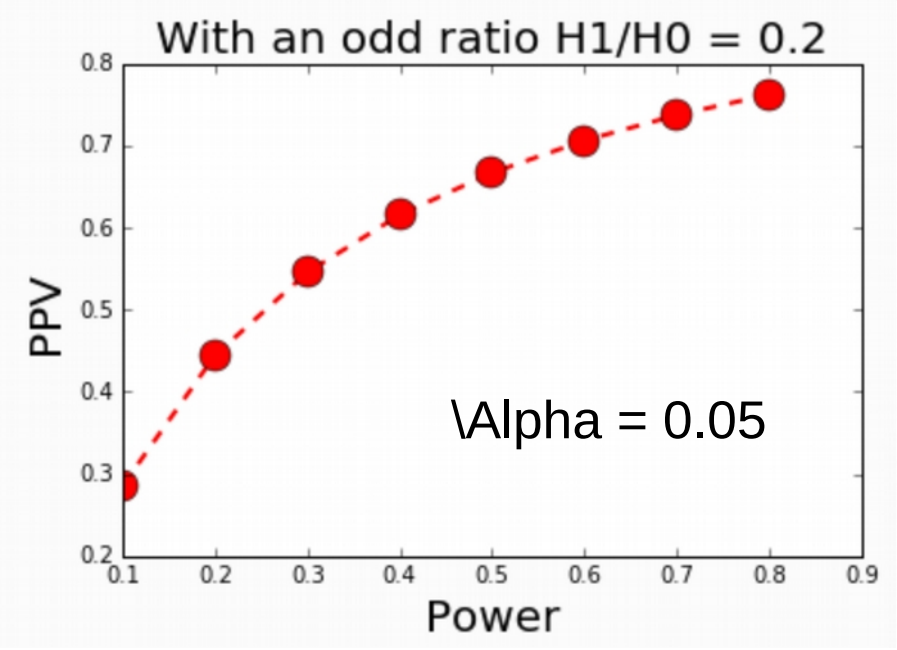
Open access, freely available online

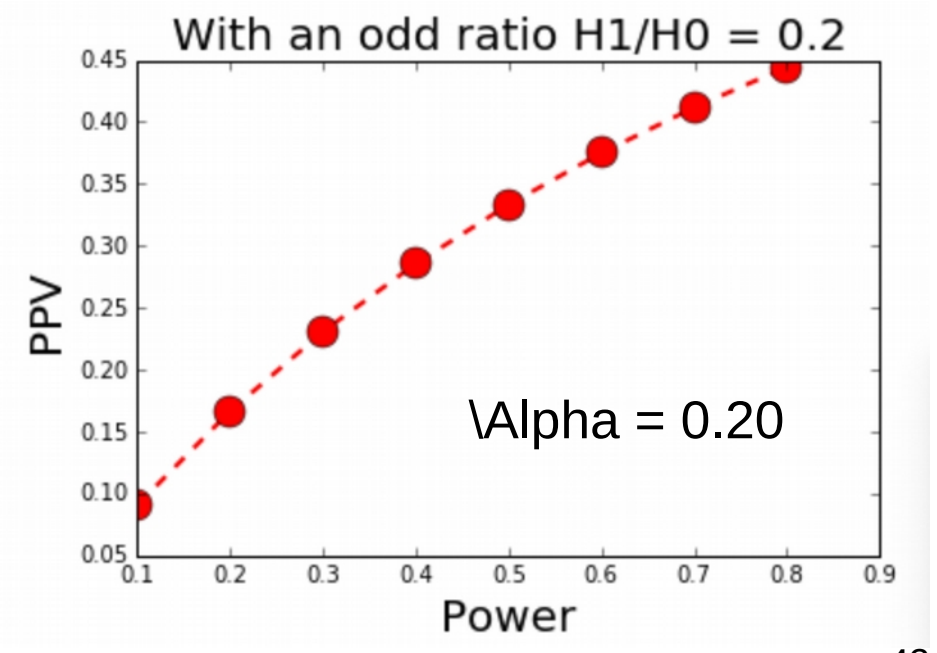
Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

$$P(H_A \mid T_S) = \frac{WR}{WR + \alpha}$$





BMJ bioethical issues in biostatistical consulting

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.

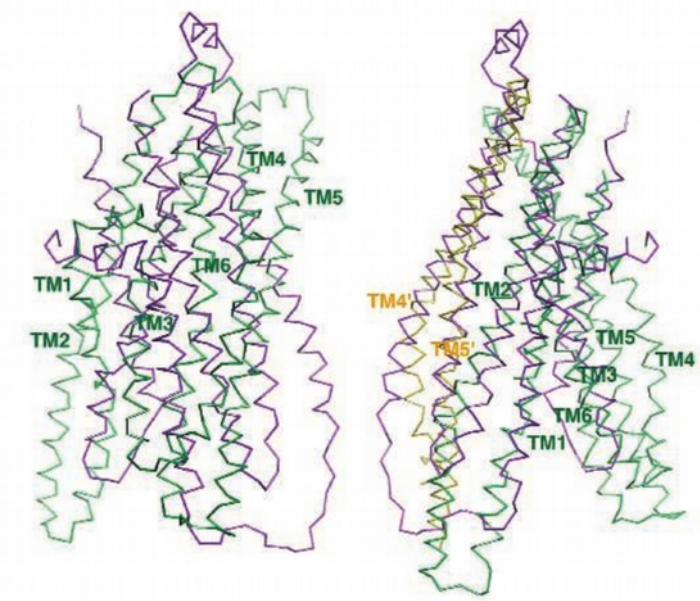
Three causes

1. Poor statistical procedures

2. Issues in data and software

3. A cultural issue: Publication practices and research incentives

Protein structure flip



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

Forensic Analysis

- 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

Same software different OS version

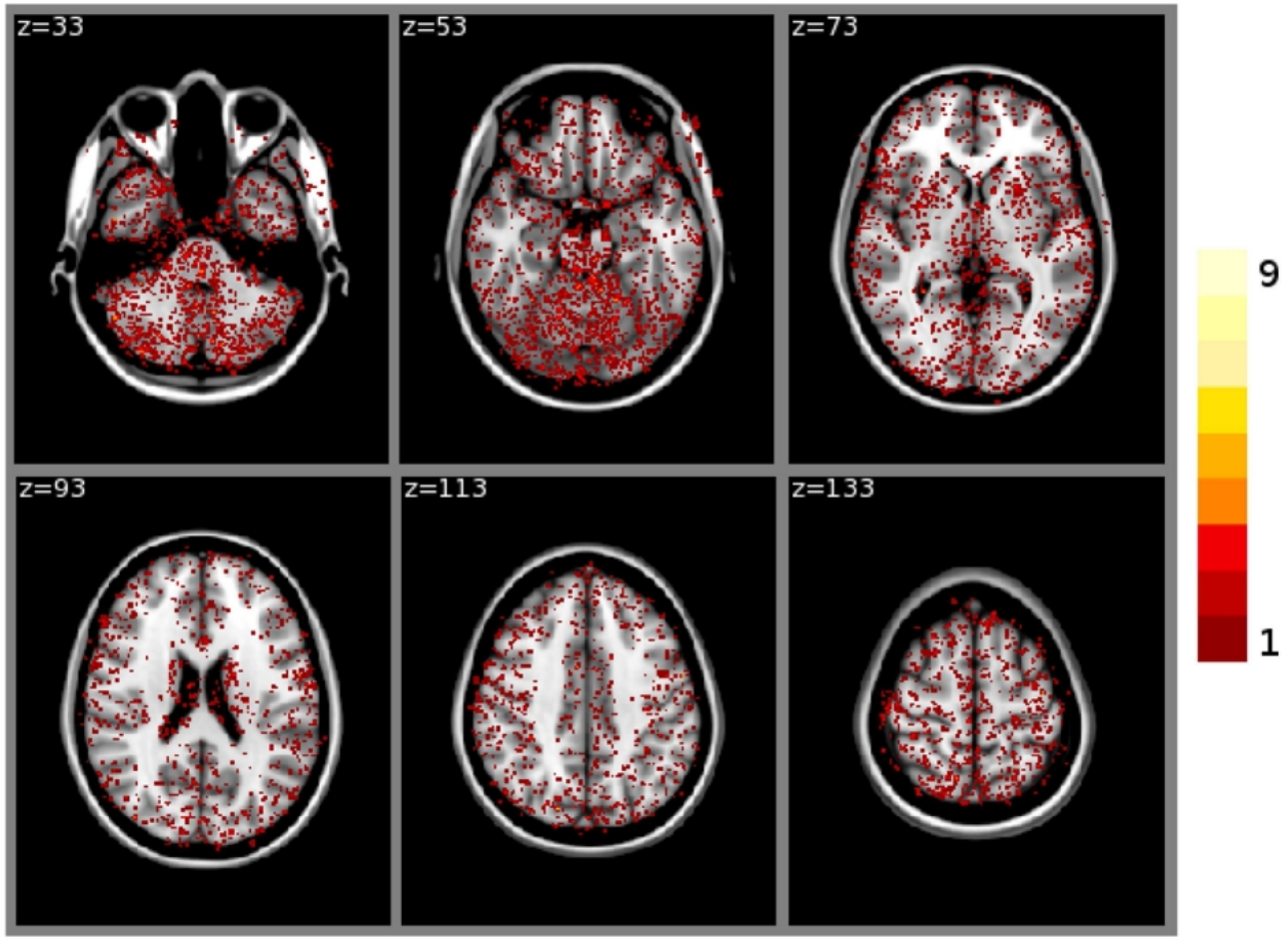
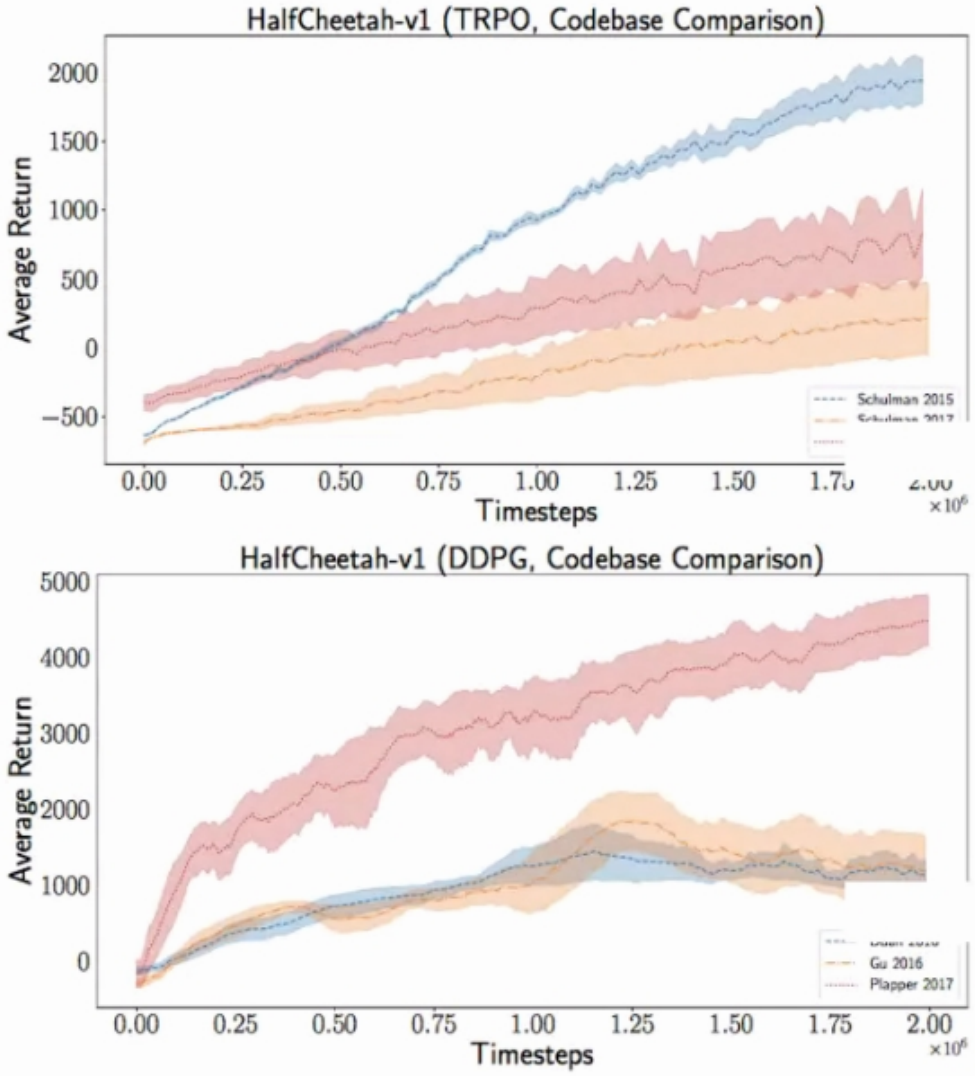


FIGURE 2 | Sum of binarized differences between cortical tissue classifications obtained on cluster A and cluster B (FSL FAST, build 1, *n* = 150 subjects). All binarized differences were resampled to the default MNI152 volume template.

Glatard et al, 2015, F. in Neuroinformatics

Example in machine learning

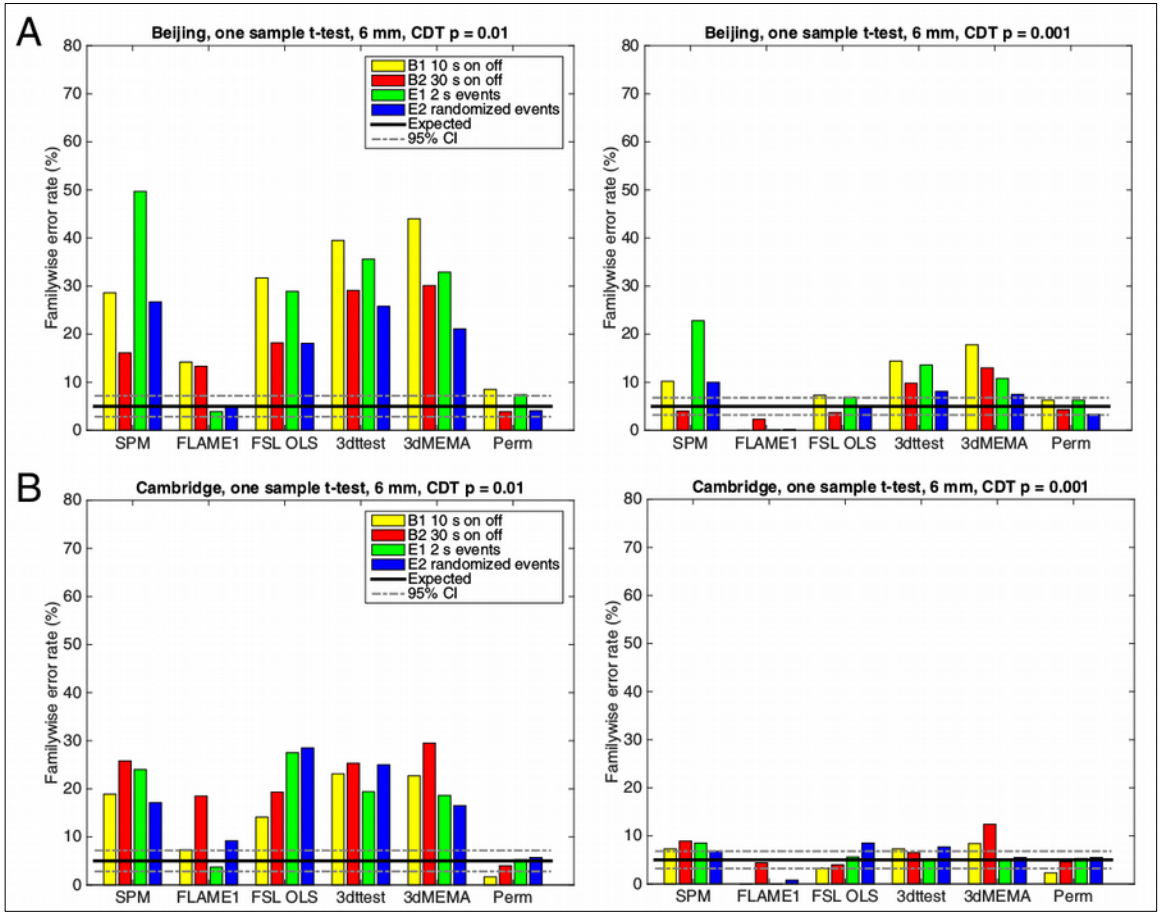
- Same algorithm
  - Top graph: TRPO
  - Bottom graph: DDPG
- Same domain
- Simulation environment
- Different implementations



Credit : Joelle Pineau

Across parametrizations of the same software (and across software)

“Cluster failure”? Or RFT misuse?



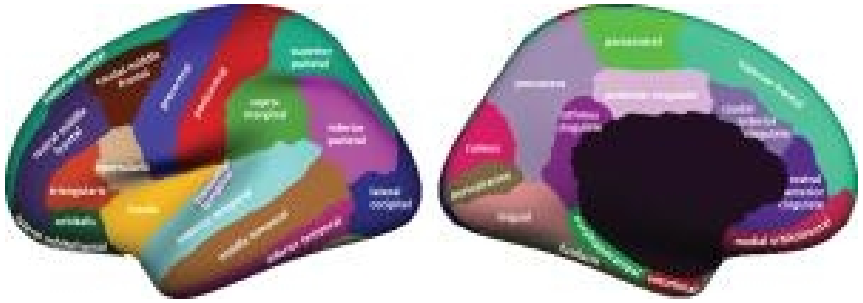
Eklund et al., PNAS, 2016 :

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

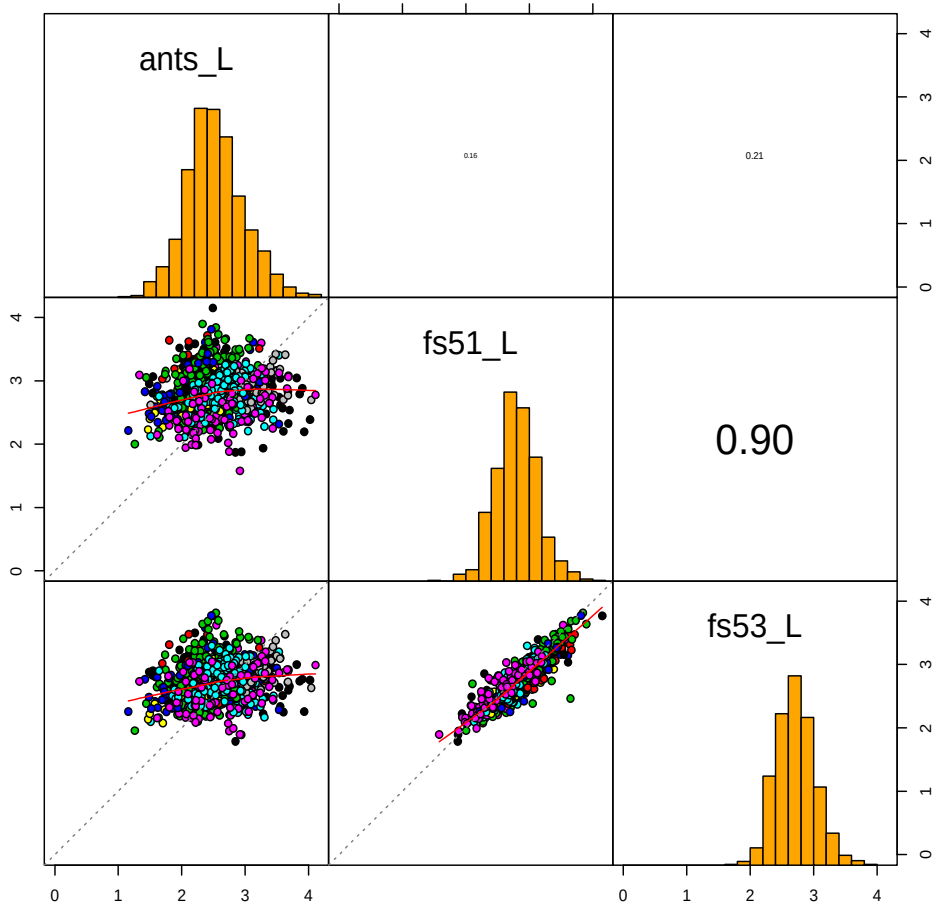
-
- Estimated 3,500 papers affected by low threshold ?
- But 13000 w/o multiple comparisons ?

Across software

ANTS – FS5.1- FS5.3



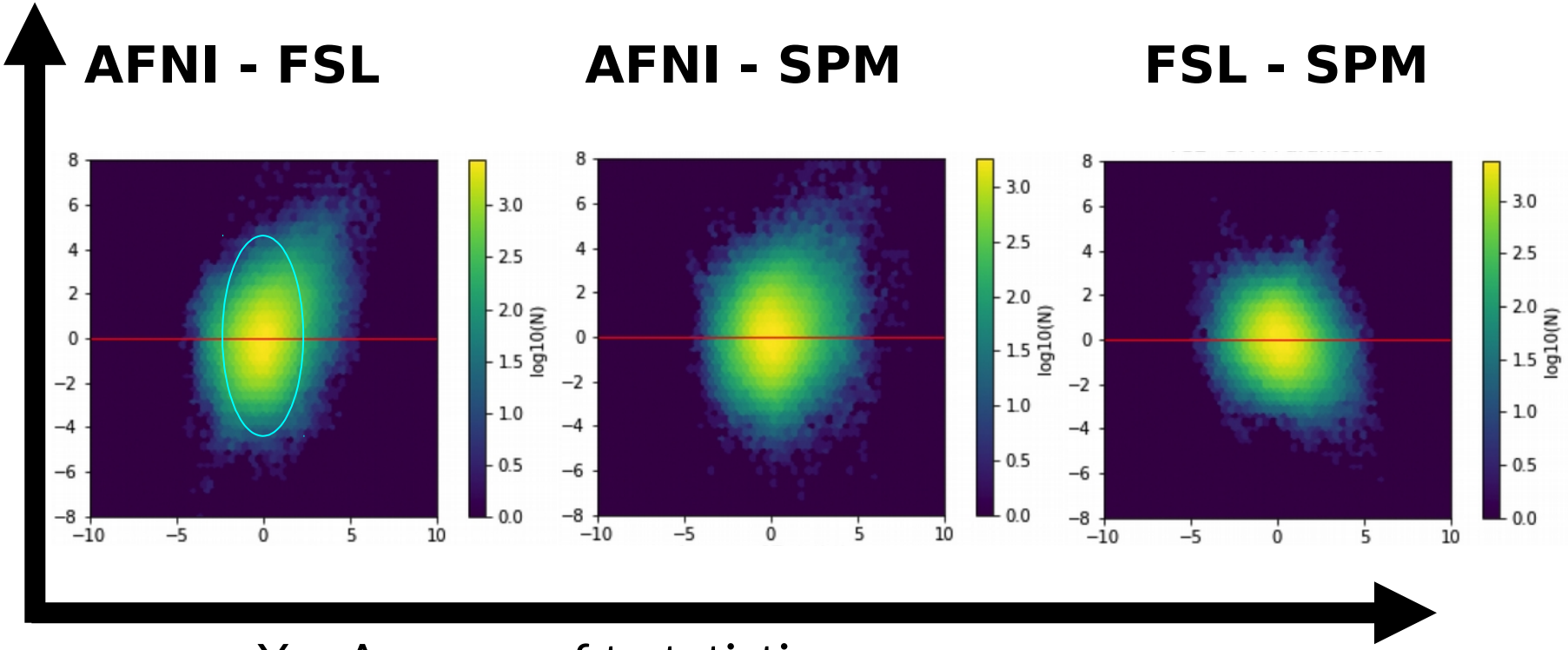
Size of the left caudal anterior Cingulate



Dickie E et al., 2017

How close are results on the same dataset?

Y = Diff. of t-statistics



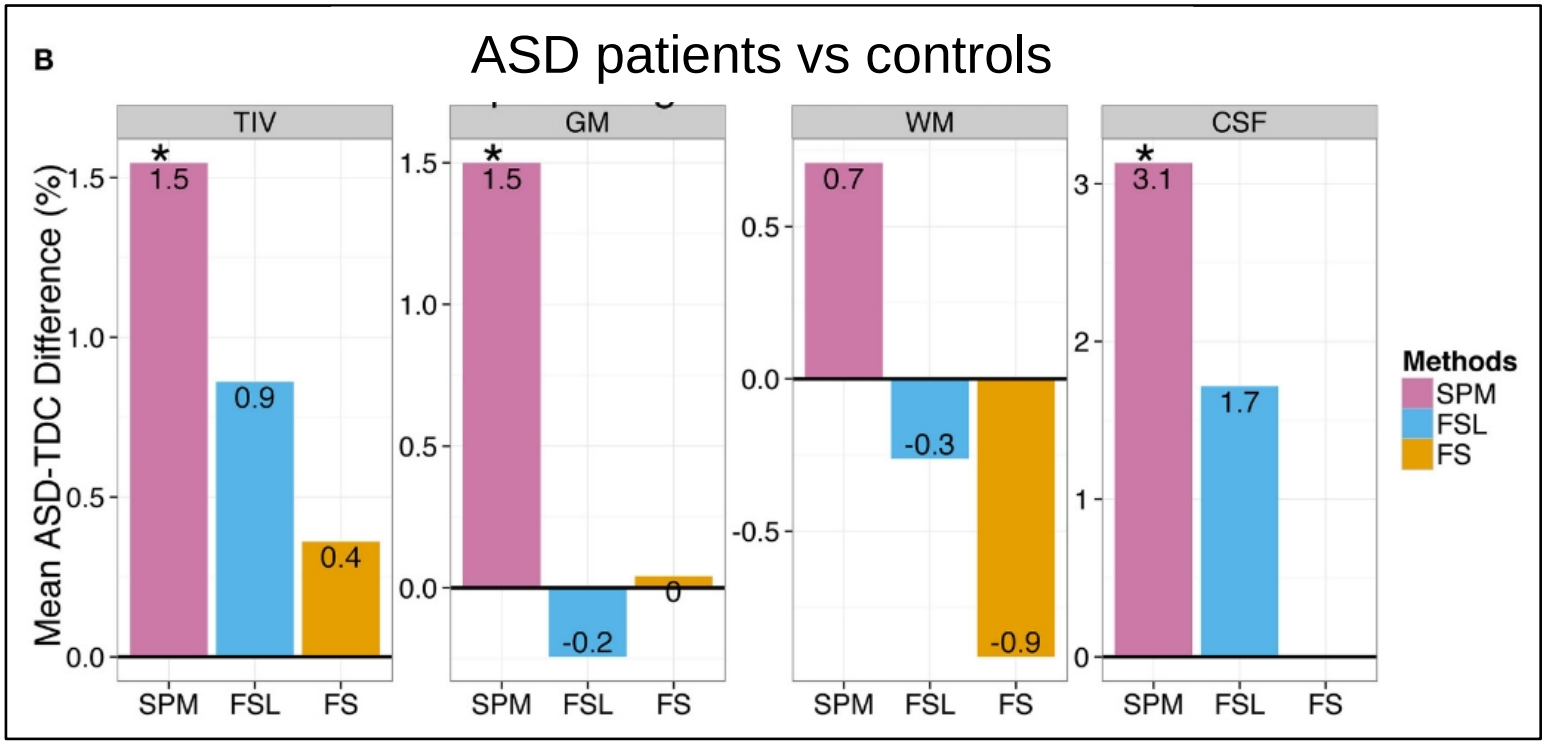
X = Average of t-statistics

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

Alex Bowring, Camille Maumet, Thomas Nichols

Software, version, OS

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



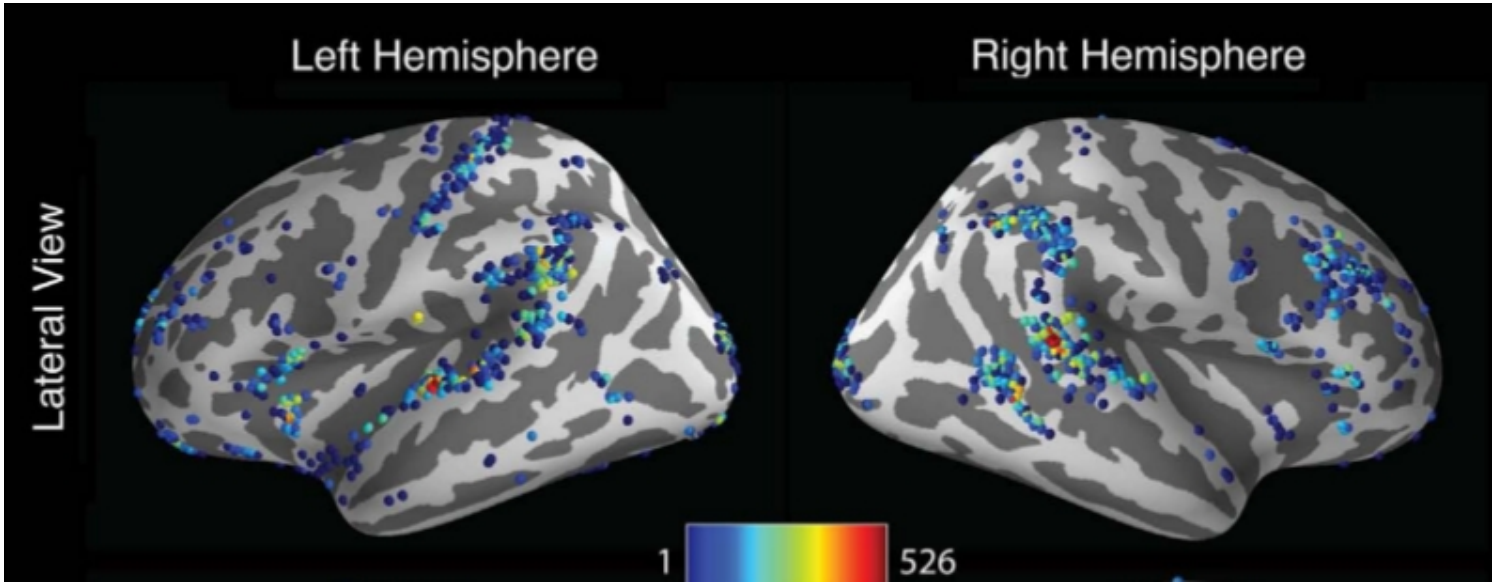
- Change from FSL to SPM?
- Change from v.1.12 to v.2.1 ?
- Change from cluster A to cluster B? Glatard et. al., finsc, 2015

Across possible pipelines

Estimating analytical flexibility of fMRI

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



Cause: bugs in data

- 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

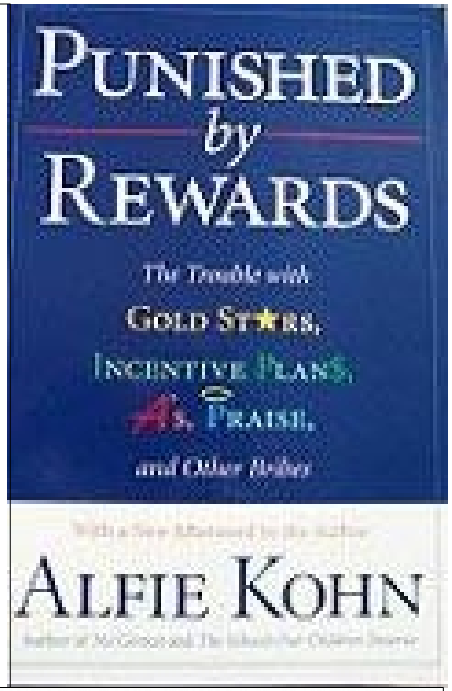
1. Poor statistical procedures

2. Issues in data and software

3. A cultural issue: Publication practices and research incentives

Misplaced 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

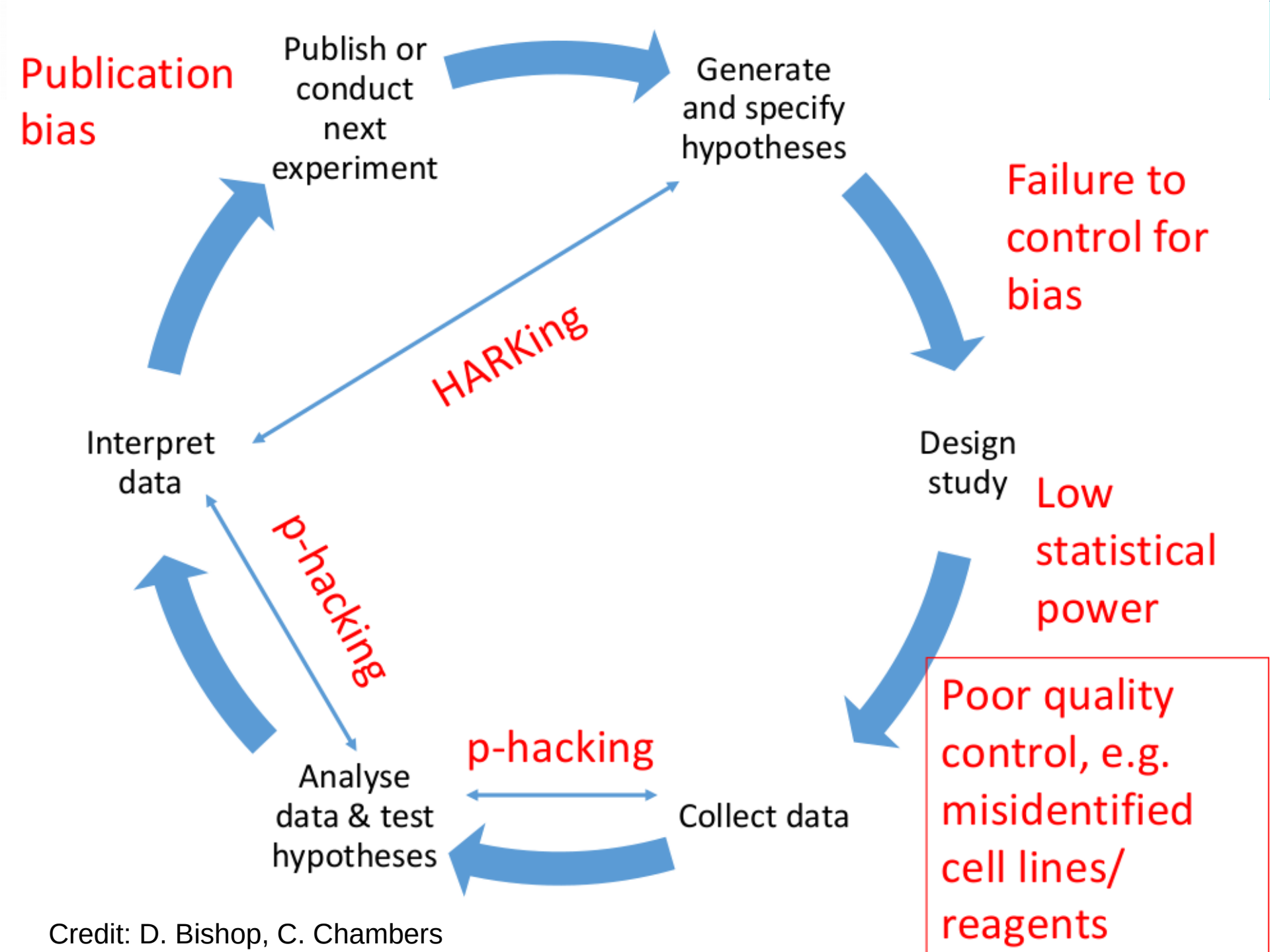
rsos.royalsocietypublishing.org

Research

The natural selection of bad science

Paul E. Smaldino[1] and Richard McElreath[2]

[1]Cognitive and Information Sciences, University of California, Merced, CA 95343, USA
[2]Department of Human Behavior, Ecology, and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany



Credit: D. Bishop, C. Chambers

Part I: Reproducibility: background

Part II : Etiology of Irreproducibility

Part III : Epistemology and statistics

Part IV :Some therapeutic proposals

Epistemiological questions

- What is the fastest way to make progress in science ? Are we asking the right questions?
- Are our current statistical procedures adapted?
- If not, which procedures should we be using?

- Disclaimer 1: I am no epistemiologist (but we need to think about epistemiology)
- Disclaimer 2: “The philosophy of science is as useful to scientists as ornithology is to birds” (Richard Feynman)
- Credit: Z. Dienes, psychology as a science

Induction versus deduction

- Induction: “Generalizing about the properties of a class of objects”
  - The black swan problem
- Deductive reasoning:
  - All men are mortal. (First premise)
  - Socrates is a man (Second premise)
  - Therefore, Socrates is mortal. (Conclusion)
- Way more inductions than deduction in neuroscience ?

Humes, Popper

- Science versus pseudo science: the demarcation problem (“free will is an illusion”)
  - Testable theory

- Induction is not the best way to make progress in science
  - We cannot reason from past to future instances
  - Has induction not already proved itself ?
    - Newton vs relativity
  - What about the probability of the theory ?
    - All future instance may disprove the current theory

Bertrand Russell vs Karl Poper

- Bertrand Russell (1872-1970): Induction is necessary to distinguish good from bad theories

- Popper argues back that induction is not needed:
  - A fact that complies with the theory does not prove the theory “All swans are white” - does a blue pen add support to the theory ?
  - A fact that does not comply with the theory disproves it
  - Falsification is the right tool

- Criteria: falsification - simplicity - precision - general
  - More falsification is better
  - More precise (eg >0 vs 1<x<5)
  - Applies to a broader range of situations

How is this related to statistical tests?

- Patients A have larger hippocampus than patients B.
  - Size(As) > Size(Bs)
  - Test H0: Size(As) == Size(Bs) ? is this reasonable?

- What is “wrong” ?
  - Test rejects the “not the theory” - asymmetry !
  - Reject” H0 (very precise) but many “alternatives” are possible
  - A Popperian approach would falsify with examples where Size(As) <= Size(Bs) but that would need to be put in a probabilistic context

How is this related to statistical tests?

- NHST is working within the induction framework
  - Rejecting the null is a probabilistic statement
  - Asymmetry issue: when is rejecting the null informative?
    When there is power

- Prediction: distinguish ...
  - Models that embed a theory (eg, a causal model) where some deductive reasoning can be tested
  - Models that do not (DL) but are still useful

- “Induction” still a problem: considered as the generalization of prediction to other populations

Part I: Reproducibility: background

Part II : Etiology of Irreproducibility

Part III : Epistemology and statistics

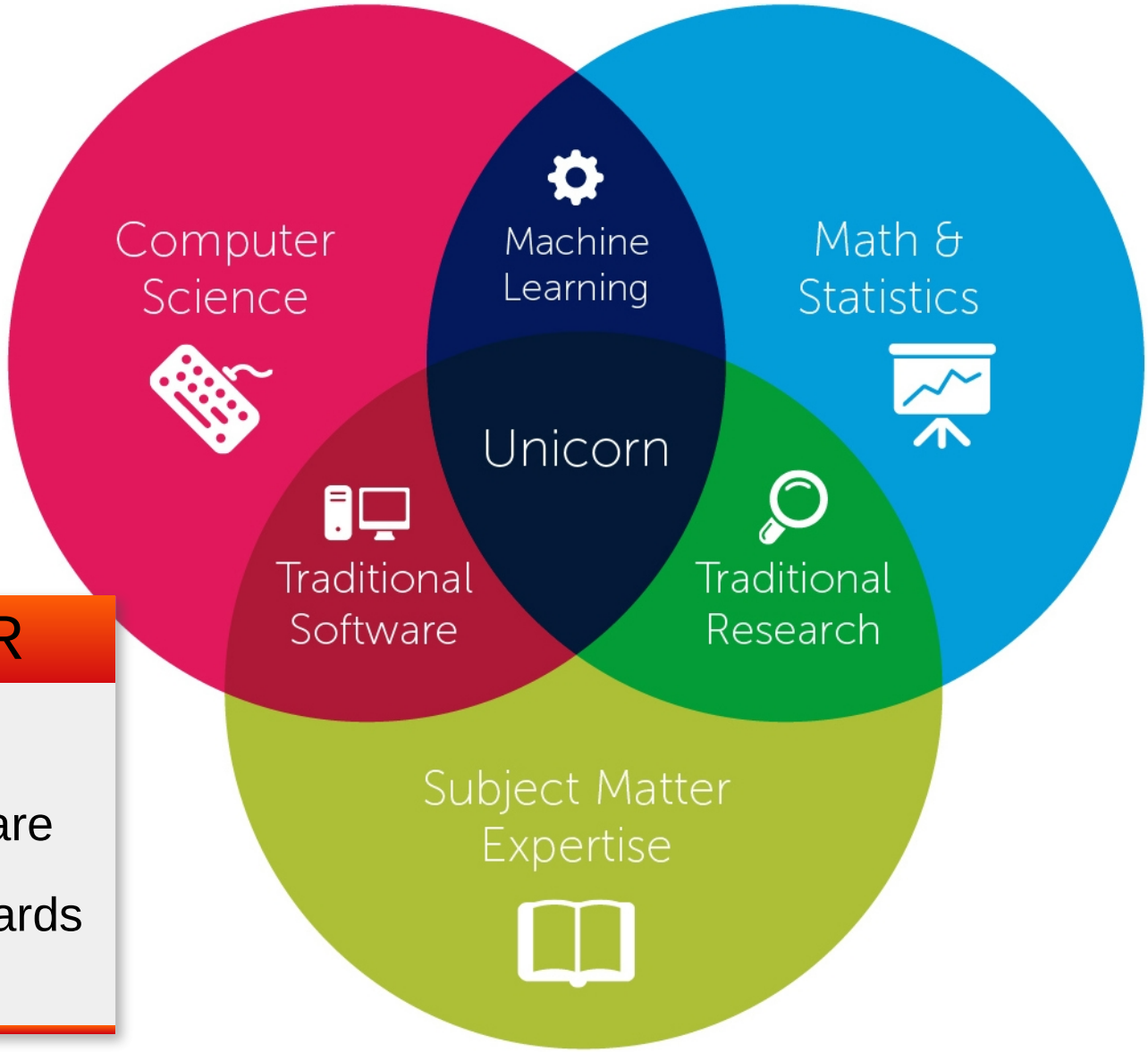
Part IV : Some therapeutic proposals

What can we do ?

- Improve Training
- Develop better tools – make these tools that could change the culture
- Change the incentives

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

Data Science



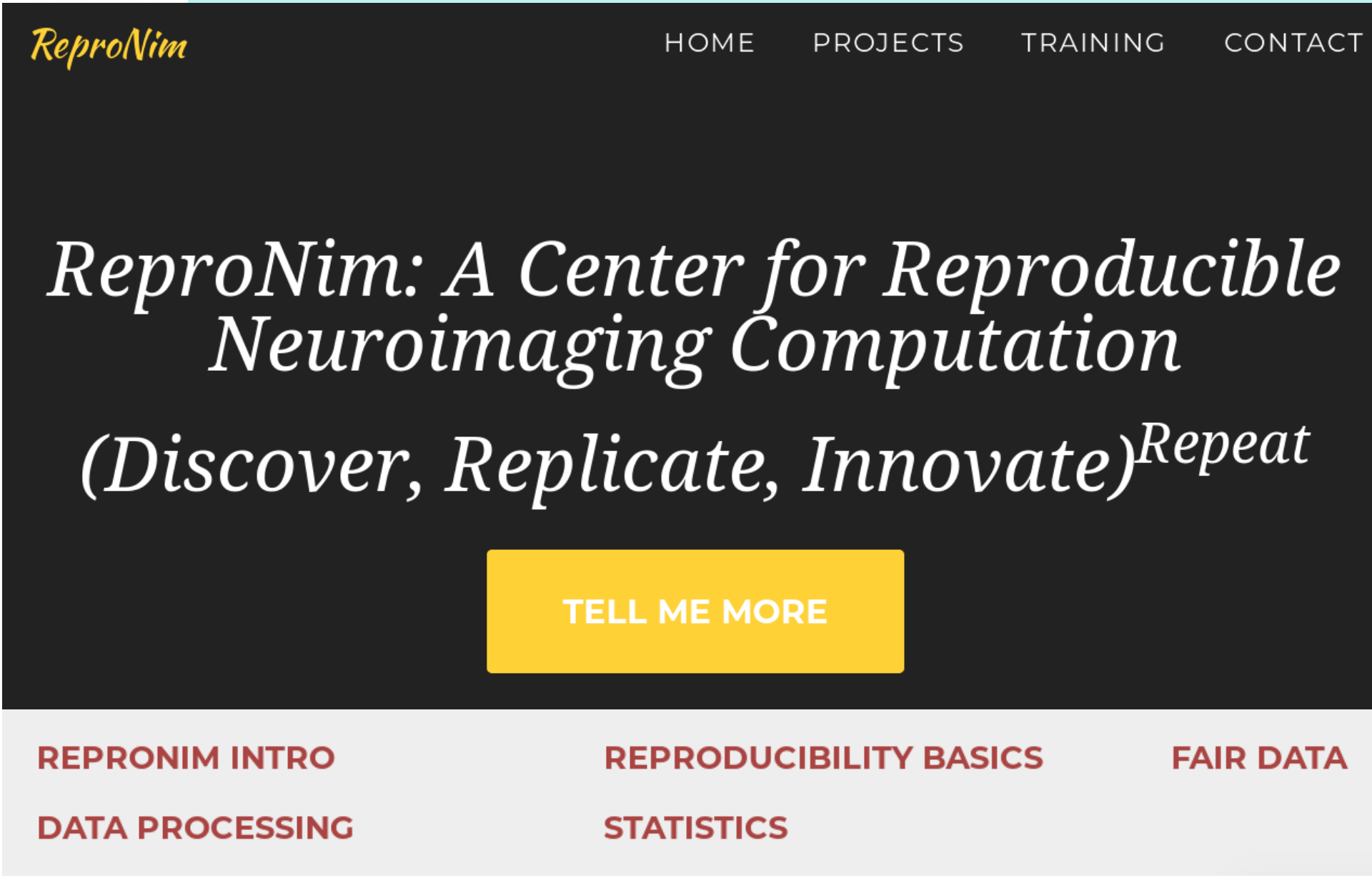
FAIR

- Data
- Software
- Standards reuse

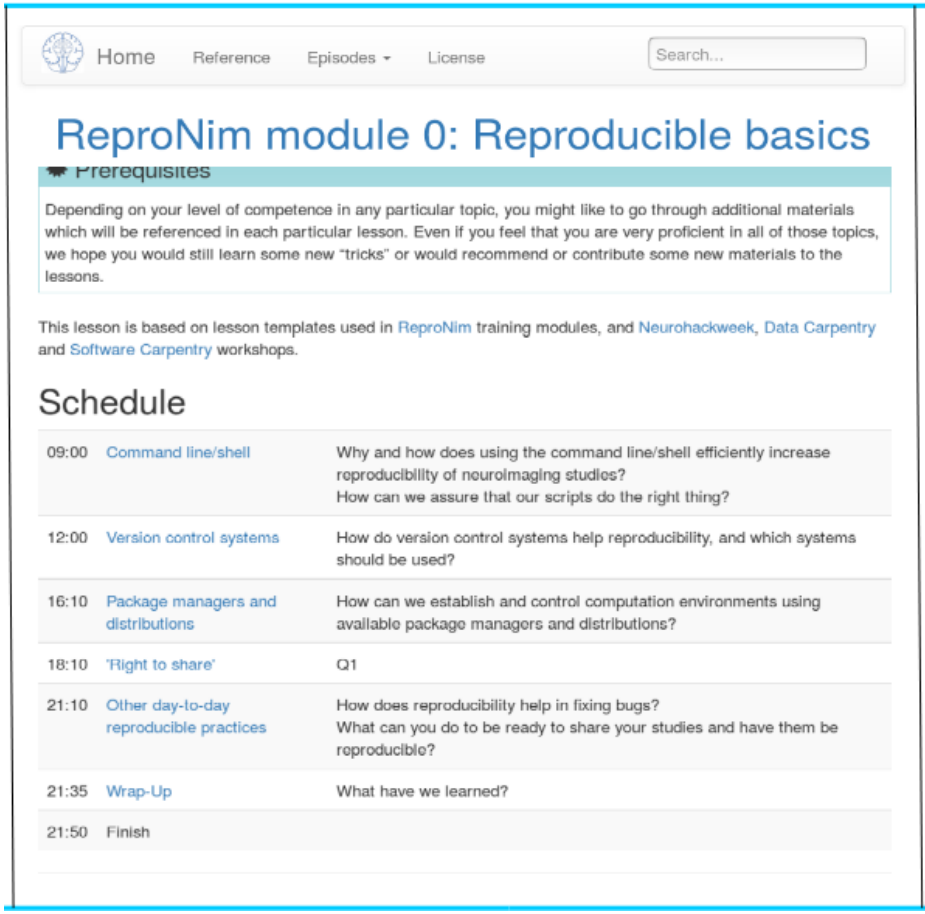
Science and Technology

Research

Training: NIH P41 ReproNim



Training



Home Reference Episodes License Search...

ReproNim module 0: Reproducible basics

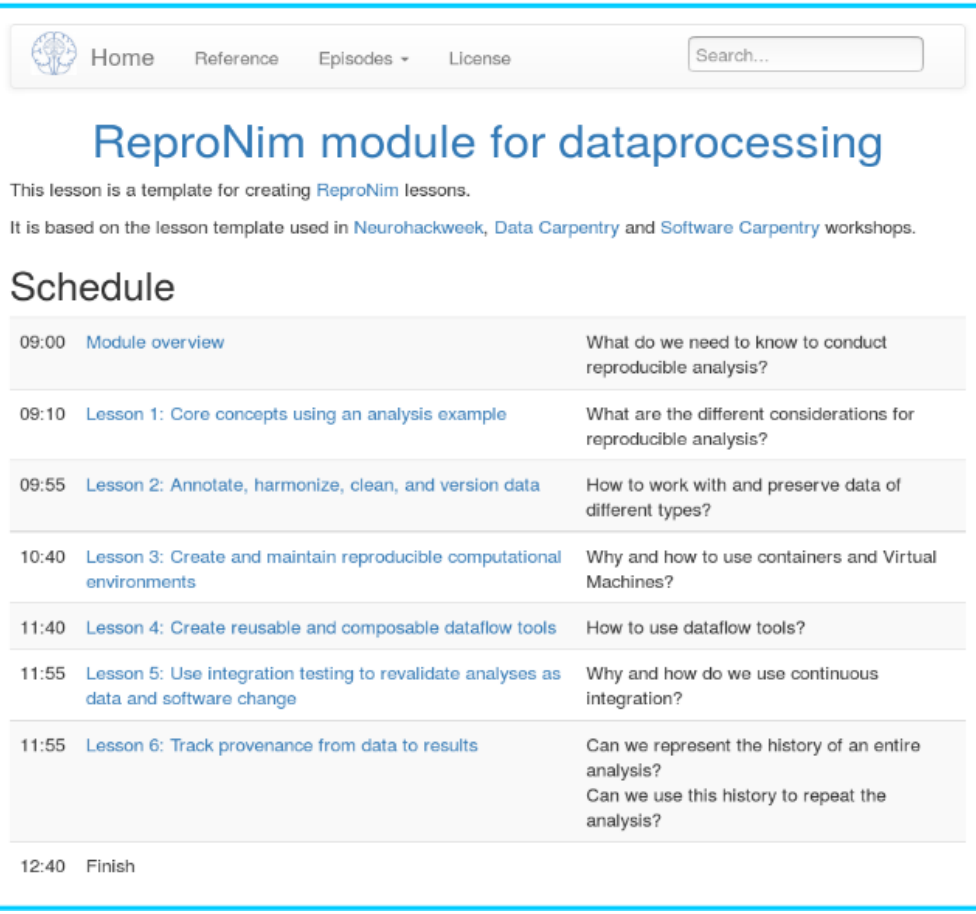
Prerequisites

Depending on your level of competence in any particular topic, you might like to go through additional materials which will be referenced in each particular lesson. Even if you feel that you are very proficient in all of those topics, we hope you would still learn some new "tricks" or would recommend or contribute some new materials to the lessons.

This lesson is based on lesson templates used in ReproNim training modules, and Neurohackweek, Data Carpentry and Software Carpentry workshops.

Schedule

09:00	Command line/shell	Why and how does using the command line/shell efficiently increase reproducibility of neuroimaging studies? How can we assure that our scripts do the right thing?
12:00	Version control systems	How do version control systems help reproducibility, and which systems should be used?
16:10	Package managers and distributions	How can we establish and control computation environments using available package managers and distributions?
18:10	'Right to share'	Q1
21:10	Other day-to-day reproducible practices	How does reproducibility help in fixing bugs? What can you do to be ready to share your studies and have them be reproducible?
21:35	Wrap-Up	What have we learned?
21:50	Finish	



Home Reference Episodes License Search...

ReproNim module for dataprocessing

This lesson is a template for creating ReproNim lessons.

It is based on the lesson template used in Neurohackweek, Data Carpentry and Software Carpentry workshops.

Schedule

09:00	Module overview	What do we need to know to conduct reproducible analysis?
09:10	Lesson 1: Core concepts using an analysis example	What are the different considerations for reproducible analysis?
09:55	Lesson 2: Annotate, harmonize, clean, and version data	How to work with and preserve data of different types?
10:40	Lesson 3: Create and maintain reproducible computational environments	Why and how to use containers and Virtual Machines?
11:40	Lesson 4: Create reusable and composable dataflow tools	How to use dataflow tools?
11:55	Lesson 5: Use integration testing to revalidate analyses as data and software change	Why and how do we use continuous integration?
11:55	Lesson 6: Track provenance from data to results	Can we represent the history of an entire analysis? Can we use this history to repeat the analysis?
12:40	Finish	

SCIENTIFIC DATA

OPEN

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

Susanna-Assunta Sansone[1], Patricia Cruse[2] & Mark Thorley[3]

Received: 26 January 2018
Accepted: 29 January 2018
Published: 27 February 2018

Resources for data management, discovery and (re)use are numerous and diverse, and more specifically we need data resources that enable the FAIR principles[1] of Findability, Accessibility, Interoperability and Reusability of data.

Changing the publication model

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

Example

My paper concludes:

- Increase in resting state connectivity between Right Superior Temporal Gyrus and the Right Superior Frontal Gyrus in subjects with autism, and this connectivity correlated with diagnostic severity.

Credit : D. Kennedy

What statistic? (covariates, corrections)

What data? (MR parameters)
What analysis? (software and parameters)

My paper concludes:

- Increase in resting state connectivity between Right Superior Temporal Gyrus and the Right Superior Frontal Gyrus in subjects with autism, and this connectivity correlated with diagnostic severity.

What subject characteristics? (age, gender, SES, genetics, environment, etc.)

What measure?

What anatomic framework? (atlas)

Credit : D. Kennedy

My paper actually concludes:

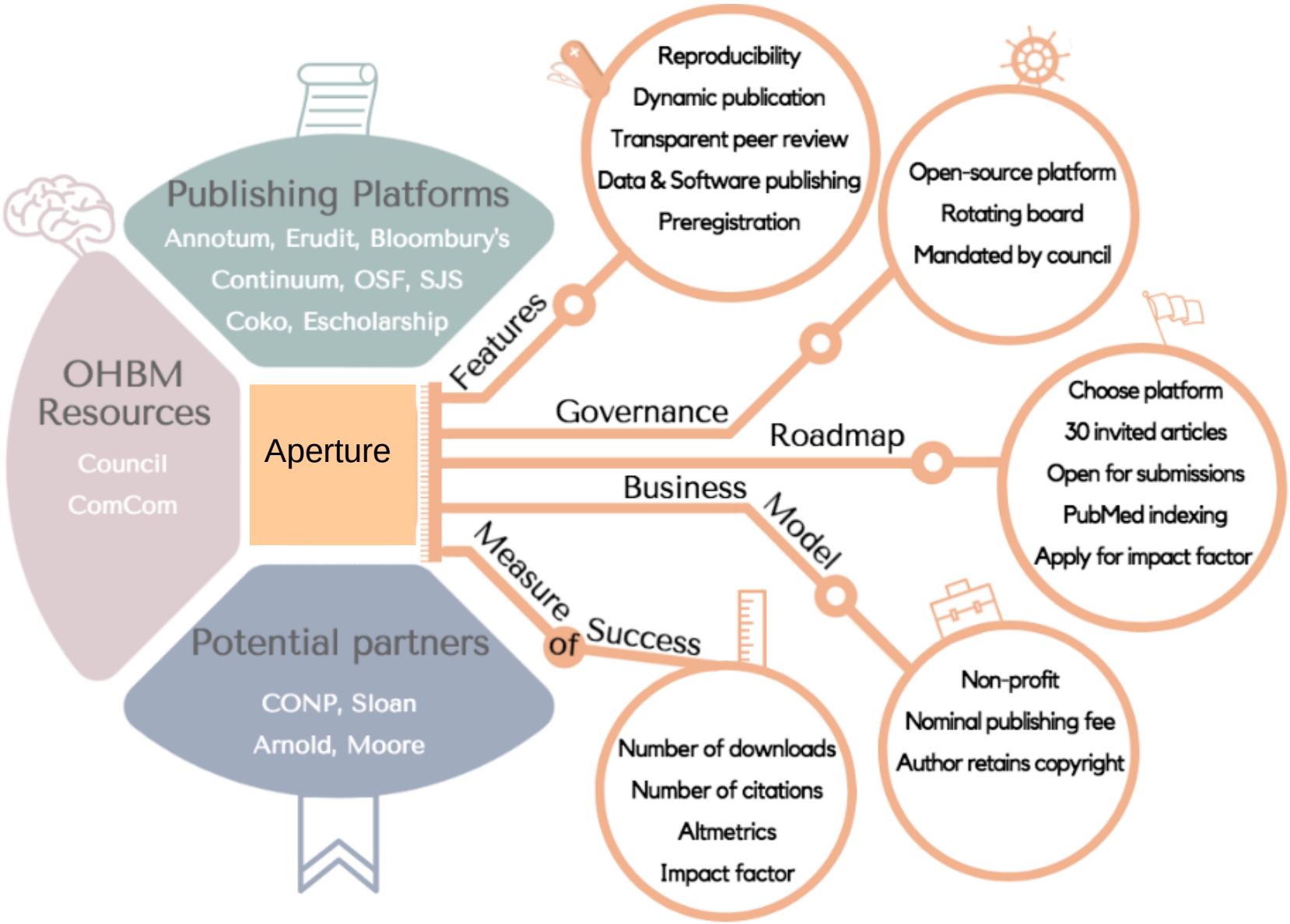
- Using a paired T-test, covarying for age, gender and handedness using cluster-size FWE correction, we saw an increase (P<.01) in regional seed-based using CONN resting state connectivity between AAL regions of the Right Superior Temporal Gyrus and the Right Superior Frontal Gyrus in 40 subjects with autism (age 14+-5, 19M/11F, IQ 90+- 10, ADOS 20+-5), and this connectivity correlated (Pearson, P<.05) with diagnostic severity as measured by the social subscale of the ADI.

Credit : D. Kennedy

Publishing: what do we need

- Publish reusable research objects
  - Data first !
  - Software, workflow, analyses
  - Jupyter notebooks, hybrid objects
  - Pre-registered report
- Vetting objects
  - By experts
  - By community based (alt)metrics
- Make published research object machine readable as much as possible

OHBM Open Publishing Initiative



Thank you

- Lab@McGill**: https://neurodatascience.github.io/**
- **McGill** colleagues: S. Brown, T. Glatard, G. Kiar, A. Evans, C. Greenwood, A. DeGuise and 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**: McGill HBHL, HBHL NeuroHub, NIH, NIMH