Barriers to reproducibility and how to overcome them

Dorothy V. M. Bishop

Professor of Developmental Neuropsychology

University of Oxford

@deevybee

wellcome trust

The EU Framework Programme for Research and Innovation

HORIZON 2020

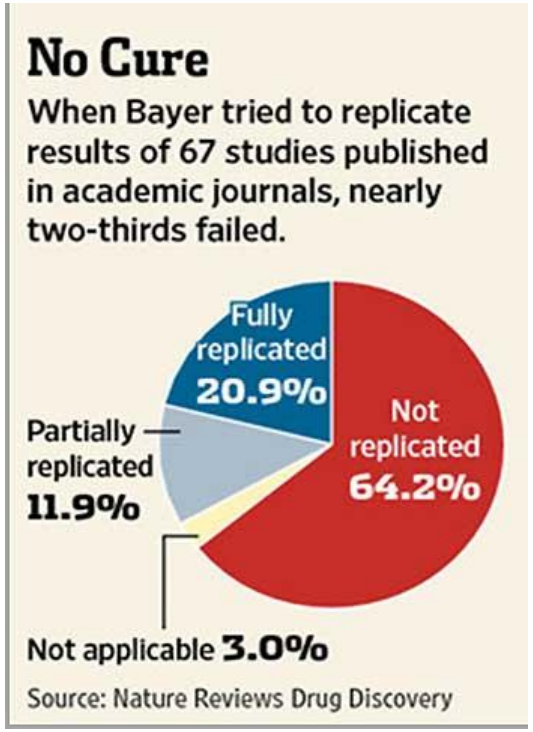
What is the replication crisis?

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

Published: January 8, 2014

nature International weekly journal of science

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

News & Comment > News > 2015 > May > Article

NATURE | NEWS

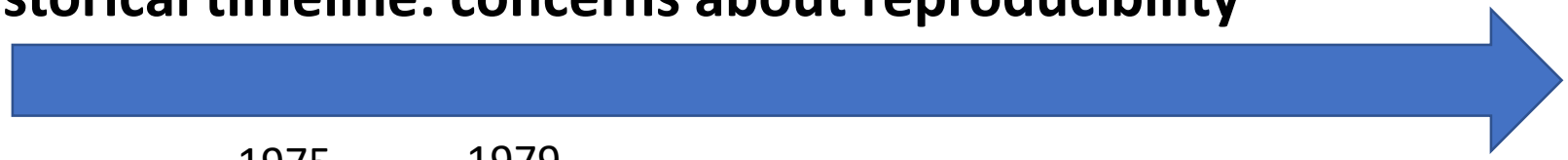
First results from psychology's largest reproducibility test

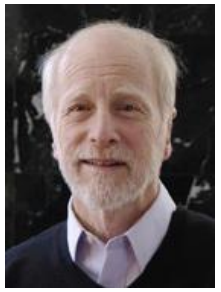
Kirstie Whitaker: “What is reproducibility?”

		Data	
		Same	Different
Code	Same	Reproducible	Replicable
	Different	Robust	Generalisable

@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.4244996

Historical timeline: concerns about reproducibility





1975
Greenwald

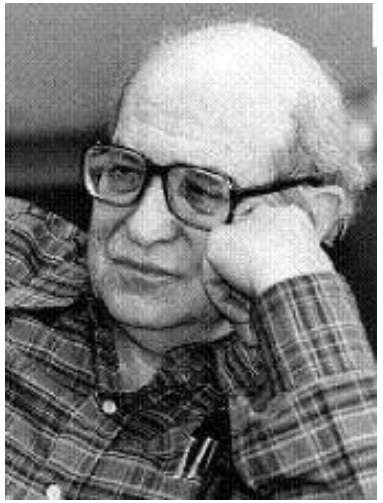
1979
Rosenthal

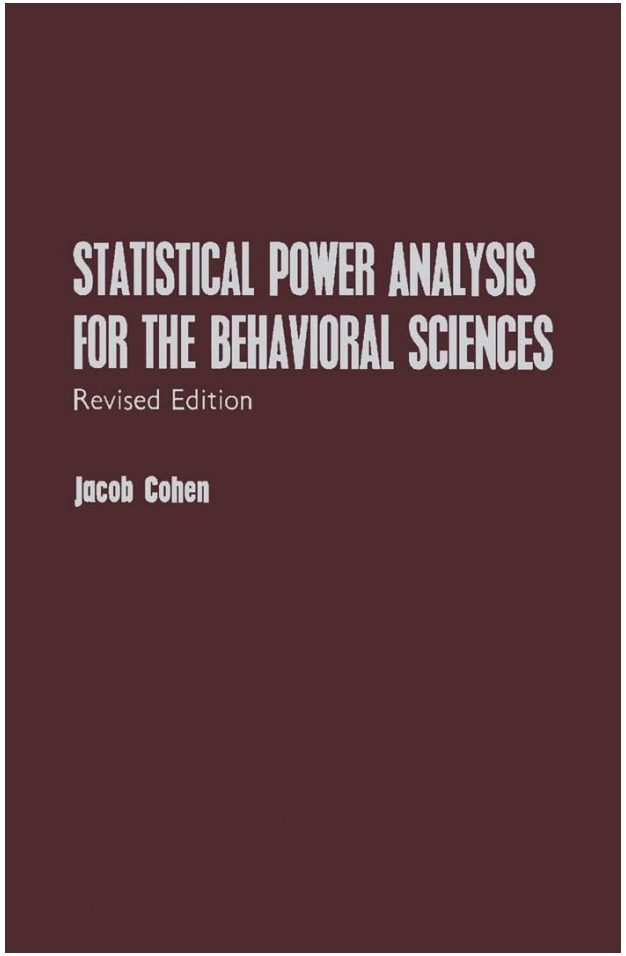
The “file drawer” problem

“As it is functioning in at least some areas of behavioral science research, the research-publication system may be regarded as a device for systematically generating and propagating anecdotal information.”

Publication bias

1969
Cohen





Low power

Sample size too small to reliably detect a true effect of interest

Historical timeline: concerns about reproducibility

1956
De Groot

Failure to distinguish between hypothesis-testing and hypothesis-generating (exploratory) research
-> misuse of statistical tests

Describes P-hacking (though that term not used)

P-hacking

Simple explainer using poker

3 of a kind

Probability from unbiased deck of cards = 1 in 50

http://deevybee.blogspot.com/2016/01/the-amazing-significo-why-researchers.html

P-hacking

Simple explainer using poker

3 of a kind

Probability from unbiased deck of cards = 1 in 50

- If magician tells you he'll deal you '3 of a kind', and he does so, you should be impressed

http://deevybee.blogspot.com/2016/01/the-amazing-significo-why-researchers.html

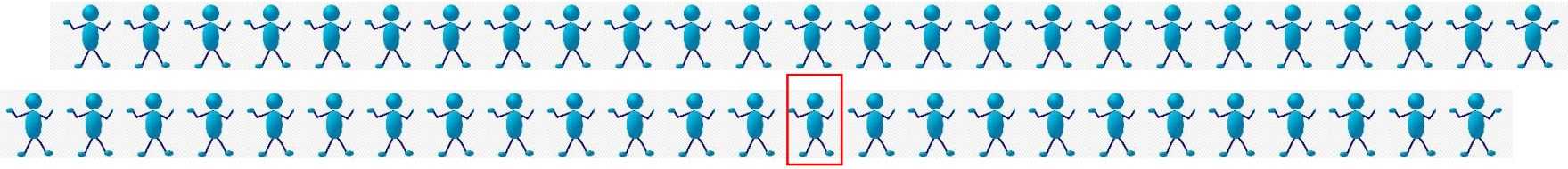
P-hacking

Simple explainer using poker

3 of a kind

Probability from unbiased deck of cards = 1 in 50

- If magician deals 50 hands, and **one** of them is '3 of a kind', you should not be impressed



'Surprisingness' of a result only interpretable in context of full dataset

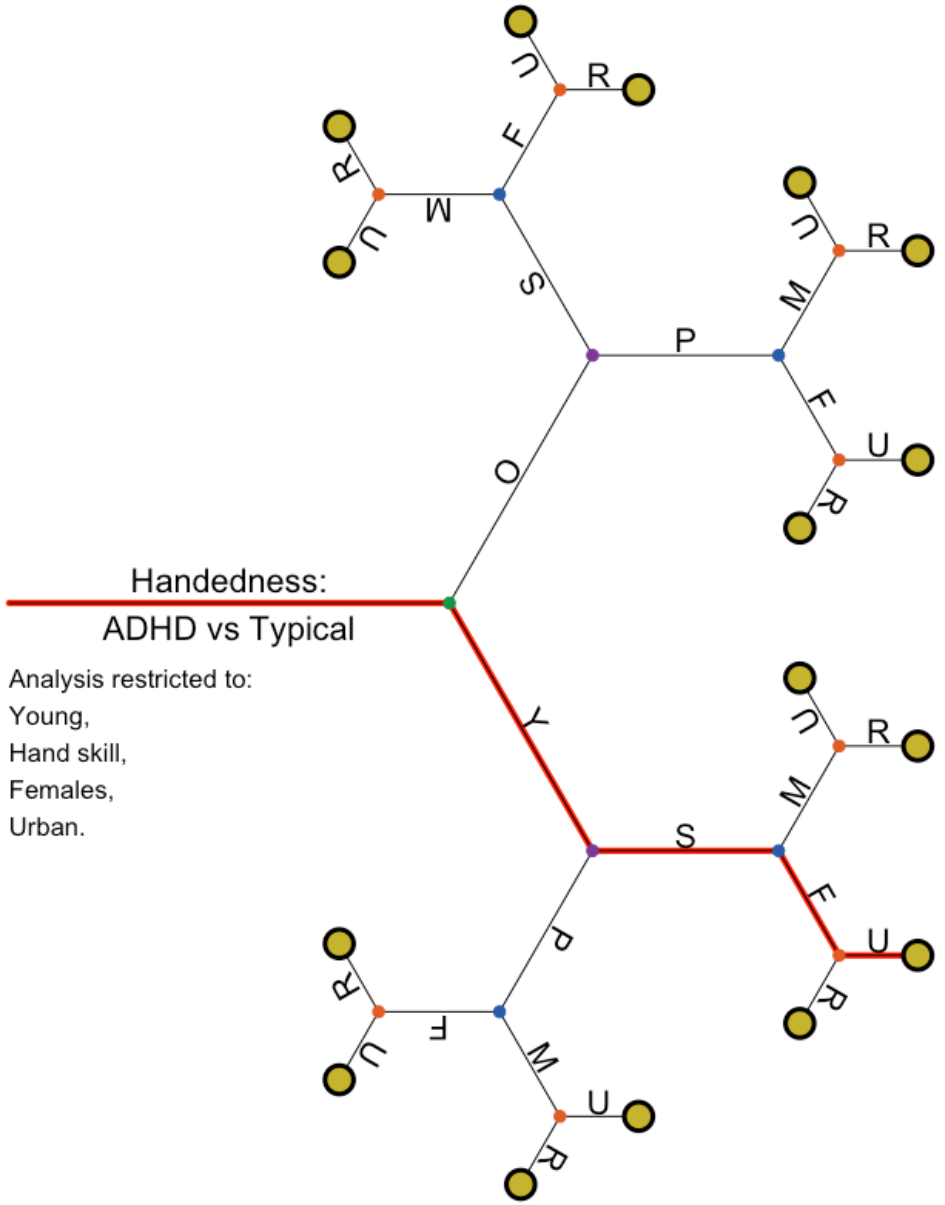
http://deevybee.blogspot.com/2016/01/the-amazing-significo-why-researchers.html

Flexible analytic pipeline -> explosion in N possible comparisons

Large population database used to explore link between ADHD and handedness

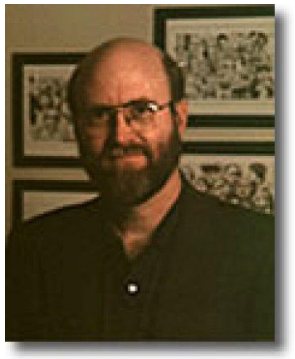
Focus just on Young, Urban, Females on measure of hand skill: 16 contrasts at this level

Probability of a ‘significant’ p-value < .05 = .56



HARKing

1956 De Groot
1969 Cohen
1975 Greenwald
1979 Rosenthal
1998 Kerr



HARKing: Hypothesizing After the Results are Known

Norbert L. Kerr
Department of Psychology
Michigan State University

- "Presenting post hoc hypotheses in a research report as if they were, in fact, a priori hypotheses."
- A way of "translating type I errors into theory"

In survey by Kerr & Harris (1998), 52% respondents said they knew of editors/reviewers encouraging HARKing

The four horsemen of the Apocalypse

HARKing

Low power

P-hacking

Publication bias

If we've known about this for decades, why haven't the problems been fixed?

No one cause:

- Cognitive biases that make it hard to do science well
- N.B. need more and better training in research design and statistics, **that takes these biases into account**
- Research environment and incentives

Cognitive biases that make it hard to do science well

- Tendency to see patterns in things
- Failure to understand sampling and probability
- Confirmation bias
- Errors of omission seen as acceptable

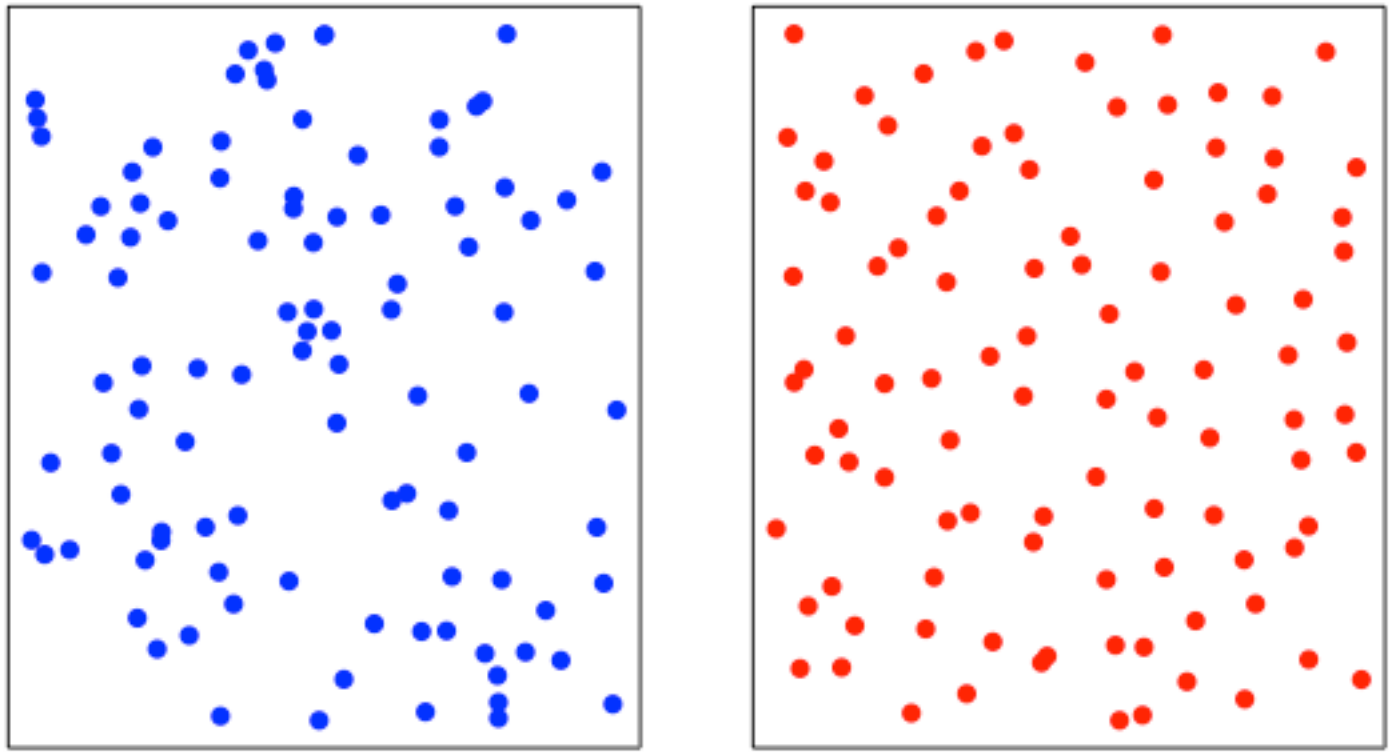
Tendency to see patterns in things

Failure to appreciate power of ‘the prepared mind’

Pareidolia

Position of bomb hits:
General has map of bomb hits and wants to know if bombs were dropped at random or whether some sites are being targeted.

Which map suggests targeting? Blue, red or neither?



Example from Lazic, S. (2016) Experimental Design for Laboratory Biologists

Failure to understand sampling and probability

Consider this problem

A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50% of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50%, sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys. Which hospital do you think recorded more such days?

1.The larger hospital
2.The smaller hospital
3.About the same (that is, within 5% of each other)

Example from Daniel Kahneman & Amos Tversky

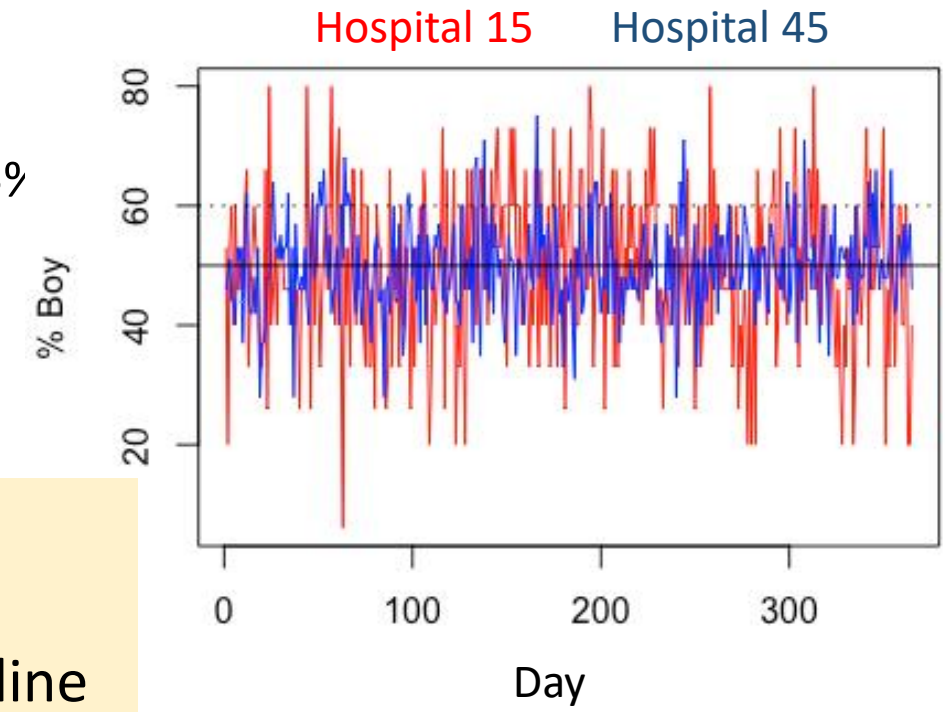
Example from Daniel Kahneman & Amos Tversky

A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50% of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50%, sometimes lower. For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys. Which hospital do you think recorded more such days?

1.The larger hospital
2.The smaller hospital
3.About the same (that is, within 5%

Expected value:
Hosp15 = 57 days
Hosp45 = 26 days

Small sample gives noisier estimates: red line bounces around much more than blue line



Insensitivity to sample size

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105-110.

- People have strong intuitions about random sampling;
- These intuitions are wrong in fundamental respects;
- These intuitions are shared by naive subjects and by trained scientists;
- Intuitions are applied with unfortunate consequences in the course of scientific inquiry

Work in progress: The experimenter game

- A pharma company has a promising drug that claims to reduce white matter loss in ageing rats by half a standard deviation

- You think there's a 50:50 chance that it really works

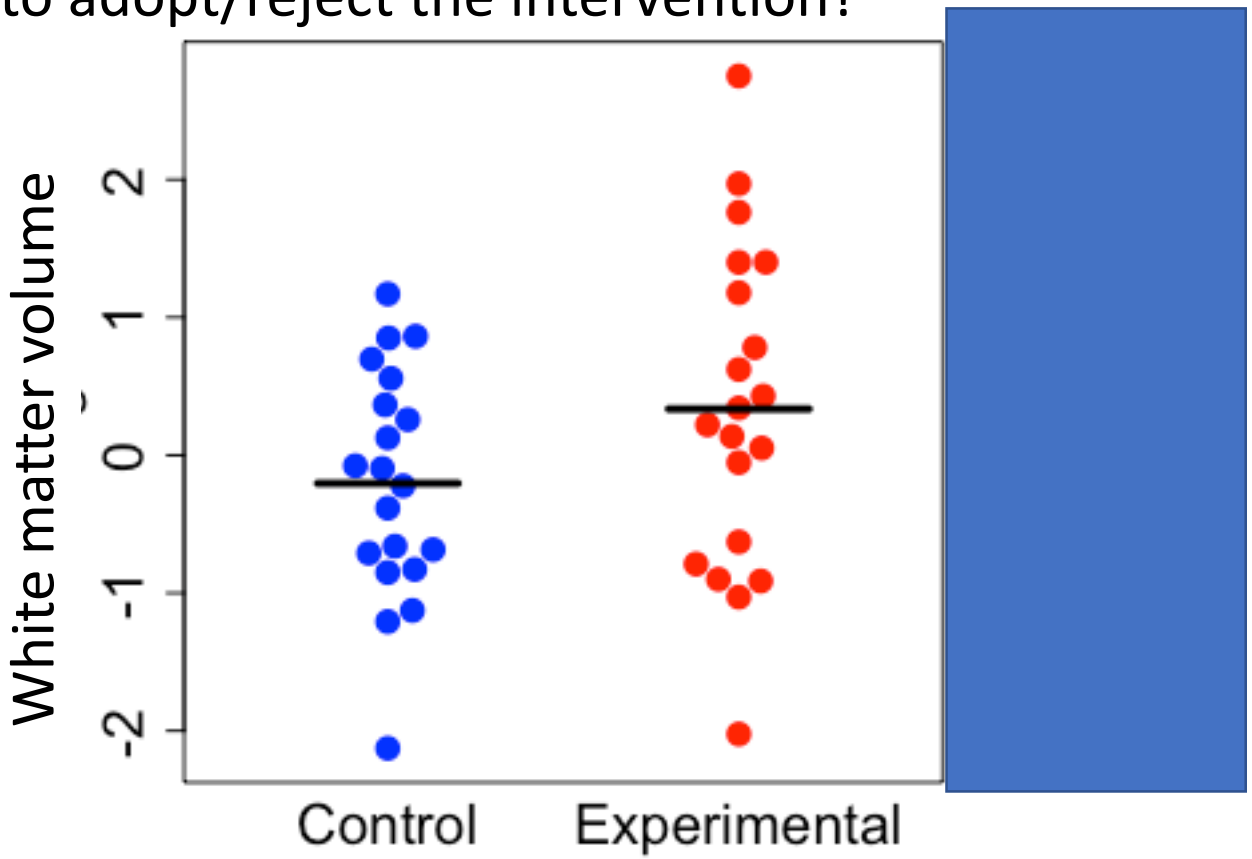
- You can run some tests on samples of rats, but it costs money – the more rats you test, the more expensive.

- You have an optimization problem!
- So what's your experimental strategy?

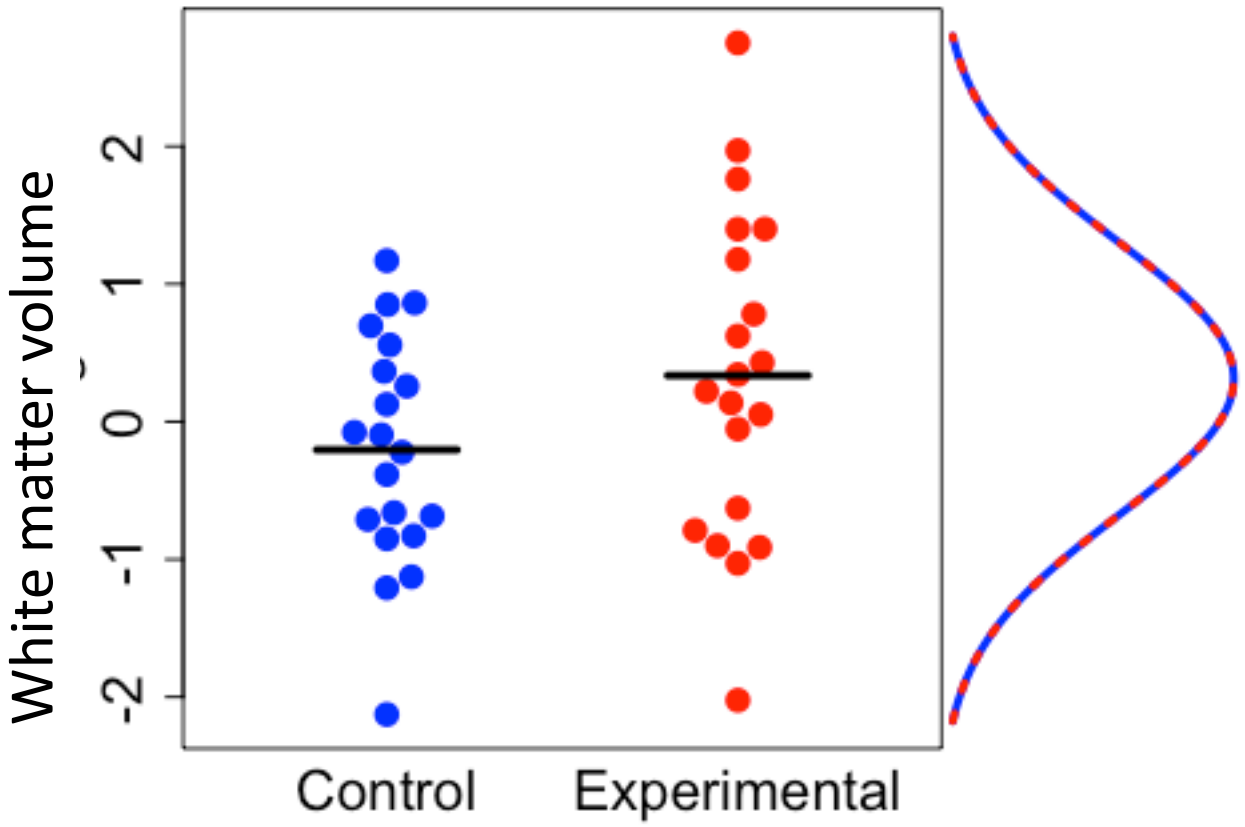
You decide to run a study with two groups of N rats

What value of N should you start with? - let's try 20 per group

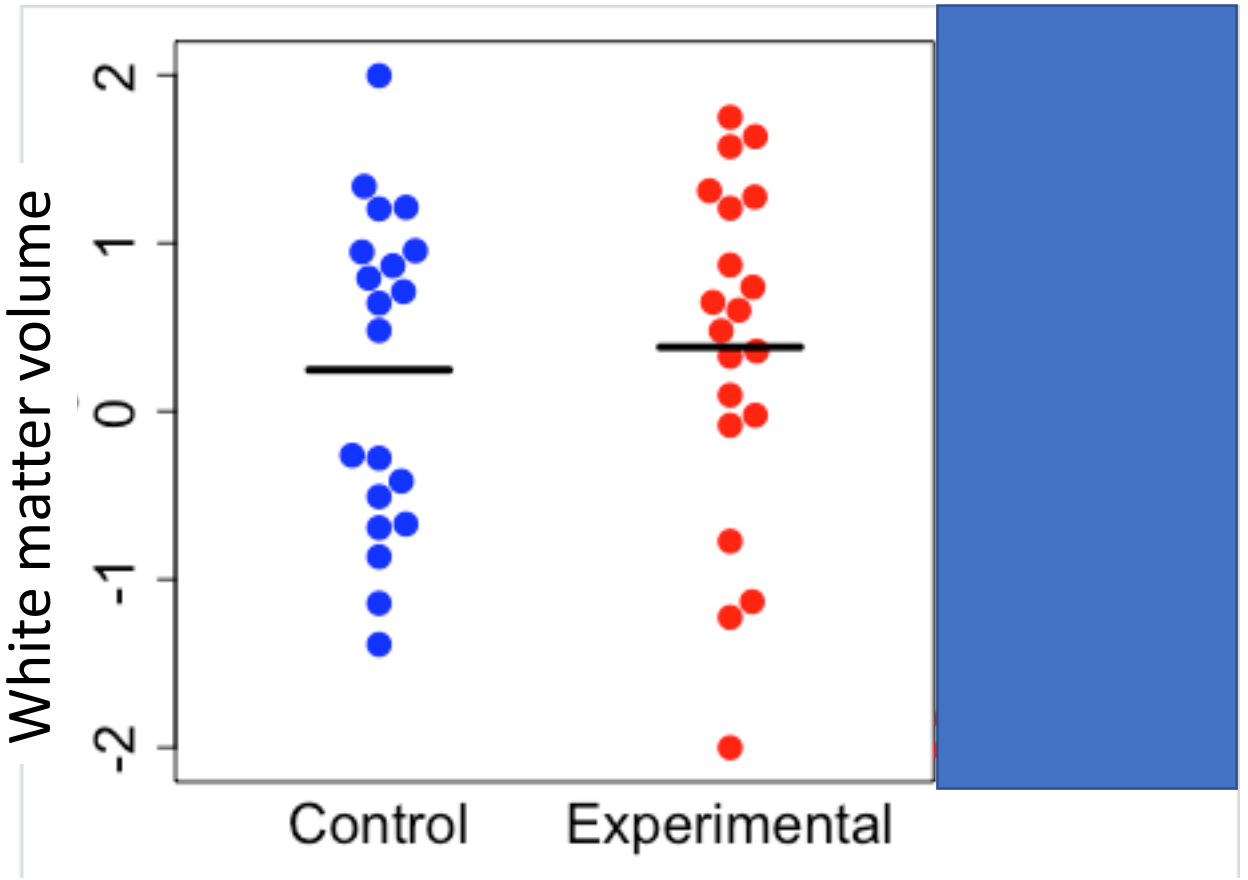
Here's a sample of data: do you think this is sufficient to decide whether to adopt/reject the intervention?



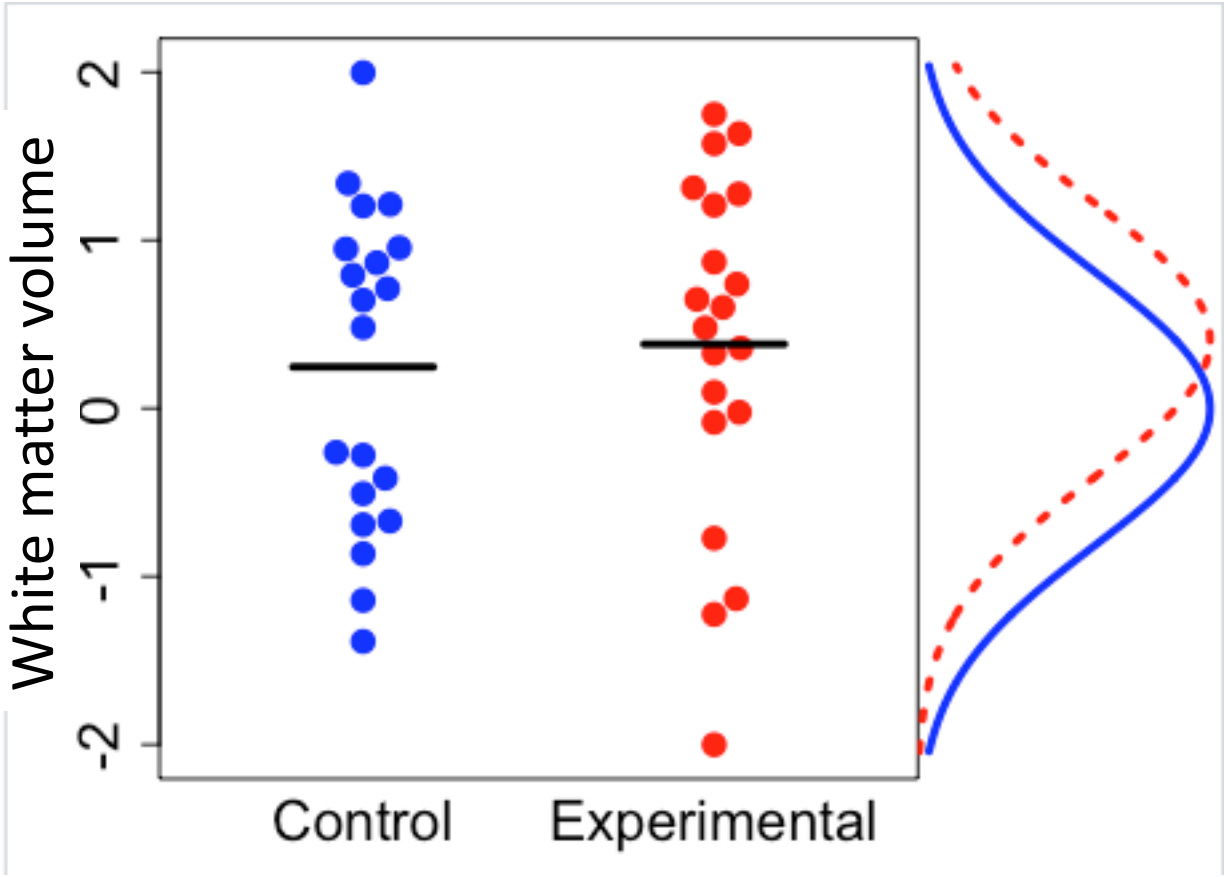
This sample was drawn from population with no real difference



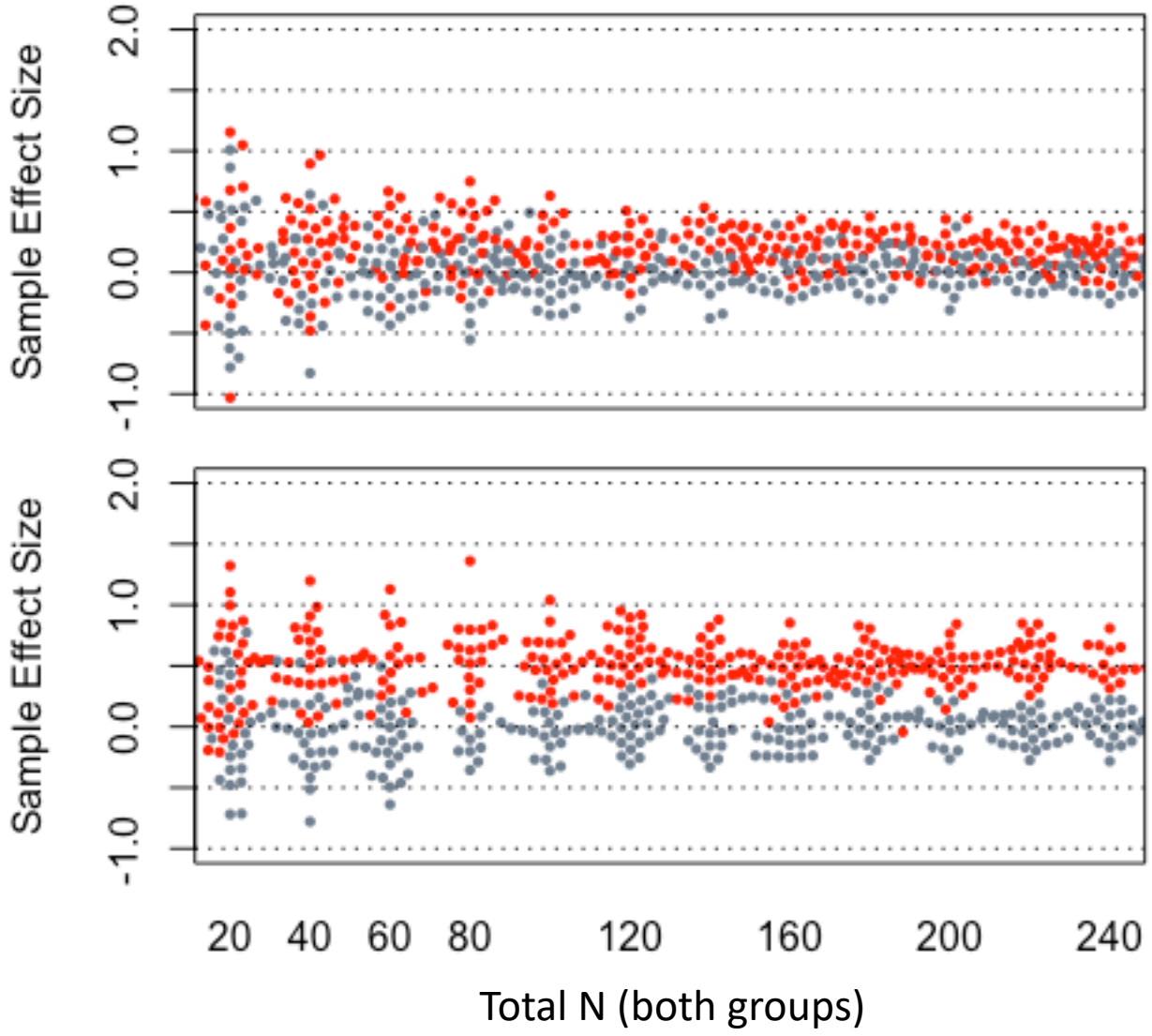
Here's another sample of data: do you think this is sufficient to decide whether to adopt/reject the intervention?



This time the sample was drawn from a sample with a true effect.
With small sample, difference can look small when there is a true effect – this illustrates problem of LOW POWER



Separation between red dots (drawn from population with true effect) and grey dots (drawn from population with no effect) shows sample size where can reliably detect a true effect



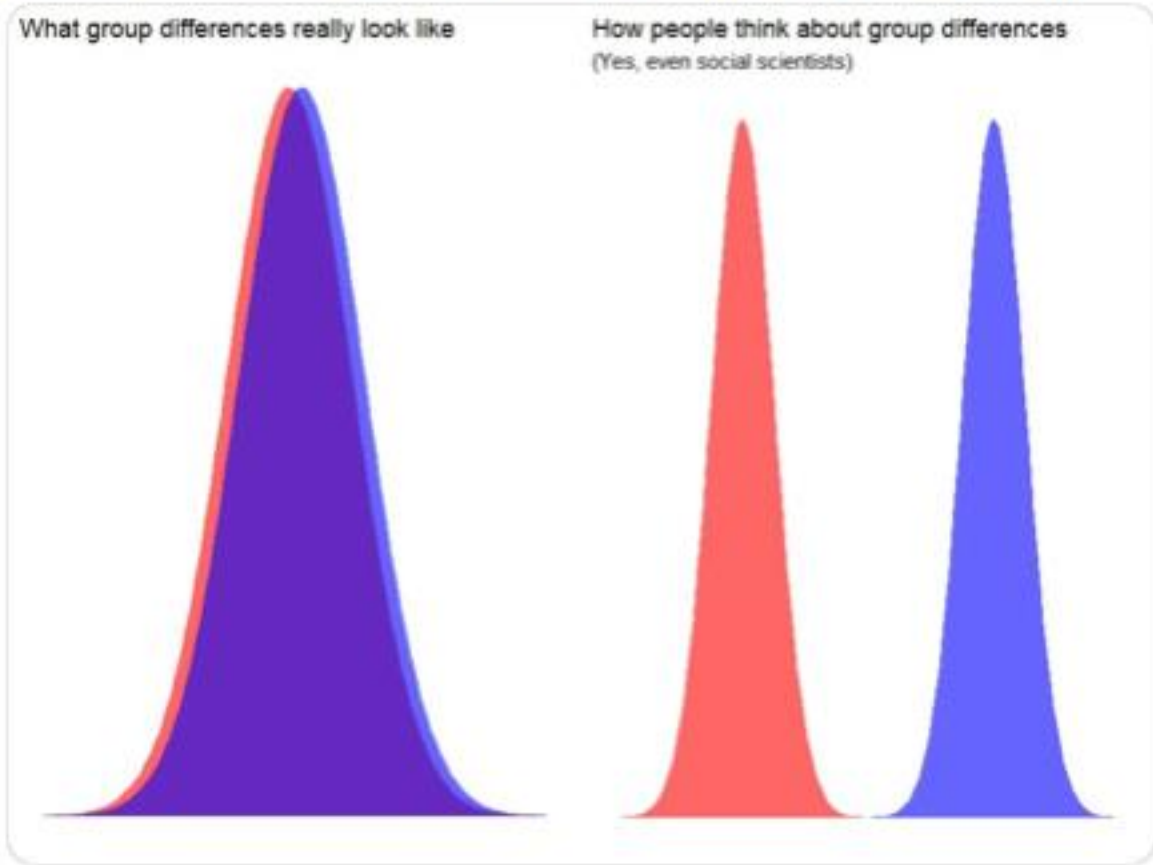
Population effect size = .2

Population effect size = .5

You Retweeted

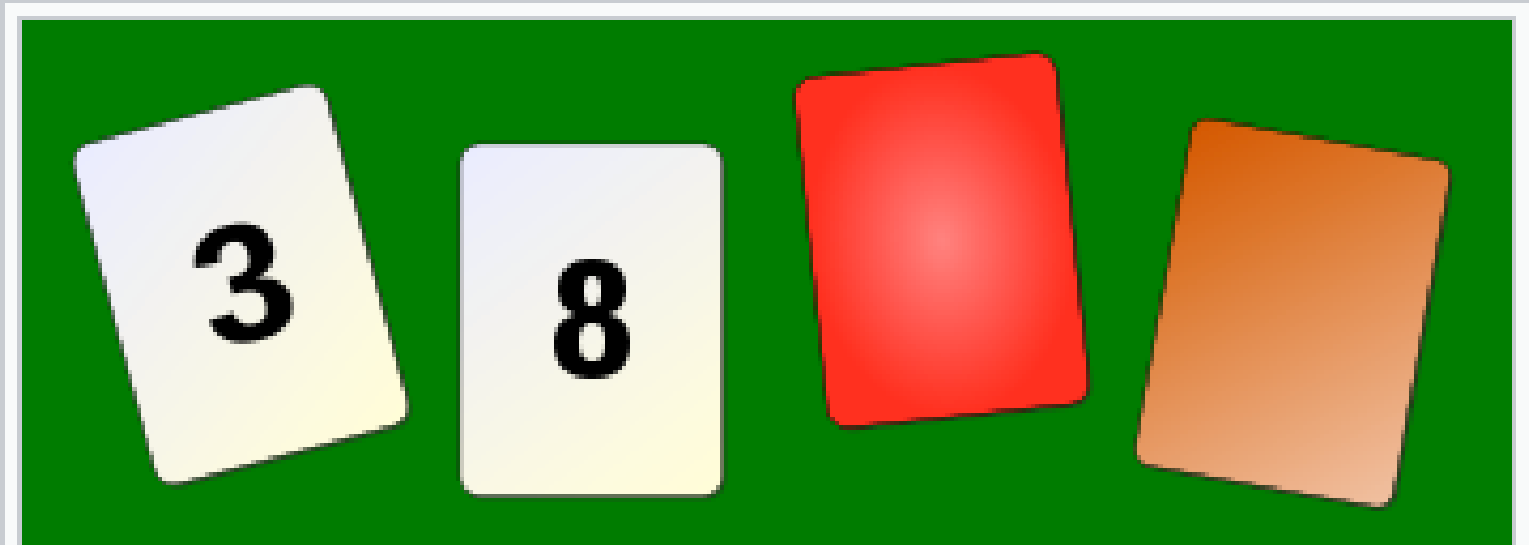
Stephen Vaisey @vaiseys · Nov 15

What group differences look like vs. how people talk and think about them.



Confirmation bias

Wason task:
a way of thinking about experimental design

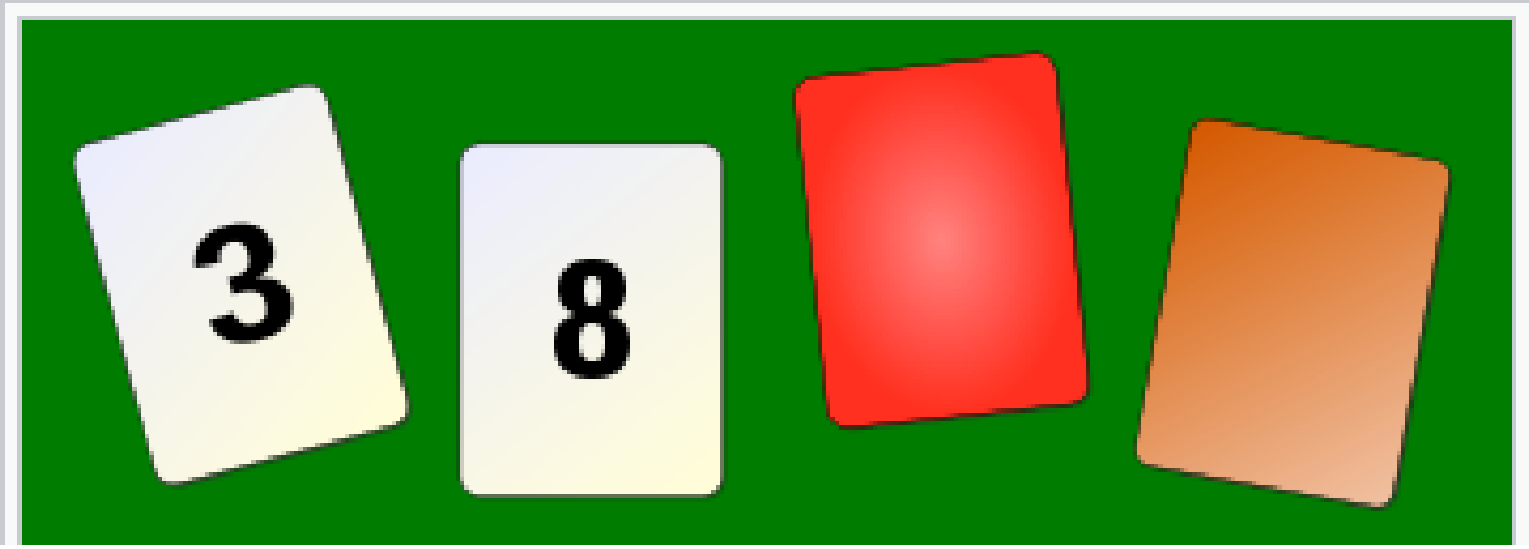


Each card has a number on one side and a patch of colour on the other.

You are asked to test the hypothesis that – for these 4 cards - if an even number appears on one side, then the opposite side is red.

- Which card(s) would you turn over to test the hypothesis?

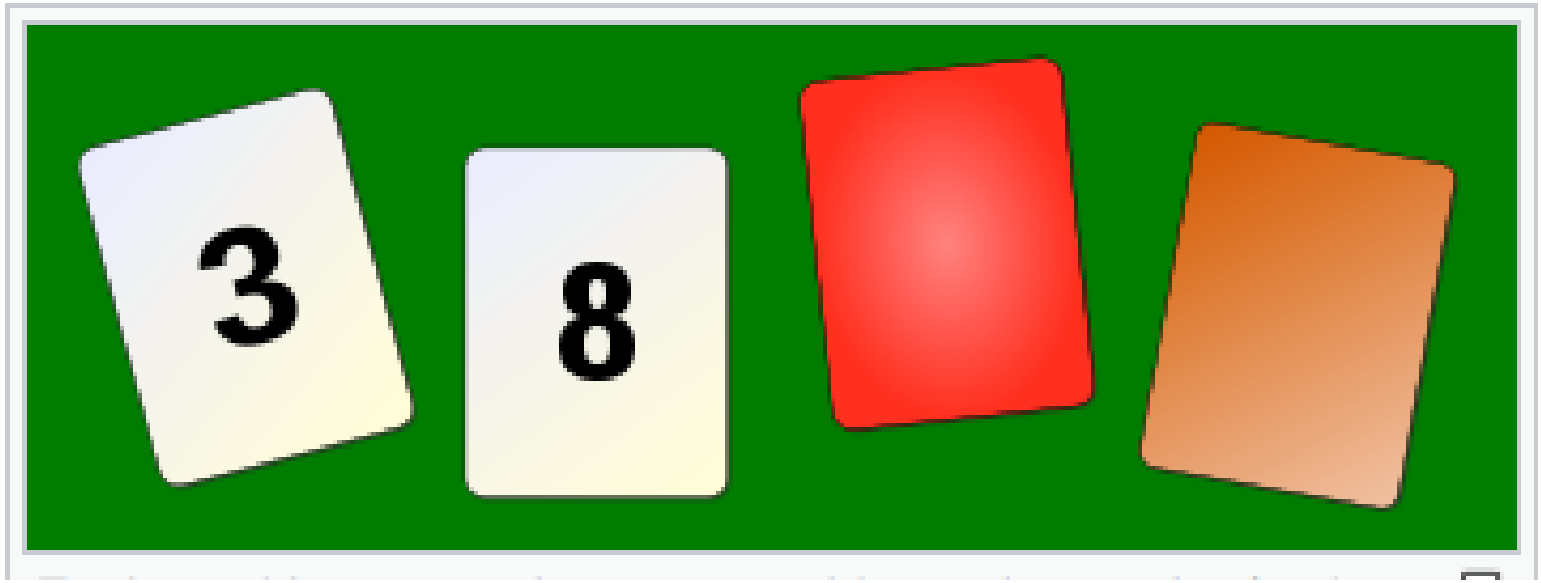
Wason task:
a way of thinking about experimental design



Each card has a number on one side and patch of colour on the other. You are asked to test the hypothesis that – for these 4 cards - if an even number appears on one side, then the opposite side is red.

- Usual response is B & C are critical.
- But C is **not** critical (we're testing 'if P then Q', not 'if Q then P')
- D **is** critical as it has potential to *disconfirm* hypothesis – but usually overlooked

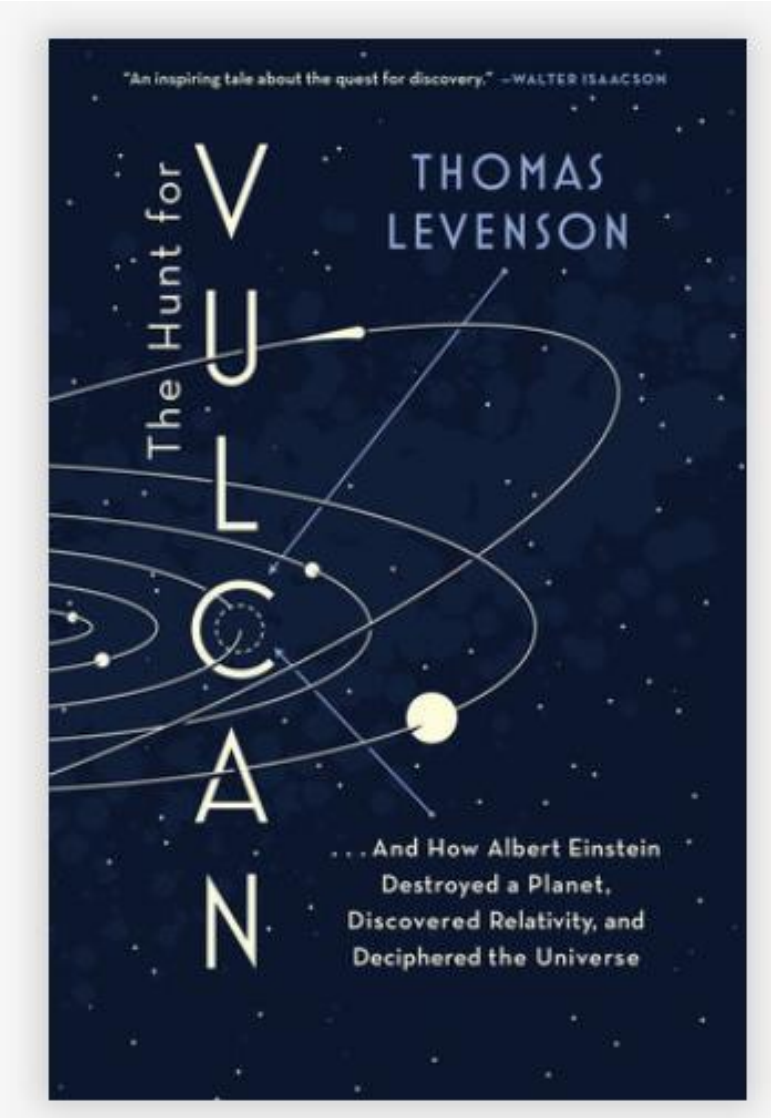
Wason task:
Shows how confirmation bias can affect **experimental design**



In survey of 84 scientists (physicists,biologists, psychologists, sociologists) Mahoney (1976) found fewer than 10% correctly identified the critical cards

We need to design experiments to look for *disconfirmation* of a theory .
In practice: "*To test a hypothesis, we think of a result that would be found if the hypothesis were true and then look for that result*" (J. Baron, 1988, p. 231).

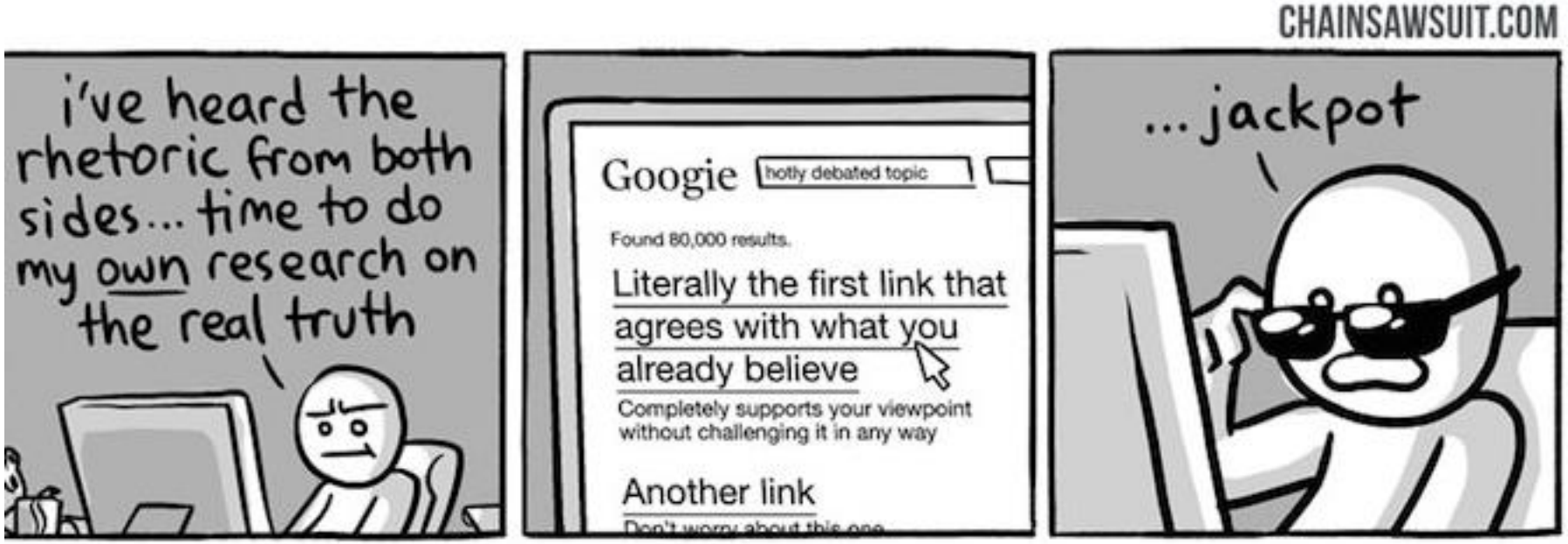
Confirmation bias **at level of observations**: Seeing what you expect to see



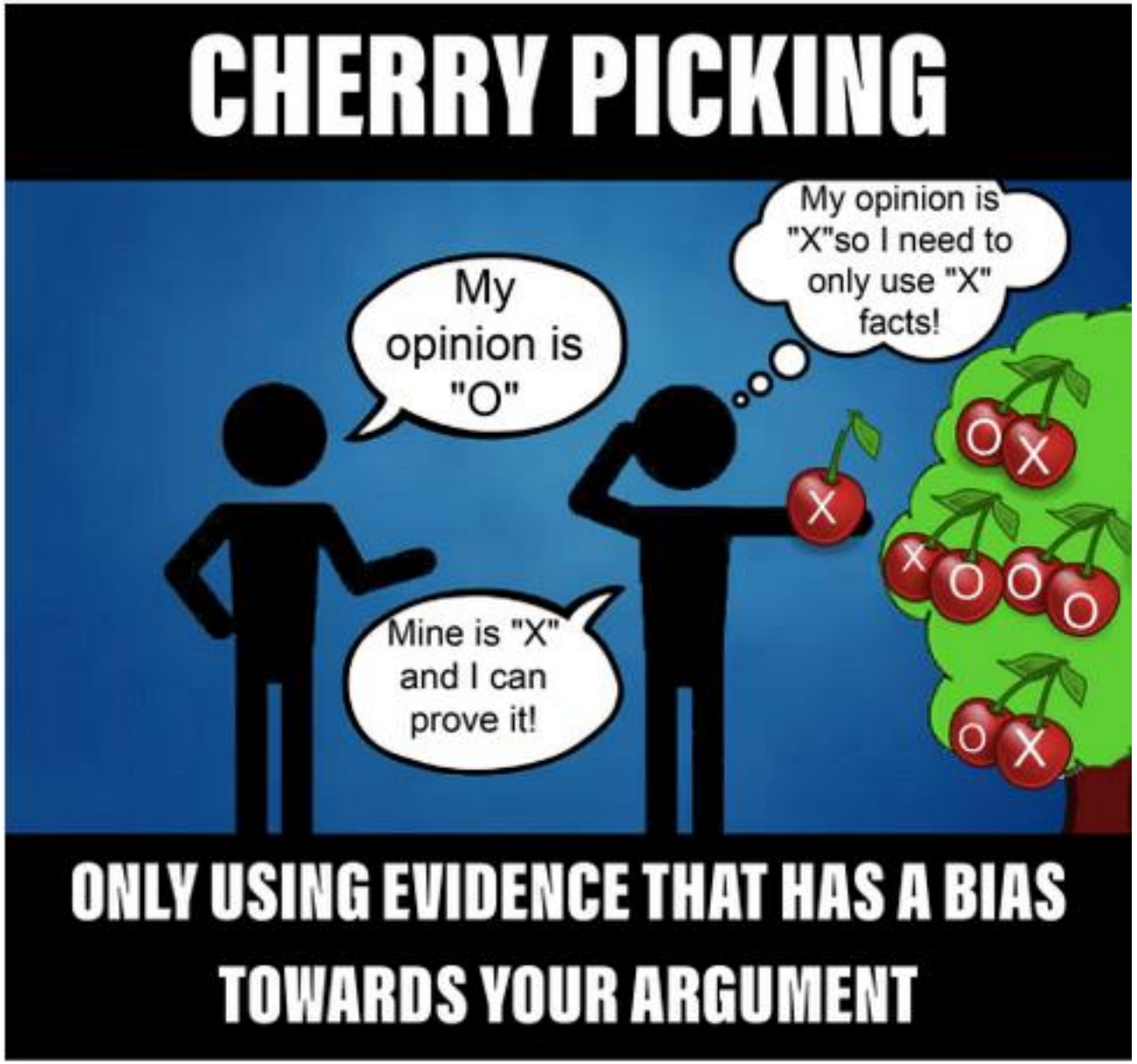
"The self-deception comes in that over the next 20 years, people believed they saw specks of light that corresponded to what they thought Vulcan should look during an eclipse: round objects crossing the face of the sun, which were interpreted as transits of Vulcan."

Confirmation bias affects **how we remember and process information**

- Cherry-picking may not be deliberate
- We find it much easier to process and remember information that agrees with our viewpoint



Confirmation bias affects **literature reviews**



Consequence of omission errors in literature reviews

- When we read a peer-reviewed paper, we tend to trust the citations that back up a point
- When we come to write our own paper, we cite the same materials
- A good scientist won't cite papers without reading them, but even this won't save you from bias – you inherit it from prior papers
- If prior papers only cite materials agreeing with a viewpoint, that viewpoint gets entrenched
- You won't know – unless you explicitly search – that there are other papers that give a different picture

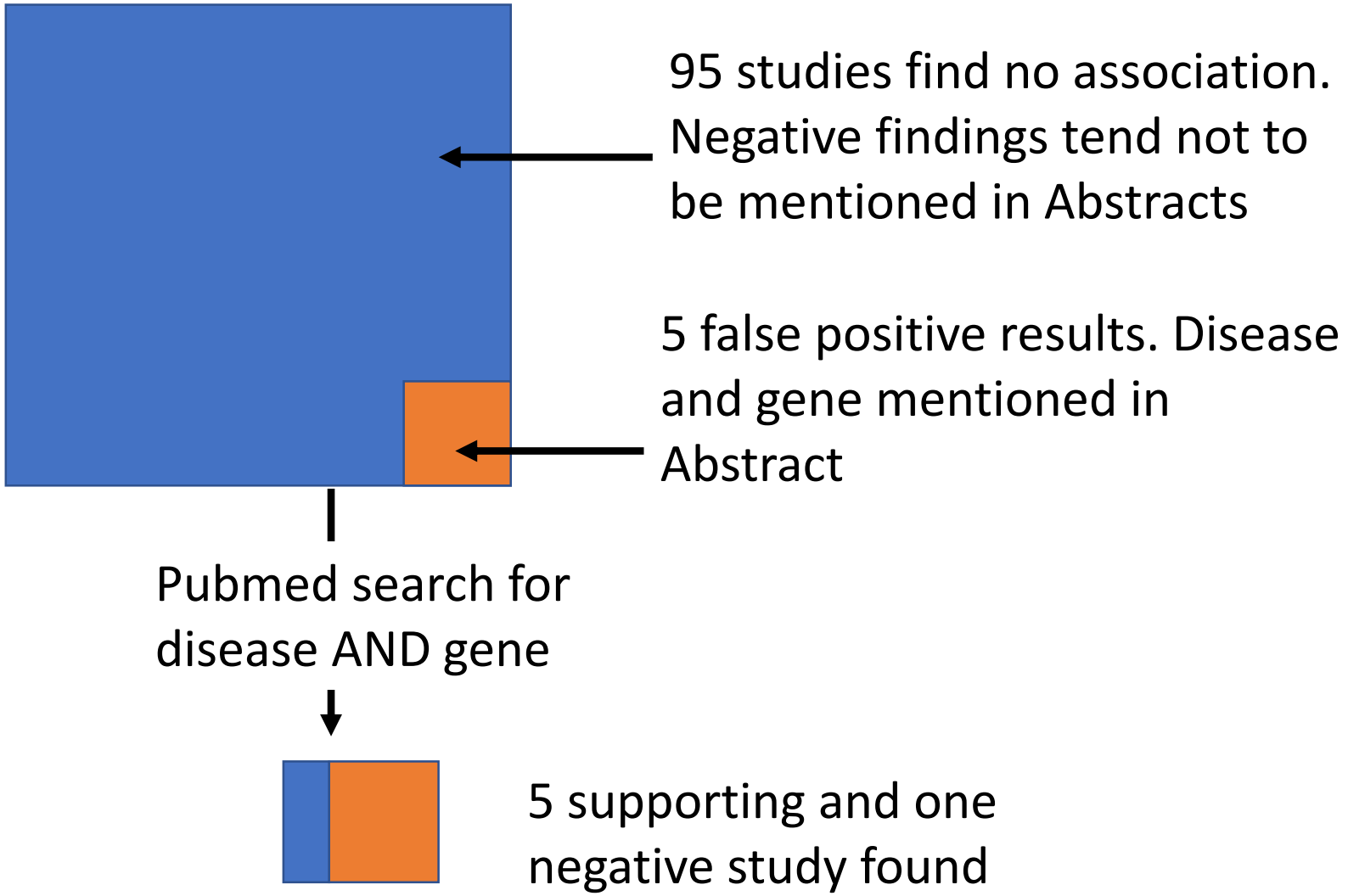
The (partial*) solution

Always start with a systematic review

- **Systematic review**
  - Collecting and summarise all empirical evidence that fits pre-specified eligibility criteria to address a specific question
- **Meta-analysis**
  - Use statistical methods to summarise the results of these studies

**But depends on finding all relevant papers*

100 relevant studies on gene/disease association



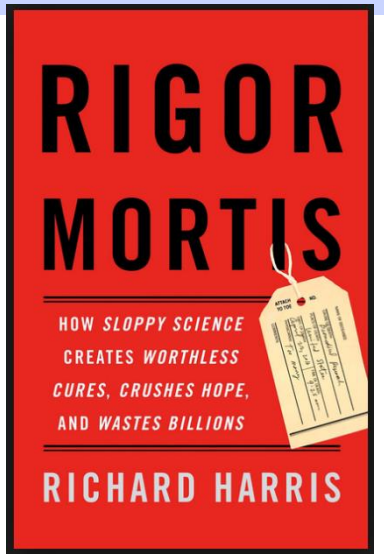
Example from Lazic, S. (2016) Experimental Design for Laboratory Biologists

Cognitive biases pervade every step of the research process

Reading literature	Confirmation bias, Omissions
Experimental design	Confirmation bias, Law of small numbers
Experimental observations	Seeing patterns, Confirmation bias
Data analysis	Confirmation bias, Seeing patterns, Law of small numbers, Omissions
Scientific reporting	Confirmation bias, Omissions

Will anything change?

"It really is striking just for how long there have been reports about the poor quality of research methodology, inadequate implementation of research methods and use of inappropriate analysis procedures as well as lack of transparency of reporting. All have failed to stir researchers, funders, regulators, institutions or companies into action". Bustin, 2014



Reasons for optimism

- Concern from those who use research:
  - Doctors and patients
  - Pharma companies
- Concern from funders
- Increase in studies quantifying the problem
- Social media

One solution

Preregistration of analyses

Science
Head quarters

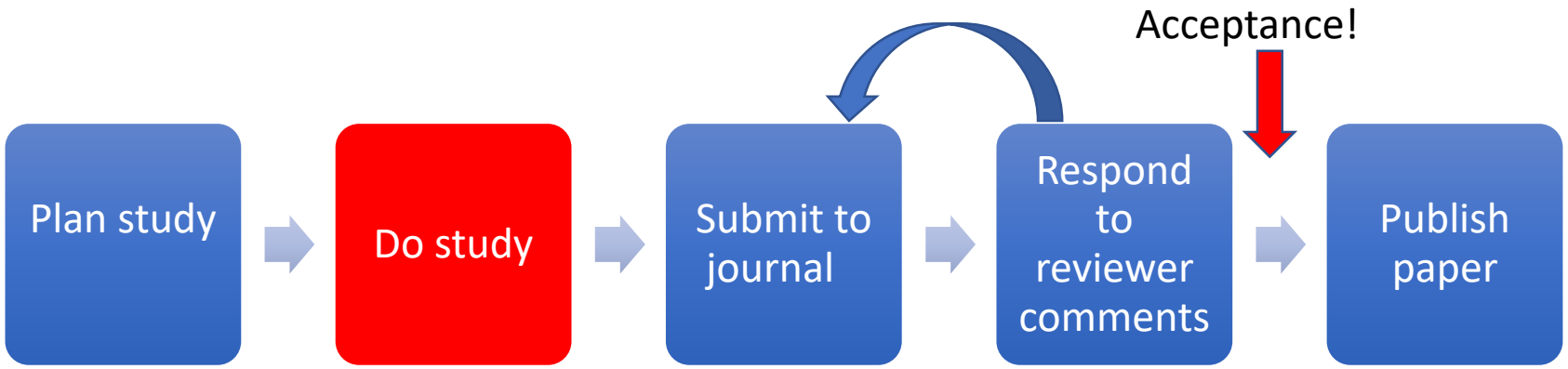
Psychology's 'registration revolution'

Moves to uphold transparency are not only making psychology more scientific – they are harnessing our knowledge of the mind to strengthen science

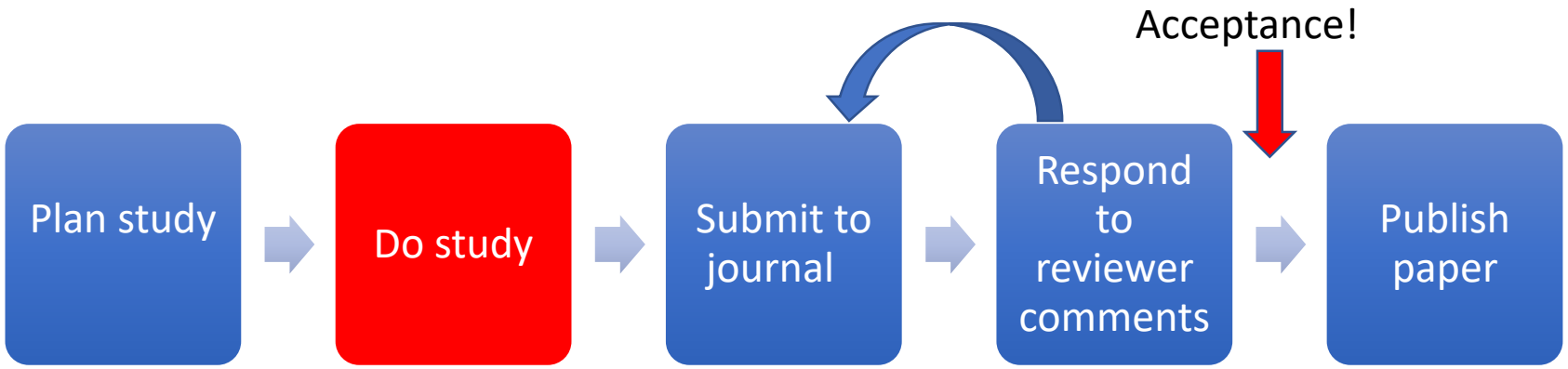
Chris Chambers

Tuesday 20 May 2014 07.15 BST

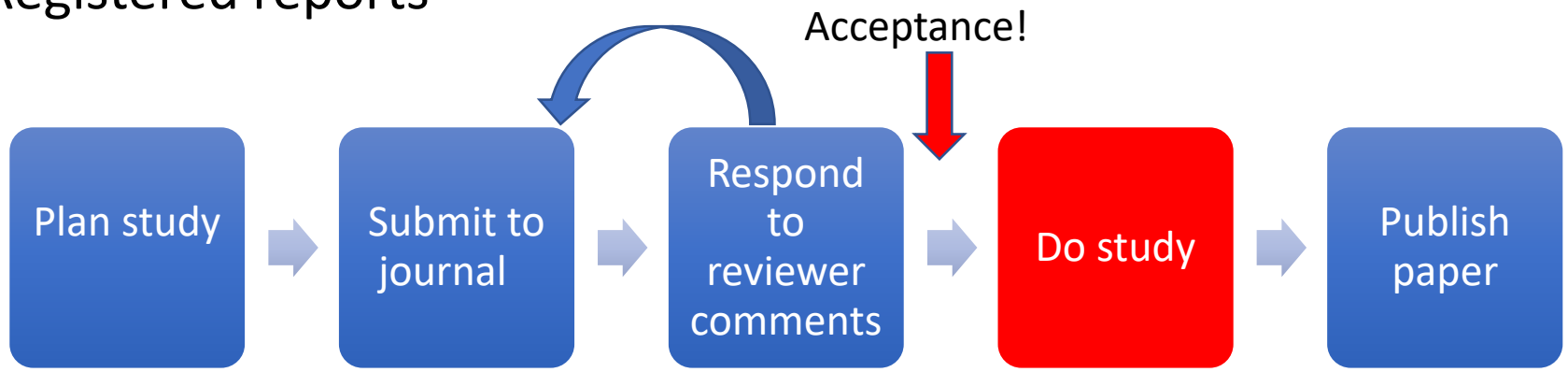
Classic publishing



Classic publishing



Registered reports



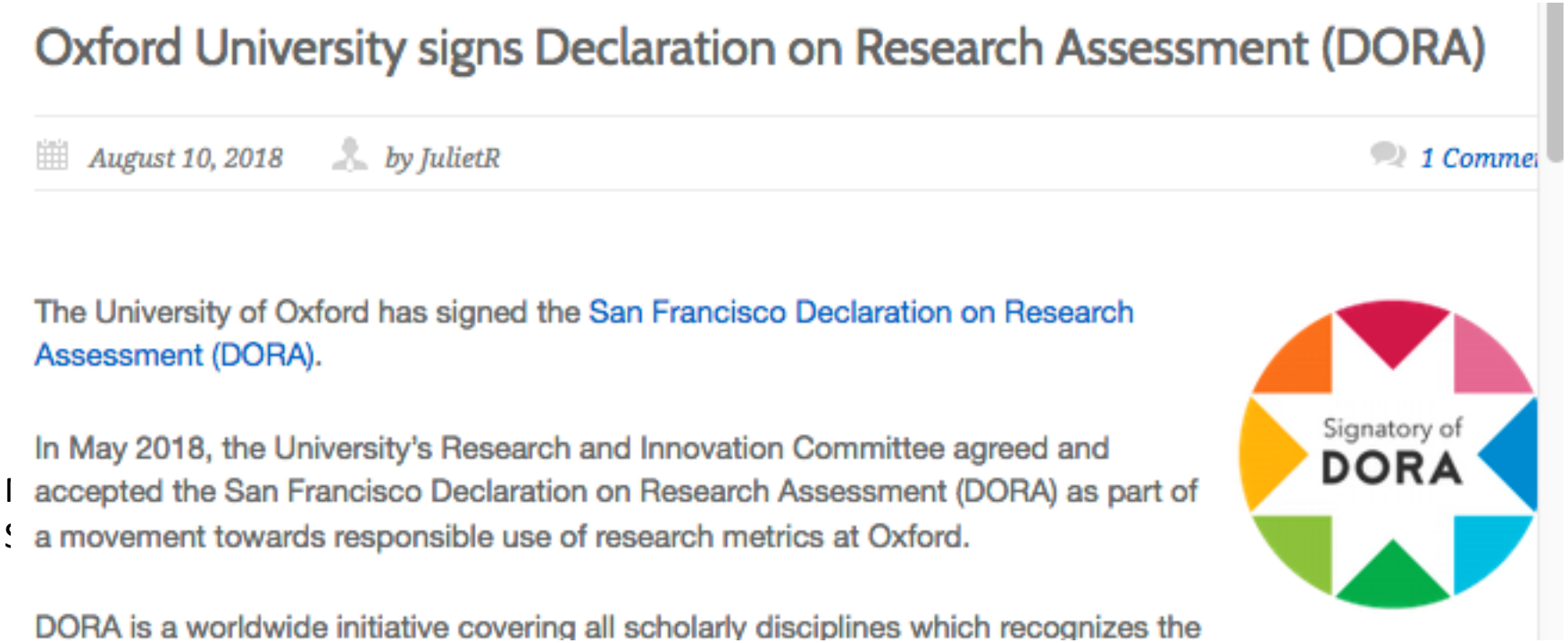
Registered reports solves issues of:

- **Publication bias**: publication decision made on the basis of quality of introduction/methods, before results are known
- **Low power**: researchers required to have 90% power
- **P-hacking**: analysis plan specified up-front
- **HARKing**: hypotheses specified up-front.

Unanticipated findings can be reported but clearly demarcated as ‘exploratory’

Solutions for institutions

- Reward those who adopt open science practices
- Reward research reproducibility over impact factor in evaluation



Oxford University signs Declaration on Research Assessment (DORA)

August 10, 2018 *by JulietR* *1 Comme*

The University of Oxford has signed the San Francisco Declaration on Research Assessment (DORA).

In May 2018, the University's Research and Innovation Committee agreed and accepted the San Francisco Declaration on Research Assessment (DORA) as part of a movement towards responsible use of research metrics at Oxford.

DORA is a worldwide initiative covering all scholarly disciplines which recognizes the

Signatory of DORA

New network launched to enhance the rigour and reliability of UK scientific research

Press release issued: 12 September 2018

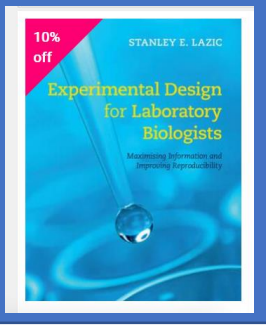
A new Reproducibility Network that aims to improve the rigour and reliability of UK-led scientific research will launch at the University of Bristol this week [Wednesday 12 September]. The Network aims to reinforce the leading position of UK science by co-ordinating shared training and best practice across research-intensive universities.

Resources for training

- http://deevybee.blogspot.com/2012/11/bishopblog-catalogue-updated-24th-nov.html (Google 'bishopblog catalogue')
- https://www.slideshare.net/deevybishop

Experimental Design for Laboratory Biologists : Maximising Information and Improving Reproducibility

Stanley E Lazic



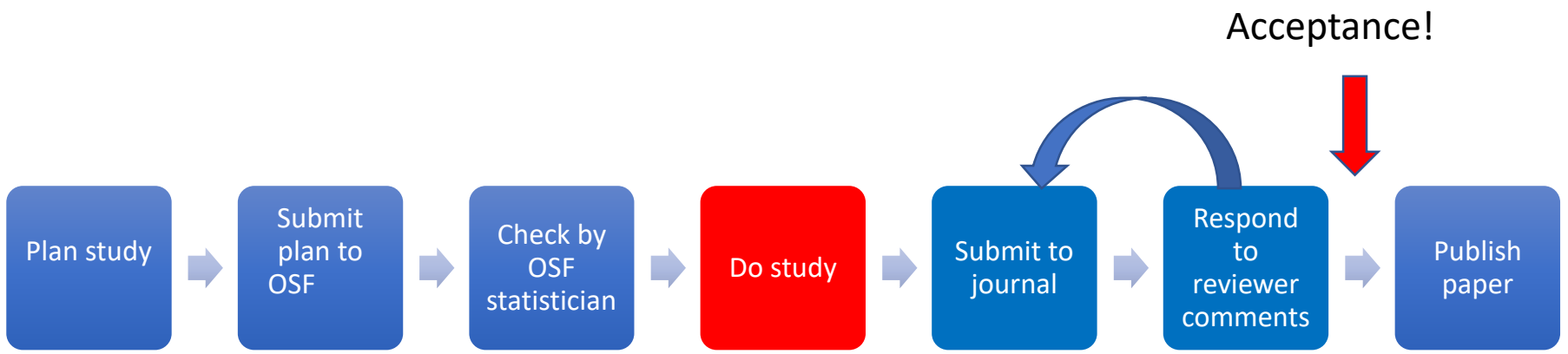
New!!!
Interested in volunteering for our training study?

Contact Jackie.Thompson@psy.ox.ac.uk

Professor Dorothy Bishop, FRS, FMedSci, FBA,
Wellcome Trust Principal Research Fellow,
Department of Experimental Psychology,
Anna Watts Building,
Woodstock Road,
Oxford,
OX2 6GG.

@deevybee

Pre-registration on Open Science Framework



- Similar to regular publication route
- No guarantee of publication
- But reviewers generally positive about preregistered papers because prevents p-hacking or HARKing
- And benefits of having well-worked out plan – less stress when it comes to making sense of data

COS CENTER FOR OPEN SCIENCE