Understanding how cognitive biases affect research

Dorothy V. M. Bishop

Emeritus Professor of Developmental Neuropsychology

University of Oxford

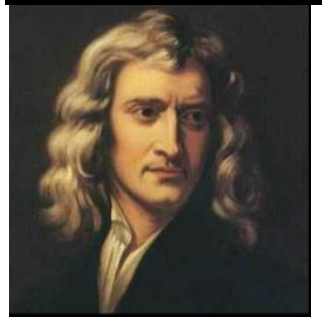
@deevybee@mastodon.social

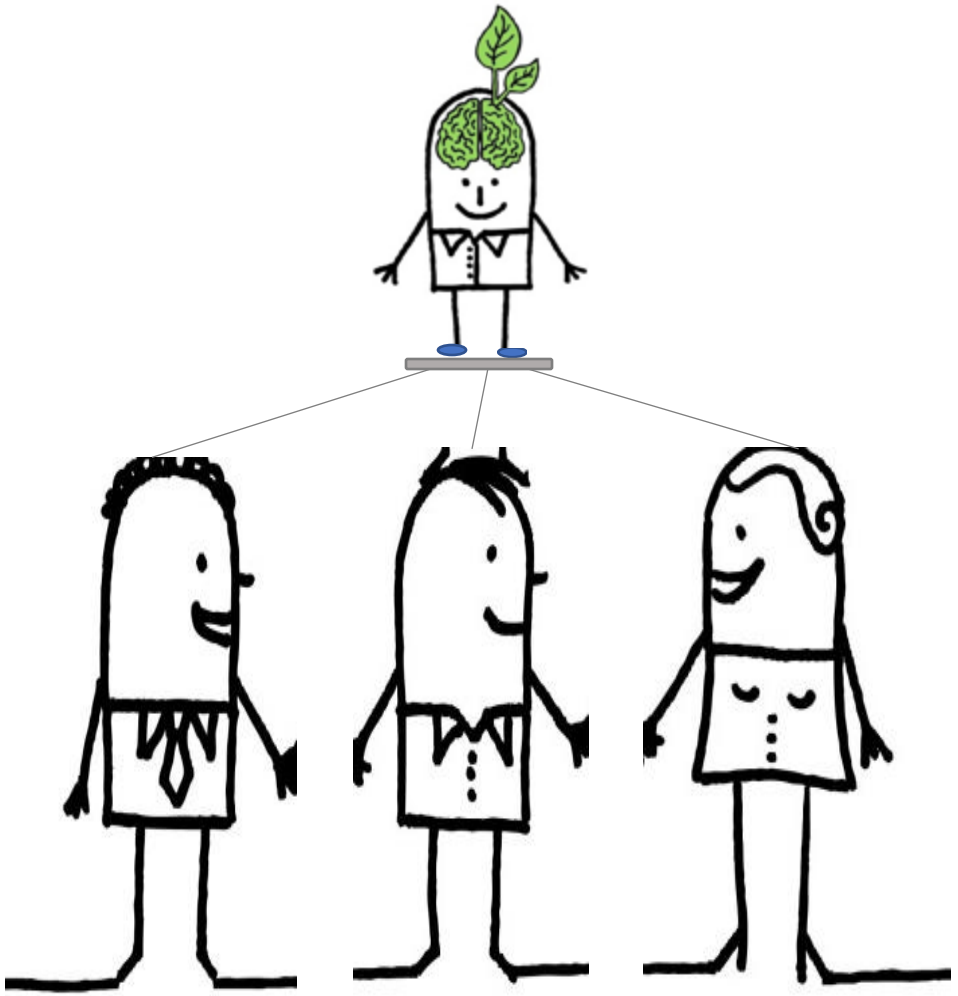
@deevybee.bsky.social

Cumulative science

"If I have seen further...
It is by standing on
the shoulders of giants...."

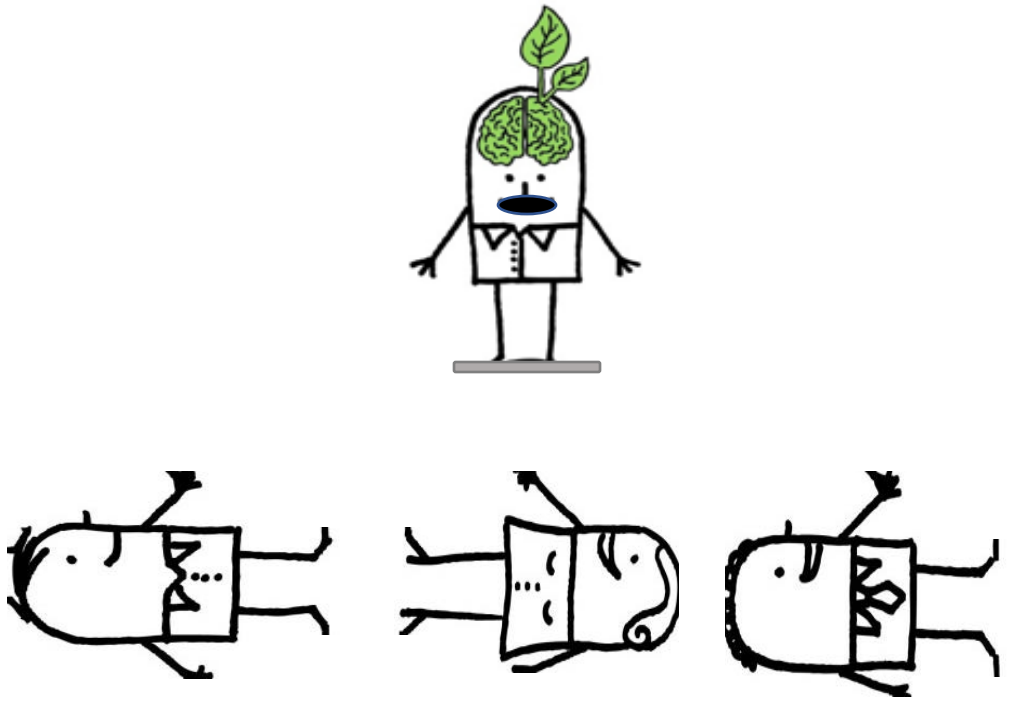
Isaac Newton
1676





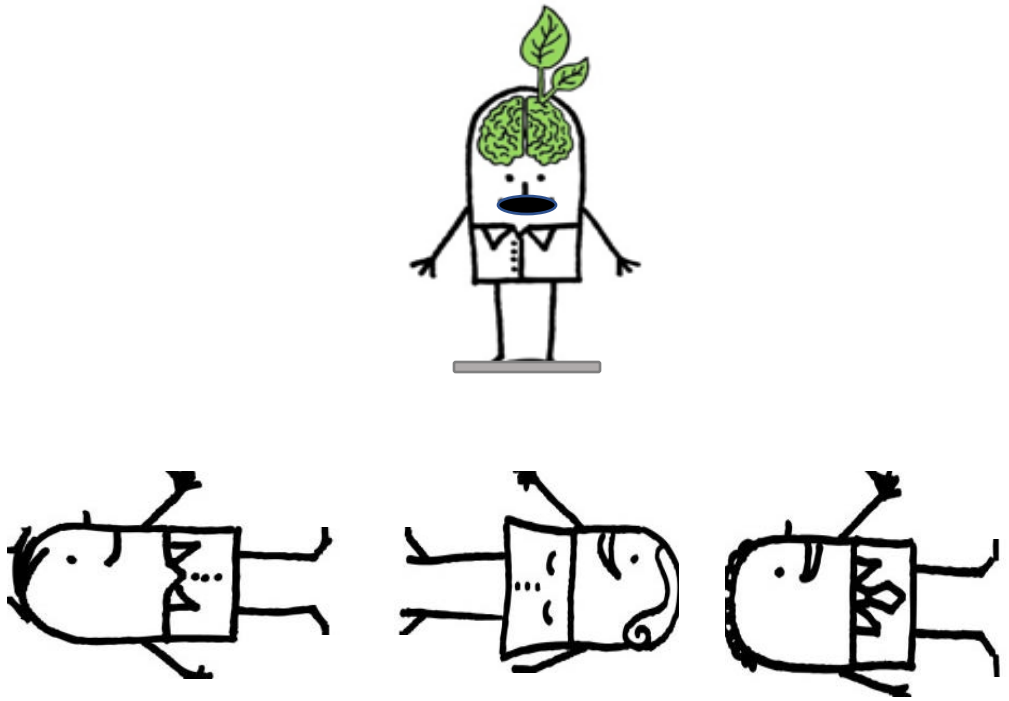
Quasi cumulative science

A common tragedy for ECRs:
Trying to build on previous literature that is wrong



Quasi cumulative science

A tragedy also for society – particularly apparent in the age of COVID19



So what’s gone wrong?

The four horsemen of the reproducibility Apocalypse

Publication bias

Low power

Data dredging (P-hacking)

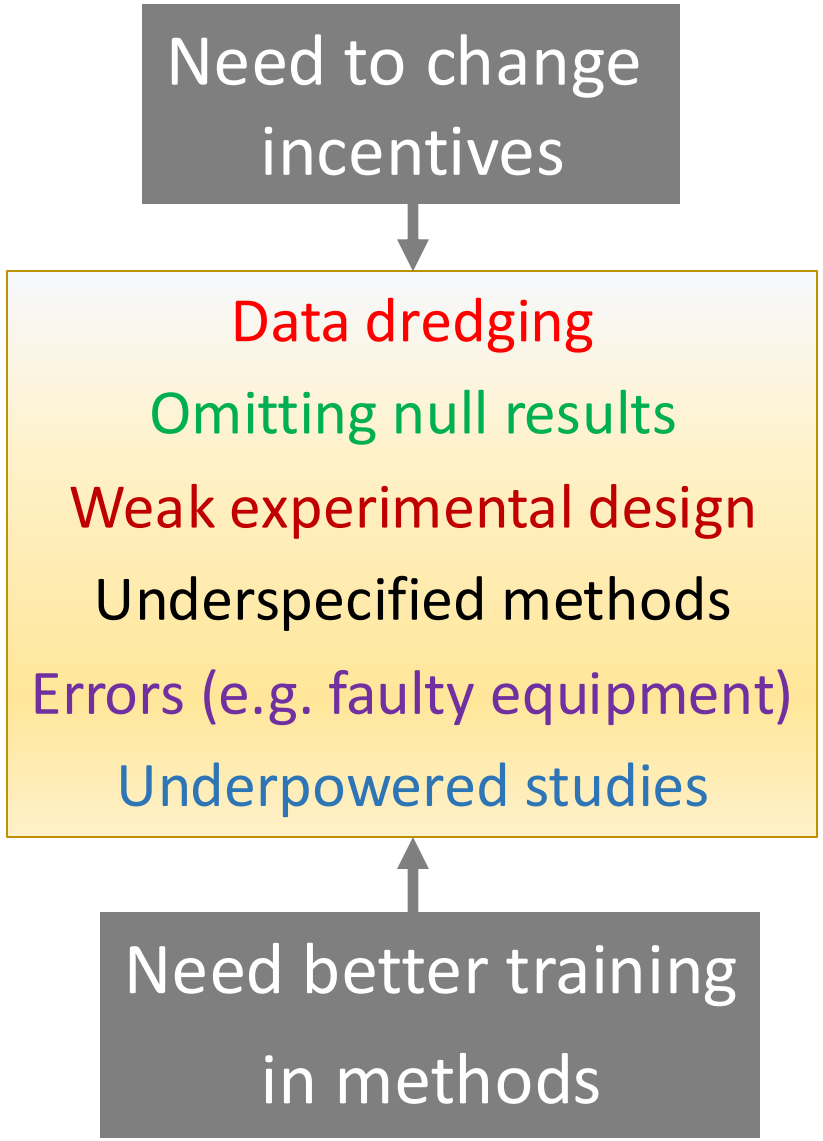
Hypothesising after results are known (HARKing)

Bishop, D. V. M. (2019). Rein in the four horsemen of irreproducibility. *Nature, 568*, 435. doi:10.1038/d41586-019-01307-2

We’ve known about all these for decades

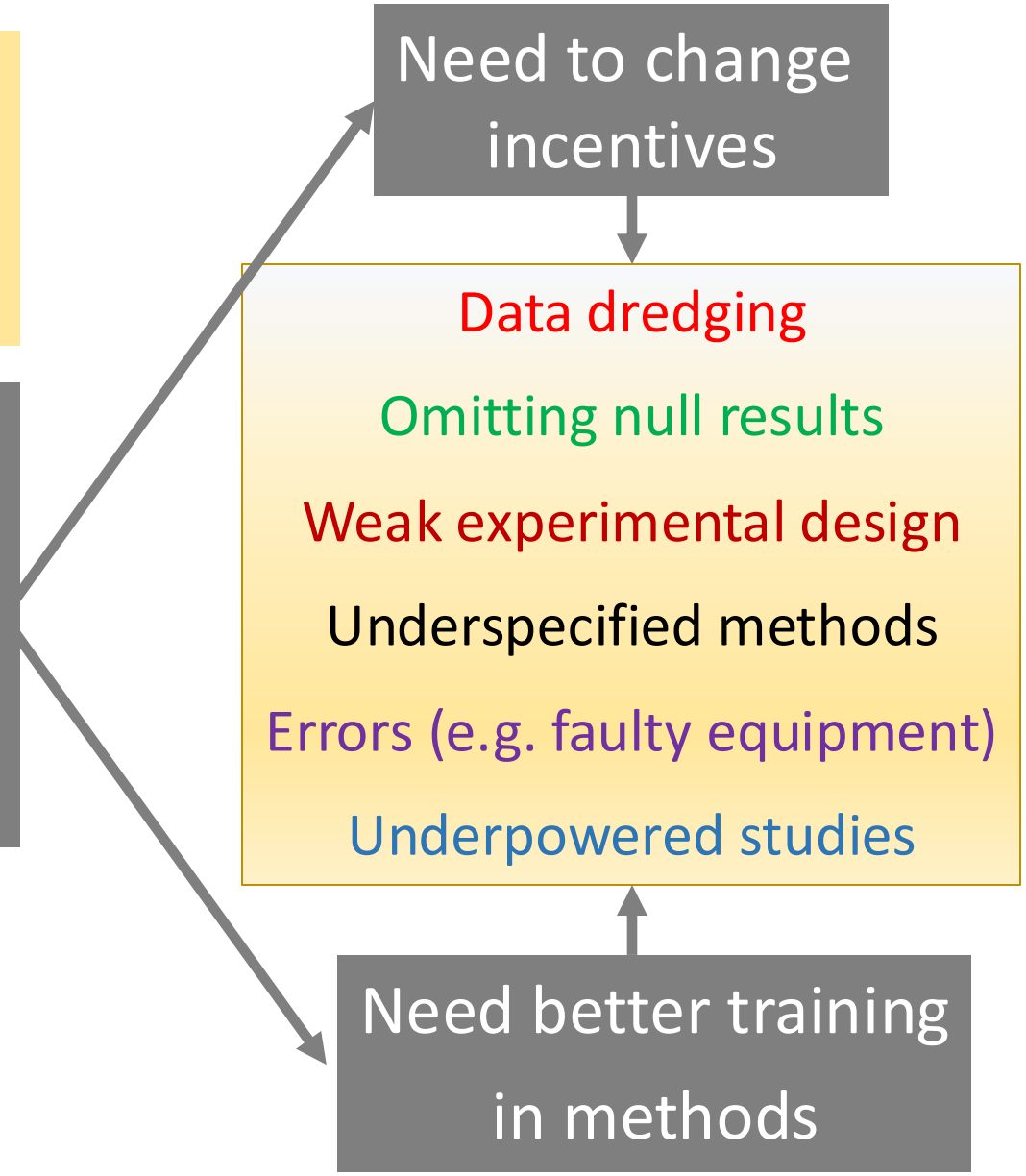
So why are we still doing things wrong?

Academy of Medical Sciences, 2015
Report on Reproducibility and Reliability of Biomedical Research



What's missing?
How humans think & reason

Find ways to counteract cognitive biases



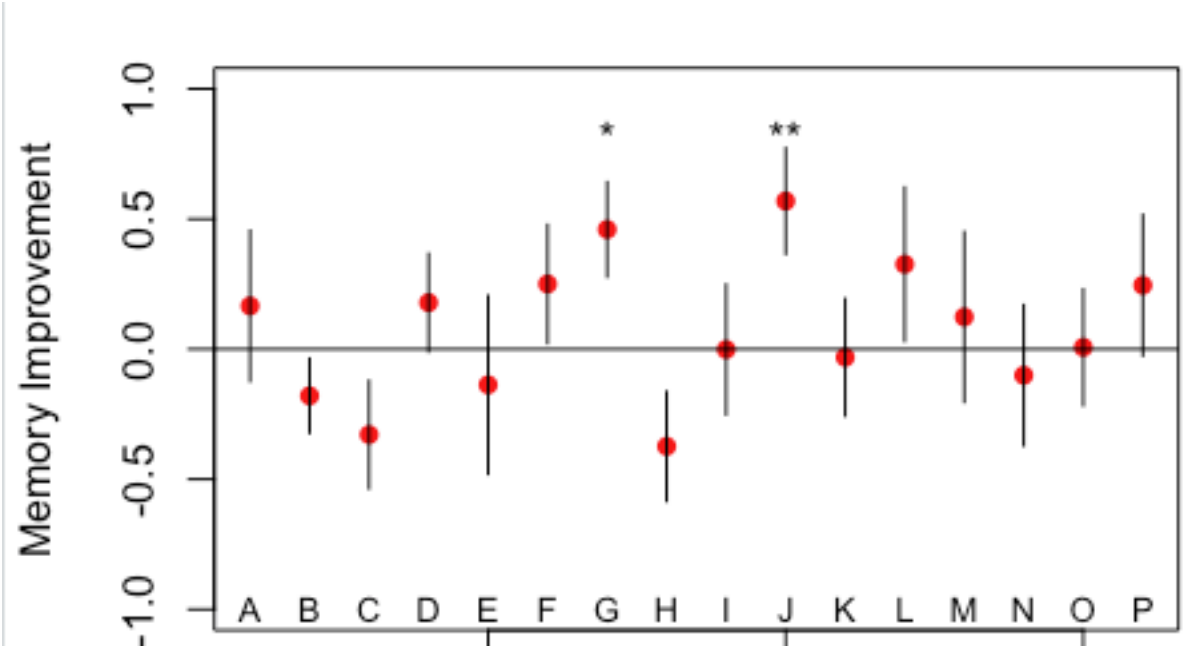
Three cognitive constraints that can make it hard to do science well

- Systematic misunderstanding of probability
- Asymmetric moral judgements
- Confirmation bias: selective attention/memory

Errors of statistical reasoning

Failure to understand p-values leads to p-hacking

I test 16 compounds to see if they improve memory in rats
Should I be excited? Error bars show SE.



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

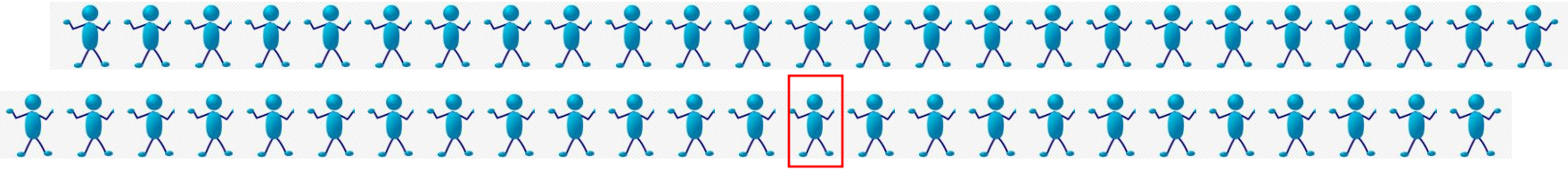
Simple explainer using poker

3 of a kind

Probability from unbiased deck of cards = 1 in 50

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

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

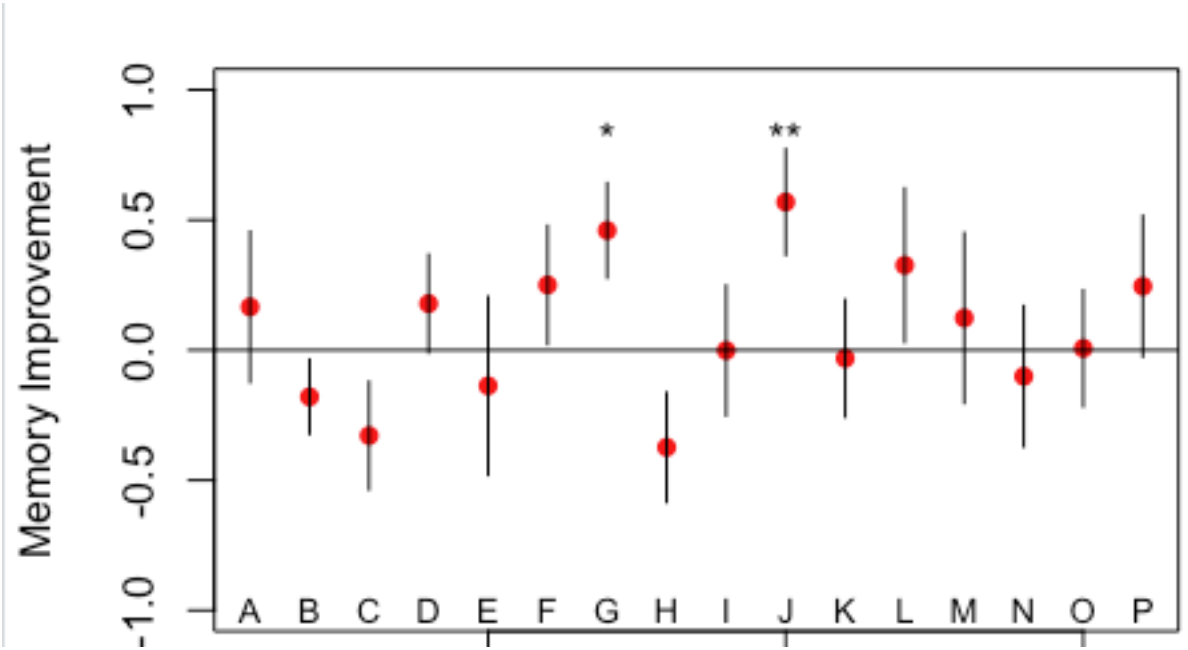


'Surprisingness' of a result only interpretable in context of full dataset

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

Key error is to treat a p-value as an indicator of importance of a finding, regardless of context

Simulated data from random normal distribution with mean = 0



- I'm not testing whether J is 'significant' here – I'm testing whether *any* of 16 compounds is 'significant'
- Suppose that in reality ALL compounds are ineffective.
- For each compound, if we use p = .05, then probability of null result is .95.
- So probability that NO test gives a significant result is .95^16 = .44
- It follows that probability that AT LEAST ONE test is significant = 1-.44 = .56

To understand p-values:

SIMULATED DATA

"Just as lab scientists are not allowed to handle dangerous substances without safety training, researchers should not be allowed anywhere near a P value or similar measure of statistical probability until they have demonstrated understanding of what it means."

Bishop DVM. Nature 583,
World view: How scientists can stop fooling themselves. July 23 2020

Registered reports

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

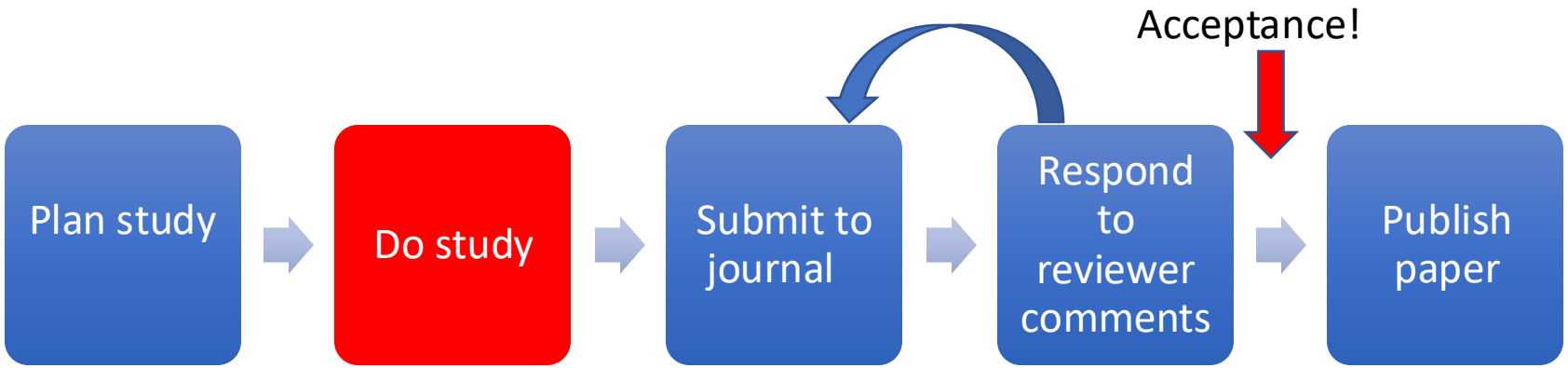
Chris Chambers

Tuesday 20 May 2014 07.15 BST

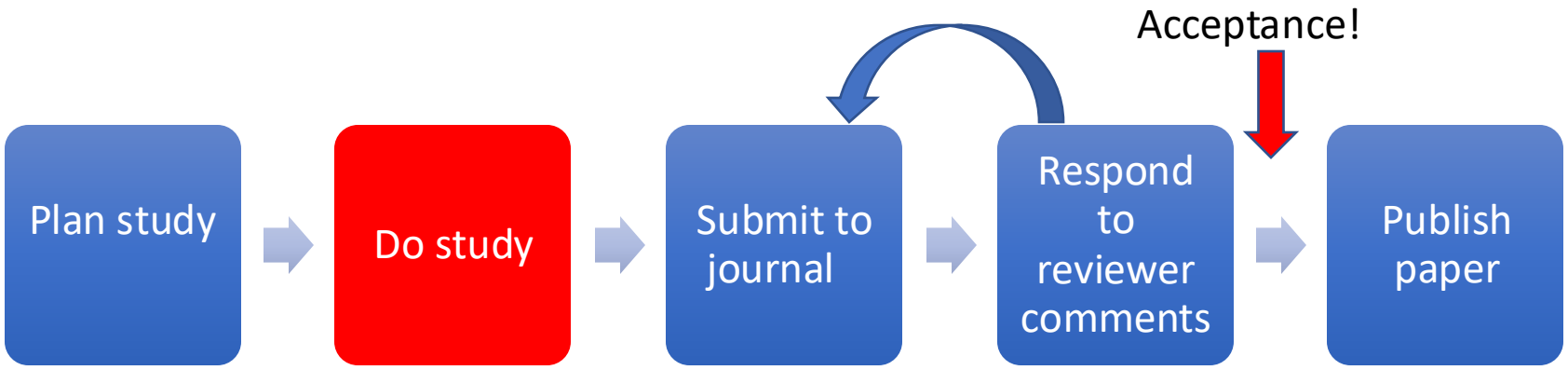
Shares 464 Comments 9

Save for later

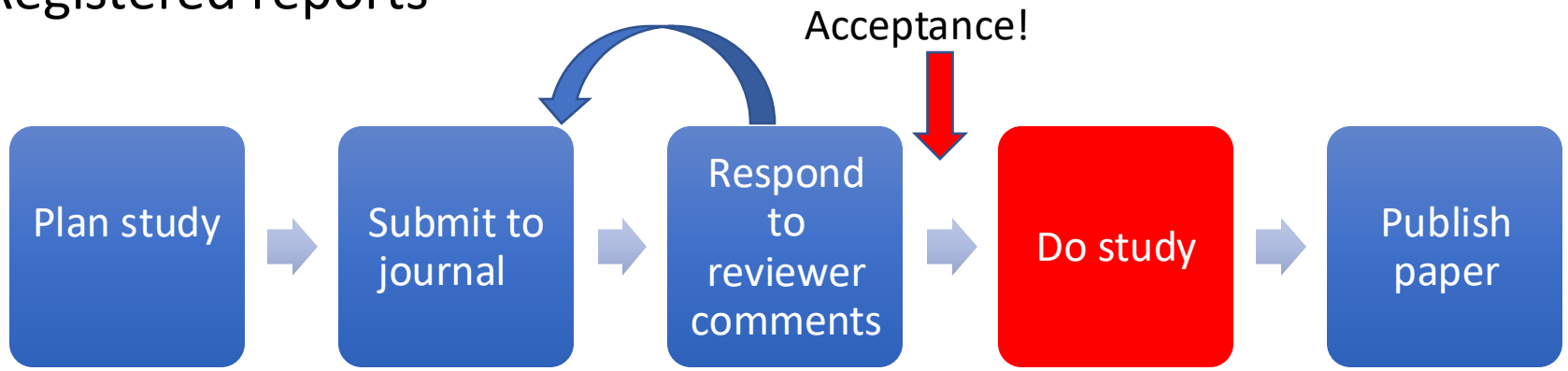
Classic publishing



Classic publishing



Registered reports



Publication decision based on whether an interesting question is addressed with strong methodology, rather than on the results

Asymmetric moral judgements

A researcher conducted a study of an intervention for dyslexia. The intervention effect is null, so the researcher swaps the group status for a subset of participants to give a significant effect. Is this:

1. Totally ethically acceptable
2. A bit questionable, but not a serious breach of ethics
3. Fairly unethical, but would not merit more than disapproval
4. Clearly unethical; researcher should be censured and sent for retraining
5. Clearly unethical; researcher should be fired

A researcher conducted a study of an intervention for dyslexia. One of four outcomes gives a significant effect at .05 but does not survive correction for multiple comparisons. The researcher decides not to report the other three outcomes

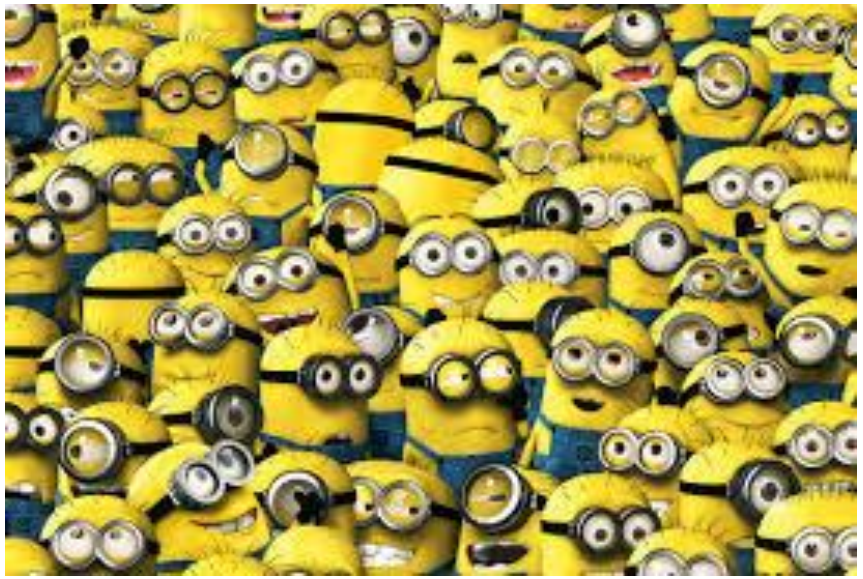
1. Totally ethically acceptable
2. A bit questionable, but not a serious breach of ethics
3. Fairly unethical, but would not merit more than disapproval
4. Clearly unethical; researcher should be censured and sent for retraining
5. Clearly unethical; researcher should be fired

Omitting to report results does not seem like fabrication/falsification to many people!

A researcher conducted a study of an intervention for dyslexia. The intervention effect falls short of significance, so the researcher decides not to publish it. Is this:

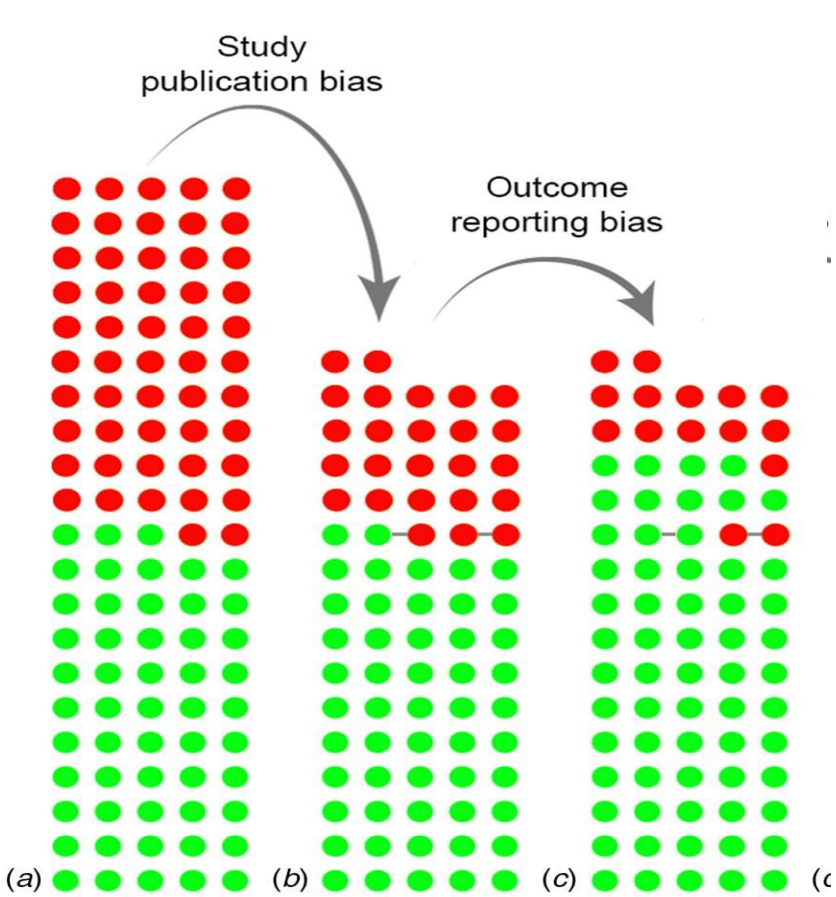
1. Totally ethically acceptable
2. A bit questionable, but not a serious breach of ethics
3. Fairly unethical, but would not merit more than disapproval
4. Clearly unethical; researcher should be censured and sent for retraining
5. Clearly unethical; researcher should be fired

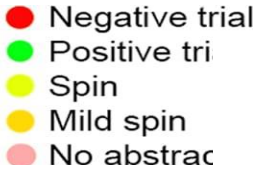
Most people regard publication bias as morally acceptable



So just how bad is publication bias?

How common is this in practice?
Even when there are trial registries, we get publication bias

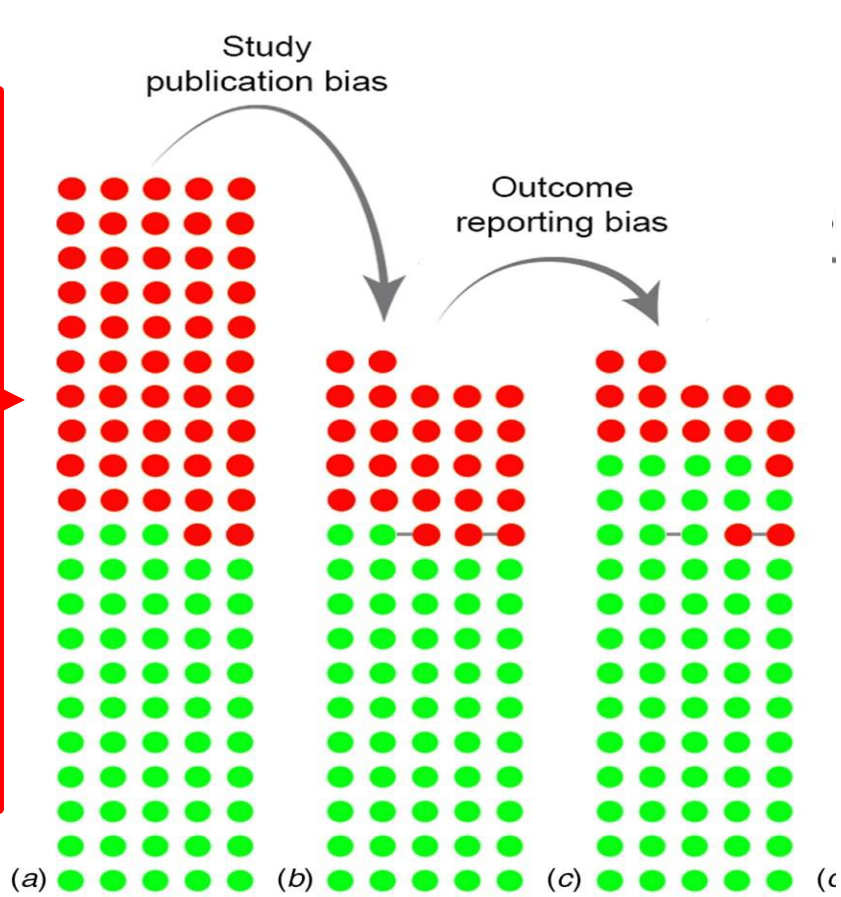


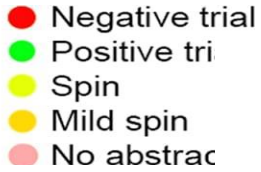


de Vries, Y. A., et al (2018). The cumulative effect of reporting and citation biases on the apparent efficacy of treatments: the case of depression. *Psychological Medicine 48*, 2453-2455. doi:10.1017/S0033291718001873

How common is this in practice?
Even when there are trial registries, we get publication bias

We only know these studies exist because they are in a registry!

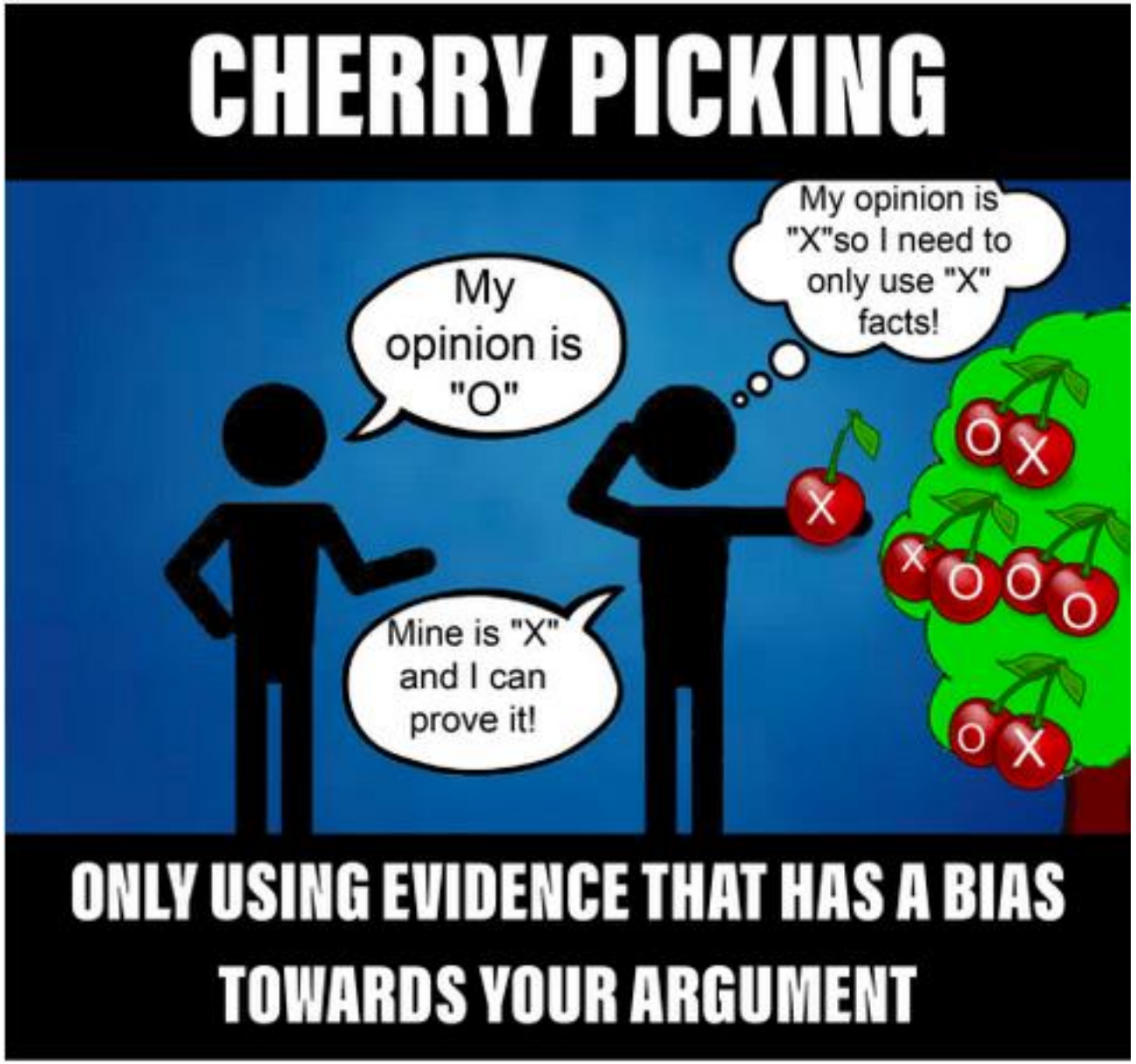




de Vries, Y. A., et al (2018). The cumulative effect of reporting and citation biases on the apparent efficacy of treatments: the case of depression. *Psychological Medicine 48*, 2453-2455. doi:10.1017/S0033291718001873

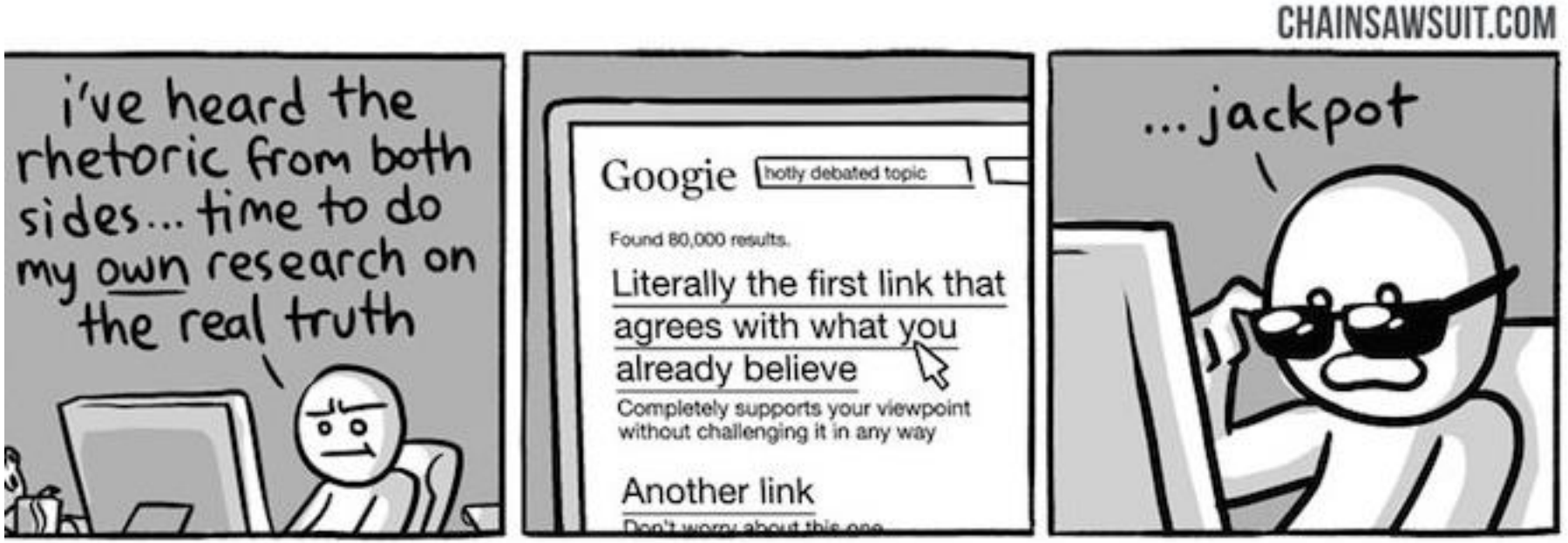
Quasi cumulative science

Picture of consistent support for the hypothesis, when reality is far more mixed, or bulk of research even goes against it


CHERRY PICKING
My opinion is "O"
My opinion is "X"so I need to only use "X" facts!
Mine is "X" and I can prove it!
ONLY USING EVIDENCE THAT HAS A BIAS TOWARDS YOUR ARGUMENT

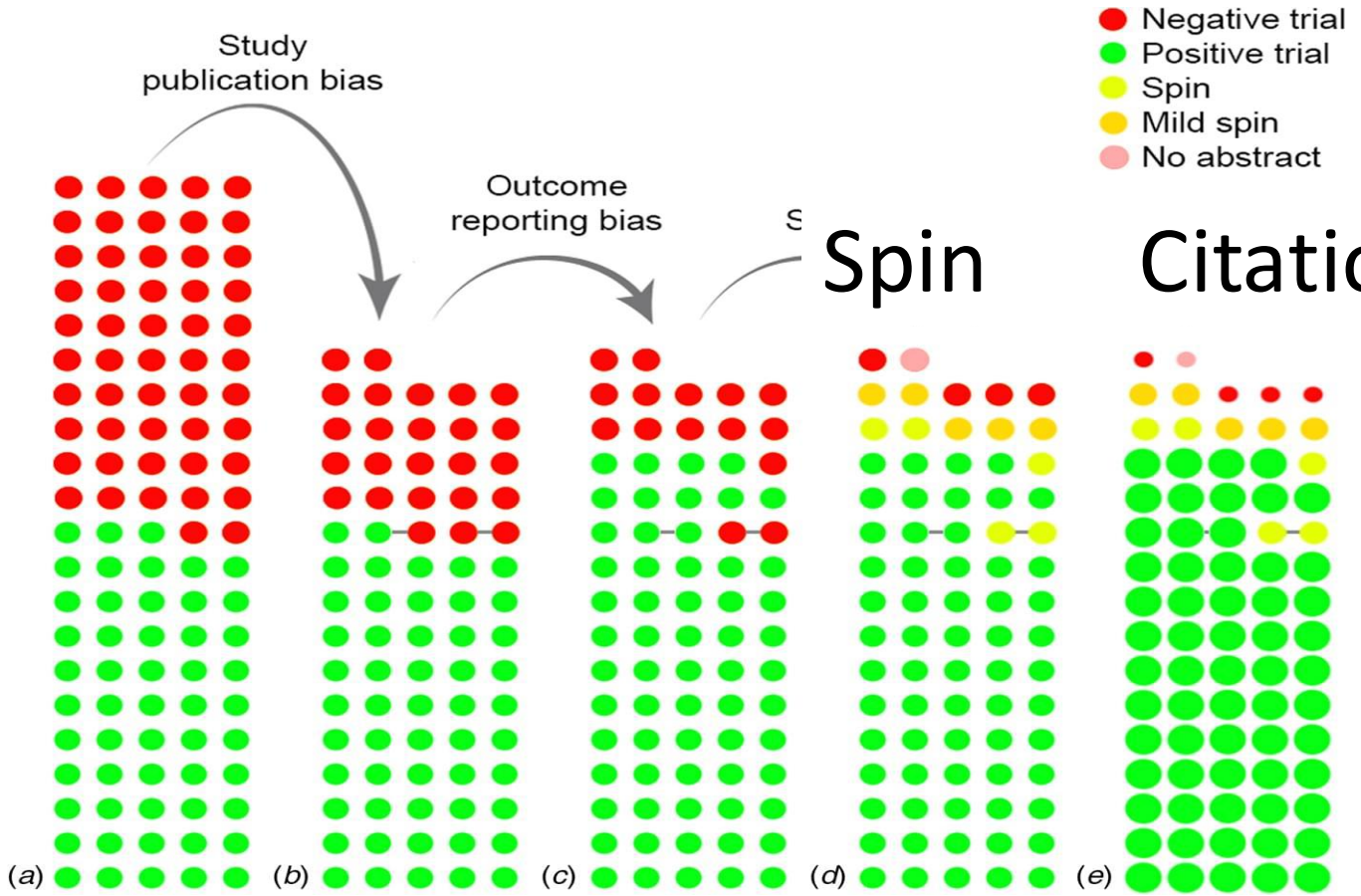
Confirmation bias

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



Confirmation bias leads to spin and citation bias

Publication bias and reporting bias conspire to mislead us



de Vries, Y. A., et al (2018). The cumulative effect of reporting and citation biases on the apparent efficacy of treatments: the case of depression. *Psychological Medicine 48*, 2453-2455. doi:10.1017/S0033291718001873

eLIFE elifesciences.org

Publication bias and the canonization of false facts

Silas Boye Nissen, Tali Magidson, Kevin Gross, Carl T Bergstrom

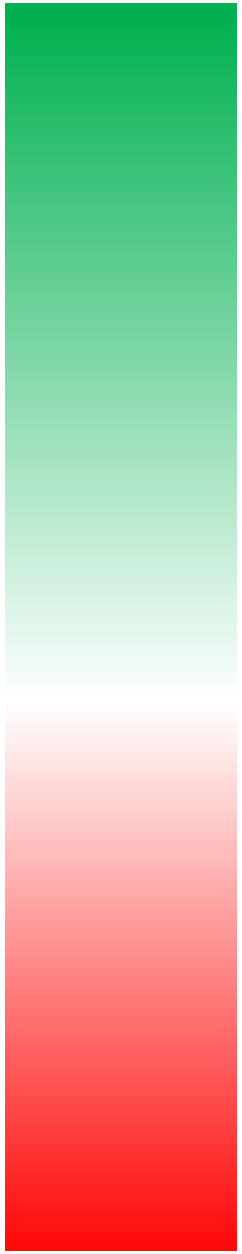
Canonization: when a claim is widely accepted as true, on the basis of multiple pieces of supporting evidence

- a canonized fact can be taken for granted rather than treated as an open hypothesis in the subsequent primary literature;

In Bayesian language, could say that canonization has occurred when our prior belief in the phenomenon is so strong that it would take overwhelming amounts of counter-evidence to shift to another view

CANONIZED

Overwhelming evidence in favour



- The earth is round
- Global temperatures are increasing
- Caffeine makes you wakeful

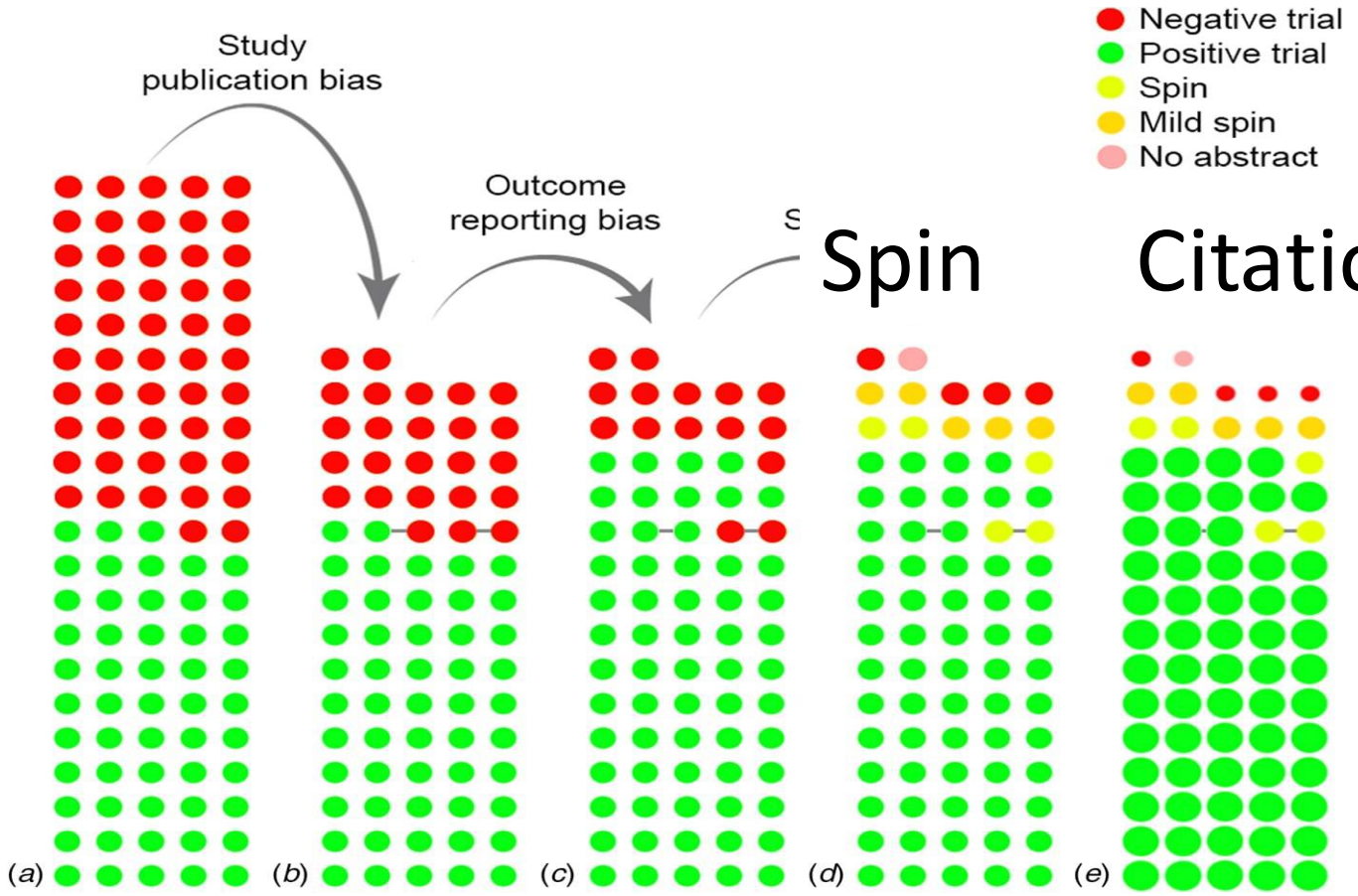
Equipoise

- Drug X is effective against COVID-19

Overwhelming evidence against

- Extrasensory perception is real
- Astrological forces determine your fate

Publication bias and reporting bias conspire to mislead us



de Vries, Y. A., et al (2018). The cumulative effect of reporting and citation biases on the apparent efficacy of treatments: the case of depression. *Psychological Medicine 48*, 2453-2455. doi:10.1017/S0033291718001873

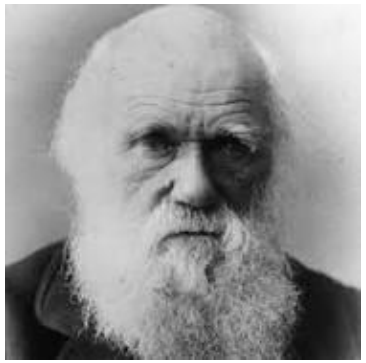
Solutions to citation bias

- Methods of systematic review as mandatory part of graduate training: realise it is important to look at all the evidence, rather than cherry-picking
  - But still problems! Spin and fraudulent research

- Increase awareness of serious consequences of biased reporting – not a 'victimless crime':
  - Affects users of research and other scientists

- Engender a mindset of 'organized skepticism' – with examples of famous scientists who promote this

I had, also, during many years, followed a golden rule, namely, that whenever a published fact, a new observation or thought came across me, which was opposed to my general results, to make a memorandum of it without fail and at once: for I had found by experience that such facts and thoughts were far more apt to escape from the memory than favourable ones. p 45



The autobiography of Charles Darwin and selected letters, edited by Francis Darwin. First published in 1892 in the United States by D. Appleton and Company

Looking ahead: What can we do?

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.

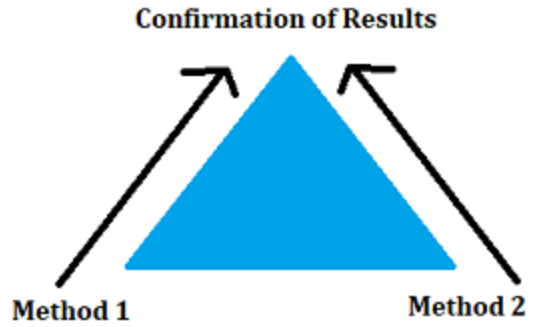
Unanticipated findings can be reported but are clearly labelled as ‘exploratory’

Also – **reviewer feedback come at a point where it can be useful**

Better study design
An approach from epidemiology: Triangulation

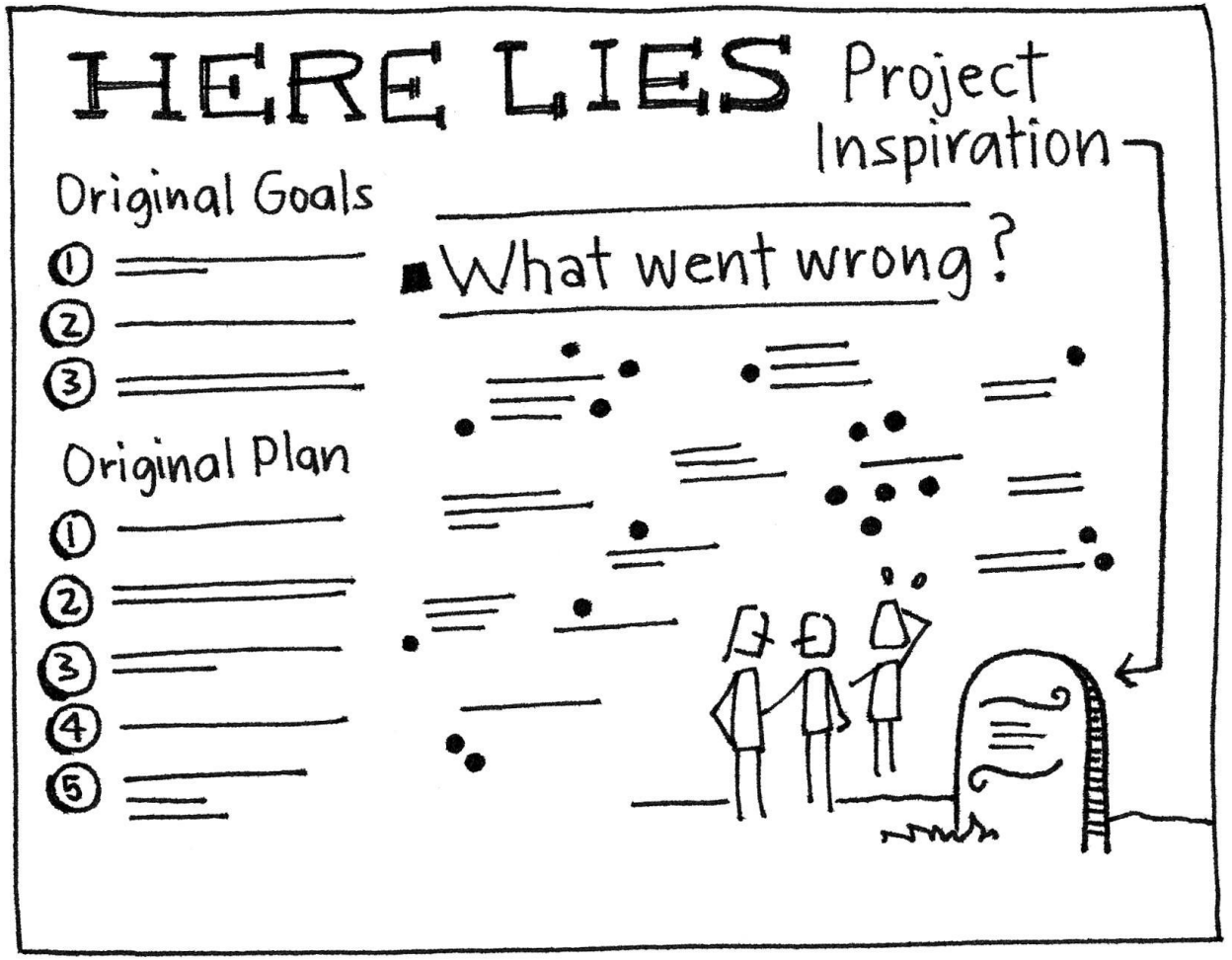
- Strategic use of multiple approaches to address one question.
- Each approach has its own unrelated assumptions, strengths and weaknesses.
- Results that agree across different methodologies are less likely to be artefacts.

Encourages researchers to specifically consider alternative causal mechanisms and control for these



Munafò, M. R., & Davey Smith, G. (2018). Robust research needs many lines of evidence. *Nature, 553*(7689), 399-401.

Better study design
An idea from business studies: the pre-mortem



Improving statistical intuitions

- Simulating data - for intuitive understanding of p–hacking and power
- Development of game-based formats for exploring analyses and data

https://shinyapps.org/apps/p-hacker/

Improving statistical intuitions

- Simulating data - for intuitive understanding of p–hacking and power
- Development of game-based formats for exploring analyses and data

Registered report

Can we shift belief in the 'Law of Small Numbers'?

D. V. M. Bishop ✉, Jackie Thompson and Adam J. Parker

Published: 02 March 2022
https://doi.org/10.1098/rsos.211028

https://shinyapps.org/apps/p-hacker/

Thank you for listening!

Bishop, D. (2019). The psychology of experimental psychologists: Overcoming cognitive constraints to improve research: The 47th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, *73*, 174702181988651. https://doi.org/10.1177/1747021819886519