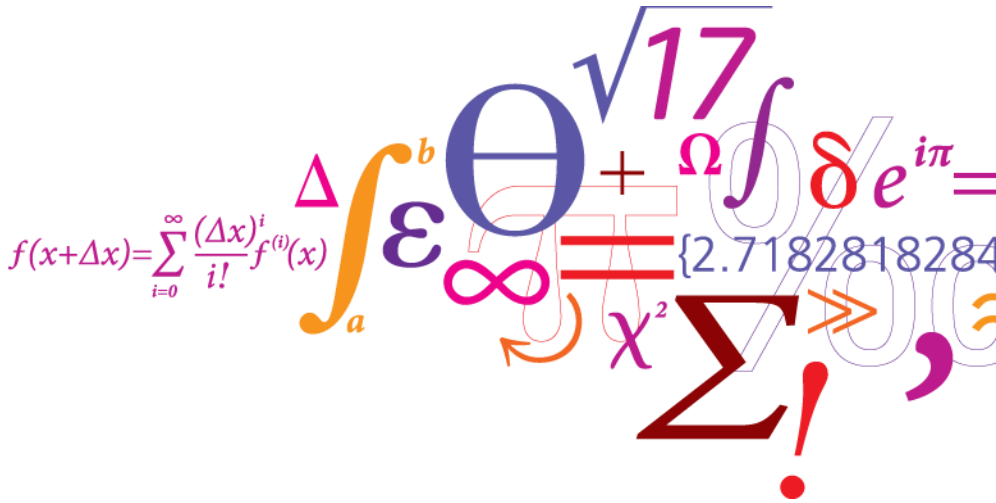
Reproducibility in science

Tommy Sonne Alstrøm
Associate Professor
Section of Cognitive Systems



DTU Compute
Department of Applied Mathematics and Computer Science

Agenda

13.00-13.50: Tommy: intro, reproducibility in science

14.00-14.20: Nick: basic git
14.20-15.00: Exercise 1, git clone,add,status,commit

15.00-15.45: Nick: remotes, pull, push, branches
15.45-16.30: Exercise 2: push/pull, typical git problem solving

16.30-16.40: Nick: .gitignore, user interfaces
16.40-17.00: Exercise 3: .gitignore

TA: Maxim Khomiakov on zoom (ping him in chat), Anders Stevnhoved Olsen, on campus

Get material here: https://gitlab.gbar.dtu.dk/02466F22/reprod_research/tree/pdf

Topics that will be covered

Topic 1 – Reproducibility issues – examples of correlations and causation

- Examples of correlation
- Publication on correlation, causation and p-values

Topic 2 – Which obstacles hinders research to be reproducible and which solutions are offered by literature?

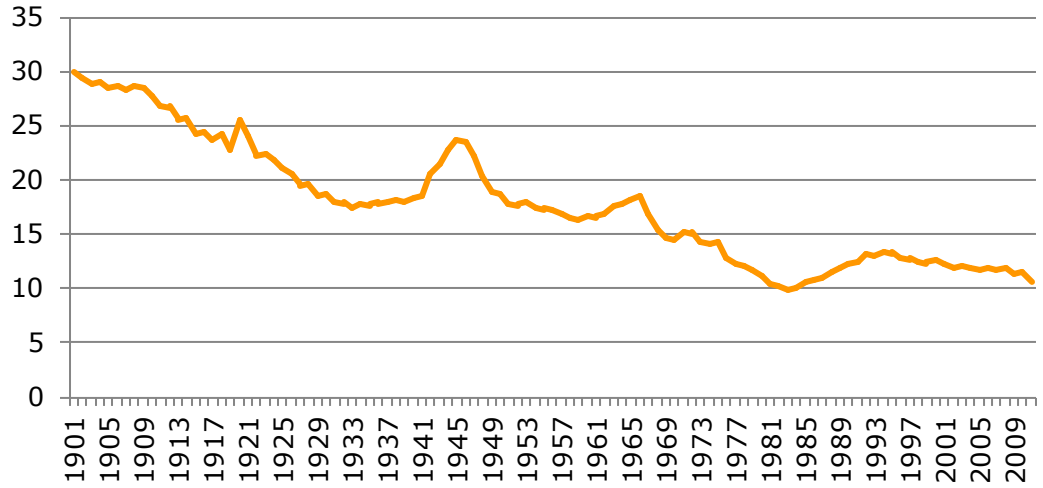
- Nature publishing take on reproducible research
- Survey from ~1500 researchers

Topic 3 – What tools can boost reproducibility?

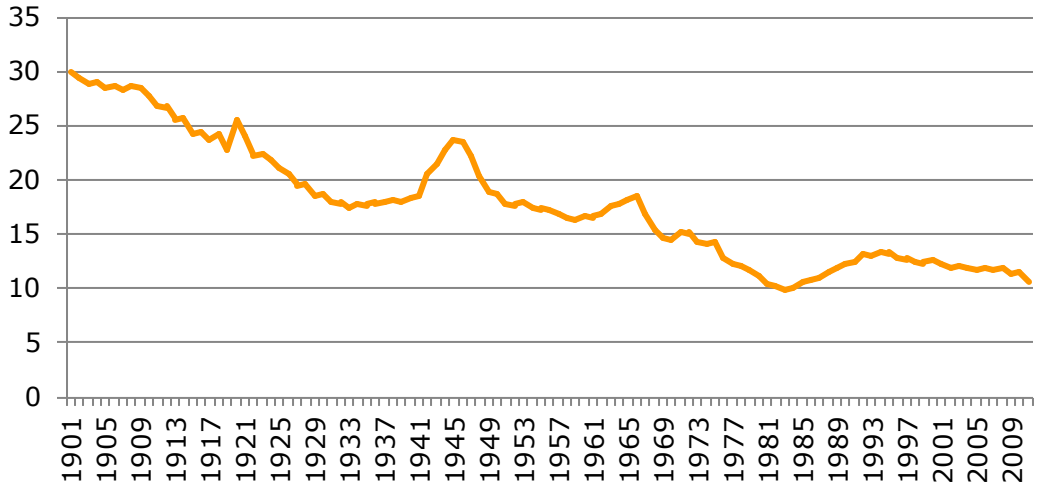
- Proper statistical thinking
- Proper research data management
- Recommendations for further reading

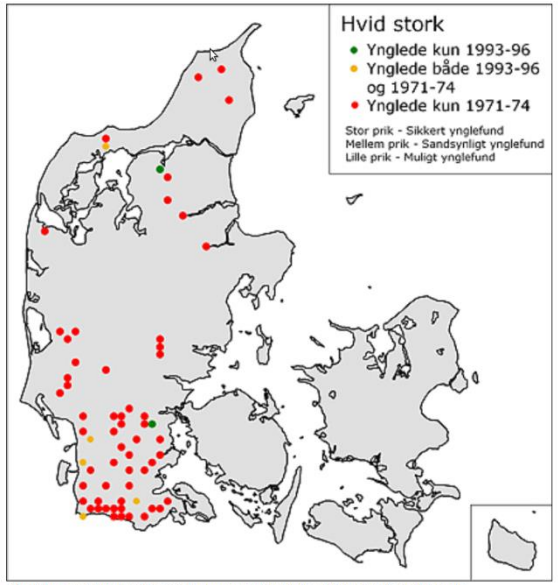
Topic 1 – Reproducibility issues – examples of correlations and causation

Children born in total /population 1.jan (in 1000)

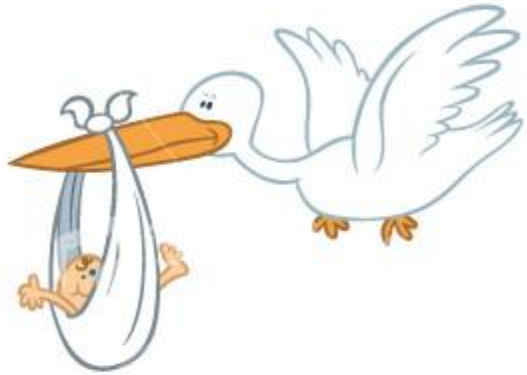


Children born in total /population 1.jan (in 1000)

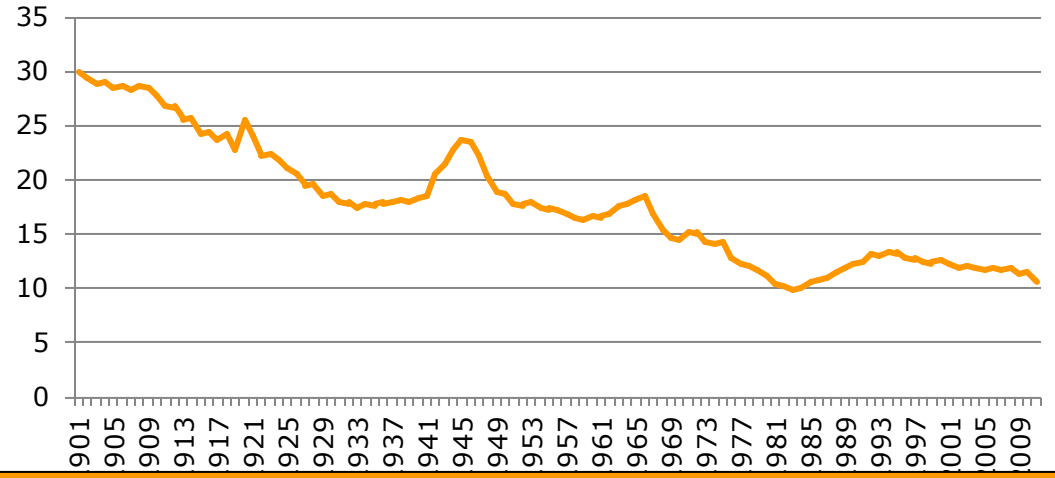




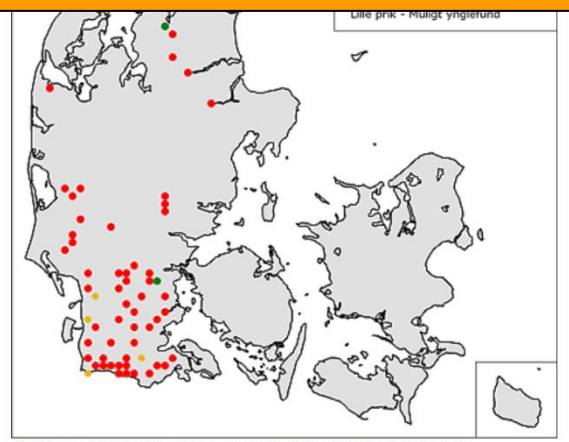
Ændringer i yngleudbredelsen opgjort ud fra DOFs landsdækkende kortlægninger i 1971-74 og 1993-96.



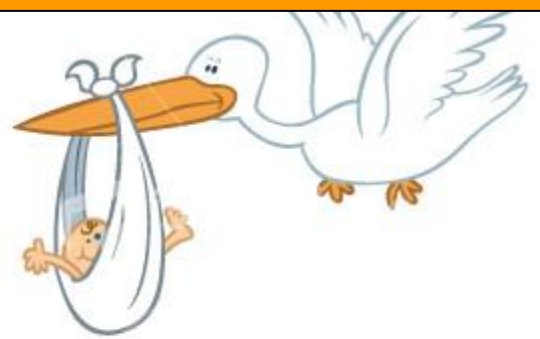
Children born in total /population 1.jan (in 1000)



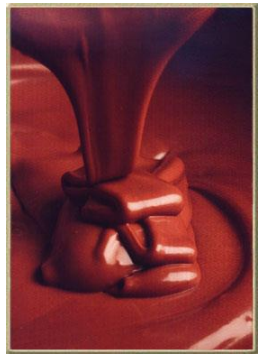
Causality is **not** the same as correlation / connection / association



Ændringer i yngleudbredelsen opgjort ud fra DOFs landsdækkende kortlægninger i 1971-74 og 1993-96.



Professor Franz H. Messerli: Chocolate Consumption, Cognitive Function and Nobel Laureates



- Science documents that chocolate intake improves cognitive function
- Number of Nobel prize winners is a surrogate to reflect a country's overall cognitive function.
- Extract data from Wikipedia and various other resources, e.g. chocolate consumption data is based on the years 20xx.

Chocolate Consumption, Cognitive Function, and Nobel Laureates. Messerli, Franz H. New England Journal of Medicine 367, 1562-1564, 2012

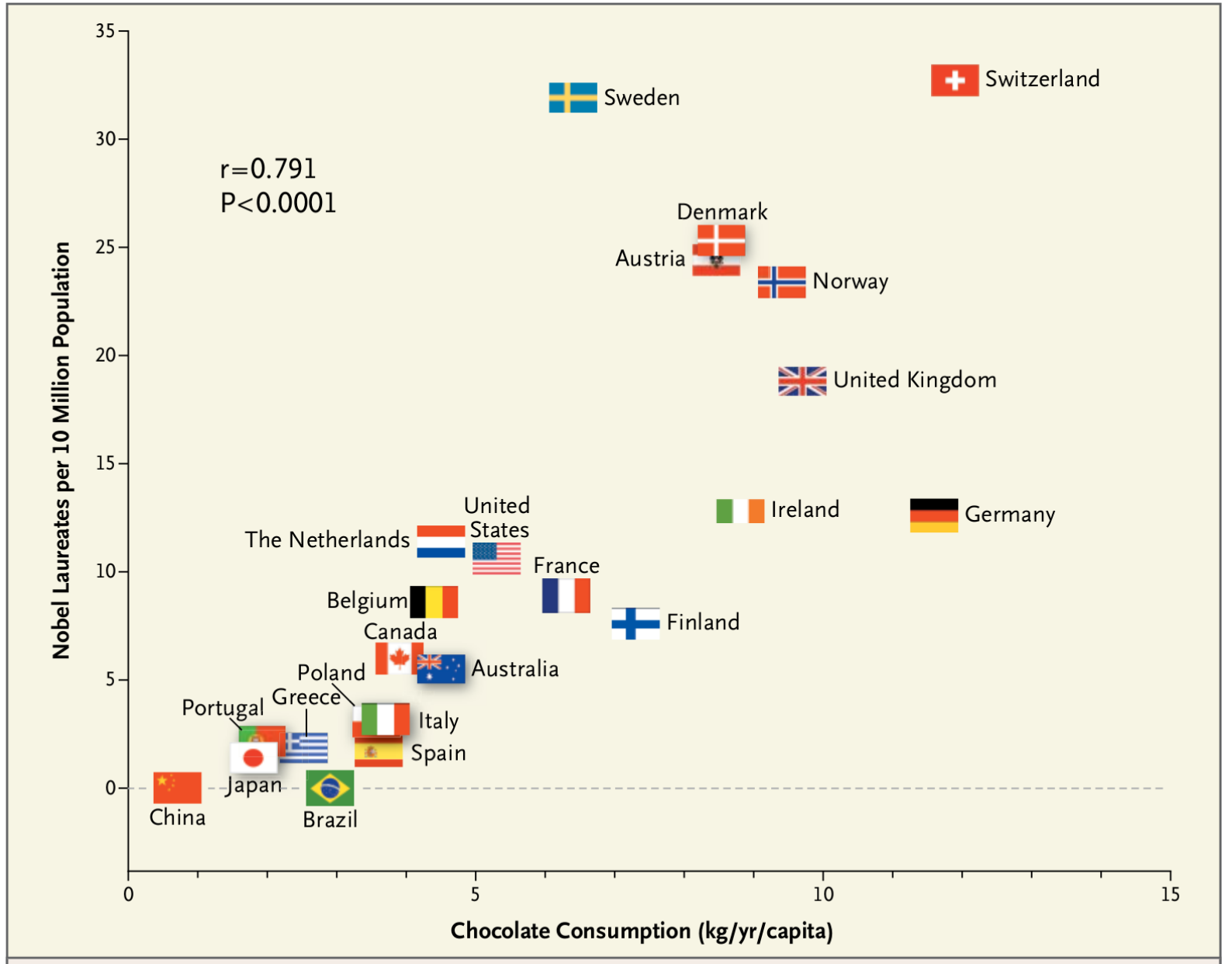


Figure 1. **Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

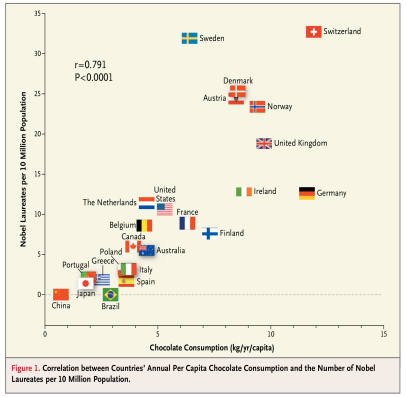
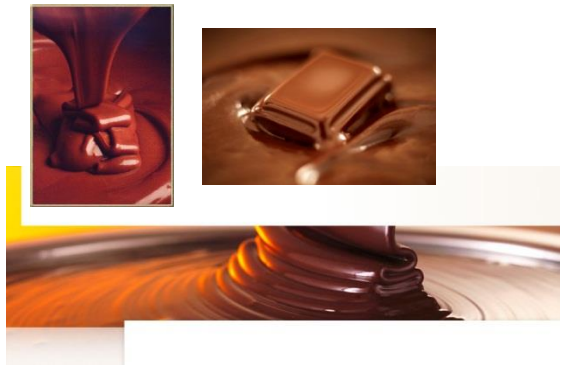


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Discussion in the paper

- Correlation between X and Y not the same as causation, however it's likely that chocolate intake provides fertile ground to spout Nobel prizes
- Reverse Causation: It's likely that a persons with superior cognitive function eat more chocolate.
- Data may by contaminated; Receiving the Nobel prize will increase the countrywide chocolate intake due to celebratory events!
- Differences in socioeconomic status, geographic factors and climate factors may play an effect, but fail short of fully explaining the close correlation observed
- 0.4 kg chocolate = 1 Nobel prize winner
- Conclusion: It remains to be determined whether chocolate intake is the underlying mechanism



Comments in media

Eric Cornell, American physicist, shared the Nobel Prize in 2001 joked:

"Personally I feel that milk chocolate makes you stupid.

"Now dark chocolate is the way to go. It's one thing if you want like a medicine or chemistry Nobel Prize, O.K.,

"but if you want a physics Nobel Prize it pretty much has got to be dark chocolate.

"Messerli said **the whole idea is absurd**, although the data are legitimate **and contain a few lessons about the fallibility of science.**"

https://www.reuters.com/article/us-eat-chocolate-win-the-nobel-prize/eat-chocolate-win-the-nobel-prize-idUSBRE8991MS20121010

Reproducibility – what does Wikipedia say

- Reproducibility is the ability of **an entire experiment** or study to be duplicated, either by the same researcher or **by someone else working independently**.

- Reproducible data - **repeatability** which is the degree of agreement of tests or measurements on replicate specimens by the same observer in the same laboratory.

- Reproducible research - the idea that the ultimate product of **academic research** is the paper along with the **full computational environment** used to produce the results in the paper such as the code, data, etc. that can be used to reproduce the results and create new work based on the research.

https://en.wikipedia.org/wiki/Reproducibility

Nature Publishing Group has declared war on irreproducible research

- There is growing alarm about results that cannot be reproduced. Explanations include increased levels of scrutiny, complexity of experiments and statistics, and pressures on researchers. Journals, scientists, institutions and funders all have a part in tackling reproducibility.
- *Nature* has taken substantive steps to improve the transparency and robustness in what we publish, and to promote awareness within the scientific community.

http://go.nature.com/huhbyr

Unreliable research - Trouble at the lab

The biotech company Amgen had a team of about 100 scientists trying to reproduce the findings (preclinical) of 53 "landmark" articles in cancer research published by reputable labs in top journals.
Only 6 of the 53 studies were reproduced (about 10%).

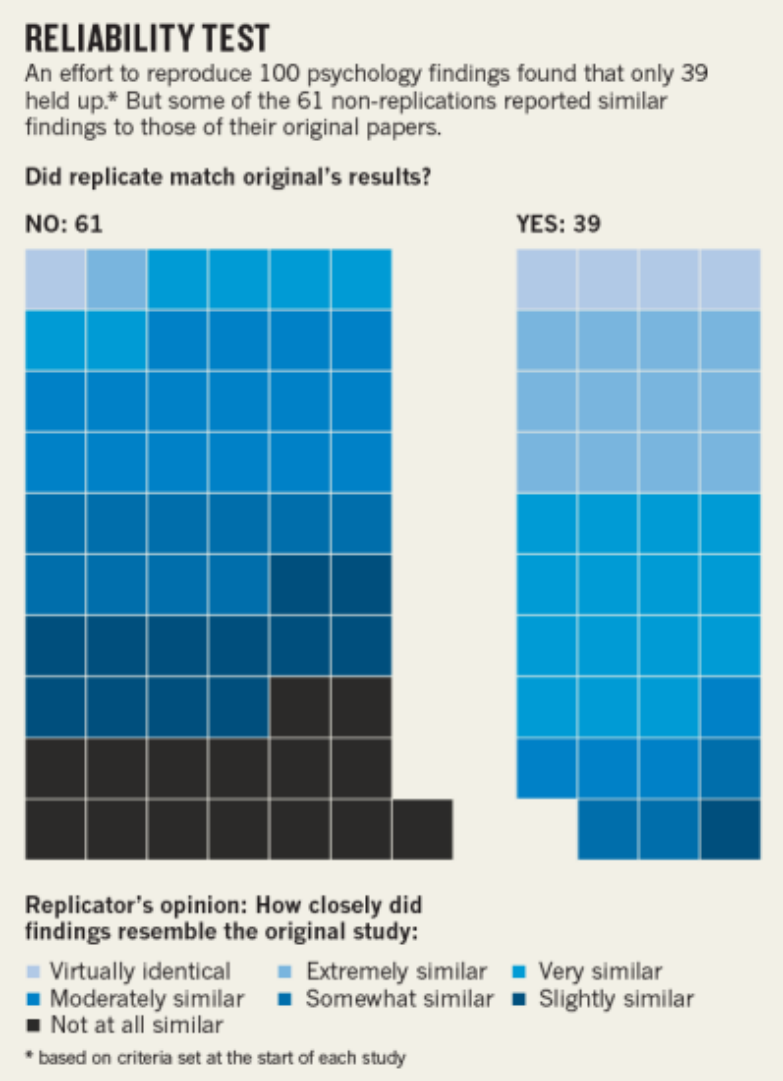
REPRODUCIBILITY OF RESEARCH FINDINGS
Preclinical research generates many secondary publications, even when results cannot be reproduced.

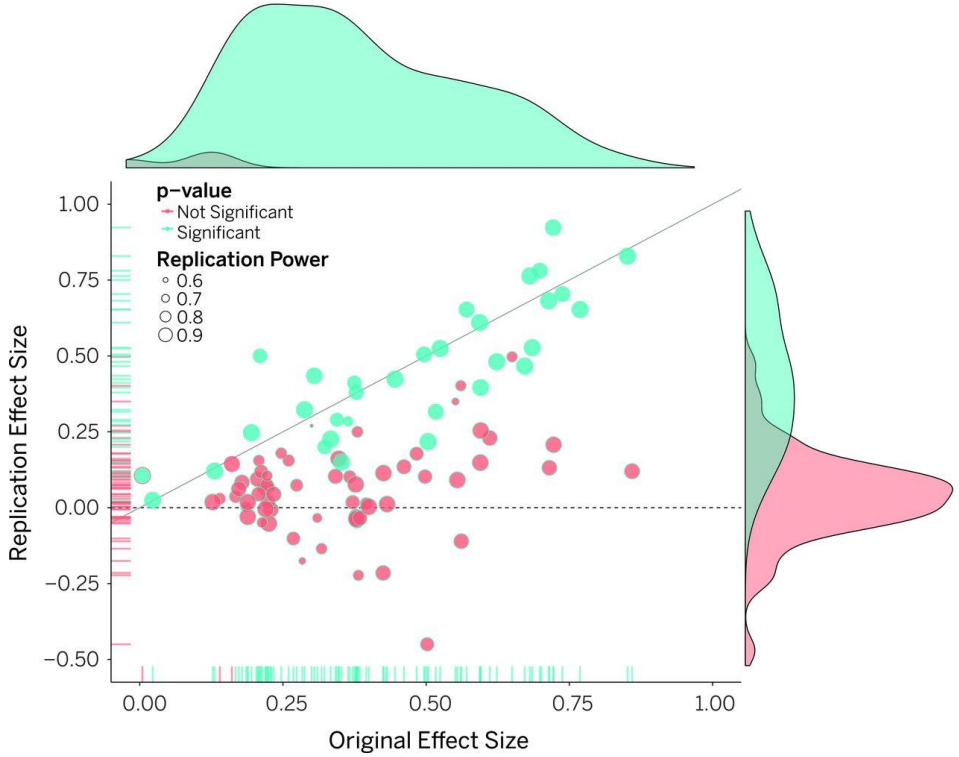
Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles*	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development programme.
*Source of citations: Google Scholar, May 2011.

Drug development: Raise standards for preclinical cancer research. C. Glenn Begley and Lee M. Ellis. Nature 483, 531–533, 2012.

Estimating the Reproducibility of Psychological Science

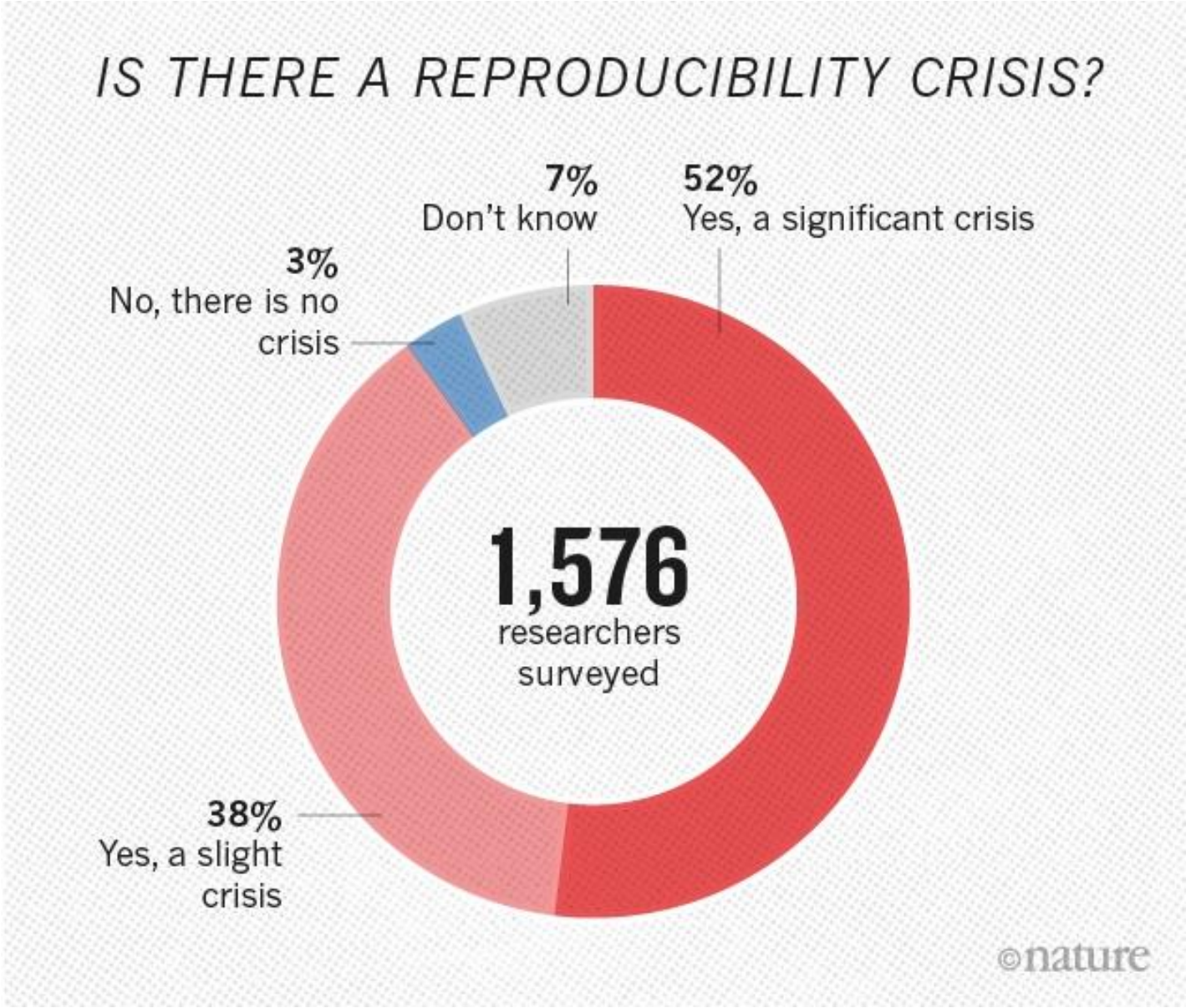




Open Science Collaboration Science 2015;349:aac4716

Science
AAAS

Topic 2 – Which obstacles hinders research to be reproducible and which solutions are offered by literature?



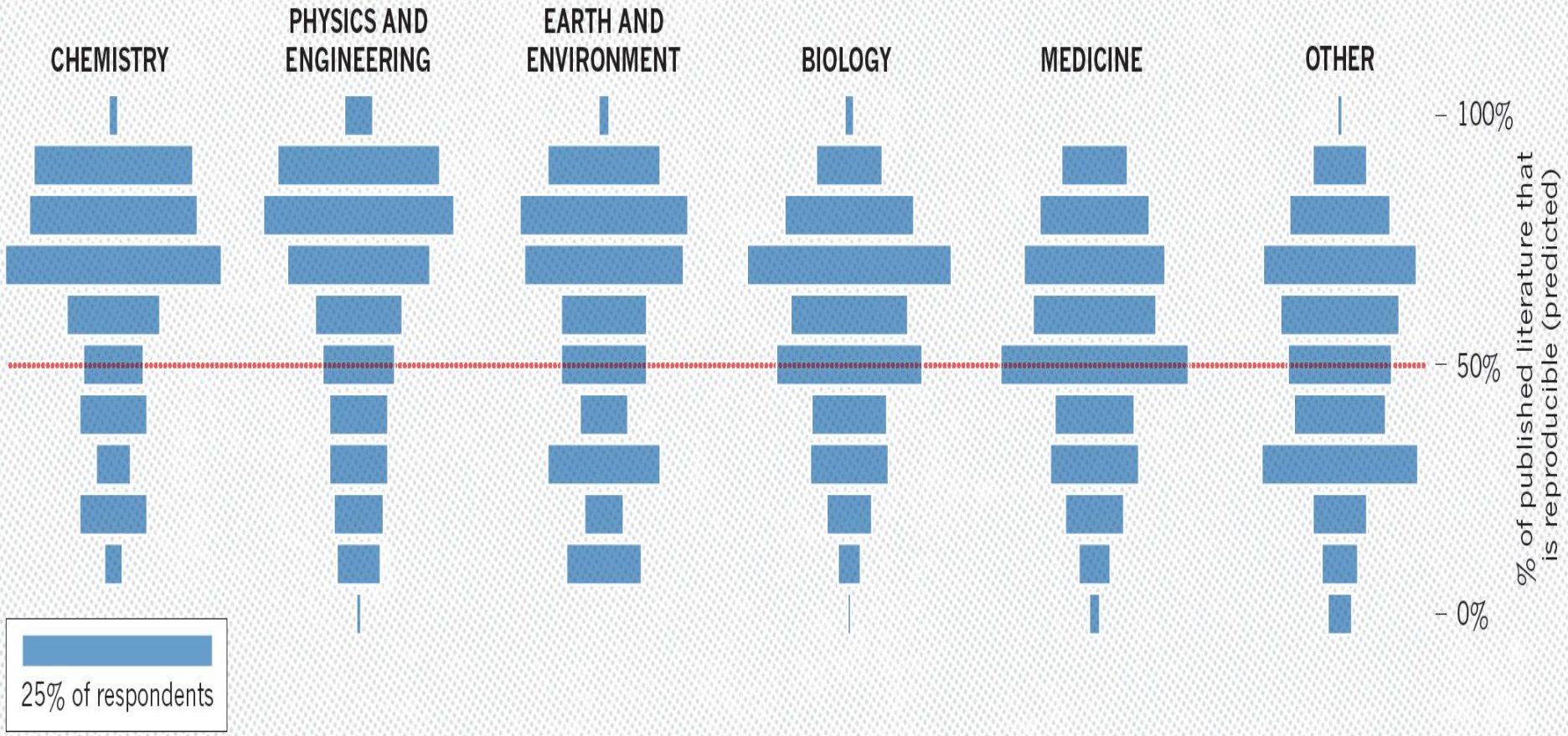
1,500 scientists lift the lid on reproducibility. Monya Baker. Nature 533, 452–454, May 2016.

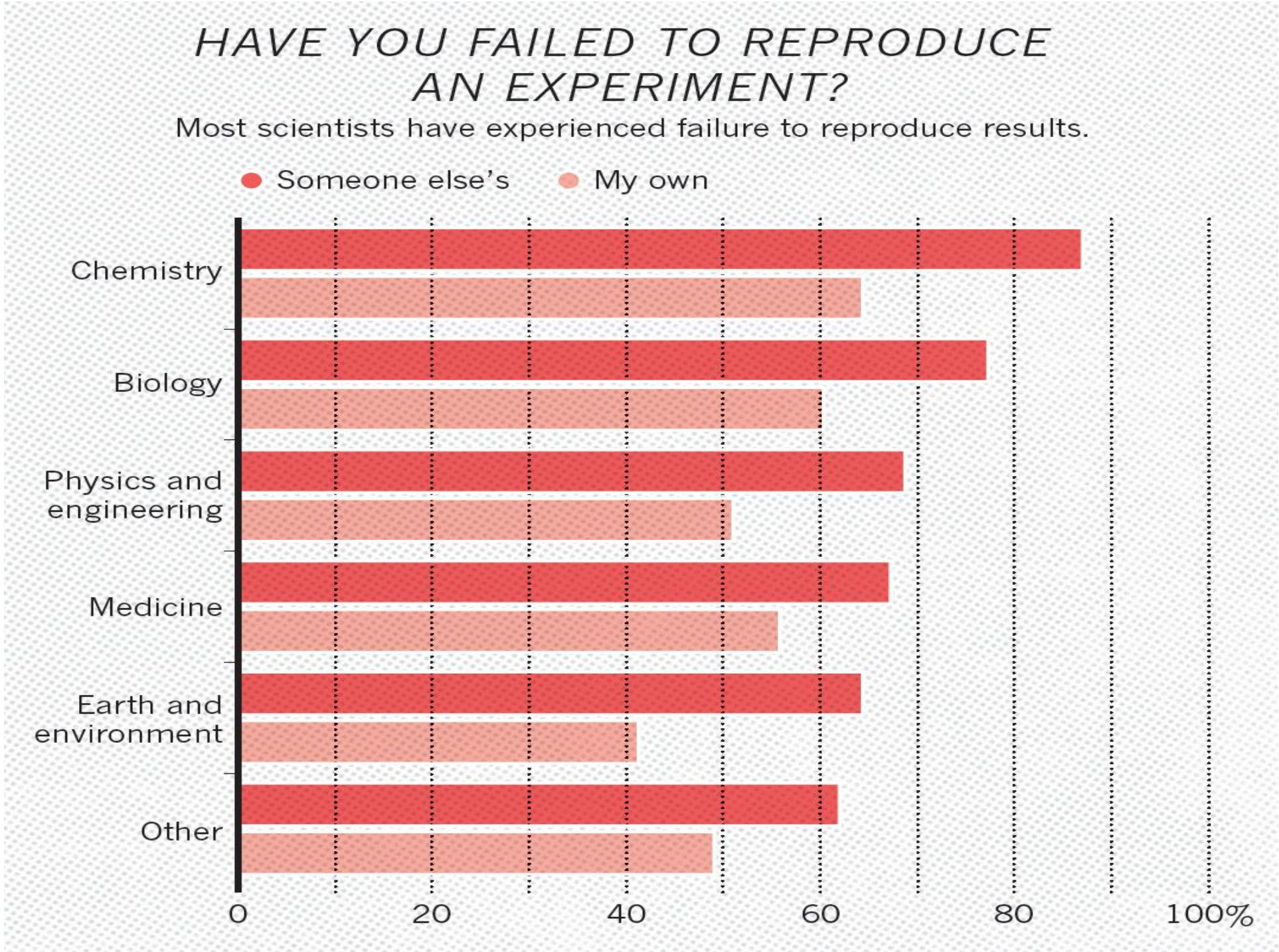
A 'CRISIS' IN NUMBERS

Nature surveyed 1,576 scientists online to get their thoughts on reproducibility in their field and in science in general. See **go.nature.com/2vjr4y** for more charts and access to the full data.

HOW MUCH PUBLISHED WORK IN YOUR FIELD IS REPRODUCIBLE?

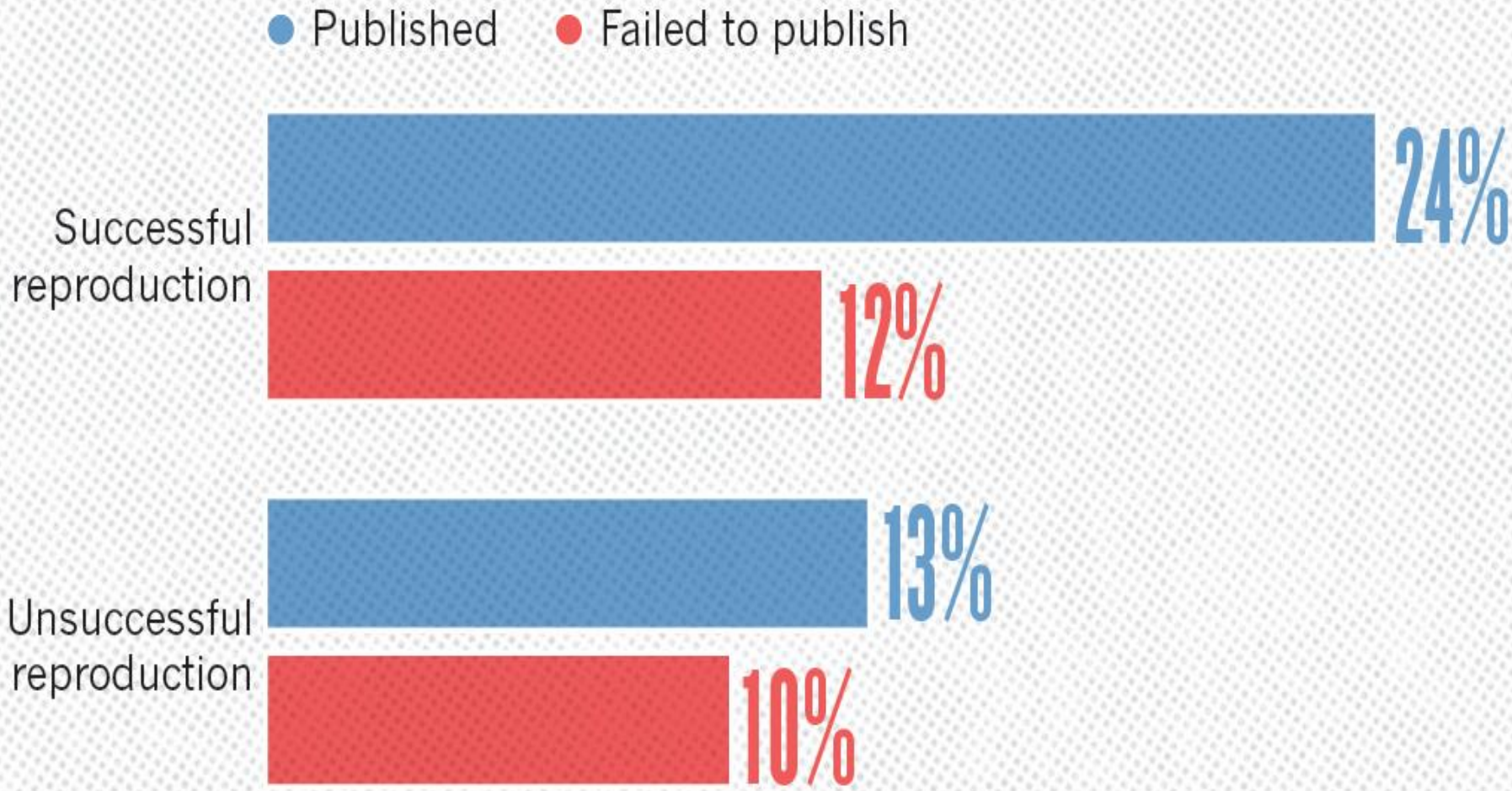
Physicists and chemists were most confident in the literature.




HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?
Most scientists have experienced failure to reproduce results.
Someone else's
My own
Chemistry
Biology
Physics and engineering
Medicine
Earth and environment
Other
0
20
40
60
80
100%

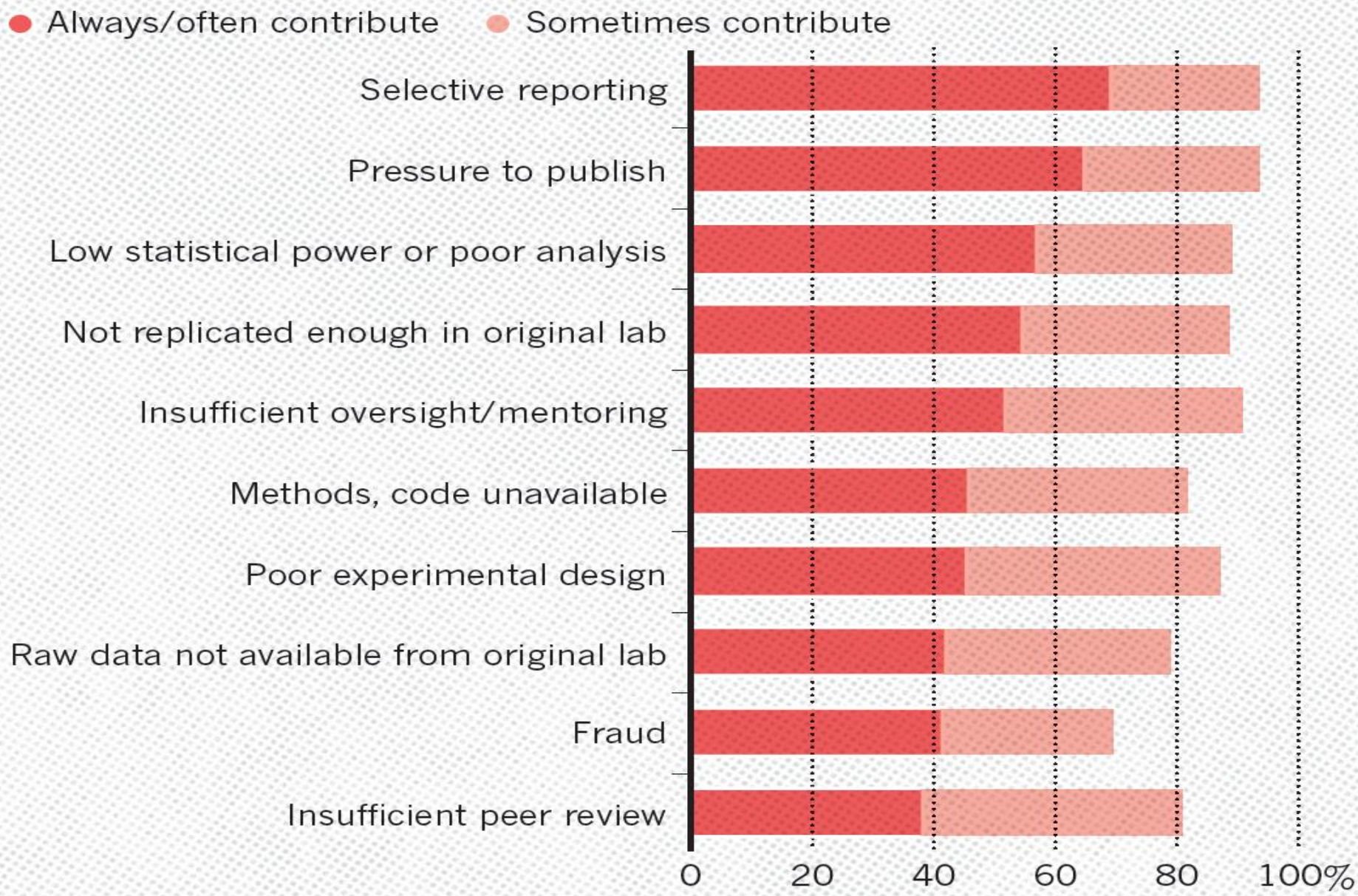
HAVE YOU EVER TRIED TO PUBLISH A REPRODUCTION ATTEMPT?

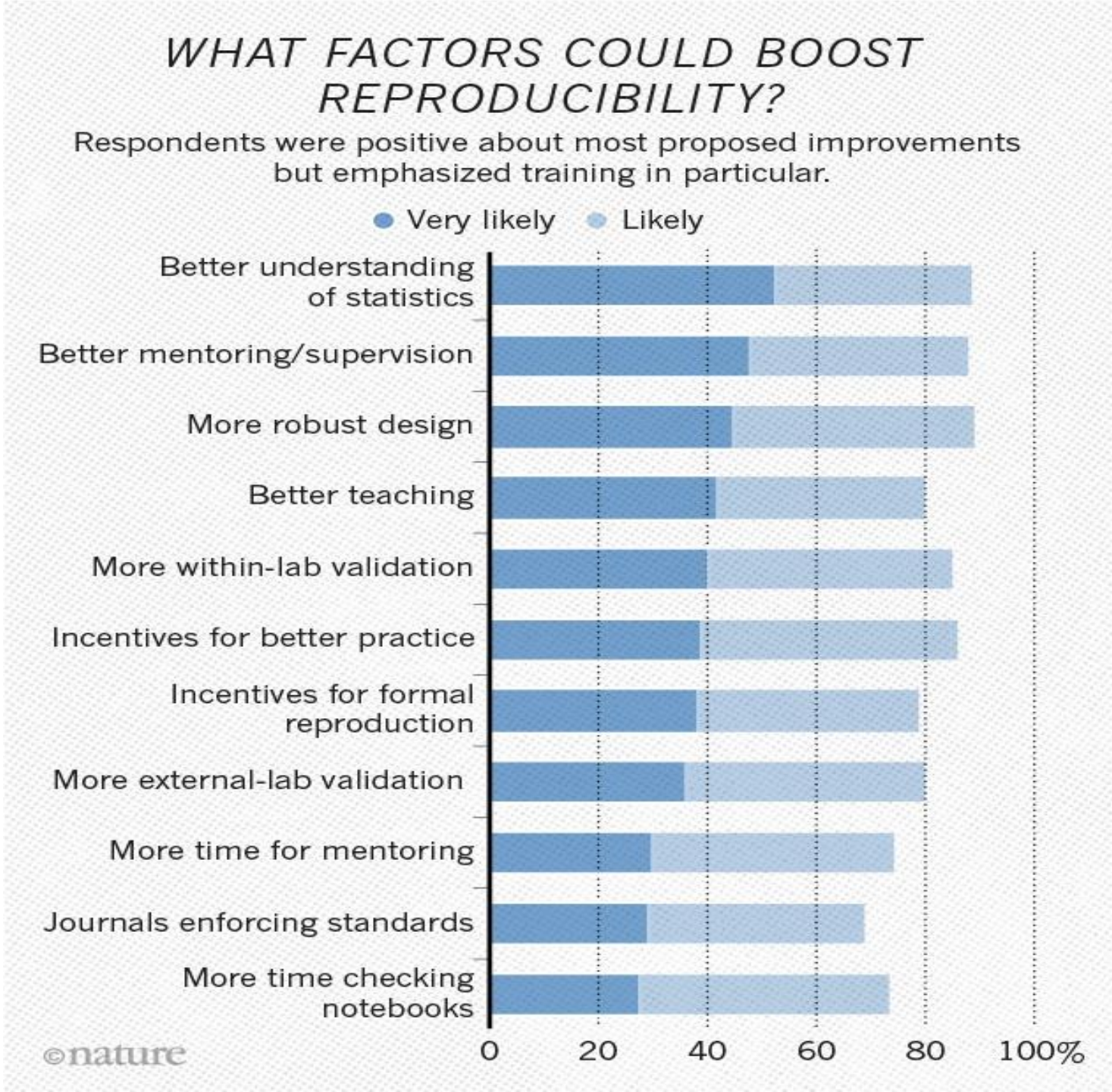
Although only a small proportion of respondents tried to publish replication attempts, many had their papers accepted.

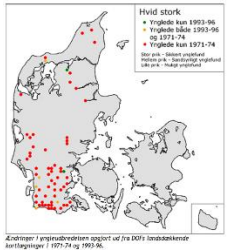


WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

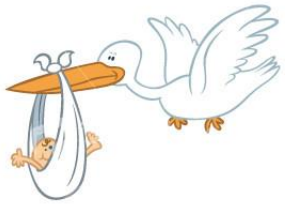
Many top-rated factors relate to intense competition and time pressure.




WHAT FACTORS COULD BOOST REPRODUCIBILITY?
Respondents were positive about most proposed improvements but emphasized training in particular.
Very likely
Likely
Better understanding of statistics
Better mentoring/supervision
More robust design
Better teaching
More within-lab validation
Incentives for better practice
Incentives for formal reproduction
More external-lab validation
More time for mentoring
Journals enforcing standards
More time checking notebooks
0
20
40
60
80
100%
©nature



Study design



Results from observational studies start to become believable when:

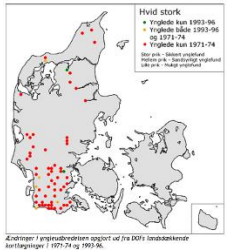
(i) the results are replicated in many studies,

(ii) each of the studies controlled for plausible confounding variables,

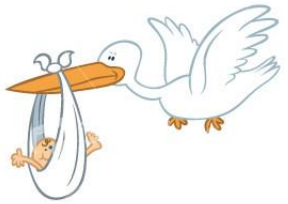
(iii) there is a plausible scientific explanation for the existence of a causal relationship.

"A good example is smoking and cancer. Numerous studies have shown a relationship between smoking and cancer even after adjusting for many **confounding variables**. Moreover, in **laboratory studies**, smoking has been shown to **damage lung cells**. Finally, a **causal link between** smoking and cancer has been found **in randomized animal studies**. It is this collection of evidence over many years that makes this a convincing case. One single observational study is not, by itself, strong evidence. Remember that when you read the newspaper."

Wasserman, All of Statistics, sec 16.3



Study design



Results from observational studies start to become believable when:
(i) the results are replicated in many studies,
(ii) each of the studies controlled for plausible confounding variables,
(iii) there is a plausible scientific explanation for the existence of a causal relationship.

"A good example is smoking and cancer. Numerous studies have shown a relationship between smoking and cancer even after adjusting for many **confounding variables**. Moreover, in **laboratory studies**, smoking has been shown to **damage lung cells**. Finally, a **causal link between** smoking and cancer has been found **in randomized animal studies**. It is this collection of evidence over many years that makes this a convincing case. One single observational study is not, by itself, strong evidence. Remember that when you read the newspaper."

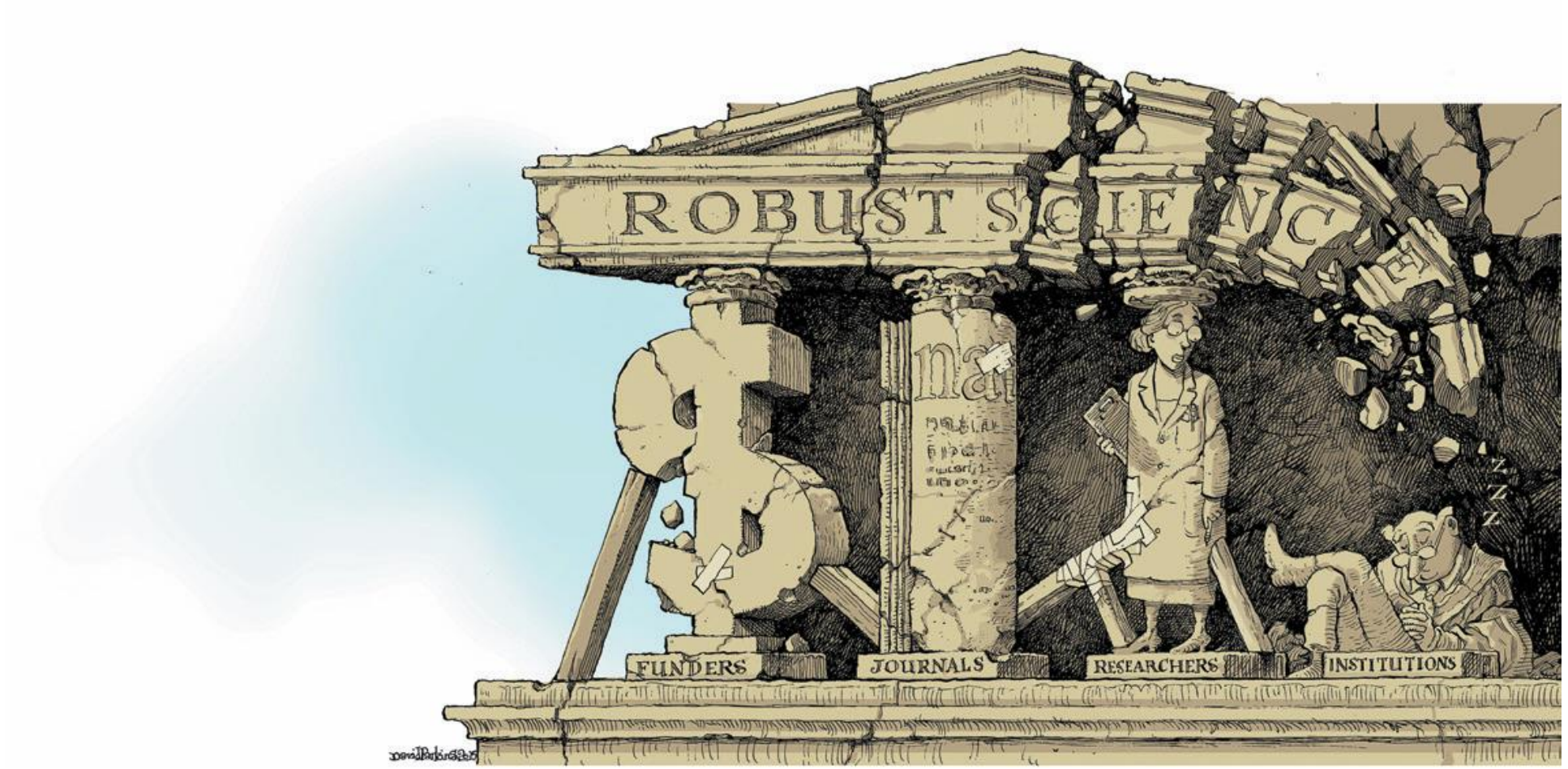
Theorem

In general, correlation is **not** causation, however, when the data is randomly collected, correlation estimates causation.

Wasserman, All of Statistics, sec 16.4

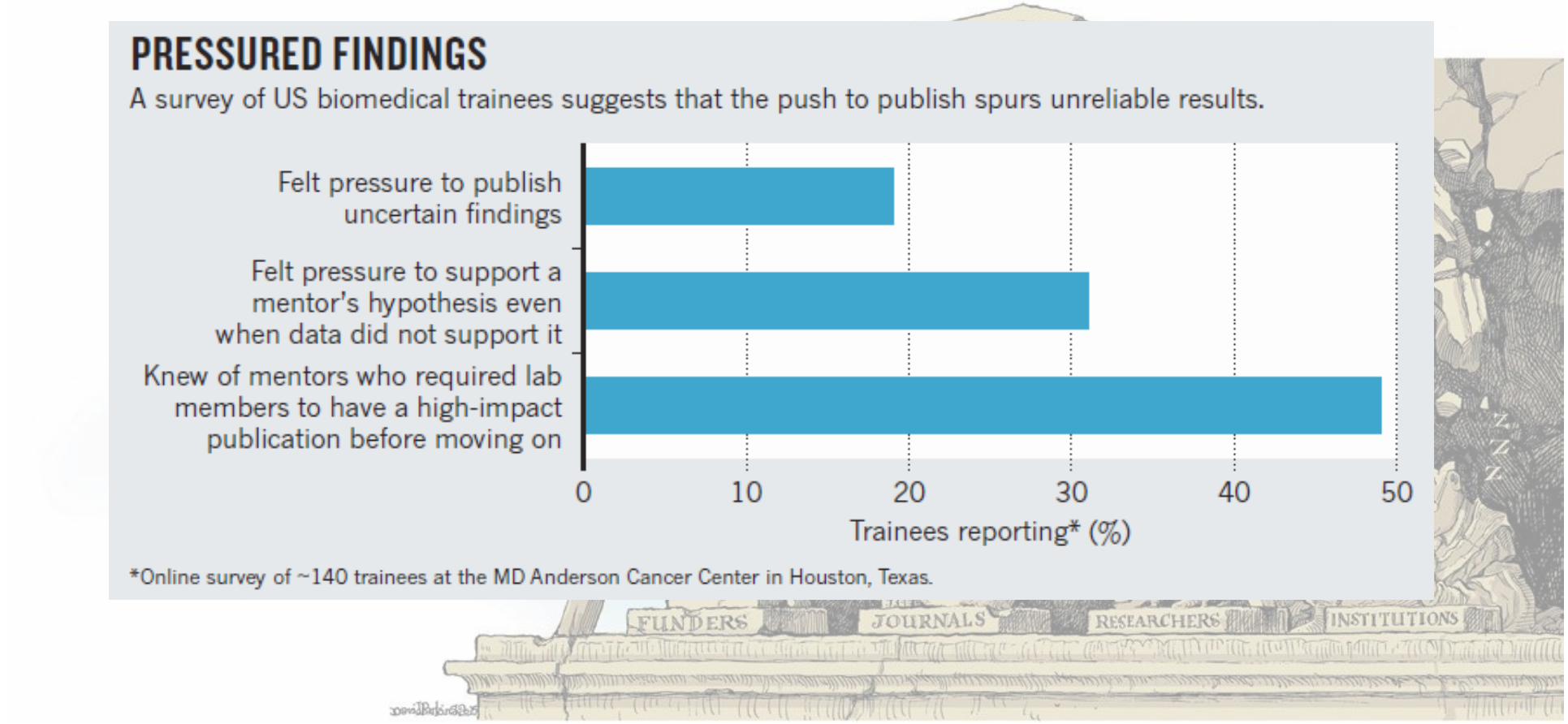
Wasserman, All of Statistics, sec 16.3

Researchers has a stake in the outcome



C. Glenn Begley, Alastair M. Buchan and Ulrich Dirnagl. "Institutions must do their part for reproducibility" *Nature* **525**, 25-27 (03 September 2015).

Researchers has a stake in the outcome



C. Glenn Begley, Alastair M. Buchan and Ulrich Dirnagl. "Institutions must do their part for reproducibility" *Nature* **525**, 25-27 (03 September 2015).

Replicates and repeats—what is the difference and is it significant?

Fundamental principle 1

Science is knowledge obtained by repeated experiment or observation: if $n = 1$, it is not science, as it has not been shown to be reproducible.

Fundamental principle 2

Experimental design, at its simplest, is the art of varying one factor at a time while controlling others…

We always need to consider plausible alternative interpretations of an observed result.

Vaux, David L., Fiona Fidler, and Geoff Cumming. "Replicates and repeats—what is the difference and is it significant?" *EMBO reports* 13.4 (2012): 291-296.

Many hands make tight work

Crowdsourcing research can balance discussions

- Published research suggesting that **noble-sounding German surnames, such as König (king) and Fürst (prince), could boost careers**.
- Psychologist, Uri Simonsohn, was skeptical and asked for the data set.
- Re-running the analysis yielded the same outcome.
- Simonsohn's different analytical approach showed **no connection**.
- In analyses run by a single team, researchers take on multiple roles: inventors. optimistic analyst, and as devil's advocates.

Raphael Silberzahn and Eric L. Uhlmann. "Many hands make tight work". *Nature* **526**, 189-191 (08 October 2015).

ONE DATA SET, MANY ANALYSTS

Twenty-nine research teams reached a wide variety of conclusions using different methods on the same data set to answer the same question (about football players' skin colour and red cards).

Dark-skinned players four times more likely than light-skinned players to be given a red card.

- Statistically significant effect
- Non-significant effect

78.7*
11.5*

Twice as likely

Equally likely

Point estimates and 95% confidence intervals. *Truncated upper bounds.

ONE DATA SET, MANY ANALYSTS

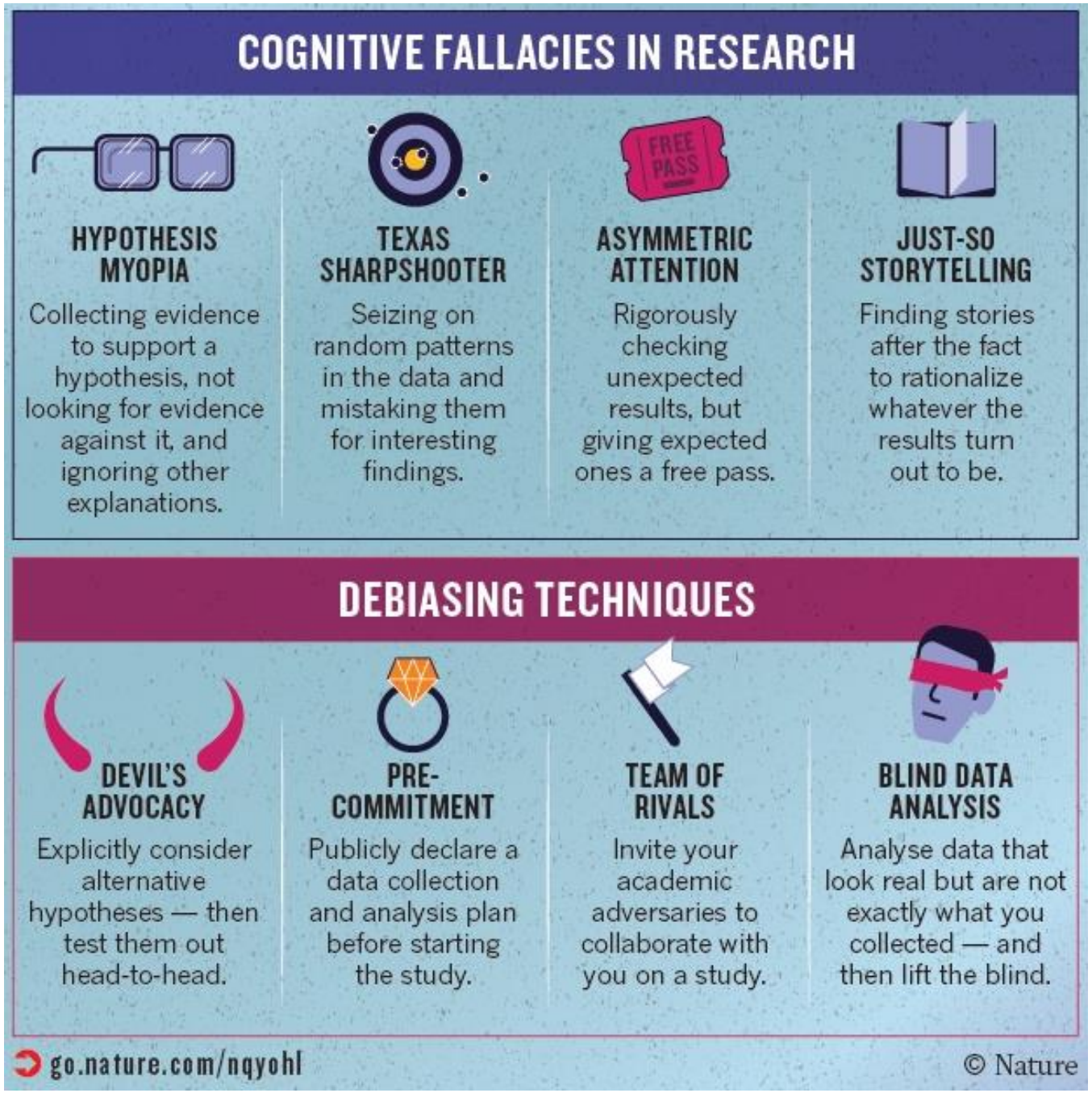
Twenty-nine research teams reached a wide variety of conclusions using different methods on the same data set to answer the same question (about football players' skin colour and red cards).

Any single team's results are **strongly influenced by subjective choices** during the analysis phase. Had any one of these 29 analyses come out as a single peer-reviewed publication, **the conclusion could have ranged from no skin color bias in referee decisions to a huge bias**.

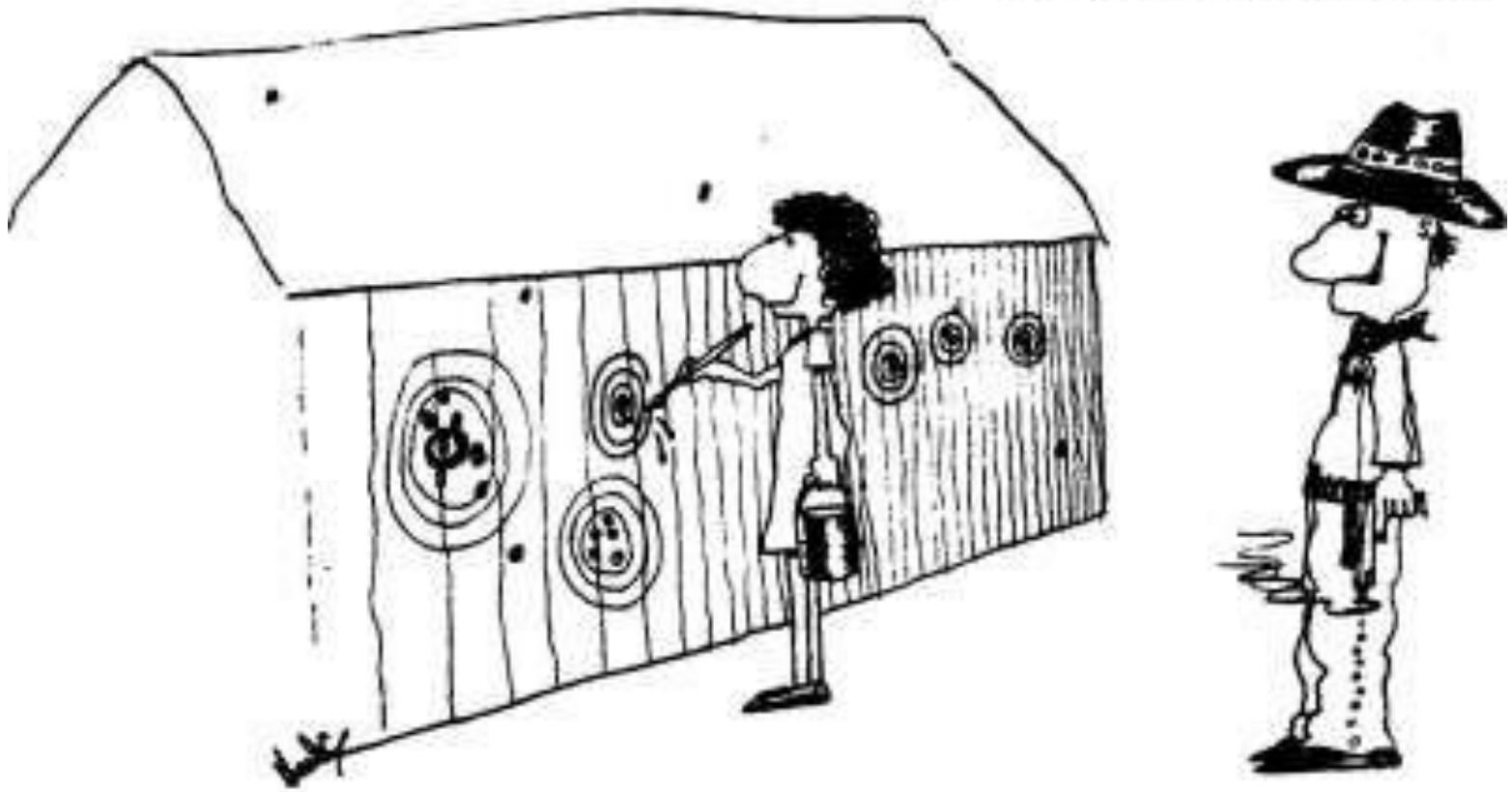
Cognitive Bias

HUMANS ARE REMARKABLY GOOD AT SELF-DECEPTION.
BUT GROWING CONCERN ABOUT REPRODUCIBILITY IS DRIVING MANY RESEARCHERS TO SEEK WAYS TO FIGHT THEIR OWN WORST INSTINCTS.

How scientists fool themselves – and how they can stop, *Nature* **526**, 182-185 (08 October 2015)

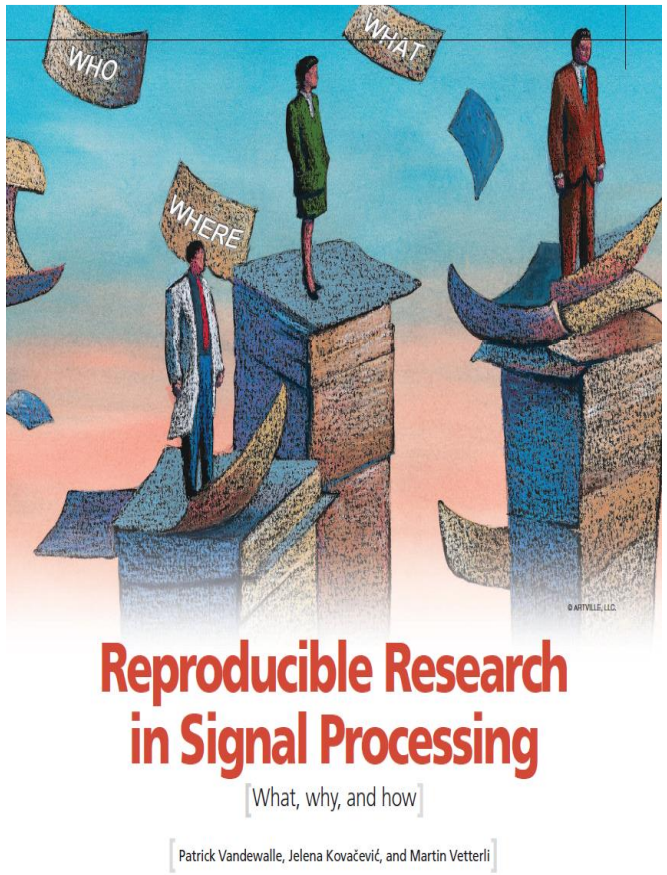

COGNITIVE FALLACIES IN RESEARCH
FREE PASS
HYPOTHESIS MYOPIA
Collecting evidence to support a hypothesis, not looking for evidence against it, and ignoring other explanations.
TEXAS SHARPSHOOTER
Seizing on random patterns in the data and mistaking them for interesting findings.
ASYMMETRIC ATTENTION
Rigorously checking unexpected results, but giving expected ones a free pass.
JUST-SO STORYTELLING
Finding stories after the fact to rationalize whatever the results turn out to be.
DEBIASING TECHNIQUES
DEVIL'S ADVOCACY
Explicitly consider alternative hypotheses — then test them out head-to-head.
PRE-COMMITMENT
Publicly declare a data collection and analysis plan before starting the study.
TEAM OF RIVALS
Invite your academic adversaries to collaborate with you on a study.
BLIND DATA ANALYSIS
Analyse data that look real but are not exactly what you collected — and then lift the blind.
go.nature.com/nqyohl
© Nature

The Texas Sharpshooter Procedure



Topic 3 – What tools can boost reproducibility?

Topic 3 – Reproducibility of computational results.



Journal of
open research software
ISSUES IN RESEARCH SOFTWARE

Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research

Victoria Stodden* and Sheila Miguez*

Science
AAAS
PERSPECTIVE

Reproducible Research in Computational Science

Roger D. Peng

DEFINITION: REPRODUCIBLE RESEARCH IN COMPUTATIONAL SCIENCE

A research work is called reproducible if all information relevant to the work, including, but not limited to, text, data and code, is made available, such that an independent researcher can reproduce the results.

The basic principle of computational reproducible research

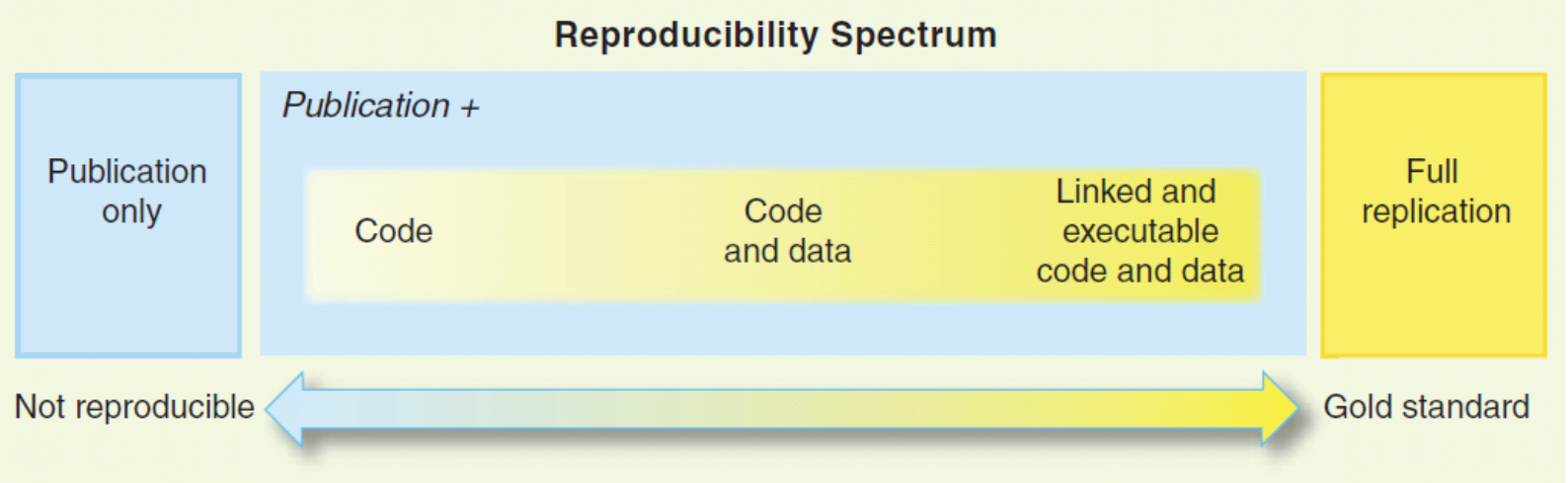


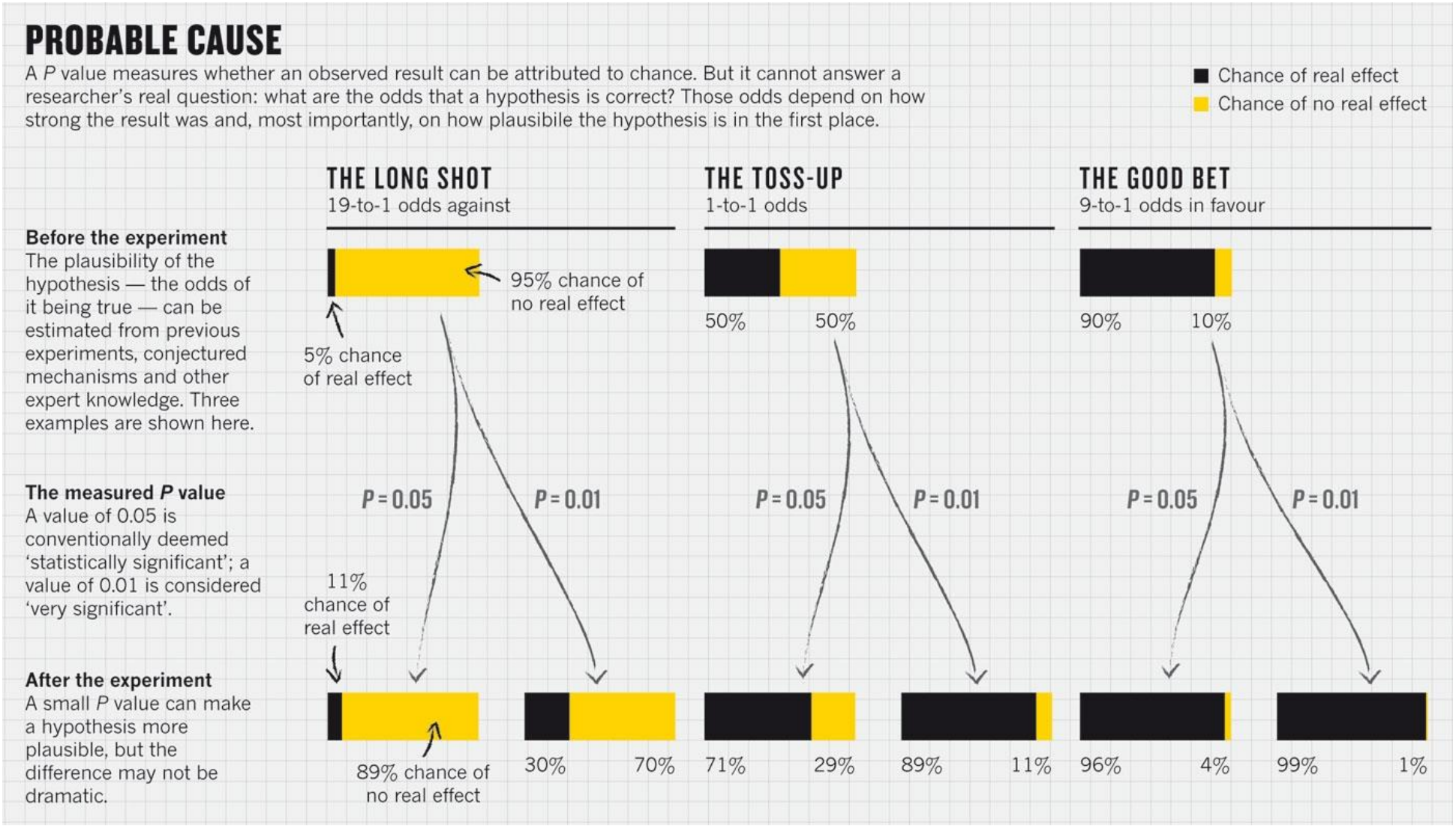
Fig. 1. The spectrum of reproducibility.

Checklist of reproducibility

- The paper
  - Is the algorithm described in sufficient detail?
  - Are exact parameter values given?
  - Is there a block diagram?
  - Is there a pseudocode?
  - Are there proofs for all the theorems?
  - Is the algorithm compared to other algorithms?

- The code
  - Are implementation details (programming language, platform, compiler flags, etc.) given?
  - Is the code available online?

- The data
  - Is there an explanation of what the data represents?
  - Is the size of the data set acceptable?
  - Is the data set available online?



Scientific method: Statistical errors. Regina Nuzzo. Nature 506, 150–152, 2014

Why Most Published Research Findings Are False. John P. A. Ioannidis. PLOS Medicine, 2005

Use version control

How to do research

1. Do awesome work
2. Write it down
3. Submit paper
4. Fame and glory
5. Move on to the next project (step 1)

http://www.ee.columbia.edu/~dpwe/e6891/slides/E6891_%20lecture%202.pdf

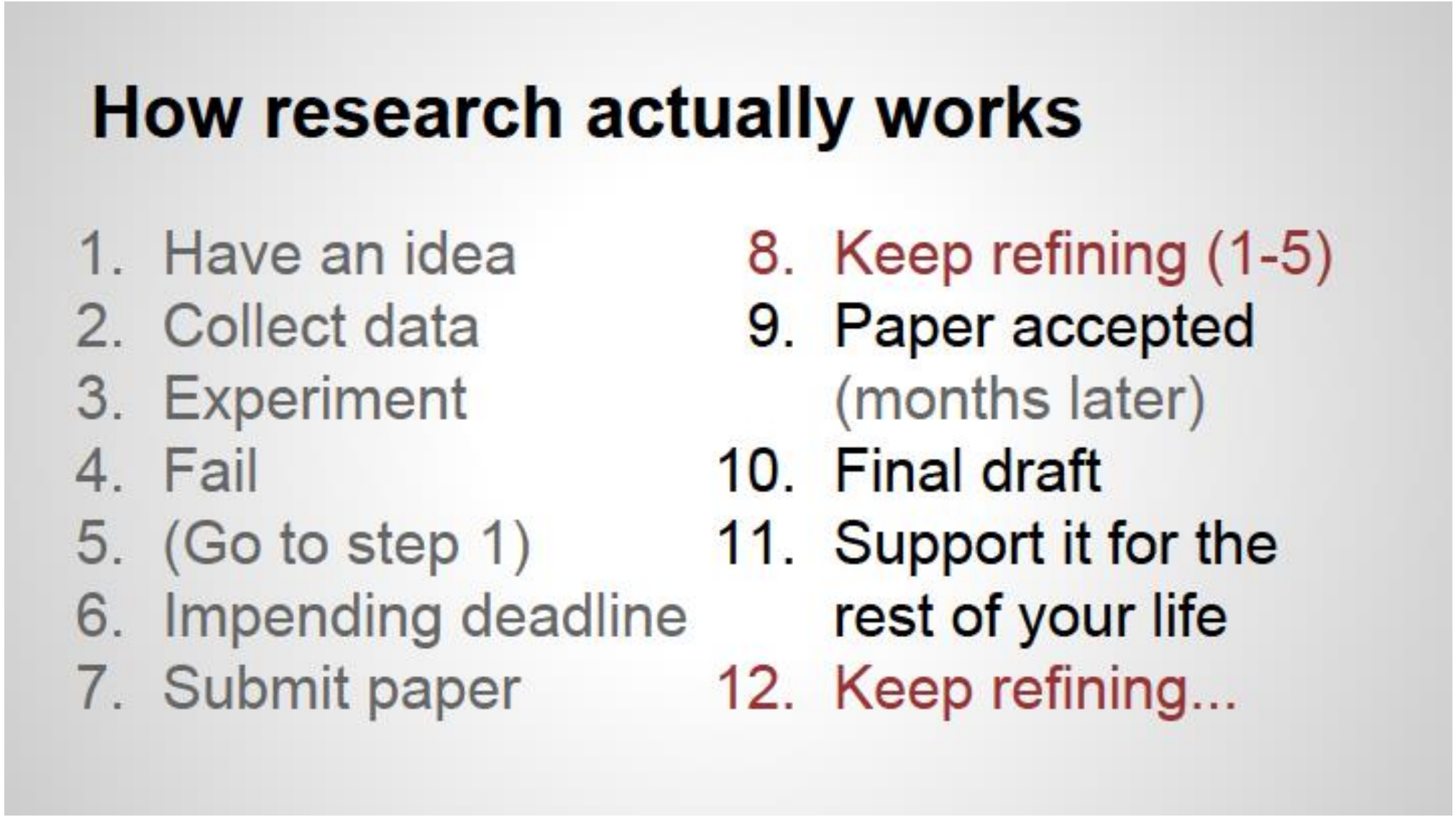
Use version control

How research actually works

1. Have an idea
2. Collect data
3. Experiment
4. Fail
5. (Go to step 1)
6. Impending deadline
7. Submit paper

http://www.ee.columbia.edu/~dpwe/e6891/slides/E6891_%20lecture%202.pdf

Use version control



http://www.ee.columbia.edu/~dpwe/e6891/slides/E6891_%20lecture%202.pdf

Mindset that ensures your research becomes “reproducible computational research”

1. Think about how to make it easy for others.

1. Think about how to make it easy for others.

1. Think about how to make it easy for others.

Personal Gain:
If the research is easy to reproduce for others, it is **easy to reproduce for you**.

Steps

1. Have a proper code and folder structure.

2. Automate as much as possible.

3. Manual steps, if any, must be well described.

4. Figures should be generated by code.

5. Figure scripts should only create figures, or very light data processing.

6. Store random seeds. On cluster, the seeds on the nodes needs to be stored as well.

Further reading

Talks from a pioneer in Data Analysis - Andrew Gelman, Professor of statistics and political science and director of the Applied Statistics Center at Columbia University

The Statistical Crisis in Science and How to Move Forward
https://youtu.be/KS3yPw91iC0

Bayes, statistics, and reproducibility (Rutgers, Foundations of Probability)
https://youtu.be/xgUBdi2wcDI

Crimes against data
https://youtu.be/fc1hkFC2c1E

Mød forskeren, der trak tæppet væk under videnskaben
https://www.zetland.dk/historie/sekdyb4n-a8dQKjjz-258ff