Crisis in statistics and in science, and their ethical implications

Andrea Saltelli
Centre for the Study of the Sciences and the Humanities (SVT), University of Bergen (UIB), and Open Evidence Research, Open University of Catalonia

Course NANO 310, August-September 2019

Where to find this talk: www.andreasaltelli.eu

The topic

From the misuse of a statistical technique to a problem in reproducibility in science; from this to an overall crisis of scientific practice and ethos, the role of technology, the impact on society.

The P-test saga

Downloaded from http://rsos.royalsocietypublishing.org/ on January 13, 2017

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Review

Cite this article: Colquhoun D. 2014 An investigation of the false discovery rate and the misinterpretation of *p*-values. *R. Soc. open sci.* **1**: 140216.
http://dx.doi.org/10.1098/rsos.140216

An investigation of the false discovery rate and the misinterpretation of *p*-values

David Colquhoun

Department of Neuroscience, Physiology and Pharmacology, University College London, Gower Street, London WC1 6BT, UK

P values by way of an example

- Two groups, one with a placebo, one with the treatment
- Random allocation to groups (+ more!)
- The difference *d* between the means of the two groups is tested (is it different from zero?)
- *p=0.05* implies that if there were no effect the probability of observing a value equal to *d* or higher would be *5%*

“At first sight, it might be thought that this procedure would guarantee that you would make a fool of yourself only once in every 20 times that you do a test”

Colquhoun D. 2014 An investigation of the false discovery rate and the misinterpretation of p-values. R. Soc. Open sci. 1: 140216. http://dx.doi.org/10.1098/rsos.140216

“The classical p-value does exactly what it says. But it is a statement about what would happen if there were no true effect. That cannot tell you about your long-term probability of making a fool of yourself, simply because sometimes there really is an effect. In order to do the calculation, **we need to know a few more things**”

Colquhoun D. 2014 An investigation of the false discovery rate and the misinterpretation of p-values. R. Soc. Open sci. 1: 140216. http://dx.doi.org/10.1098/rsos.140216

A classic exercise in screening

You test positive for AIDS (one test only). Time for despair?

Only one 1 in 100,000 has AIDS in your population

The test has a 5% false positive rate

Already one can say: in a population of say 100,000 one will have AIDS and 5,000 (5% of 100,000) will test positive

➔ Don't despair (yet)

Another exercise in screening (Colquhoun 2014)

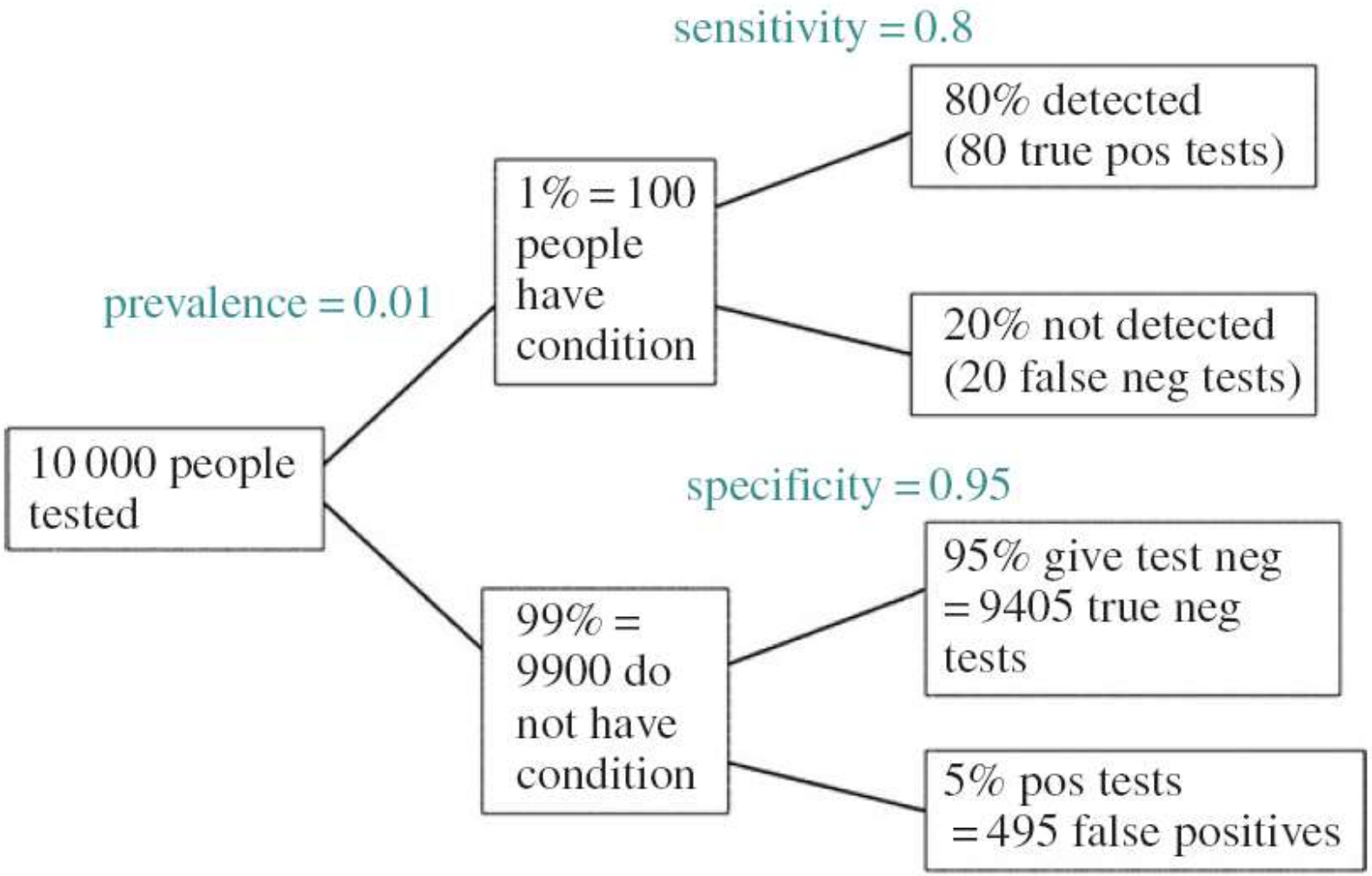
You test positive for mild cognitive impairment (MCI) (one test only). Time to retire? MCI prevalence in the population 1%, i.e. in a sample of 10,000 then 100 have MCI and 9,900 don't

The test has a 5% false positive rate; of the 9,900 who don't have MCI 495 test (false) positive and the remaining 9,405 (true) negative

The test does not pick all the 100 MCI but only 80; there will be 20 false negative. So we see 80+ 495=575 positive of which only 80 (a 14%) are true and the remaining 86% false

➔ It does not make sense to screen the population for MCI!

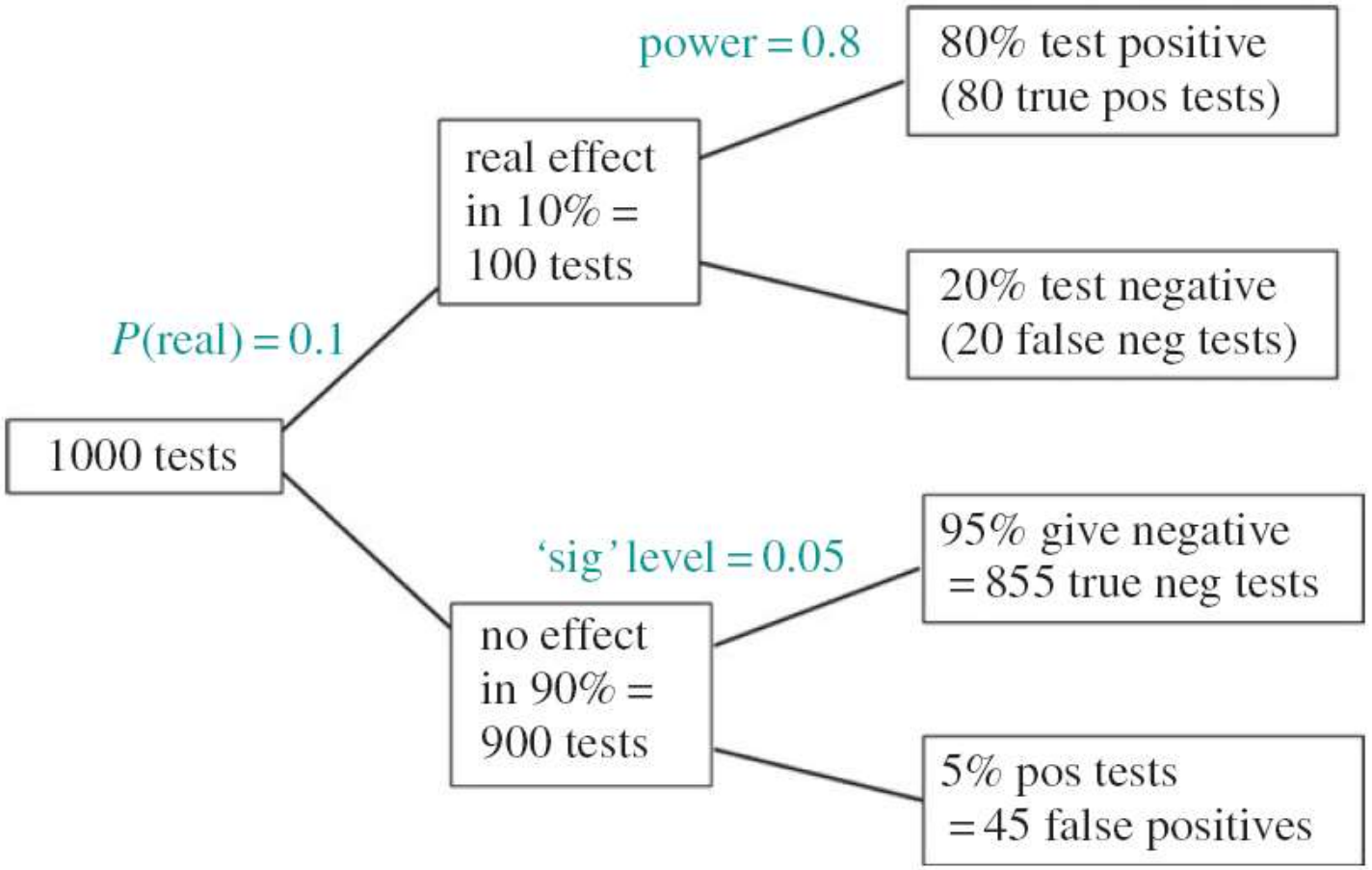
The number 86% = 495/(495+ 80) is our false discovery rate



Colquhoun D. 2014 An investigation of the false discovery rate and the misinterpretation of p-values. R. Soc. Open sci. 1: 140216. http://dx.doi.org/10.1098/rsos.140216

The same concept of false discovery rate applies to the problem of significance test

We now consider tests instead of individuals



Colquhoun D. 2014 An investigation of the false discovery rate and the misinterpretation of p-values. R. Soc. Open sci. 1: 140216. http://dx.doi.org/10.1098/rsos.140216

➔ We see 125 hypotheses as true 45 of which are not; the false discovery rate is 45/125 = 36%

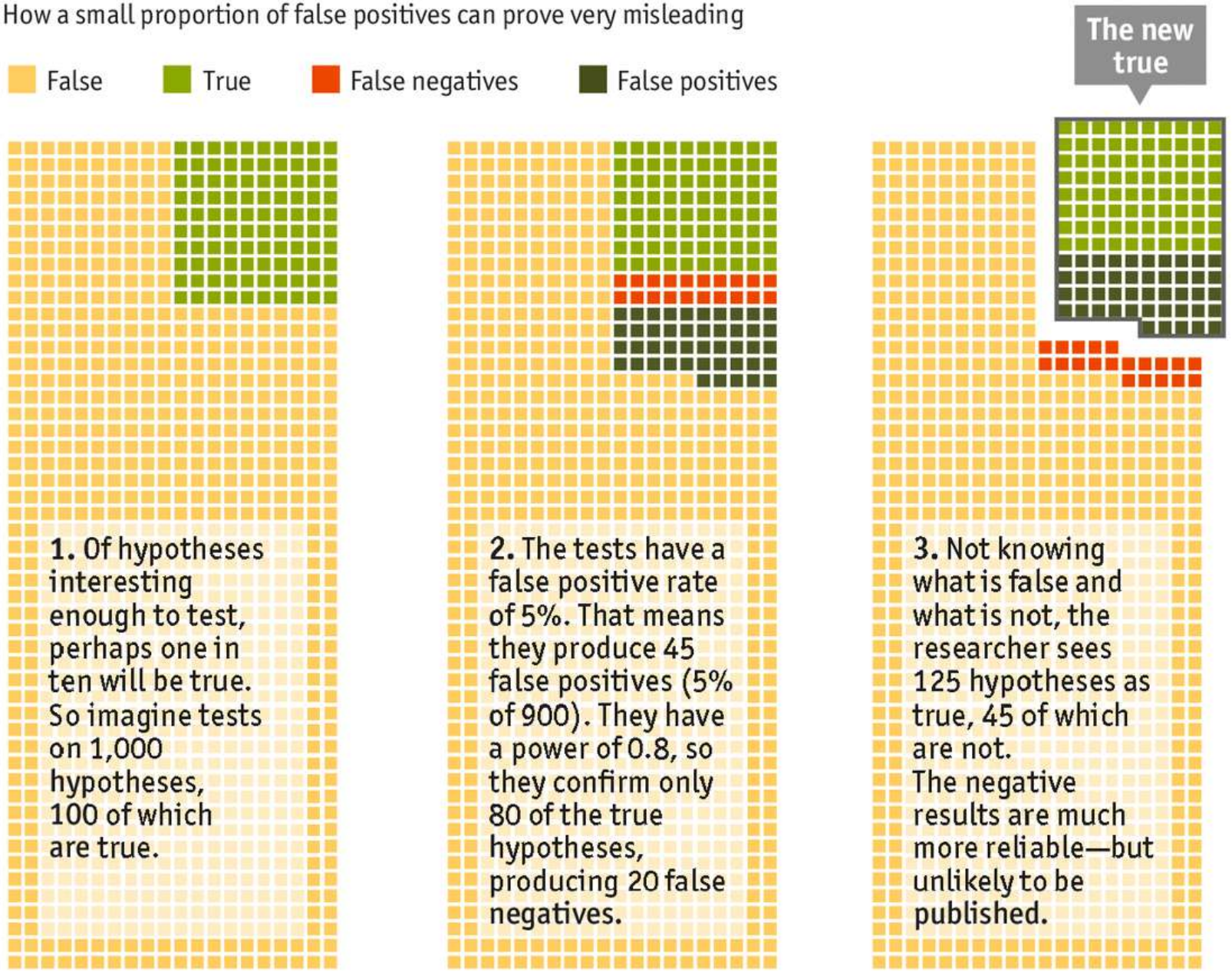
Significance p=0.05 ➔ false discovery rate of 36%

We now know that p=0.05 did not correspond to a chance in twenty of being wrong but in one in three

How many numbers did we need to know to reach this conclusion?

Unlikely results

How a small proportion of false positives can prove very misleading



Source: *The Economist*

The false discovery rate is the **black** area divided by the (**green** + **black**) one

Confused?

Is it that statistics is difficult, like baseball, or knitting?

“applied statistics is hard. Doing a statistical analysis is like playing basketball, or knitting a sweater. You can get better with practice”

Andrew Gelman

https://andrewgelman.com/2016/03/11/statistics-is-like-basketball-or-knitting/

Perhaps the real question is why do we assume that statistics can be taught at all, when we are aware that

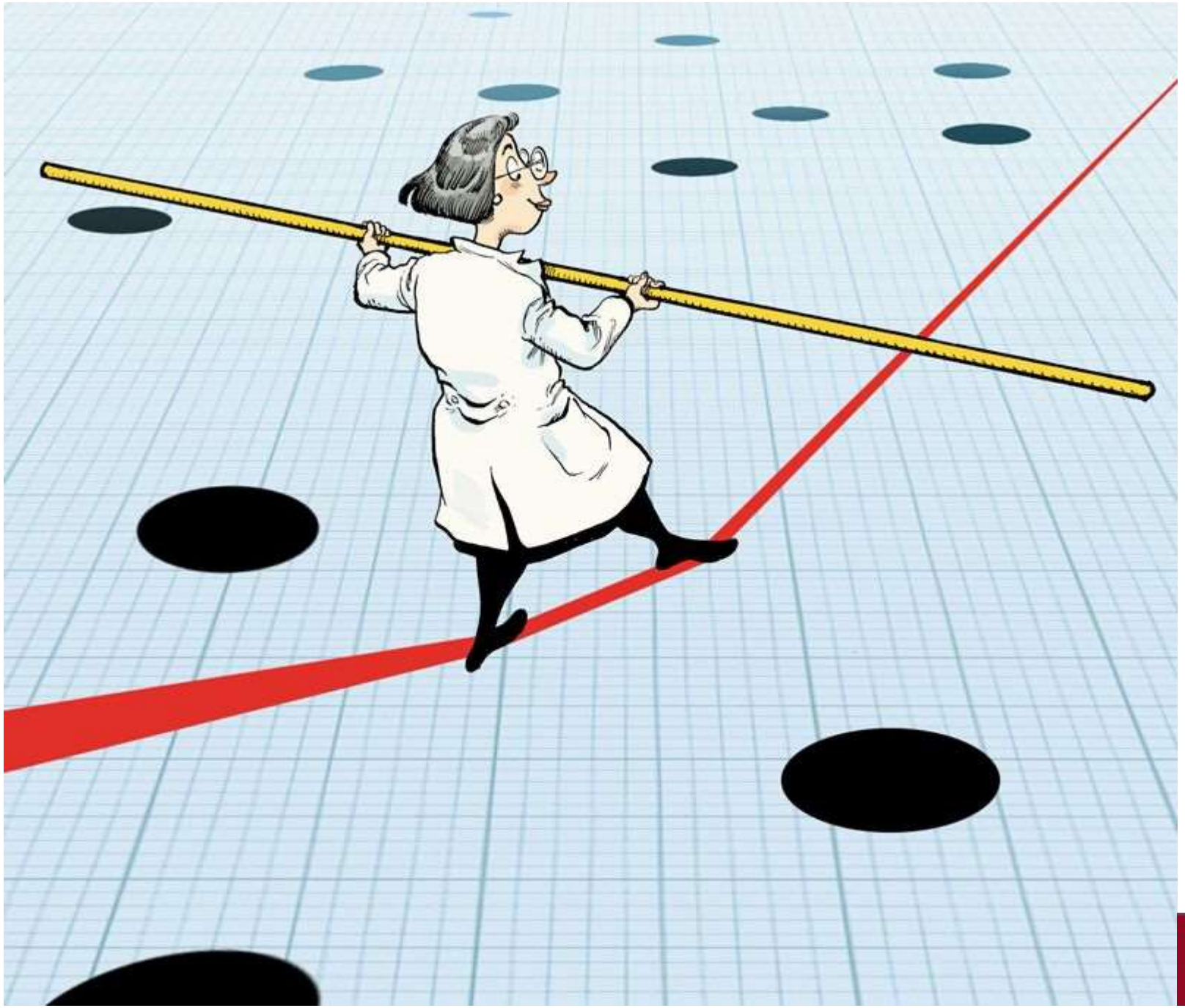
“passing information over disciplinary barriers implied dumbing it down” (E. Millgram)

“20% of the faculty teaching statistics in psychology, 39% of the professors and lecturers, and 66% of the students” don’t understand what the P-test is about

Gigerenzer, G., 2018, Statistical Rituals: The Replication Delusion and How We Got There, Advances in Methods and Practices in Psychological Science, 1-21

Crisis in statistics?

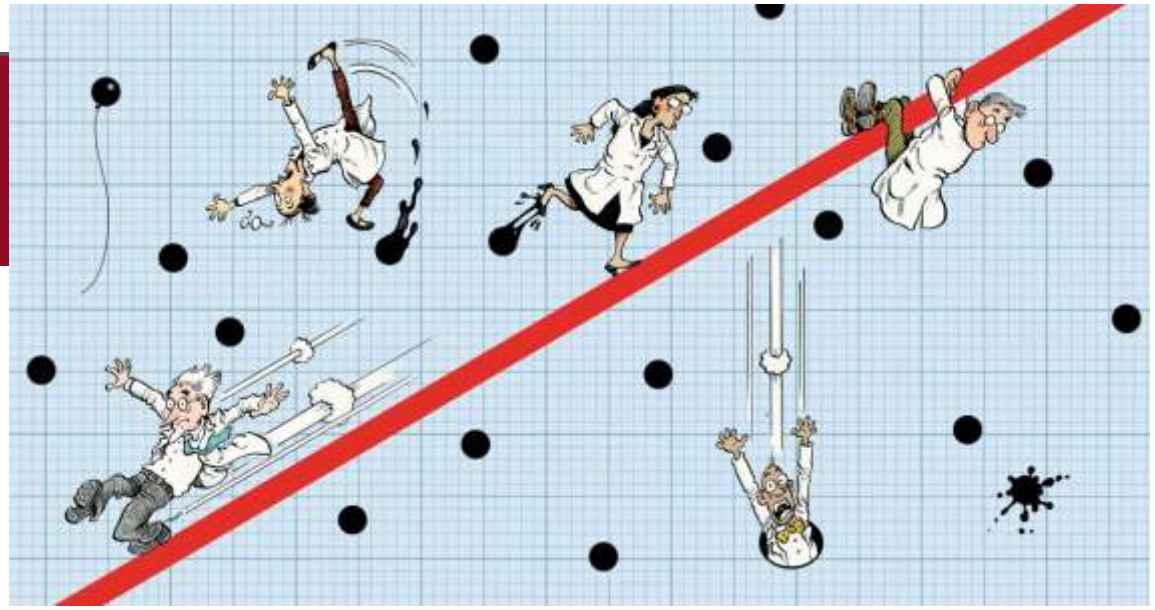
Statistics is experiencing a quality control crisis



Effect or no effect?

nature
International journal of science

nature
International journal of science



COMMENT · 28 NOVEMBER 2017

Five ways to fix statistics

As debate rumbles on about how and how much poor statistics is to blame for poor reproducibility, Nature asked influential statisticians to recommend one change to improve science. The common theme? The problem is not our maths, but ourselves.

Jeff Leek ✉, Blakeley B. McShane, Andrew Gelman ✉, David Colquhoun ✉, Michèle B. Nuijten ✉ & Steven N. Goodman ✉

nature
International journal of science

CORRESPONDENCE · 16 JANUARY 2018

Fixing statistics is more than a technical issue

Andrea Saltelli ✉ & Philip Stark

https://www.nature.com/articles/d41586-018-00647-9

nature
International journal of science

CORRESPONDENCE · 16 JANUARY 2018

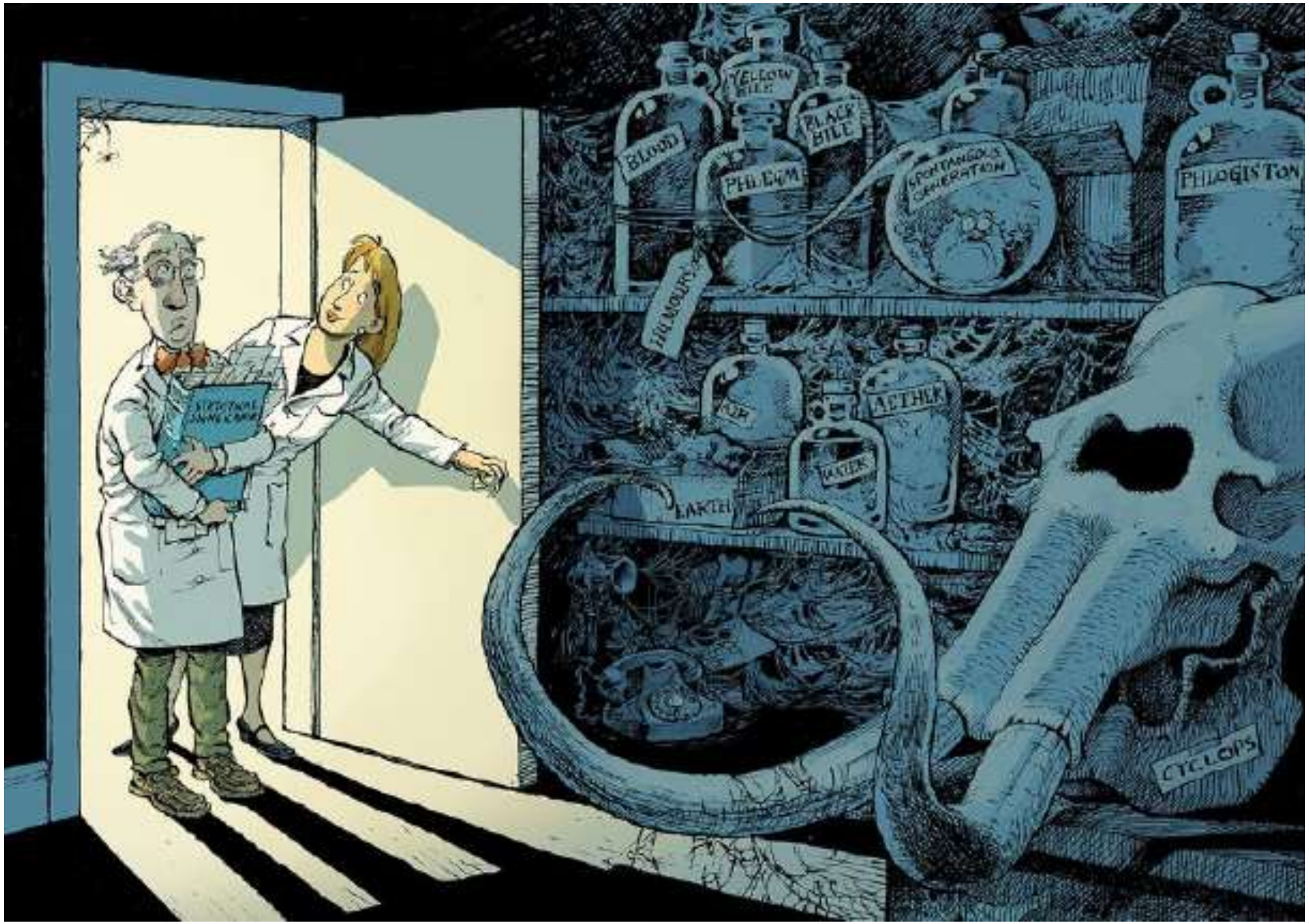
Integrity must underpin quality of statistics

Jerome Ravetz ✉

https://www.nature.com/articles/d41586-018-00648-8

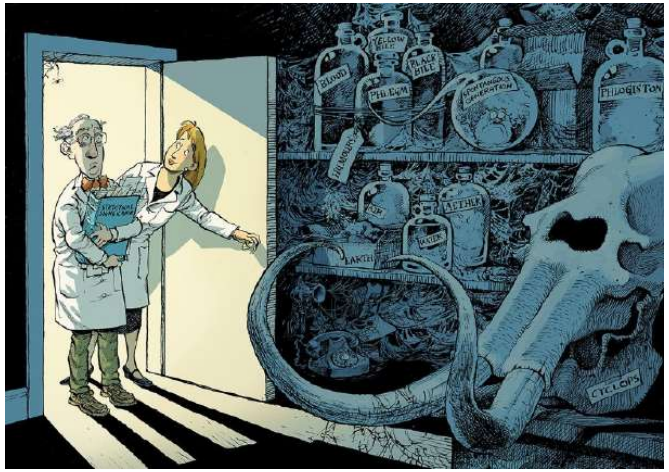
The great paradox of science is that passionate practitioners must carefully produce dispassionate facts (J. Ravetz *Scientific Knowledge and its Social Problems* Oxford Univ. Press; 1971). Meticulous technical and normative judgement, as well as morals and morale, are necessary to navigate the forking paths of the statistical garden (Saltelli and Stark, 2018)

All users of statistical techniques, as well as those in other mathematical fields such as modelling and algorithms, need an effective societal commitment to the maintenance of quality and integrity in their work. If imposed alone, technical or administrative solutions will only breed manipulation and evasion (Ravetz, 2018)



Throw away the concept of statistical significance?

nature
International journal of science



COMMENT · 20 MARCH 2019

Scientists rise up against statistical significance

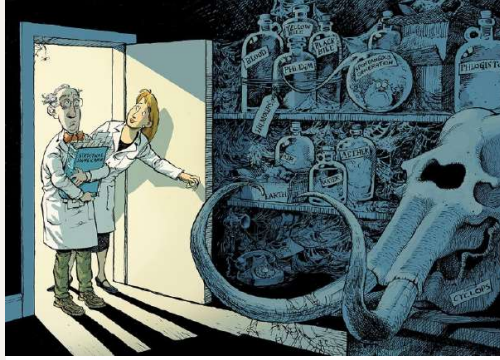
Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

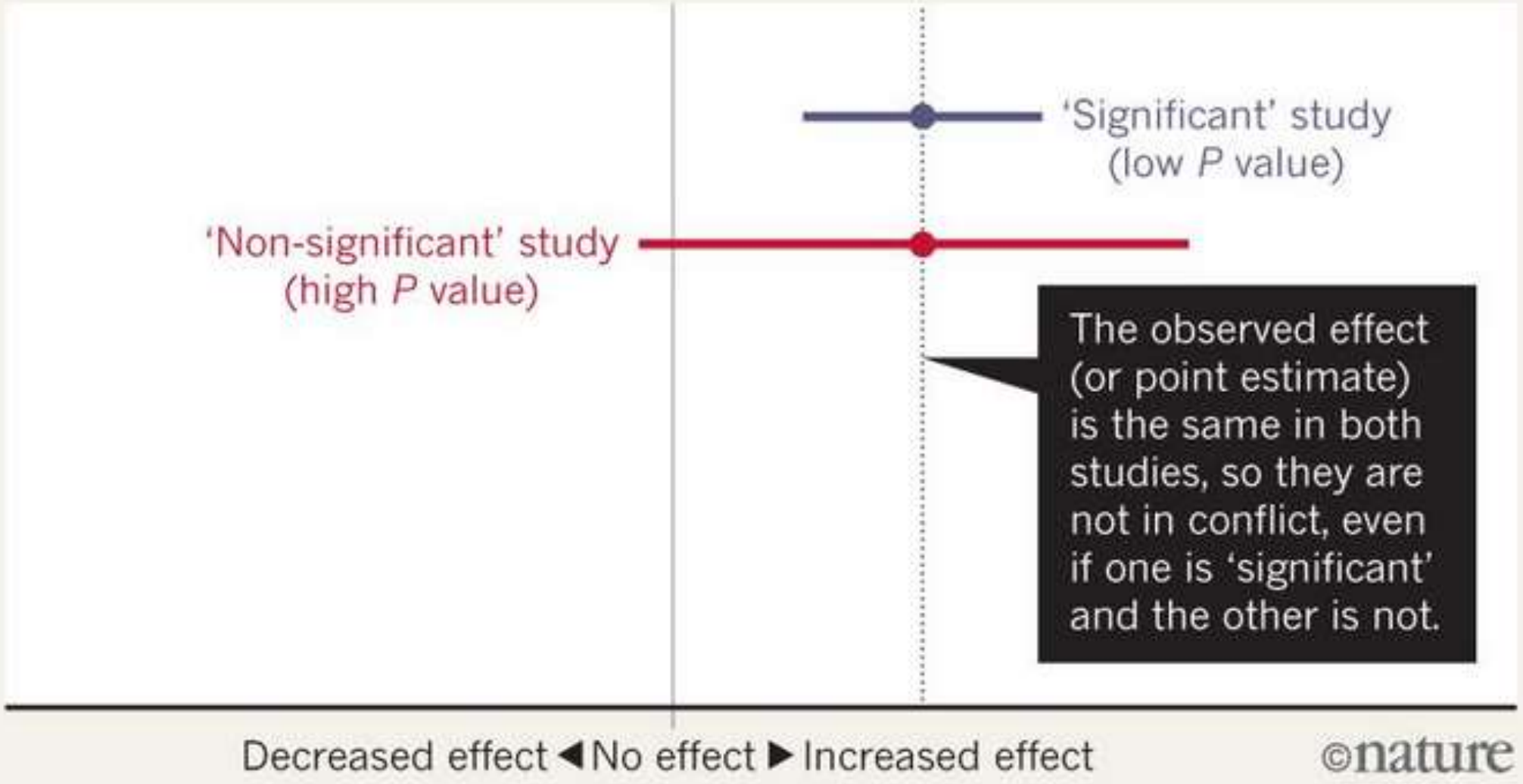
Valentin Amrhein ✉, Sander Greenland & Blake McShane

See the discussion on the blog of Andrew Gelman https://statmodeling.stat.columbia.edu/

BEWARE FALSE CONCLUSIONS

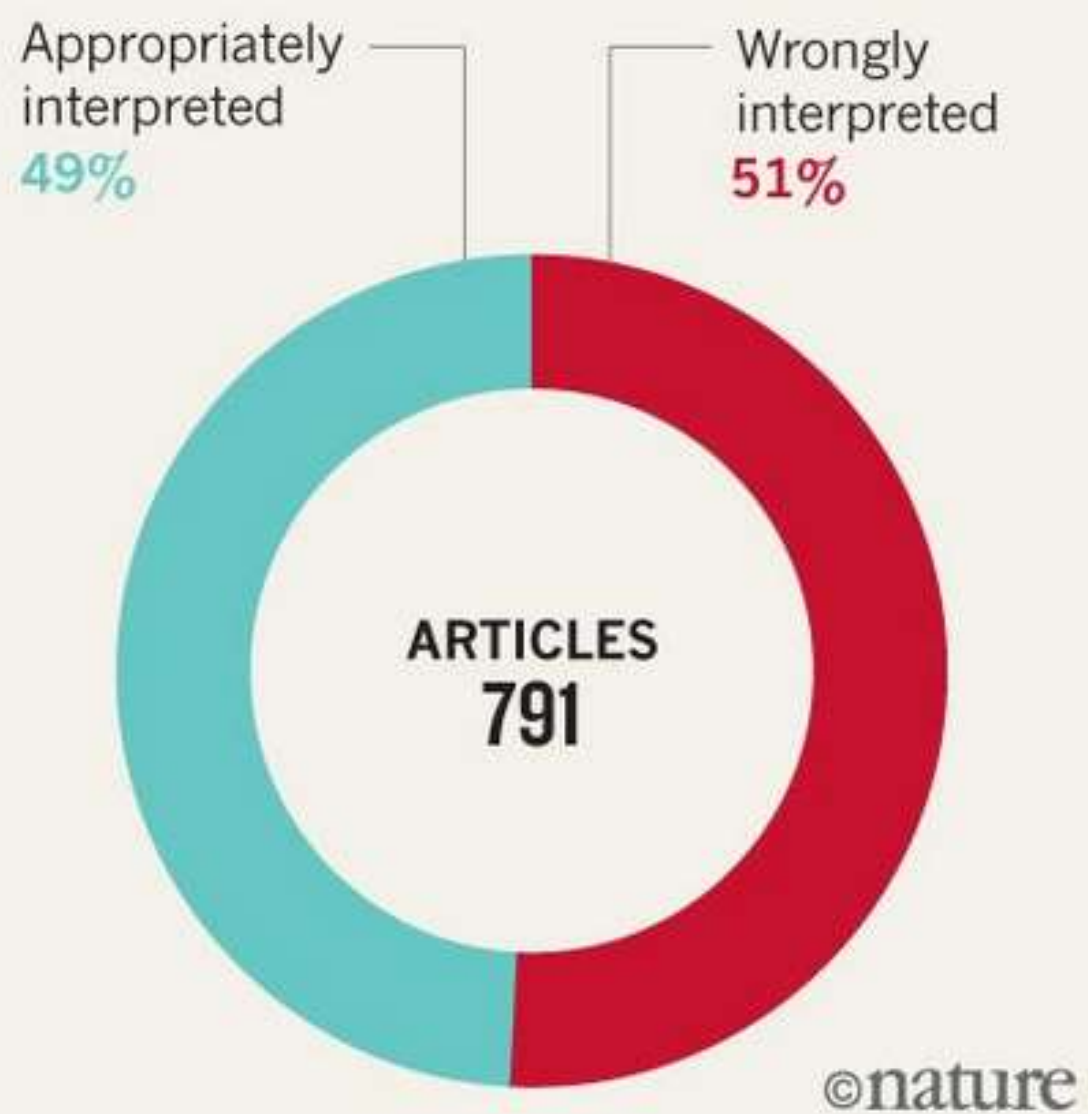
Studies currently dubbed 'statistically significant' and 'statistically non-significant' need not be contradictory, and such designations might cause genuine effects to be dismissed.



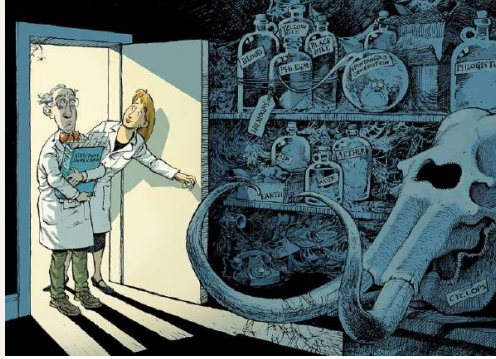


WRONG INTERPRETATIONS

An analysis of 791 articles across 5 journals* found that around half mistakenly assume non-significance means no effect.



*Data taken from: P. Schatz *et al. Arch. Clin. Neuropsychol.* **20**, 1053–1059 (2005); F. Fidler *et al. Conserv. Biol.* **20**, 1539–1544 (2006); R. Hoekstra *et al. Psychon. Bull. Rev.* **13**, 1033–1037 (2006); F. Bernardi *et al. Eur. Sociol. Rev.* **33**, 1–15 (2017).



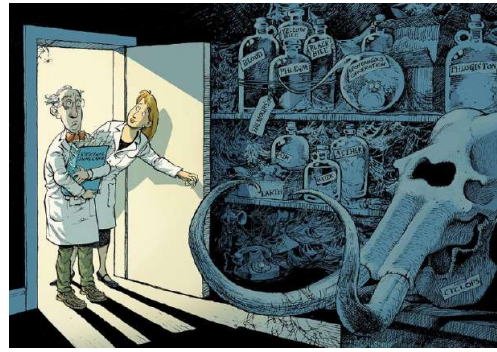
Is it appropriate to get hundreds of people to sign a letter of support for a scientific editorial?

> “I am afraid that what you are doing at this point is not science, but campaigning. ” (John Ioannidis)
>
> “YES we are campaigning and it’s long overdue . . . because YES this is an issue of politics, ethics, and injustice!” (Sander Greenland)

https://statmodeling.stat.columbia.edu/2019/03/20/retire-statistical-significance-the-discussion/

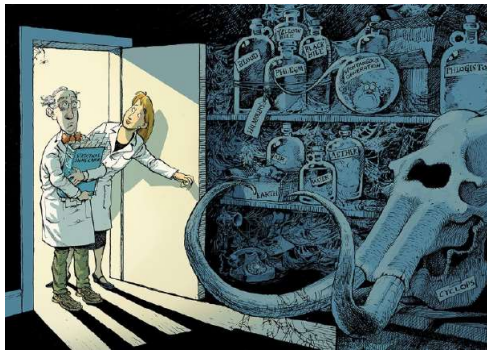
P<0.05 does not mean that H_0 is false any more than it means that H_a is true

"we are calling for a stop to the use of P values in the conventional, dichotomous way — to decide whether a result refutes or supports a scientific hypothesis"



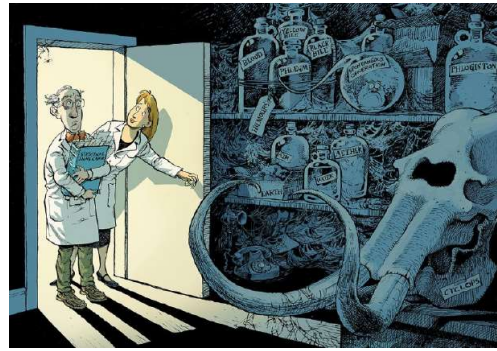
Valentin Amrhein et al., 20 March 2019, Scientists rise up against statistical significance, Nature, https://www.nature.com/articles/d41586-019-00857-9

“Unfortunately, the false belief that crossing the threshold of statistical significance is enough to show that a result is ‘real’ has led scientists and journal editors to privilege such results, thereby distorting the literature”



Valentin Amrhein et al., 20 March 2019, Scientists rise up against statistical significance, Nature, https://www.nature.com/articles/d41586-019-00857-9

“the rigid focus on statistical significance encourages researchers to choose data and methods that yield statistical significance for some desired (or simply publishable) result, or that yield statistical non-significance for an undesired result”



Valentin Amrhein et al., 20 March 2019, Scientists rise up against statistical significance, Nature, https://www.nature.com/articles/d41586-019-00857-9

Ethical P?

One might be worried that important effects are overlooked because the test of significance fails ($p>0.05$)

Another might be worried of bogus results passed on to the academic literature thanks to a low value of the p-test ($p<0.05$).

Ethical P?

The contention is normative, not to say political. To make an example, some may fear the introduction on the marked of ineffectual pharmaceutical products, others that important epidemiological effects of a pollutant on health may be overlooked.

The first group would thus have a more restrictive value for the test, the second group a less restrictive one.

Ethical P?

Philosopher Richard Rudner had already written in 1953 that it is impossible to use a test of significance without knowing to what it is being applied, i.e. without making a value judgment.

R. Rudner, “The Scientist Qua Scientist Makes Value Judgments,” Philosophy of Science, vol. 20. The University of Chicago Press Philosophy of Science Association, pp. 1–6, 1953.
http://www.andreasaltelli.eu/file/repository/00_Rudnerphs53.pdf

Ethical P?

Interestingly, Rudner used this example to make the point that scientists do need to make value judgments

R. Rudner, “The Scientist Qua Scientist Makes Value Judgments,” Philosophy of Science, vol. 20. The University of Chicago Press Philosophy of Science Association, pp. 1–6, 1953. http://www.andreasaltelli.eu/file/repository/00_Rudnerphs53.pdf

“How sure we need to be before we accept a hypothesis will depend on how serious a mistake would be”

THE SCIENTIST *QUA* SCIENTIST MAKES VALUE JUDGMENTS*

RICHARD RUDNER

Statistics reacts

The discipline of statistics has been going through a phase of critique and self-criticism, due to mounting evidence of poor statistical practice of which misuse and abuse of the P-test is the most visible sign

AMERICAN STATISTICAL ASSOCIATION RELEASES STATEMENT ON STATISTICAL SIGNIFICANCE AND *P*-VALUES

Provides Principles to Improve the Conduct and Interpretation of Quantitative Science

March 7, 2016

+ twenty 'dissenting' commentaries

Wasserstein, R.L. and Lazar, N.A., 2016. 'The ASA's statement on p-values: context, process, and purpose', The American Statistician, DOI:10.1080/00031305.2016.1154108.

See also Christie Aschwanden at http://fivethirtyeight.com/features/not-even-scientists-can-easily-explain-p-values/

P-hacking (fishing for favourable p-values) and **HARKing** (formulating the research **H**ypothesis **A**fter the **R**esults are **K**nown);
Desire to achieve a sought for – or simply publishable – result leads to fiddling with the data points, the modelling assumptions, or the research hypotheses themselves

Leamer, E. E. Tantalus on the Road to Asymptopia. J. Econ. Perspect. 24, 31–46 (2010).

Kerr, N. L. HARKing: Hypothesizing After the Results are Known. Personal. Soc. Psychol. Rev. 2, 196–217 (1998).

A. Gelman and E. Loken, "The garden of forking paths: Why multiple comparisons can be a problem, even when there is no 'fishing expedition' or 'p-hacking' and the research hypothesis was posited ahead of time," 2013.

Has statistics become a ritual?

Journal of Management

Surrogate Science

The Idol of a Universal Method for Scientific Inference

Gerd Gigerenzer, Julian N. Marewski,

First Published September 2, 2014 | Research Article | Check for updates
https://doi.org/10.1177/0149206314547522

Advances in Methods and Practices in Psychological Science aps

Statistical Rituals: The Replication Delusion and How We Got There

Gerd Gigerenzer,

First Published June 14, 2018 | Research Article | Check for updates
https://doi.org/10.1177/2515245918771329

For Gigerenzer & Marewski statistics has changed the nature all disciplines …

… Creating a persistent surrogate science based on worshipping P-values

Better to have no beliefs than to embrace falsehoods… (➔ F. Bacon's idols)

G. Gigerenzer and J. N. Marewski, "Surrogate Science," J. Manage., vol. 41, no. 2, pp. 421–440, Feb. 2015.

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

ASA
AMERICAN STATISTICAL ASSOCIATION

significance

The mechanical, ritualistic application of statistics is contributing to a crisis in science. Education, software and peer review have encouraged poor practice – and it is time for statisticians to fight back. By **Philip B. Stark** and **Andrea Saltelli**

The statistical garden of the forking paths

Jorge Luis Borges

Andrew Gelman

http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Crisis in science?

There have recently been alarms as to the scientific quality arrangement is several disciplines. The most visible symptom of this possible dysfunction is the so-called reproducibility crisis

Futures
Volume 91, August 2017, Pages 5-11

What is science's crisis really about?

Andrea Saltelli [a, b] ✉, Silvio Funtowicz [a]

Futures
Volume 104, December 2018, Pages 85-90

Why science's crisis should not become a political battling ground

Andrea, Saltelli ✉

On the radar: October 2013

The Economist

HOW SCIENCE GOEs WRONG.

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

2005

John P. A. Ioannides

J. P. A. Ioannidis, Why Most Published Research Findings Are False, PLoS Medicine, August 2005, 2(8), 696-701.

Failed replications, entire subfields going bad, fraudulent peer reviews, predatory publishers, perverse metrics, statistics on trial …

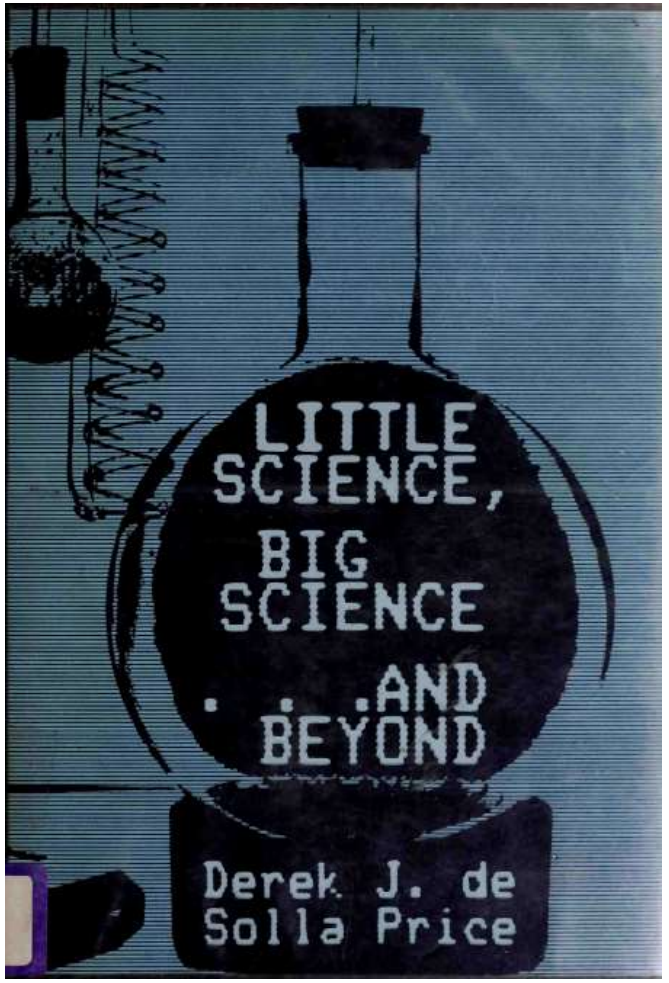
… misleading science advice, institutions on denial, a new breed of science wars

The crisis is methodological, epistemological, ethical and metaphysical

Scholars who saw it coming …

and how they were vindicated

In 1963 Derek J. de Solla Price prophesized that Science would reach saturation (and in the worst-case senility) under its own weight, victim of its own success and exponential growth (pp 1-32)



Derek J. de Solla Price

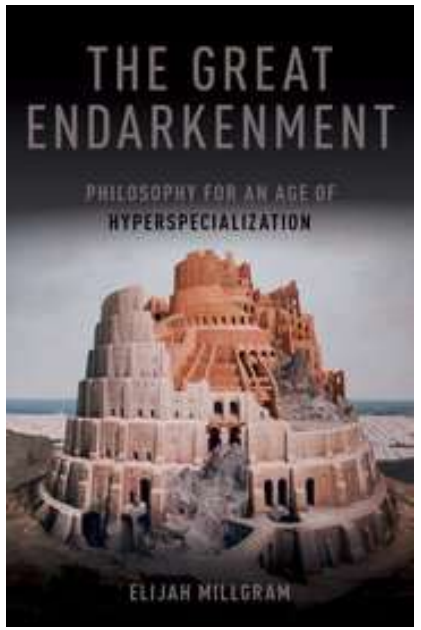
de Solla Price, D.J., 1963, Little science big science, Columbia University Press.

~2.2 million articles a year (2016) over ~30,000 journals

https://www.aje.com/en/arc/scholarly-publishing-trends-2016/

http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html

Derek de Solla Price ←→ Elijah Millgram



The Great Endarkenment.
Philosophy for an Age of Hyperspecialization
By Elijah Millgram

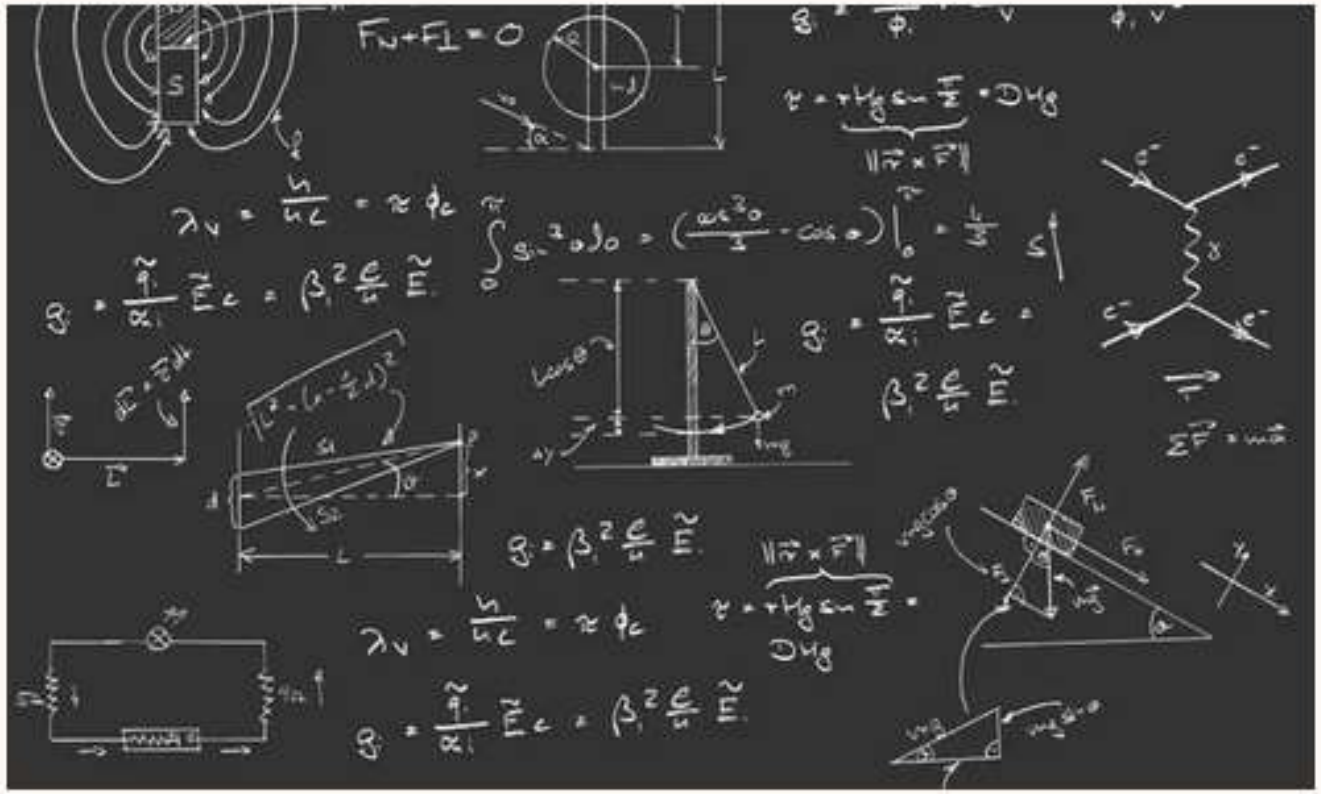
Describes a world in which all knowledge and products are the result of some form of extremely specialized expertise, and in which expertise is itself highly circumscribed, since experts depend in turn on other experts whose knowledge claims and styles of argumentation cannot be exported from one discipline to the next. ➔ “serial hyperspecializers” (p. 26)
Experts thus become “logical aliens” (p. 32)

Science from ‘Gemeinschaft’ to ‘Gesellschaft’, for Jerome R. Ravetz

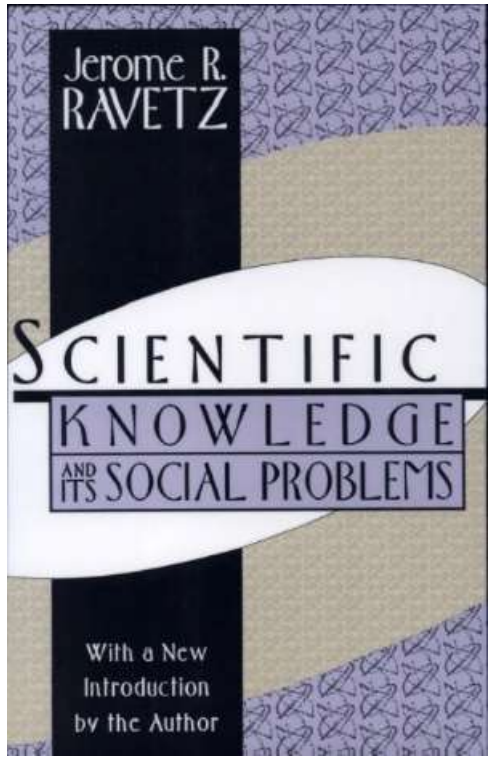
https://www.theguardian.com/science/political-science/2016/jun/08/how-should-we-treat-sciences-growing-pains

How should we treat science's growing pains?

Jerome Ravetz has been one of the UK's foremost philosophers of science for more than 50 years. Here, he reflects on the troubles facing contemporary science. He argues that the roots of science's crisis have been ignored for too long. Quality control has failed to keep pace with the growth of science.



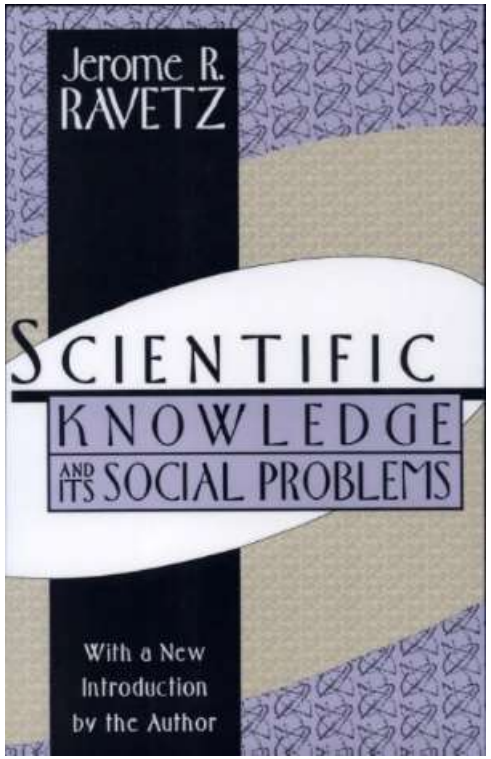
Two separate factors are necessary for the achievement of worthwhile scientific results: a community of scholars with a shared knowledge of the standards of quality appropriate for their work and a shared commitment to enforce those standards by the informal sanctions the community possesses; and individuals whose personal integrity sets standards at least as high as those required by their community.



Jerome R. Ravetz

Ravetz, J., 1971, Scientific Knowledge and its Social Problems, Oxford University Press.

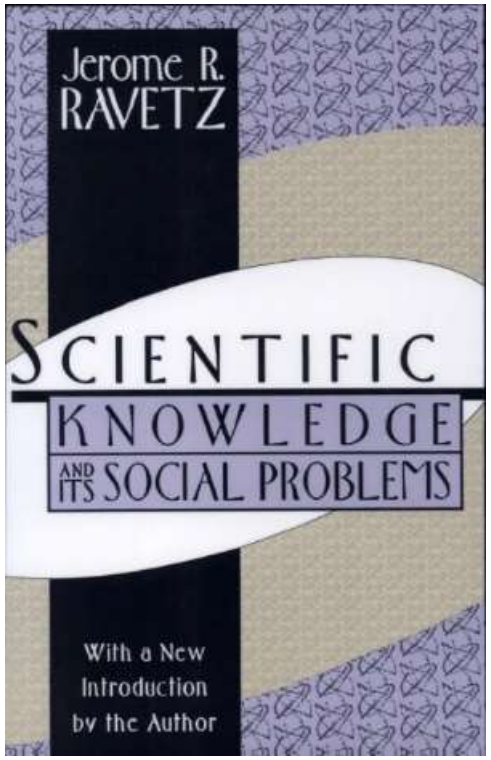
p.22: [...] The problem of quality control in science is at the centre of the social problems of the industrialized science of the present period."



Jerome R. Ravetz

Ravetz, J., **1971**, Scientific Knowledge and its Social Problems, Oxford University Press.

“If [science] fails to resolve this problem [⋯] then the immediate consequences for morale and recruitment will be serious; and those for the survival of science itself, grave”



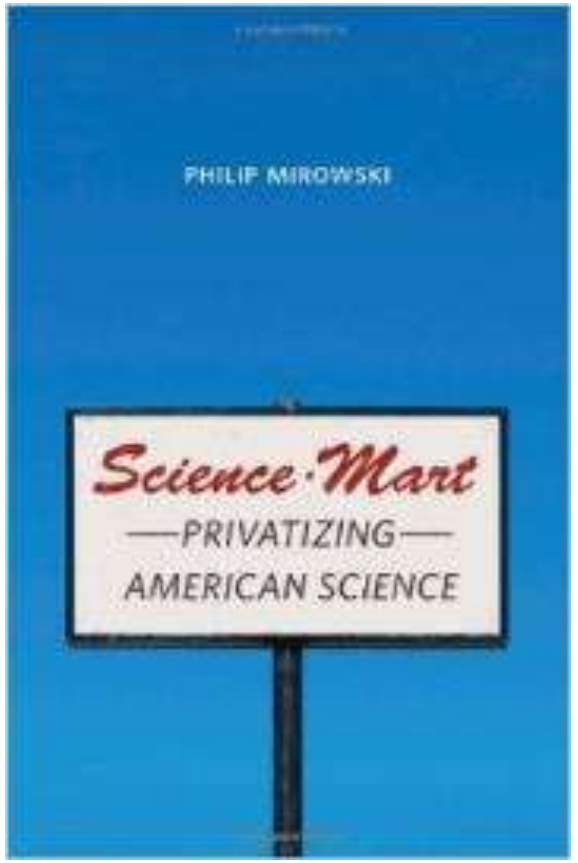
Jerome R. Ravetz

Ravetz, J., **1971**, Scientific Knowledge and its Social Problems, Oxford University Press.

… neoliberal ideologies lead to decreasing state funding of science, which becomes privatized … knowledge as a monetized commodity replaces knowledge as a public good ➔ collapse of quality

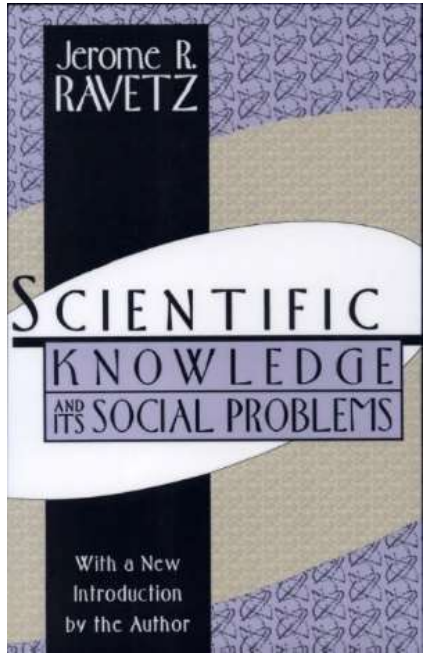
Philip Mirowski

Mirowski, P. 2011. Science-Mart: Privatizing American Science, Harvard University Press.



p. 179. For it is possible for a field to be diseased [...] reforming a diseased field is a task of great delicacy [...] not even an apparatus of institutional structures can do anything to maintain or restore the health of a field **in the absence of an essential ethical element operating through the interpersonal channel of communication.**

Ravetz, J., 1971, Scientific Knowledge and its Social Problems, Oxford University Press.



Jerome R. Ravetz

PLOS | MEDICINE

Brow

THE ECONOMIC JOURNAL

The Economic Journal, **127** (*October*), F236–F265. Doi: 10.1111/ecoj.12461 © 2017 Royal Economic Society. Published by John Wiley & Sons, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

OPEN ACCESS

ESSAY

June 21, 2017

Why Most Clinical Research Is Not Useful

John P. A. Ioannidis

Published: June 21, 2016 • https://doi.org/10.1371/journal.pmed.1002049

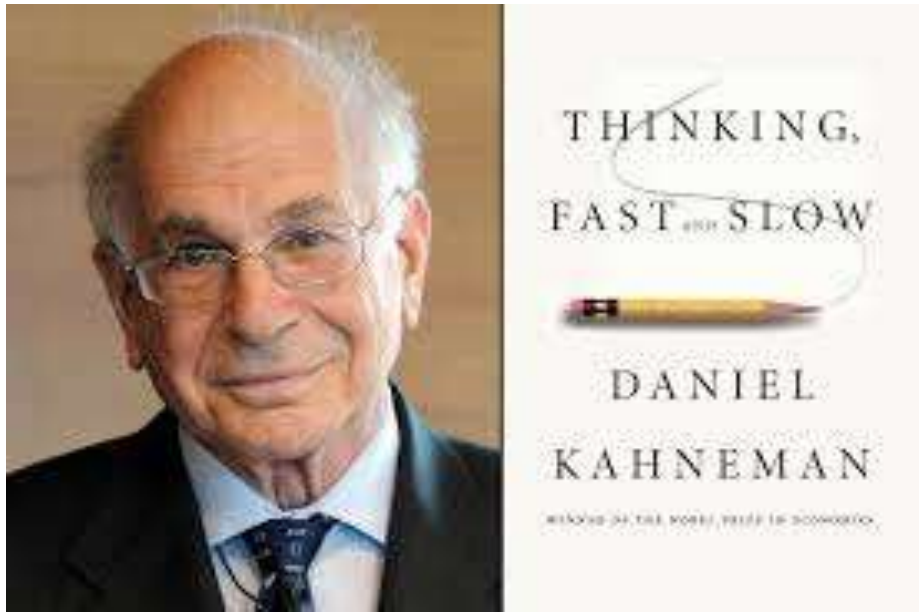
THE POWER OF BIAS IN ECONOMICS RESEARCH*

John P. A. Ioannidis, T. D. Stanley and Hristos Doucouliagos

October 27, 2017

Rather than isolated instances of corruption now entire fields of research are found diseased

Reconstruction of a Train Wreck: How Priming Research Went off the Rails



"[…]questions have been raised about the robustness of priming results … your field is now the poster child for doubts about the integrity of psychological research…"

https://replicationindex.wordpress.com/2017/02/02/reconstruction-of-a-train-wreck-how-priming-research-went-of-the-rails/comment-page-1/

An existential crisis?

Most observers have noted that the crisis has technical as well as ethical and behavioural elements which interact with one another – e.g. the 'publish or perish' obsession has an impact on selection bias – the tendency to favour positive over negative results

Bad science reproduces better than the good sort

Downloaded from http://rsos.royalsocietypublishing.org/ on September 23, 2016

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Cite this article: Smaldino PE, McElreath R. 2016 The natural selection of bad science. *R. Soc. open sci.* **3**: 160384.
http://dx.doi.org/10.1098/rsos.160384

Received: 1 June 2016
Accepted: 17 August 2016

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

PES, 0000-0002-7133-5620; RME, 0000-0002-0387-5377

Poor research design and data analysis encourage false-positive findings. Such poor methods persist despite perennial calls for improvement, suggesting that they result from something more than just misunderstanding. The persistence of poor methods results partly from incentives that favour them, leading to the natural selection of bad science. This dynamic requires no conscious strategizing—no deliberate cheating nor loafing—by scientists, only that publication is a principal factor for

As in the real world, successful labs produce more 'progeny,' such that their methods are more often copied and their students are more likely to start labs of their own. Selection for high output leads to poorer methods and increasingly high false discovery rates.

Improving the quality of research requires change at the institutional level.

Smaldino PE, McElreath R., 2016 The natural selection of bad science. R. Soc. open sci. 3: 160384. http://dx.doi.org/10.1098/rsos.160384

Bad science is ‘sticky’

ECOSPHERE
AN ESA OPEN ACCESS JOURNAL

Article | Open Access

Do rebuttals affect future science?

Jeannette A. Banobi, Trevor A. Branch, Ray Hilborn

First published: 30 March 2011 | https://doi.org/10.1890/ES10-00142.1 | Cited by: 13

"We examined seven high-profile original articles and their rebuttals, finding that original articles were cited 17 times more than rebuttals, and that annual citation numbers were unaffected by rebuttals"

Bad science in bad journals?

The 1st most cited journal in Psychology

frontiers in Human Neuroscience

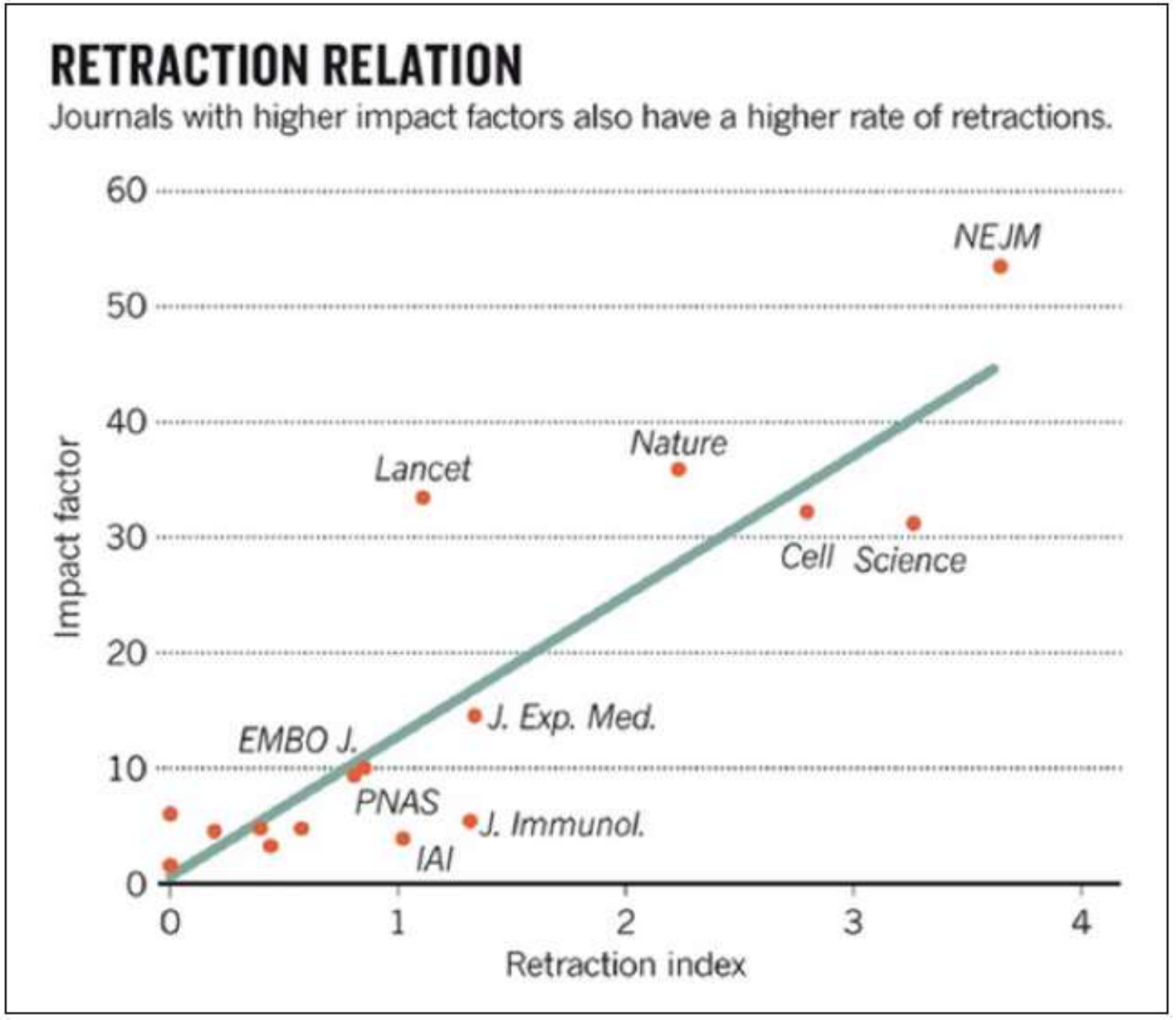
Prestigious Science Journals Struggle to Reach Even Average Reliability

"…an accumulating body of evidence suggests that methodological quality & reliability of published research works in several fields may be decreasing with increasing journal rank" (20 February, 2018)

Cutting corners effect?

Björn Brembs*

Institute of Zoology—Neurogenetics, Universität Regensburg, Regensburg, Germany



Fang FC, Casadevall A and Morrison R (2011) Retracted science and the retraction index. *Infection and Immunity* 79(10): 3855–3859

Retraction Watch

April 20, 2017

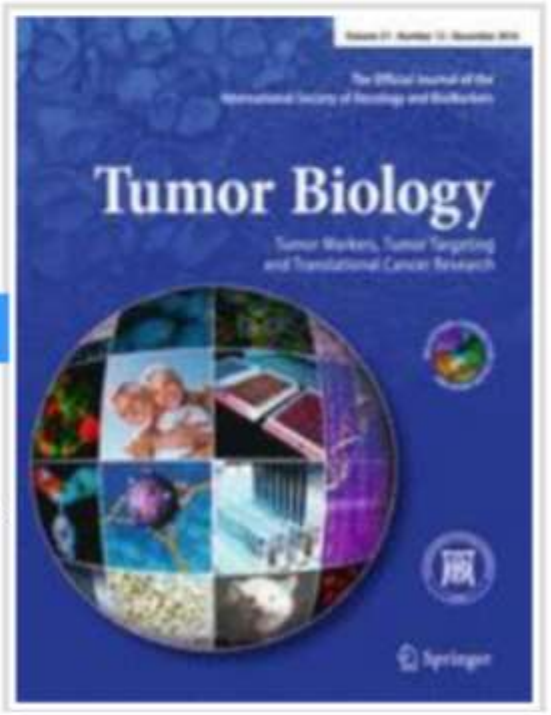
Tracking retractions as a

A new record: Major publisher retracting more than 100 studies from cancer journal over fake peer reviews

with 11 comments

Springer is retracting 107 papers from one journal after discovering they had been accepted with fake peer reviews. Yes, 107.

To submit a fake review, someone (often the author of a paper) either makes up an outside expert to review the paper, or suggests a real researcher — and in both cases, provides a fake email address that comes back to someone who will invariably give the paper a glowing review. In this case, Springer, the publisher of *Tumor Biology* through 2016, told us that an investigation produced "clear evidence" the reviews were submitted under the names of real researchers with faked emails. Some of the authors may have used a third-party editing service, which may have supplied the reviews. The journal is now published by SAGE.



Unintended effects of reforms

Good intentions going bad

TABLE 1. GROWING PERVERSE INCENTIVES IN ACADEMIA

Incentive	*Intended effect*	*Actual effect*
"Researchers rewarded for increased number of publications."	"Improve research productivity," provide a means of evaluating performance.	"Avalanche of" substandard, "incremental papers"; poor methods and increase in false discovery rates leading to a "natural selection of bad science" (Smaldino and Mcelreath, 2016); reduced quality of peer review
"Researchers rewarded for increased number of citations."	Reward quality work that influences others.	Extended reference lists to inflate citations; reviewers request citation of their work through peer review
"Researchers rewarded for increased grant funding."	"Ensure that research programs are funded, promote growth, generate overhead."	Increased time writing proposals and less time gathering and thinking about data. Overselling positive results and downplay of negative results.
Increase PhD student productivity	Higher school ranking and more prestige of program.	Lower standards and create oversupply of PhDs. Postdocs often required for entry-level academic positions, and PhDs hired for work MS students used to do.
Reduced teaching load for research-active faculty	Necessary to pursue additional competitive grants.	Increased demand for untenured, adjunct faculty to teach classes.
"Teachers rewarded for increased student evaluation scores."	"Improved accountability; ensure customer satisfaction."	Reduced course work, grade inflation.
"Teachers rewarded for increased student test scores."	"Improve teacher effectiveness."	"Teaching to the tests; emphasis on short-term learning."
"Departments rewarded for increasing U.S. News ranking."	"Stronger departments."	Extensive efforts to reverse engineer, game, and cheat rankings.
"Departments rewarded for increasing numbers of BS, MS, and PhD degrees granted."	"Promote efficiency; stop students from being trapped in degree programs; impress the state legislature."	"Class sizes increase; entrance requirements" decrease; reduce graduation requirements.
"Departments rewarded for increasing student credit/contact hours (SCH)."	"The university's teaching mission is fulfilled."	"SCH-maximization games are played": duplication of classes, competition for service courses.

Modified from Reseh (pers. comm., 2015) with permission

Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hyper-competition, Marc A. Edwards and Siddhartha Roy, ENVIRONMENTAL ENGINEERING SCIENCE, 34(1), 2017

Incentive	*Intended effect*	*Actual effect*
"Researchers rewarded for increased number of publications."	"Improve research productivity," provide a means of evaluating performance.	"Avalanche of" substandard, "incremental papers"; poor methods and increase in false discovery rates leading to a "natural selection of bad science" (Smaldino and Mcelreath, 2016); reduced quality of peer review

Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hyper-competition, Marc A. Edwards and Siddhartha Roy, ENVIRONMENTAL ENGINEERING SCIENCE, 34(1), 2017

Incentive	
"Researchers rewarded for increased number of citations."	*Actual effect*
Intended effect	Extended reference lists to inflate citations; reviewers request citation of their work through peer review
Reward quality work that influences others.	

Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hyper-competition, Marc A. Edwards and Siddhartha Roy, ENVIRONMENTAL ENGINEERING SCIENCE, 34(1), 2017

Incentive	*Actual effect*
"Researchers rewarded for increased grant funding."	Increased time writing proposals and less time gathering and thinking about data. Overselling positive results and downplay of negative results.
Intended effect	
"Ensure that research programs are funded, promote growth, generate overhead."	

Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hyper-competition, Marc A. Edwards and Siddhartha Roy, ENVIRONMENTAL ENGINEERING SCIENCE, 34(1), 2017

Incentive	*Intended effect*	*Actual effect*
Increase PhD student productivity	Higher school ranking and more prestige of program.	Lower standards and create oversupply of PhDs. Postdocs often required for entry-level academic positions, and PhDs hired for work MS students used to do.

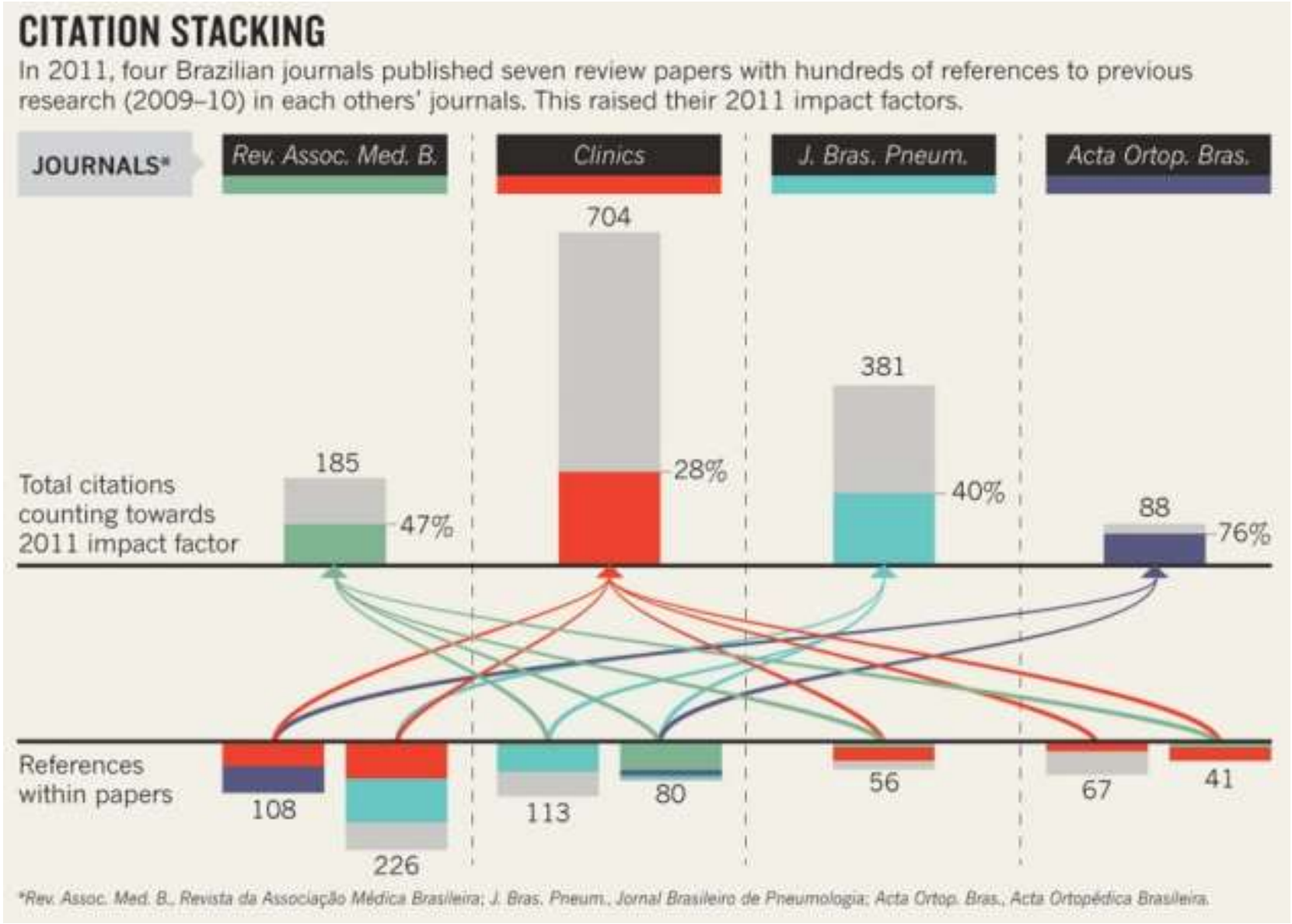
Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hyper-competition, Marc A. Edwards and Siddhartha Roy, ENVIRONMENTAL ENGINEERING SCIENCE, 34(1), 2017

Incentive	*Actual effect*
"Teachers rewarded for increased student test scores."	"Teaching to the tests; emphasis on short-term learning."
"Departments rewarded for increasing U.S. News ranking."	Extensive efforts to reverse engineer, game, and cheat rankings.
Intended effect	
"Improve teacher effectiveness."	
"Stronger departments."	

Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hyper-competition, Marc A. Edwards and Siddhartha Roy, ENVIRONMENTAL ENGINEERING SCIENCE, 34(1), 2017

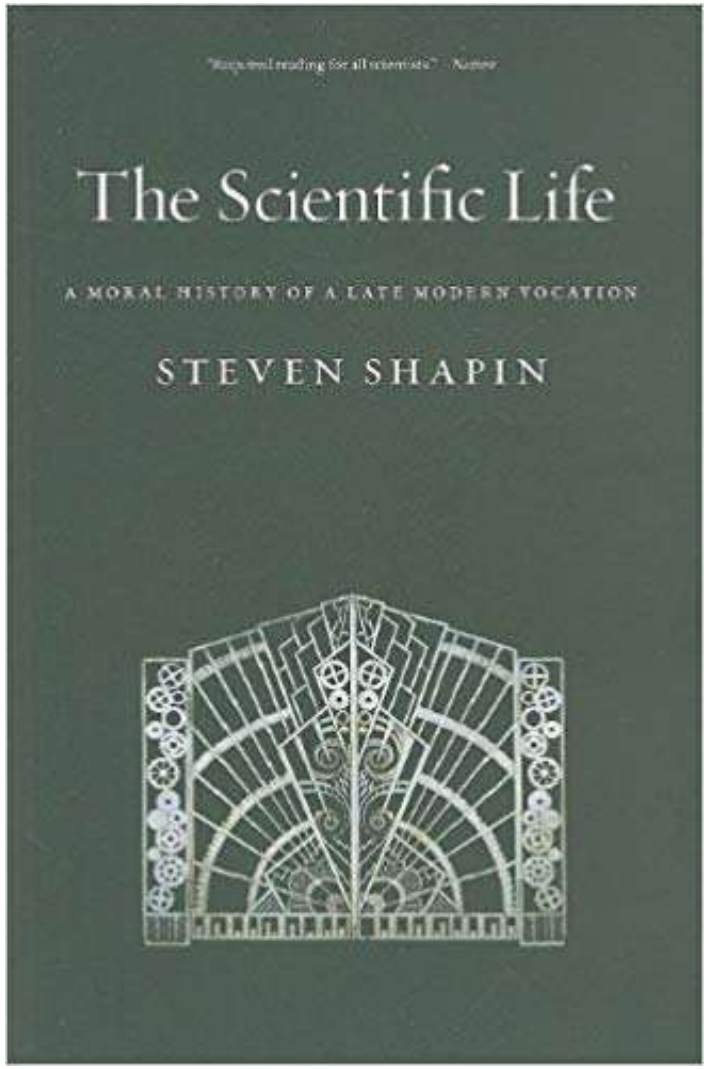
Gaming the system

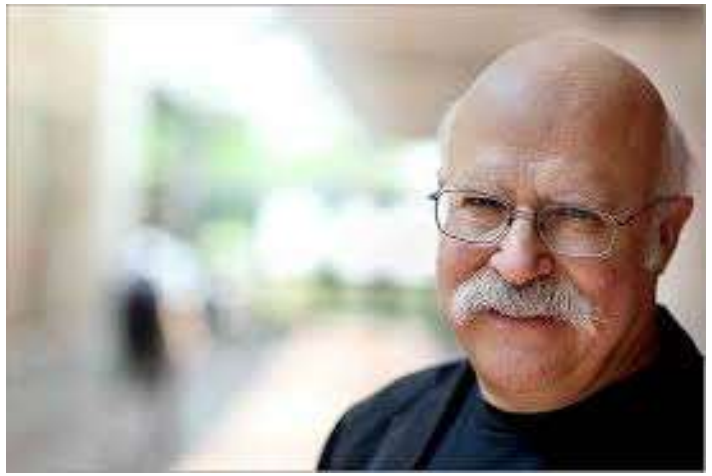
Use and abuse of metrics: from self-citation to citation cartels to citation stacking



Richard Van Noorden, 2017, Brazilian citation scheme outed. Thomson Reuters suspends journals from its rankings for 'citation stacking'. Nature, 27 August 2013

Lost ethos?





Steven Shapin

Is scientists' civility to each other what holds the venture together?

But someone disagrees: J.R. Ravetz, Morals and manners in modern science, Nature, 457(5), 662-663.

Renewable sources 100% of energy in US by 2050, says Jacobson…

…and sues for $10-million a dissenter

Los Angeles Times

A Stanford professor drops his ridiculous defamation lawsuit against his scientific critics

http://www.latimes.com/business/hiltzik/la-fi-hiltzik-jacobson-lawsuit-20180223-story.html

Climate change scepticism

World court should rule on climate science to quash sceptics, says Philippe Sands

International Court of Justice ruling would settle the scientific dispute and pave the way for future legal cases on climate change, says high-profile lawyer

Adam Vaughan

@adamvaughan_uk

Friday 18 September 2015 10.34 BST

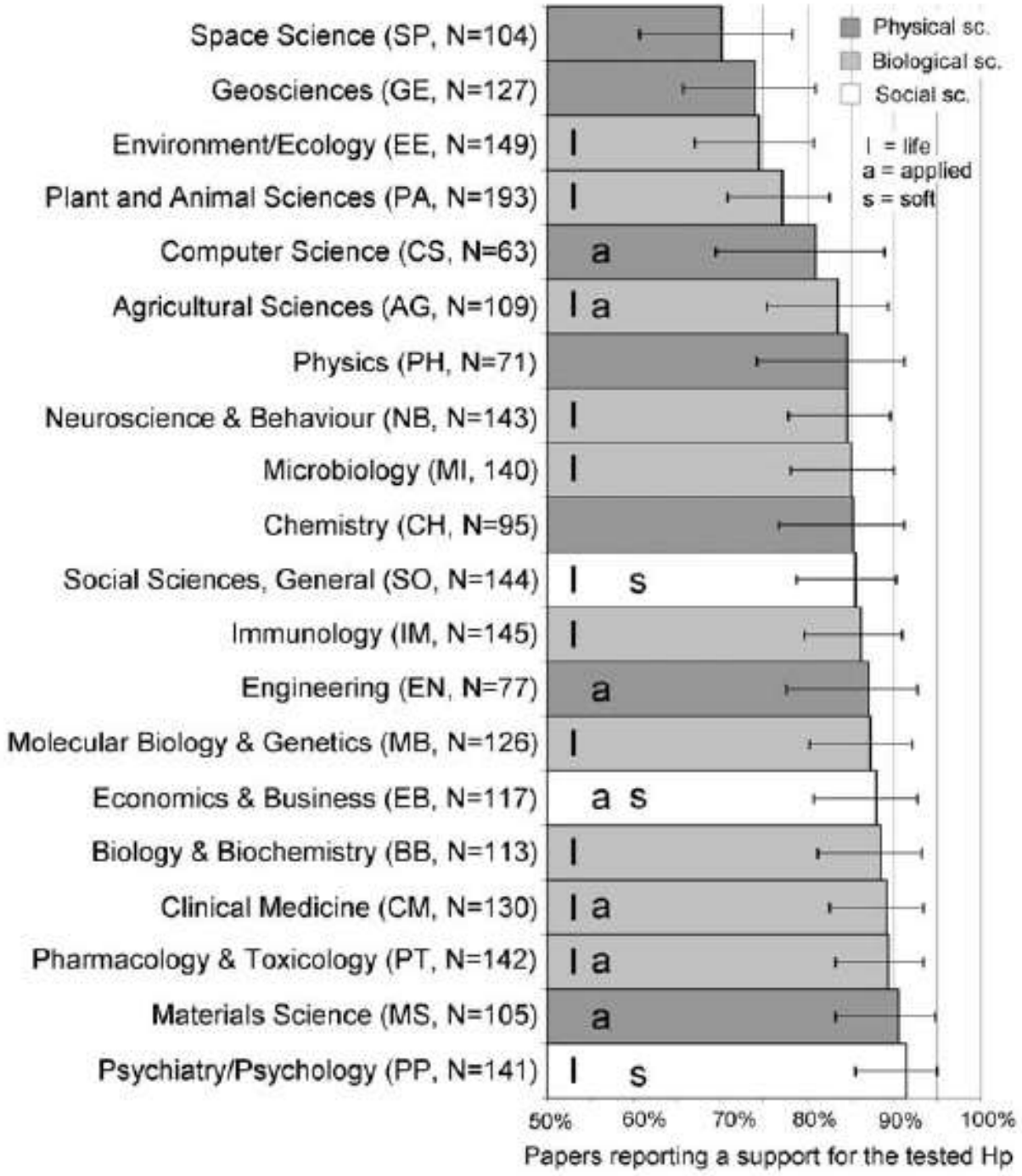
Shares 1,805

Comments 710

Save for later

Philippe Sands QC says a court ruling would carry more weight with public opinion than science alone. Photograph: Antonio Zazueta Olmos/Antonio Olmos

Not all disciplines the same



OPEN ACCESS Freely available online

PLoS one

"Positive" Results Increase Down the Hierarchy of the Sciences

Daniele Fanelli*

INNOGEN and ISSTI-Institute for the Study of Science, Technology & Innovation, The University of Edinburgh, Edinburgh, United Kingdom

"odds of reporting a positive result ~5 times higher among papers in the disciplines of Psychology and Psychiatry and Economics and Business than Space Science"

April 7, 2010

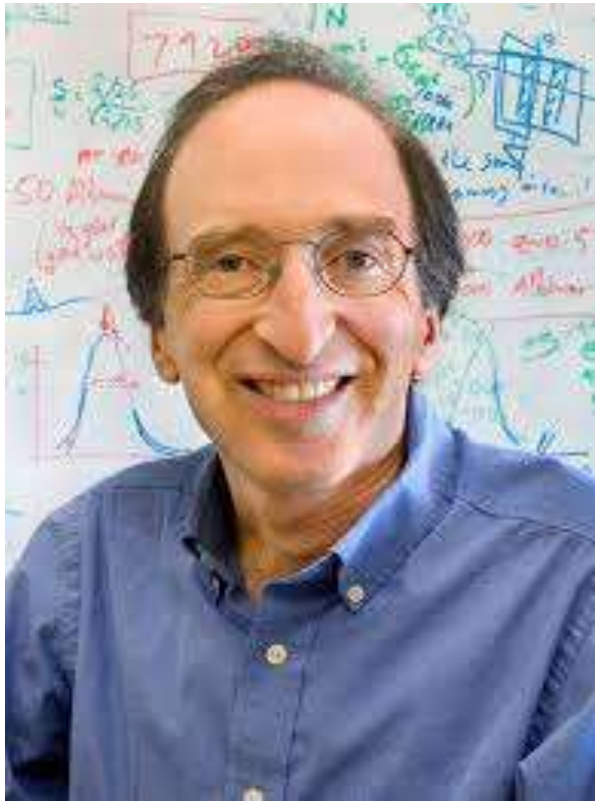
Physics as a model:

Following several high-profile errors, the particle physics community now invests great effort into intensive checking and re-checking of data prior to publication. By filtering results through independent working groups, physicists are encouraged to criticise.

R. Horton, “Offline: What is medicine’s 5 sigma?,” **Lancet**, vol. 385, p. 1380, 2015.

Saul Perlmutter, an astrophysicist at the University of California, Berkeley.

"Science is an ongoing race between our inventing ways to fool ourselves, and our inventing ways to avoid fooling ourselves.

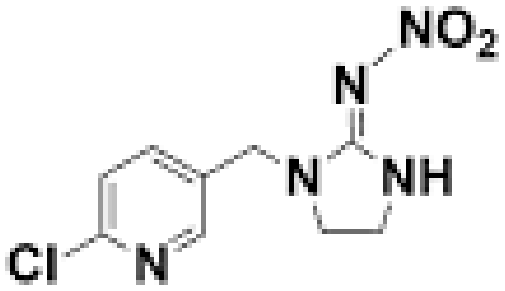


Saul Perlmutter

R. Nuzzo, "How scientists fool themselves – and how they can stop," Nature, vol. 526, no. 7572, pp. 182–185, Oct. 2015.

More ethical doubts: what is science? who is a scientist?

Can science be shoddy, entrepreneurial, reckless, or dirty?



Ravetz, J., **1971**, Scientific Knowledge and its Social Problems, Oxford University Press.

Who is a scientist?

Mark Edwards,
Aleksandr Kogan

Paolo Macchiarini,
Rick Mishkin

Does history repeat itself?

(Love canal, Flint…)

Lois Gibbs

Marc Edwards

http://www.andreasaltelli.eu/file/repository/LOVE_CANAL.pdf
https://en.wikipedia.org/wiki/Flint_water_crisis; http://flintwaterstudy.org/;
http://www.nytimes.com/2016/08/21/magazine/flints-water-crisis-and-the-troublemaker-scientist.html

Different cultures, different reactions

Yoshiki Sasai 1962 – 2014

http://www.nature.com/news/stem-cell-pioneer-blamed-media-bashing-in-suicide-note-1.15715

Different cultures, different reactions

Aaron Swartz, 1986 - 2013

https://www.rollingstone.com/culture/news/the-brilliant-life-and-tragic-death-of-aaron-swartz-20130215

The End

@andreasaltelli

MBI: Magnitude-based inference: persistent bad stats in sports research

MBI false positive rate two to six time higher than in NHST (Null hypothesis significance testing)

Christie Aschwanden and Mai Nguyen, How Shoddy Statistics Found A Home In Sports Research, Fivethirtyeight, May 16, 2018, https://fivethirtyeight.com/features/how-shoddy-statistics-found-a-home-in-sports-research/

K. L. Sainani, The Problem with 'Magnitude-Based Inference,' Medicine & Science in Sports & Exercise (MSSE), p. 1, Apr. 2018.