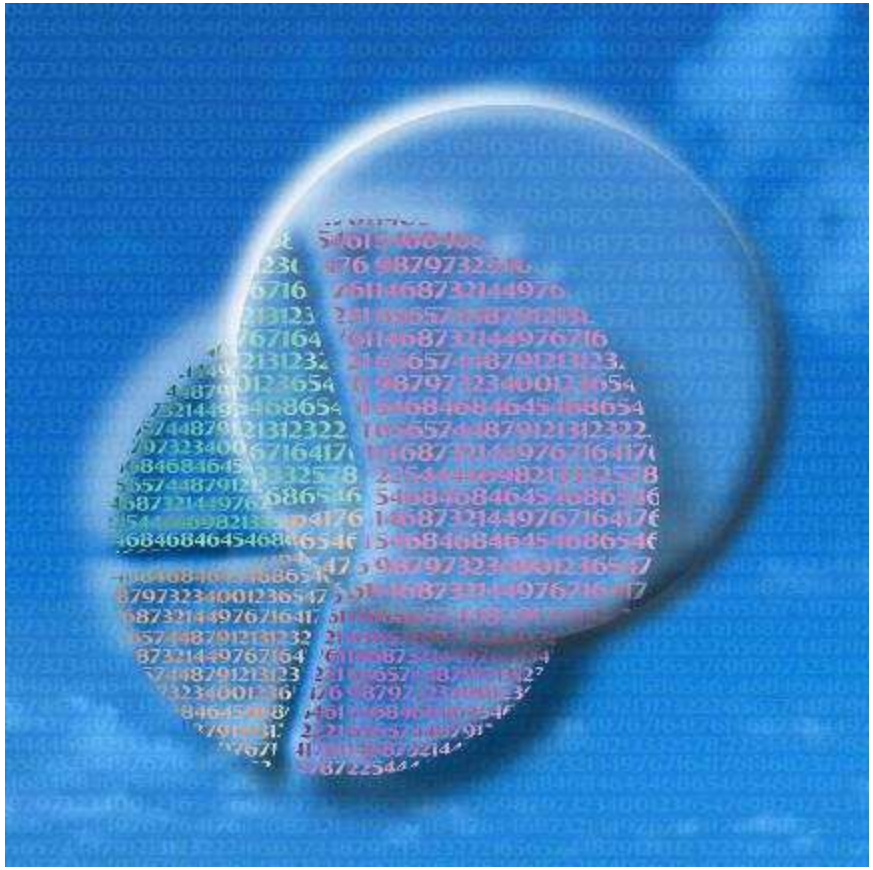


Monday 13: The p-test imbroglio and science's crisis

Andrea Saltelli
Centre for the Study of the Sciences and the Humanities (SVT) - University of Bergen (UIB)
Institut de Ciència i Tecnologia Ambientals (ICTA) -Universitat Autonoma de Barcelona (UAB)

PhD-course

Numbers for policy: Practical problems in quantification

Bergen, March 13-17, 2017

Where to find this talk: www.andreasaltelli.eu

sensitivity analysis, sensitivity auditing, science for policy, impact assessment

= more material on my web site

= discussion time

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

"If you are foolish enough to define 'statistically significant' as anything less than p=0.05 then… you have a 29% chance (at least) of making a fool of yourself.

Who would take a risk like that? Judging by the medical literature, most people would. No wonder there is a problem"

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

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?

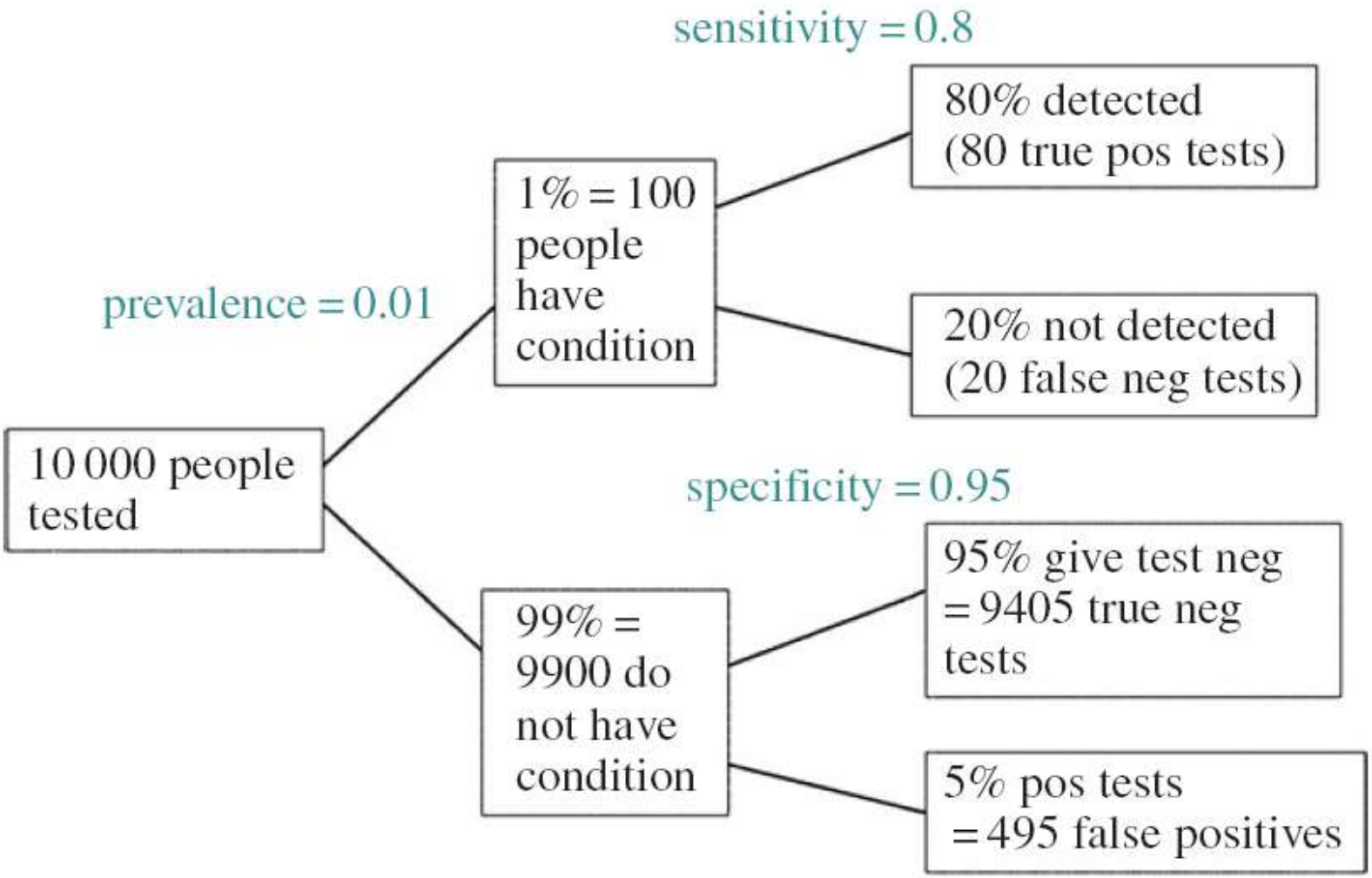
MCI prevalence in the population 1%, i.e. in a sample of 10,000 then 100 have MCI and 9,900 don't

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!

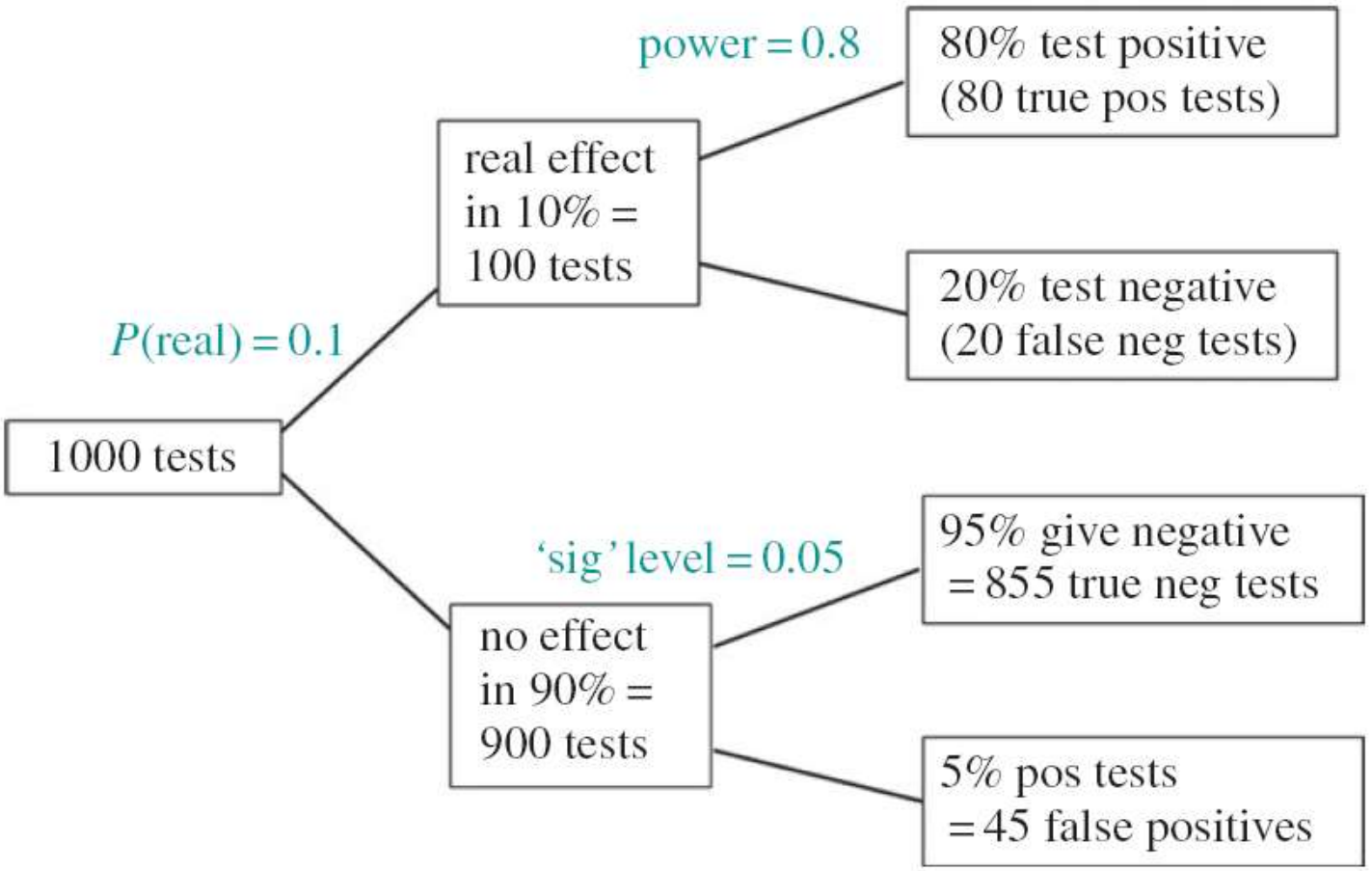
The number 86% = 495/(495+80) is our false discovery rate



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

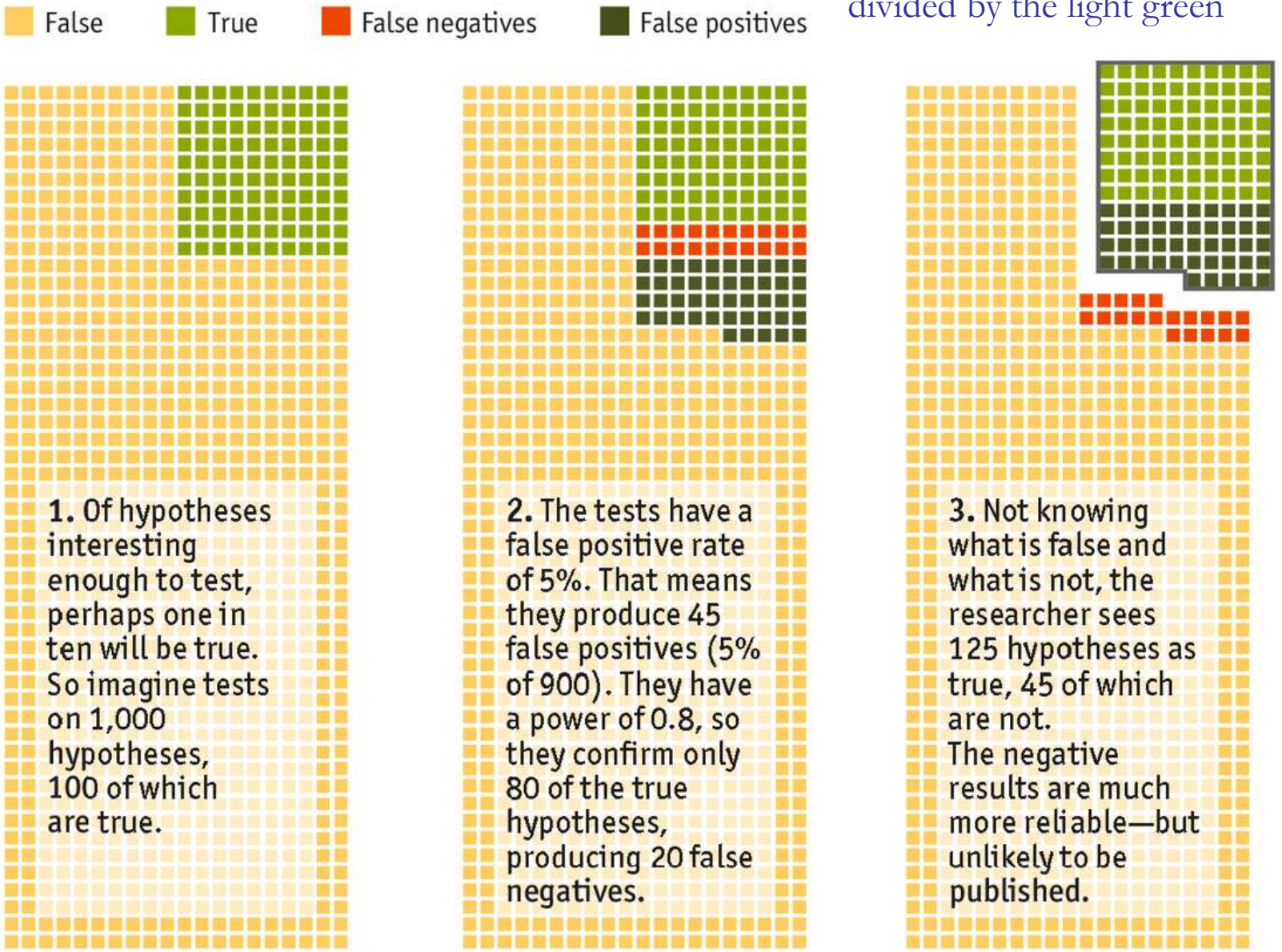
We now consider tests instead of individuals



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

Unlikely results

How a small proportion of false positives can prove very misleading



Source: *The Economist*

The false discovery rate is ~the dark divided by the light green

➔ 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?

Unreliable research

Trouble at the lab

Scientists like to think of science as self-correcting. To an alarming degree, it is not

Oct 19th 2013 | From the print edition

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance.

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

IN FOCUS NEWS

REPRODUCIBILITY

Statisticians issue warning on *P* values

Statement aims to halt missteps in the quest for certainty.

"Misuse of the P value — a common test for judging the strength of scientific evidence — is contributing to the number of research findings that cannot be reproduced"

Baker, M., 2016, Statisticians issue warning on P values, Nature, 531, 151

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

… and 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/

Special Issue:
Bayesian Probability and Statistics in Management Research

Journal of Management
Vol. 41 No. 2, February 2015 421–440
DOI: 10.1177/0149206314547522

Editorial Commentary

Surrogate Science: The Idol of a Universal Method for Scientific Inference

Gerd Gigerenzer
Max Planck Institute for Human Development
Julian N. Marewski
University of Lausanne

There is no universal method of scientific inference …

…it is better to have no beliefs than to embrace falsehoods…

Statistical methods are not simply applied to a discipline; they change the discipline itself, …

Special Issue:
Bayesian Probability and Statistics in Management Research

Journal of Management
Vol. 41 No. 2, February 2015 421–440
DOI: 10.1177/0149206314547522

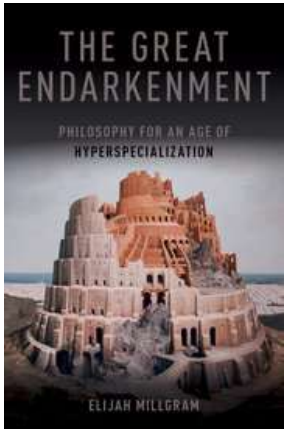
Editorial Commentary

Surrogate Science: The Idol of a Universal Method for Scientific Inference

Gerd Gigerenzer
Max Planck Institute for Human Development
Julian N. Marewski
University of Lausanne

How was it possible that this important statistical tool was misused for several decades with grave consequences for science?

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)

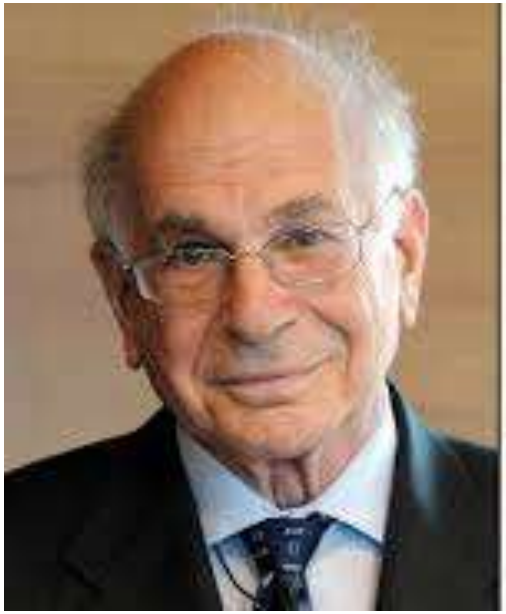
Replicability-Index
Improving the replicability of empirical research

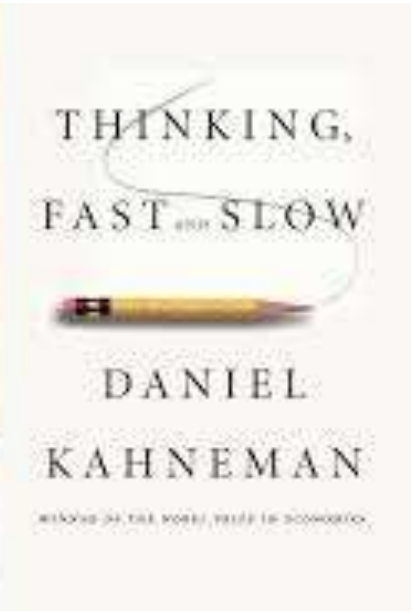
Reconstruction of a Train Wreck: How Priming Research Went off the Rails

February 2, 2017 Kahneman, Priming, r-index, Statistical Power, Thinking Fast and Slow

Authors: Ulrich Schimmack, Moritz Heene, and Kamini Kesavan

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

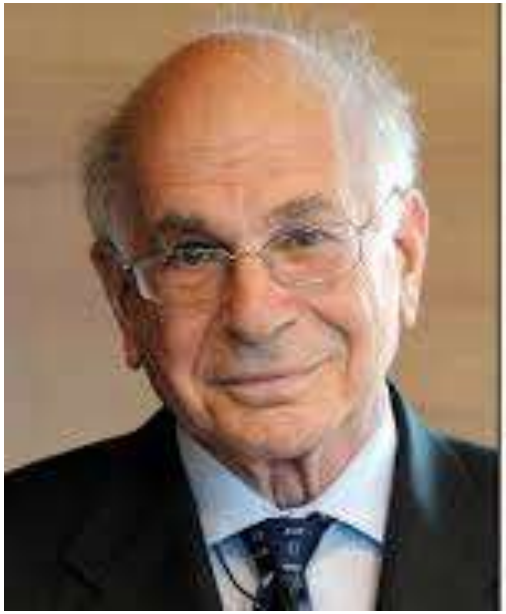


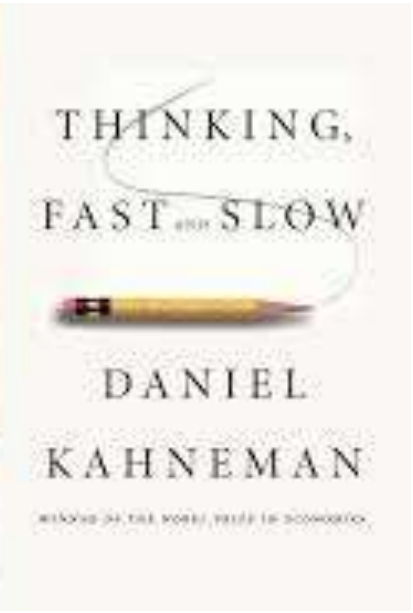


"[…]questions have been raised about the robustness of priming results … your field is now the poster child for doubts about the integrity of psychological research…

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

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

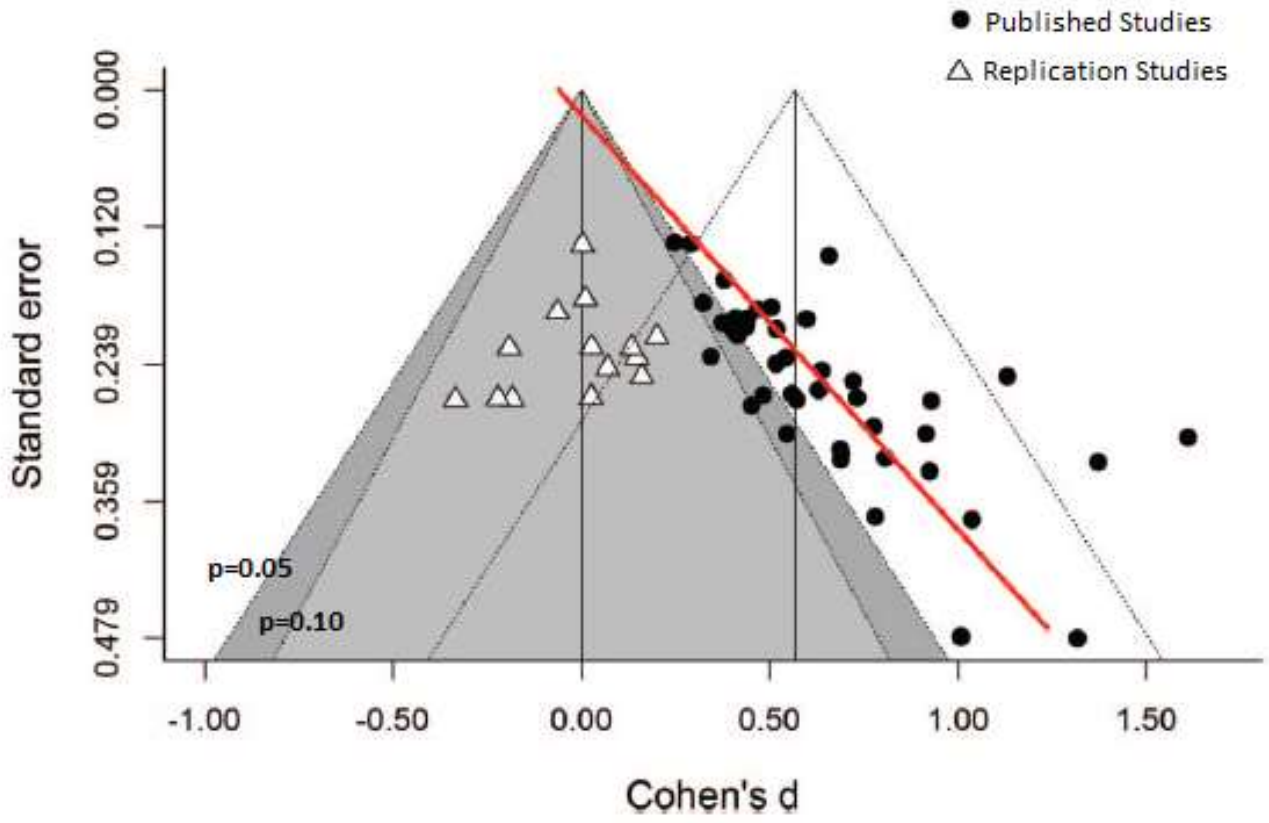




"… people have now attached a question mark to the field, and it is your responsibility to remove it… I recently wrote a book that emphasizes priming research … My reason for writing this letter is that I see a train wreck looming" (Kahneman, 2012)

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

P-hacking; a smoking gun?



Shanks et al. (2015) JEP:General

J Exp Psychol Gen. 2015 Oct 26. "Romance, Risk, and Replication: Can Consumer Choices and Risk-Taking Be Primed by Mating Motives?", Shanks DR, Vadillo MA, Riedel B, Clymo A, Govind S, Hickin N, Tamman AJ, Puhlmann LM.: http://www.ncbi.nlm.nih.gov/pubmed/26501730

LEADER 13 April 2016

Science isn't as solid as it should be – but science can fix it

Unconscious biases and data-torturing are weakening our knowledge base – but unlike politicians and bankers, scientists aren't covering up their failings

Crisis? Fix it...
Stanislav Chernivchan/EyeEm

FEATURE 13 April 2016

Why so much science research is flawed – and what to do about it

Dodgy results are fuelling flawed policy decisions and undermining medical advances. They could even make us lose faith in science. **New Scientist** investigates

An alarming amount of research is flawed
Brett Ryder

New Scientists talks of "statistical sausage factory"

Just it just about statistics?

What if even she is wrong?

On TV series over series where lab-based forensics (science) adjudicates cases

Forensics [as well as medicine, biology, economics, health, nutrition …] has produced serious misdiagnoses

National Academy of Sciences (NAS) report "Strengthening Forensic Science in the United States: A Path Forward", https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf

THE CONVERSATION

Academic rigour, journalistic flair

Arts + Culture | Business + Economy | Education | Environment + Energy | Health + Medicine | Politics + Society | Science + Technology

Science in crisis: from the sugar scam to Brexit, our faith in experts is fading

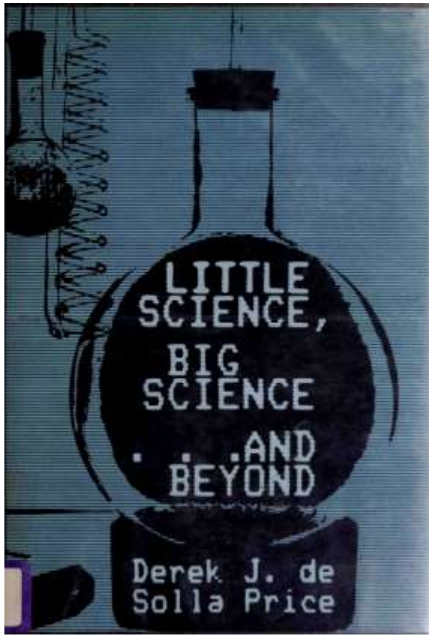
September 27, 2016 4.43pm AEST

https://theconversation.com/science-in-crisis-from-the-sugar-scam-to-brexit-our-faith-in-experts-is-fading-65016

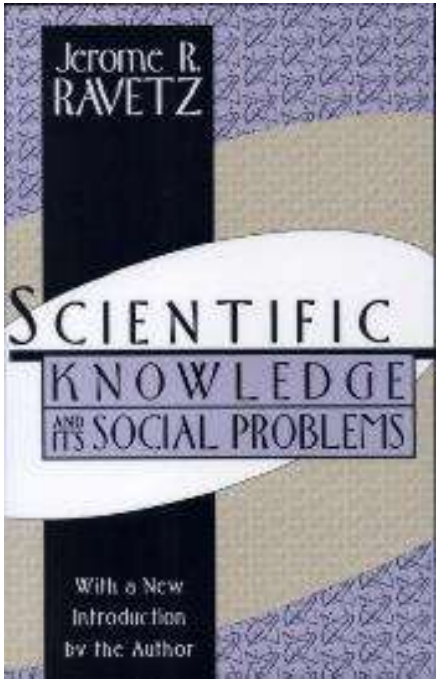
Context: A matter of opinion

1. Science is in a deep existential crisis which has ethical, epistemological, methodological and even metaphysical dimensions
2. Likewise democracy which has with science a legitimacy arrangement
3. Science and its institutions are committed to the status quo & attempt to evade a critical reflection
4. Solutions aren't forthcoming anytime soon
5. There are yet few areas of 'Reformation' where science and society work together

First thesis: Science is in a deep existential crisis which has ethical, epistemological, methodological and even metaphysical dimensions. This was neatly predicted by E. de Solla Price, Jerome R. Ravetz and others five decades ago



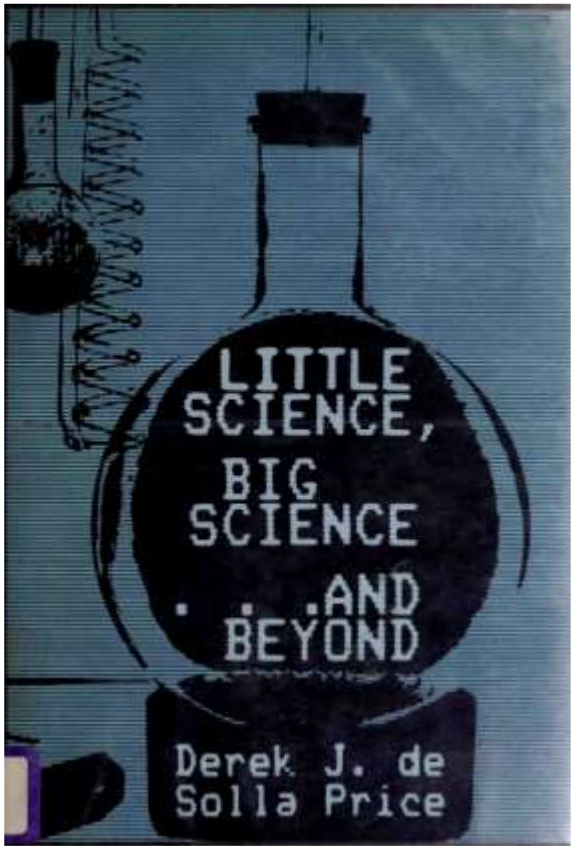
Derek J. de Solla Price



Jerome R. Ravetz

de Solla Price, D.J., 1963, Little science big science, Columbia University Press.
Ravetz, J., 1971, Scientific Knowledge and its Social Problems, Oxford University Press.

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.

Science/knowledge degenerates when it becomes a commodity for Ravetz (1971), Lyotard (1979) and Mirowski (2011).

Ravetz, J., 1971, Scientific Knowledge and its Social Problems, Oxford University Press, p. 22.

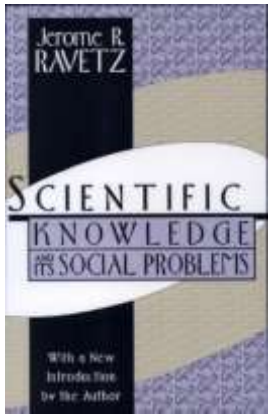
Lyotard, J.-F. 1979. La Condition postmoderne. Rapport sur le savoir, Paris : Minuit, Chapter 10.

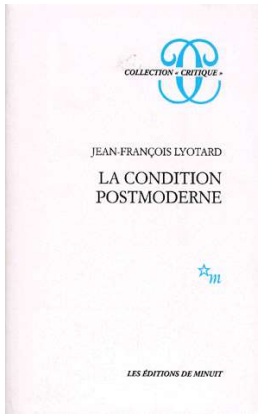
Mirowski, P. 2011. Science-Mart: Privatizing American Science, Harvard University Press.

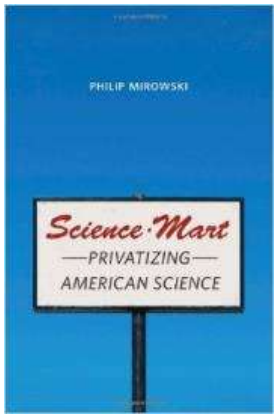
Jerome R. Ravetz

Jean-François Lyotard

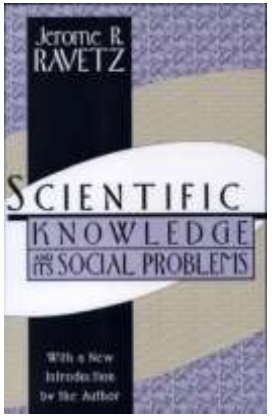
Philip Mirowski







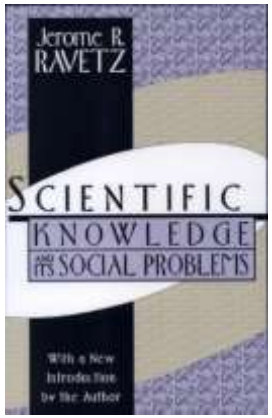
p.22: “with the industrialization of science, certain changes have occurred which weaken the operation of the traditional mechanism of quality control and direction at the highest level.



Jerome R. Ravetz

Ravetz, J., 1971, Scientific Knowledge and its Social Problems, Oxford University Press, p.22.

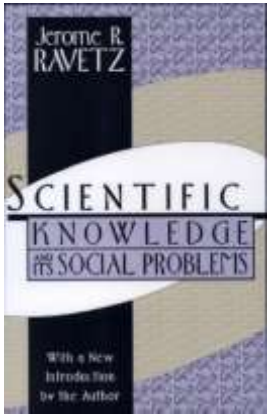
p.22: […] The problem of quality control in science is thus at the centre of the social problems of the industrialized science of the present period."



Jerome R. Ravetz

Ravetz, J., 1971, Scientific Knowledge and its Social Problems, Oxford University Press, p.22.

p.22: “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”

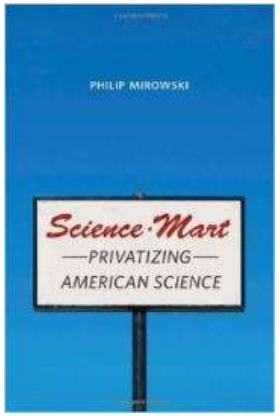


Jerome R. Ravetz

Ravetz, J., 1971, Scientific Knowledge and its Social Problems, Oxford University Press, p.22.

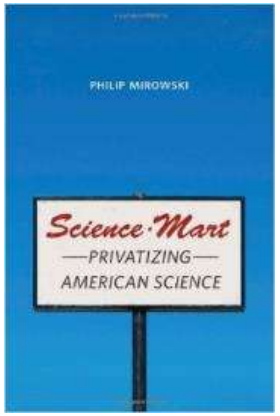
After the eighties neoliberal ideologies succeeded in decreasing state intervention in the funding of science, which became increasingly privatized … Knowledge as a monetized commodity replaces knowledge as public good…

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



Philip Mirowski

In house science labs of major corporation were closed and research outsourced to universities which … became more and more looking as profit seeking organization (technology transfer offices in every campus) … then research ended up outsourced again to contract-based research organizations (CRO's)…



Philip Mirowski

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

Summary Points

- Currently, many published research findings are false or exaggerated, and an estimated 85% of research resources are wasted.

Ioannidis, J. P. (2014). How to Make More Published Research True. PLoS medicine, 11(10), e1001747

John P. A. Ioannides

For Lancet (2015) an estimated US$200 billion were wasted in the US in 2010.

Lancet, Editorial, 2015, Rewarding true inquiry and diligence in research, 385, p. 2121.

Ioannidis JPA, 2016, Why Most Clinical Research Is Not Useful, PLoS Med 13(6): e1002049. doi:10.1371/journal.pmed.1002049

nature International weekly journal of science

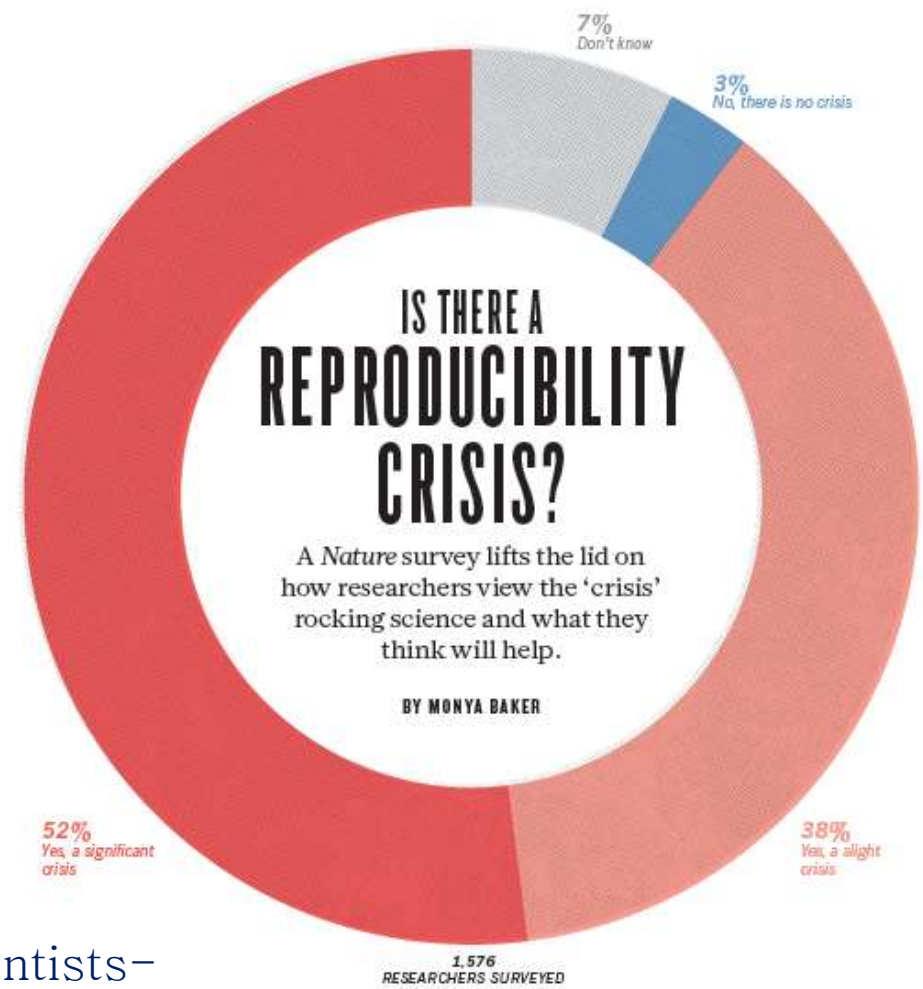
NATURE | NEWS FEATURE

1,500 scientists lift the lid on reproducibi

Survey sheds light on the 'crisis' rocking research.

Monya Baker

25 May 2016 | Corrected: 28 July 2016



http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970

nature International weekly journal of science

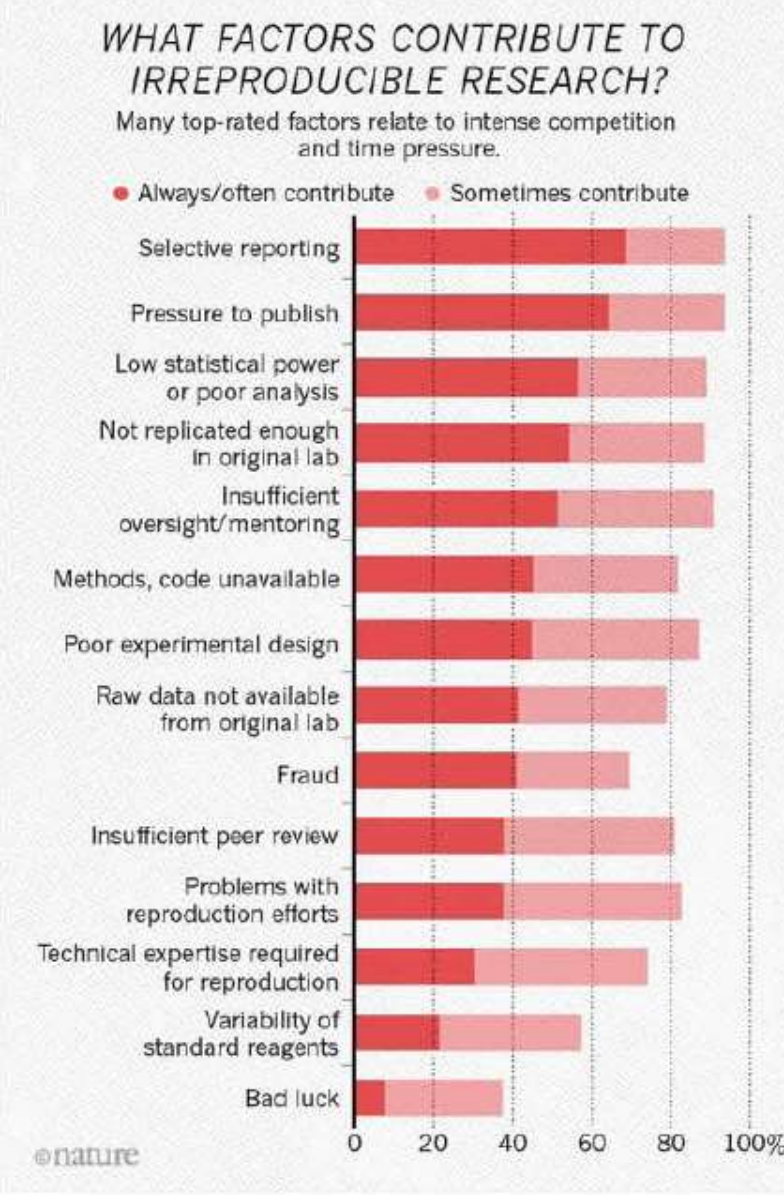
NATURE | NEWS FEATURE

1,500 scientists lift the lid on reproducibility

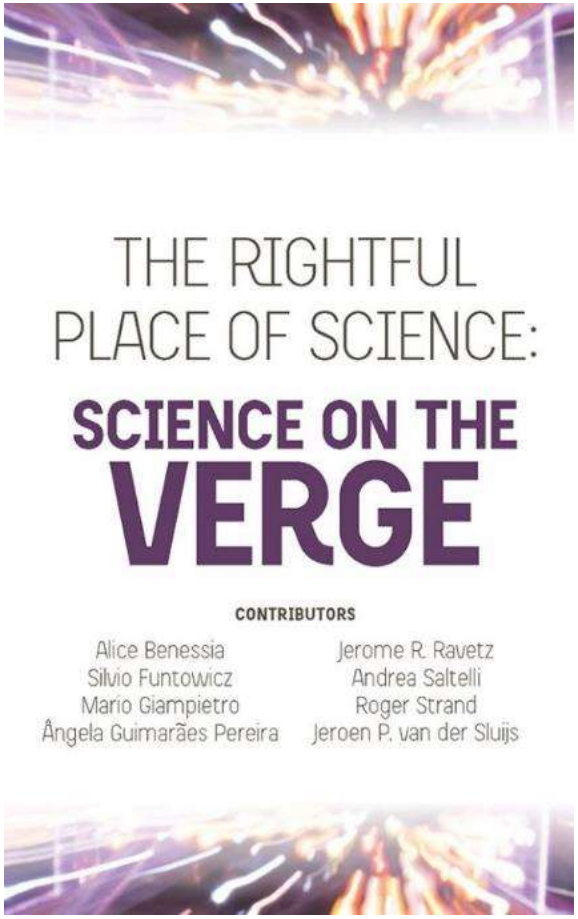
Survey sheds light on the 'crisis' rocking research.

Monya Baker

25 May 2016 | Corrected: 28 July 2016



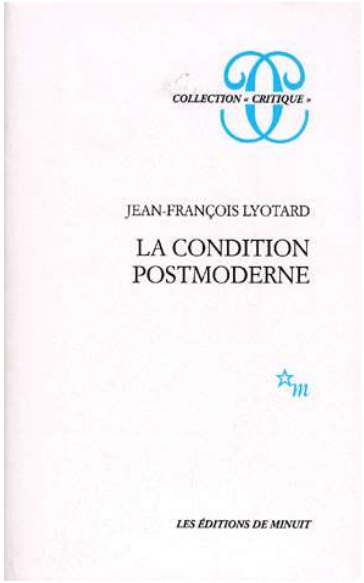
http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970



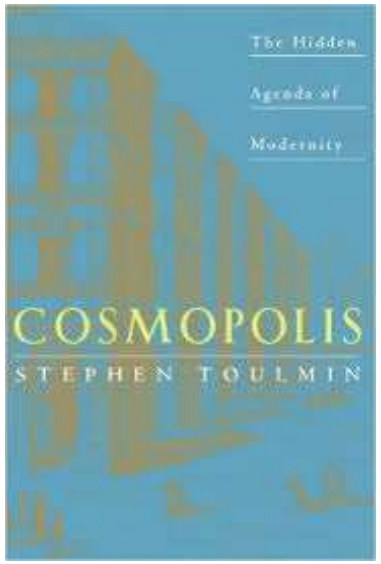
A crisis looms over the scientific enterprise. Not a day passes without news of retractions, failed replications, fraudulent peer reviews, or misinformed science-based policies

Thesis 2: Likewise in crisis is democracy which has with science a legitimacy arrangement

Michael Polanyi



Jean-François Lyotard



Stephen Toulmin

➔ today post-BREXIT, post-Trump, post-truth brouhaha, the demise of expertise …

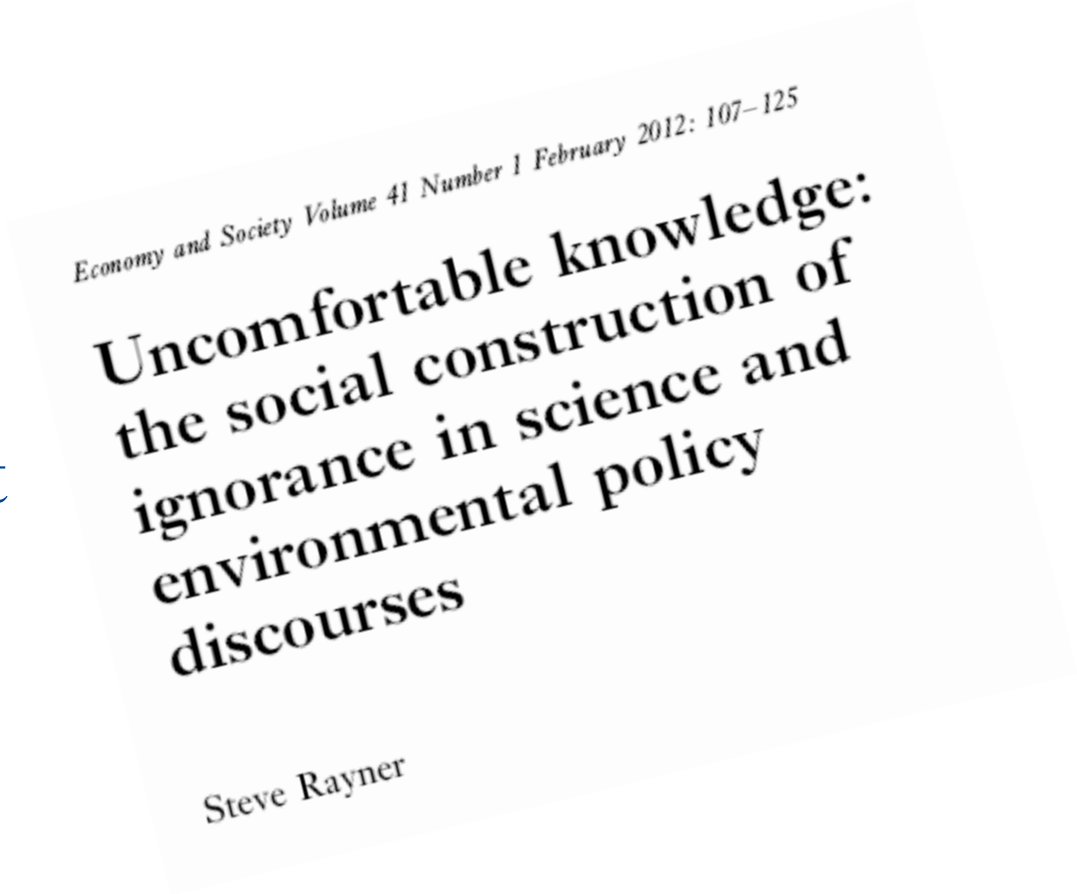
Thesis 3: Science and its institutions are committed to the status quo & attempt to evade a critical reflection with:

Denial

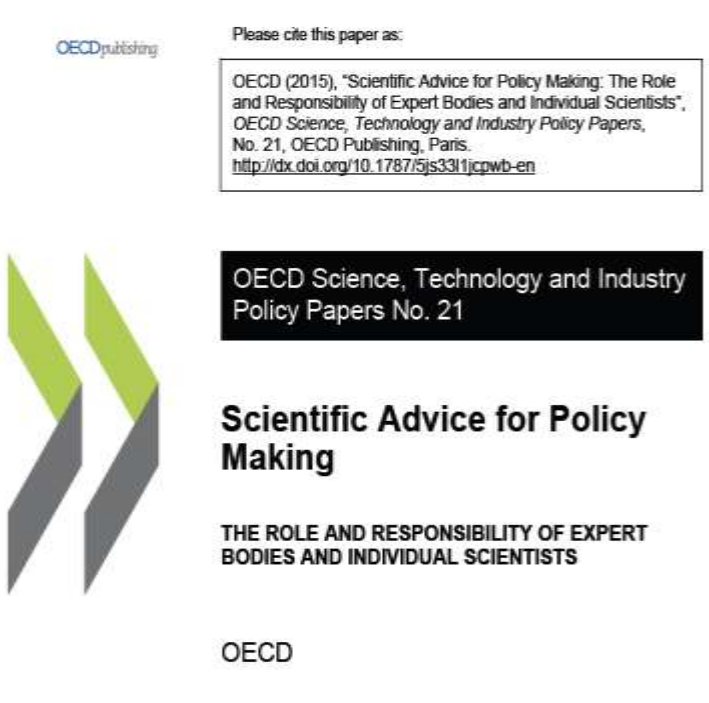
Dismissal

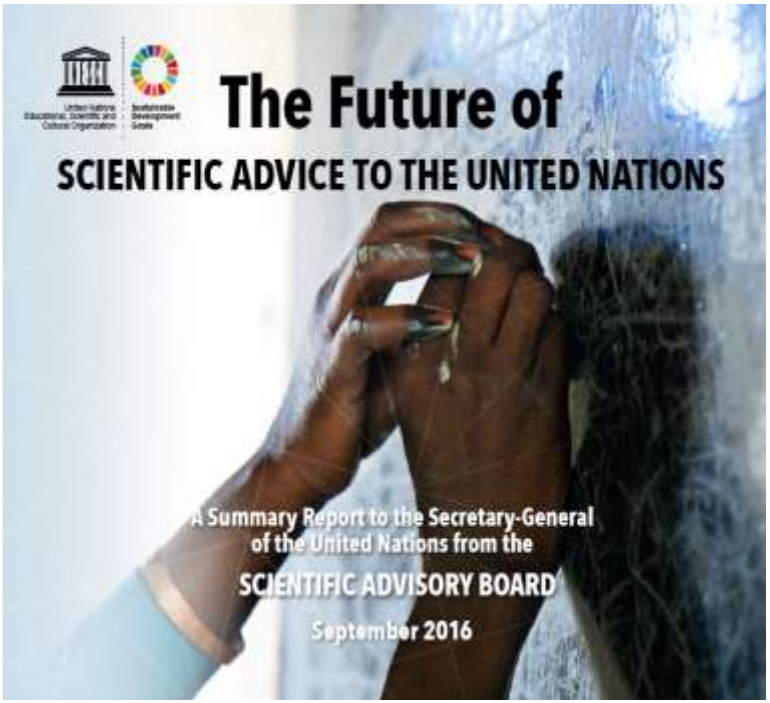
Diversion

Displacement



Denial





2015

2016

We can solve it!

nature.com > nature human behaviour > perspectives > article

MENU

nature human behaviour

Altmetric: 1,978 Views: 40,227 More detail »

Perspective | OPEN

A manifesto for reproducible science

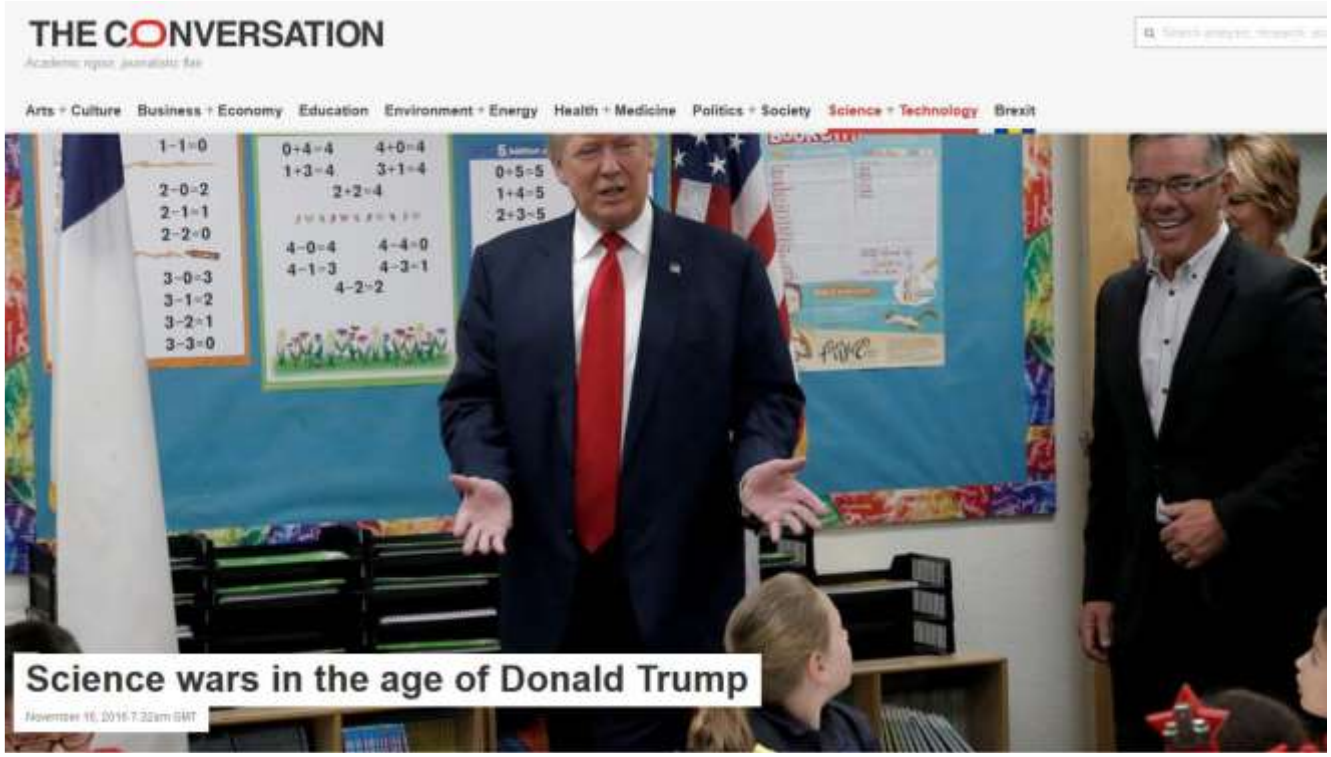
Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware & John P. A. Ioannidis

Nature Human Behaviour **1**, Article number: 0021 (2017) doi:10.1038/s41562-016-0021

Published online: 10 January 2017

"[…] measures [to] improving the transparency, reproducibility and efficiency of scientific research"

Diversion (There is a problem, and this is due to an ongoing war on science between the educated liberal left and the ignorant conservative right)



Authors

Andrea Saltelli
Adjunct professor, University of Bergen

Silvio Oscar Funtowicz
Adjunct Professor Centre for the Study of the Sciences and the Humanities, University of Bergen

https://theconversation.com/science-wars-in-the-age-of-donald-trump-67594

Displacement (This is the post-truth era)

THE CONVERSATION

Academic rigour, journalistic flair

Arts + Culture | Business + Economy | Education | Environment + Energy | Health + Medicine | Politics + Society | Science + Technology | Brexit

To tackle the post-truth world, science must reform itself

January 27, 2017 7.33am GMT

Authors

Andrea Saltelli
Adjunct professor, University of Bergen

Silvio Oscar Funtowicz
Adjunct Professor Centre for the Study of the Sciences and the Humanities, University of Bergen

https://theconversation.com/to-tackle-the-post-truth-world-science-must-reform-itself-70455

Thesis 4: Solutions aren't forthcoming anytime soon, but:

Church / indulgencies =

Science / [predatory publishers, citation cartels, trade in authorship, sugar-cholesterol scandal …]

Martin Luther

Johann Tetzel

A combination of corruption, rage and new technology could mobilise major social change (Silvio Funtowicz)

Martin Luther

Johann Tetzel

Thesis 5: Areas of resistance and 'Reformation' where science and society work together – emergence of a new polity of science, including citizen scientists and scientist-citizens

Jeffrey Beall

Lois Gibbs

Timothy Gowers

Marc Edwards

http://scholarlyoa.com/2015/01/02/bealls-list-of-predatory-publishers-2015/#more-4719
https://www.bu.edu/lovecanal/canal/
http://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0127502
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

An even newer sort of heroes?

John and Laura Arnold

Brian Nosek, the Reproducibility Project.

John Ioannidis, Meta-research innovation centre at Stanford

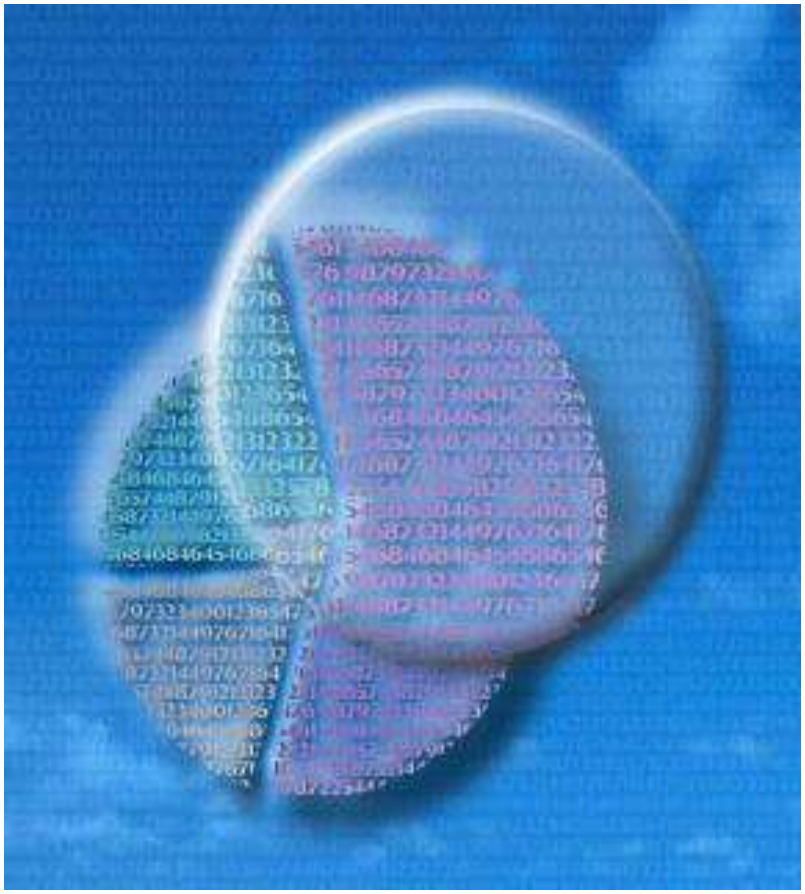
Ben Goldacre, alltrials.net

Gary Taubes, The case against sugar

https://www.wired.com/2017/01/john-arnold-waging-war-on-bad-science/

Yoshiki Sasai

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



END

Twitter:
@andreasaltelli