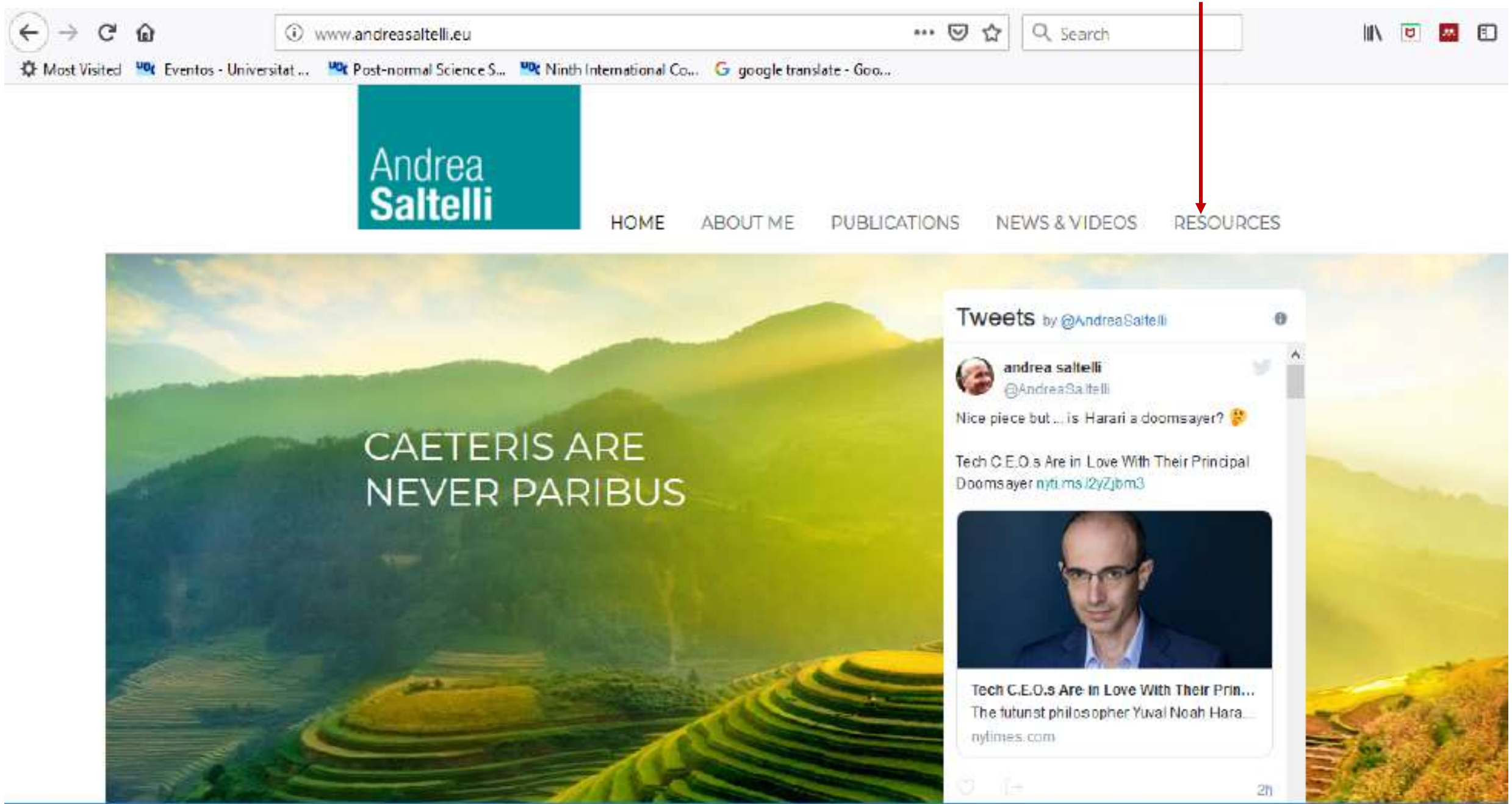
Expertise, participation, and evidence at an age of crisis

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

COST Course “Mobilising data for energy poverty research and action”, UAB, ICTA, room z022-023

Where to find this talk: www.andreasaltelli.eu



= more material on my web site

The topic of this first part

From the misuse of a statistical technique to a problem in reproducibility in science; from this to an overall crisis of expertise, scientific evidence, practice and ethos.

What about evidence based policy?

Numbers and trust

Frames

The P-test saga

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?

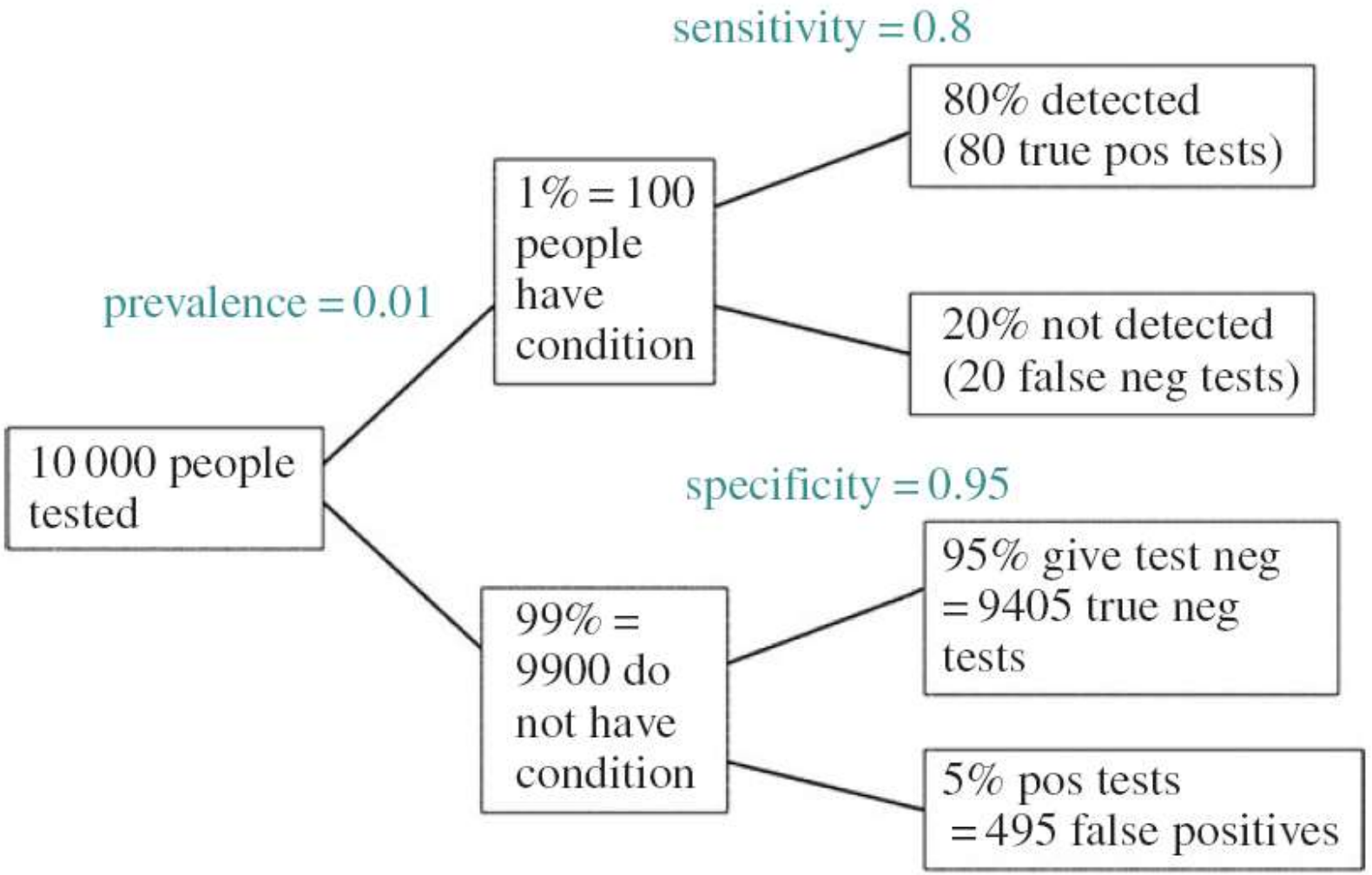
MCI prevalence in the population 1%, i.e. in a sample of 10,000 then 100 have MCI and 9,900 don't

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!

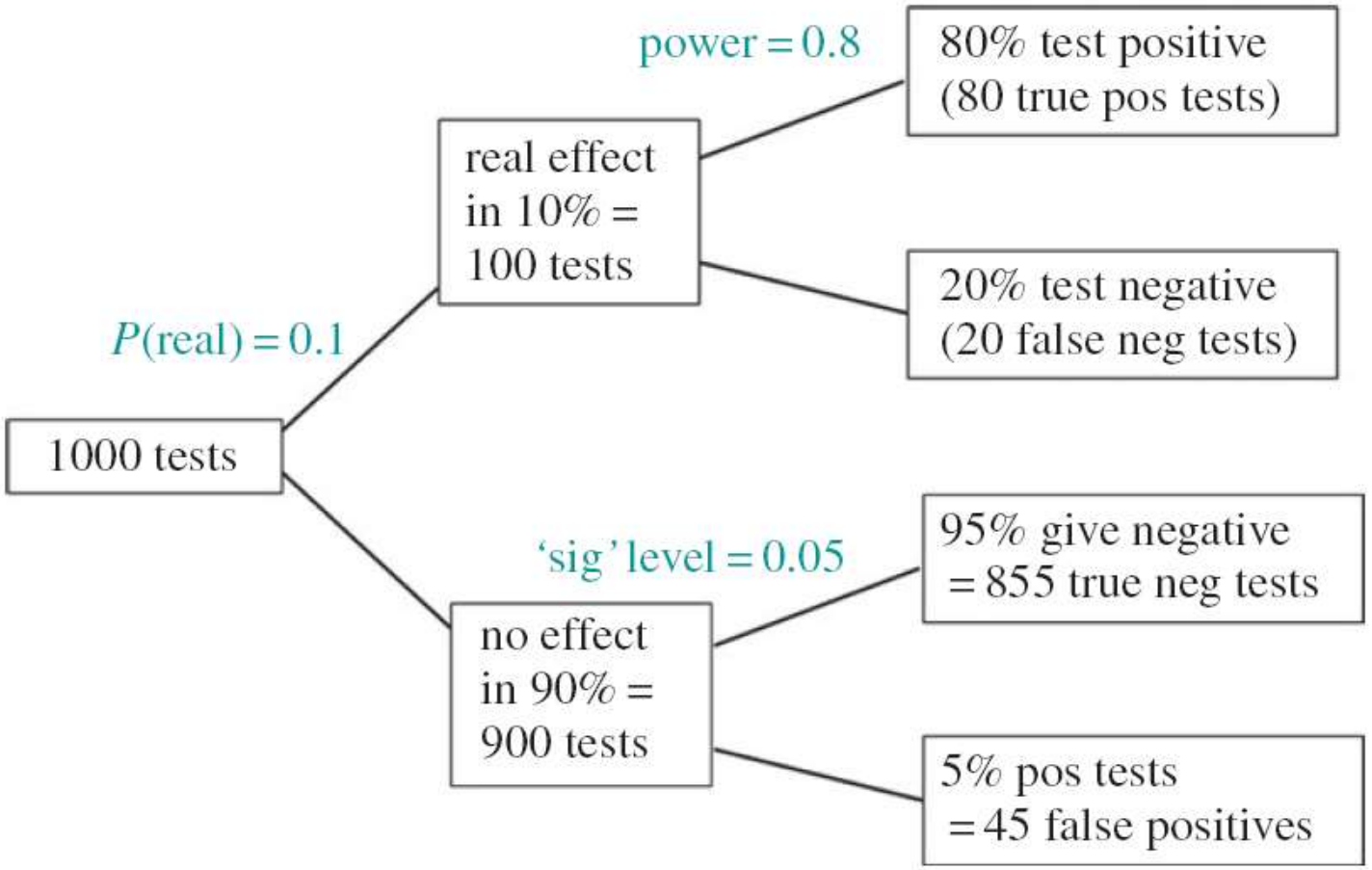
The number 86% = 495/(495+80) is our false discovery rate



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%

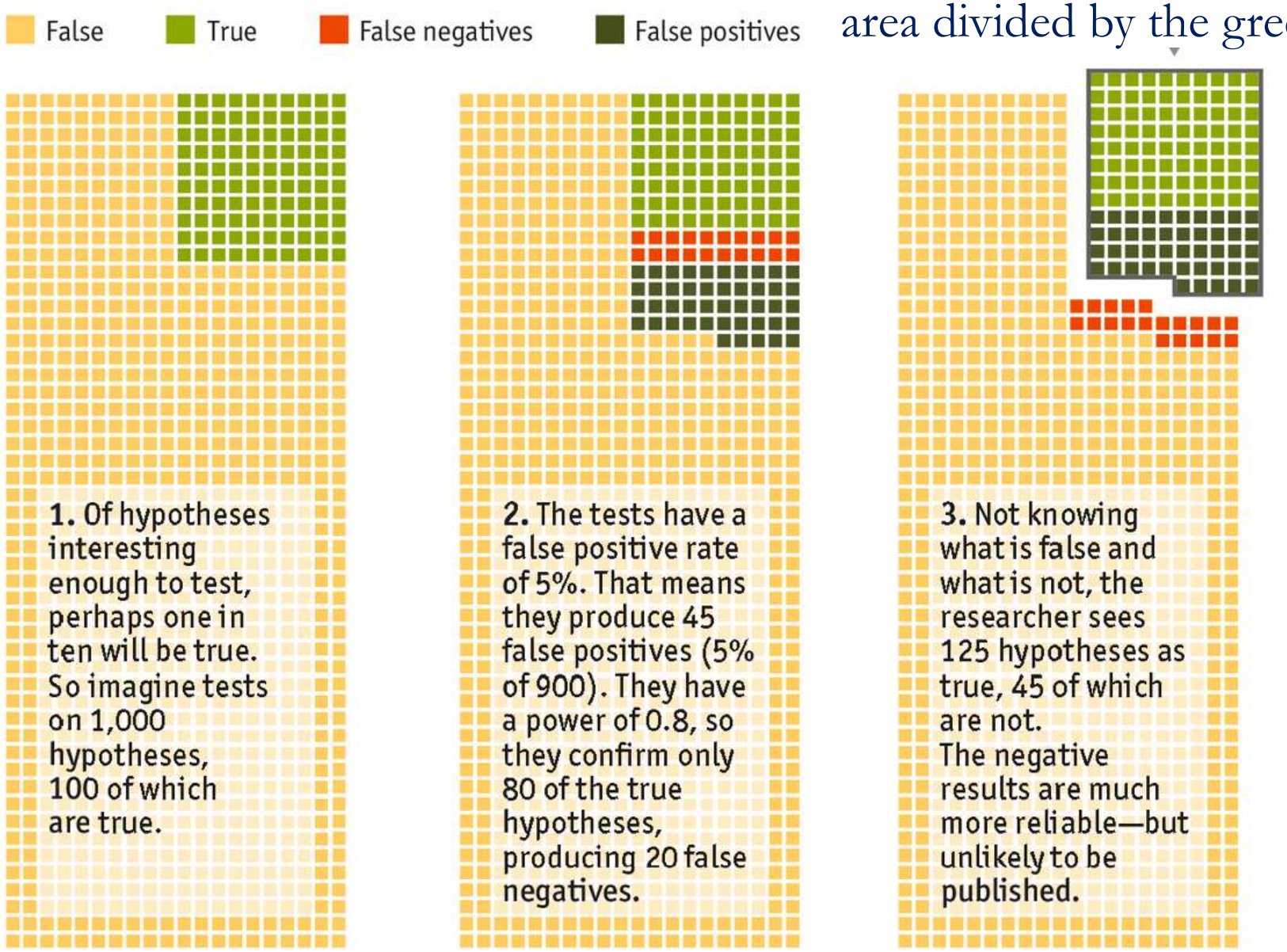
Significance p=0.05 ➔ false discovery rate of 36%

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*

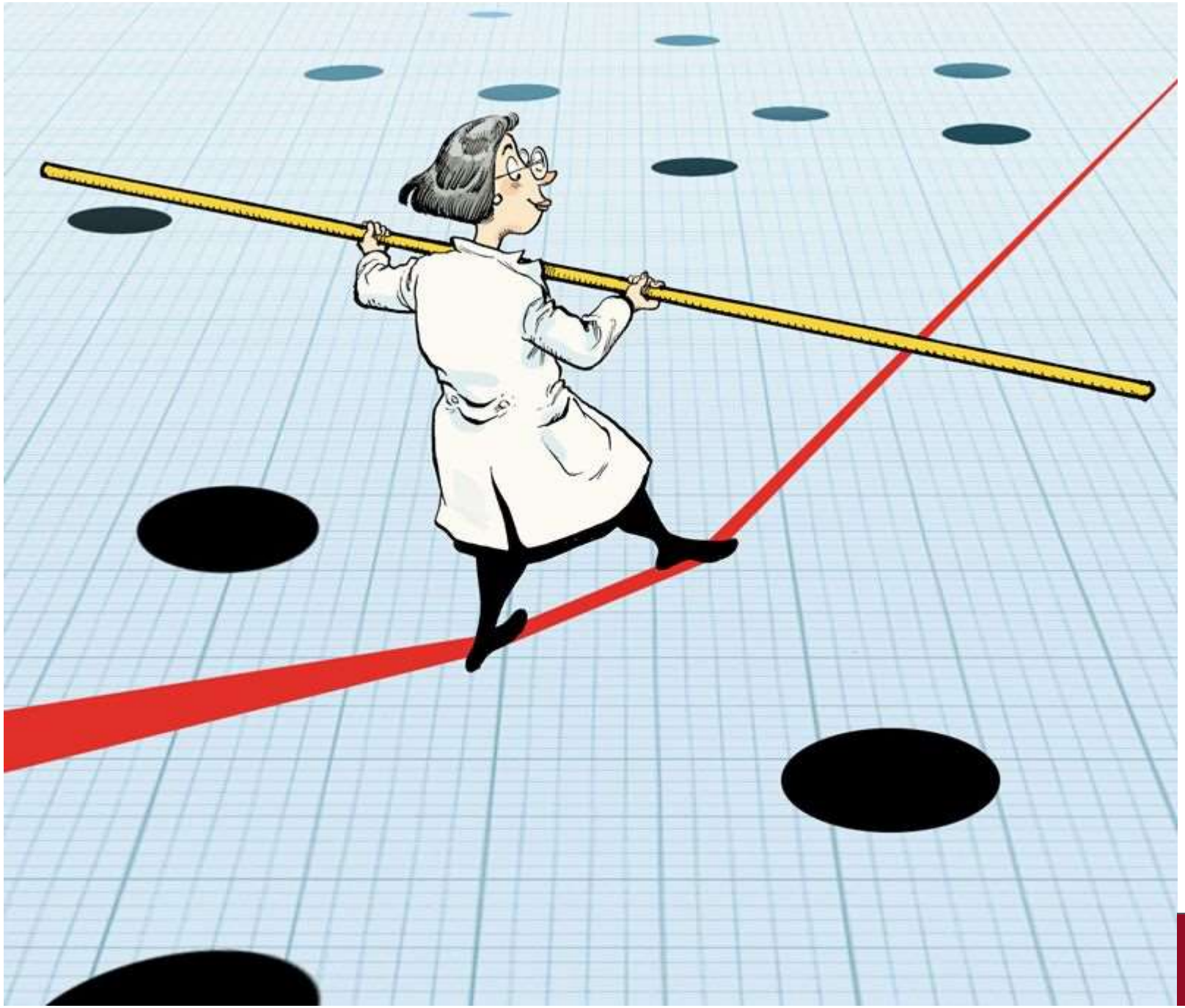
The false discovery rate is ~the dark area divided by the green+dark one

“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?

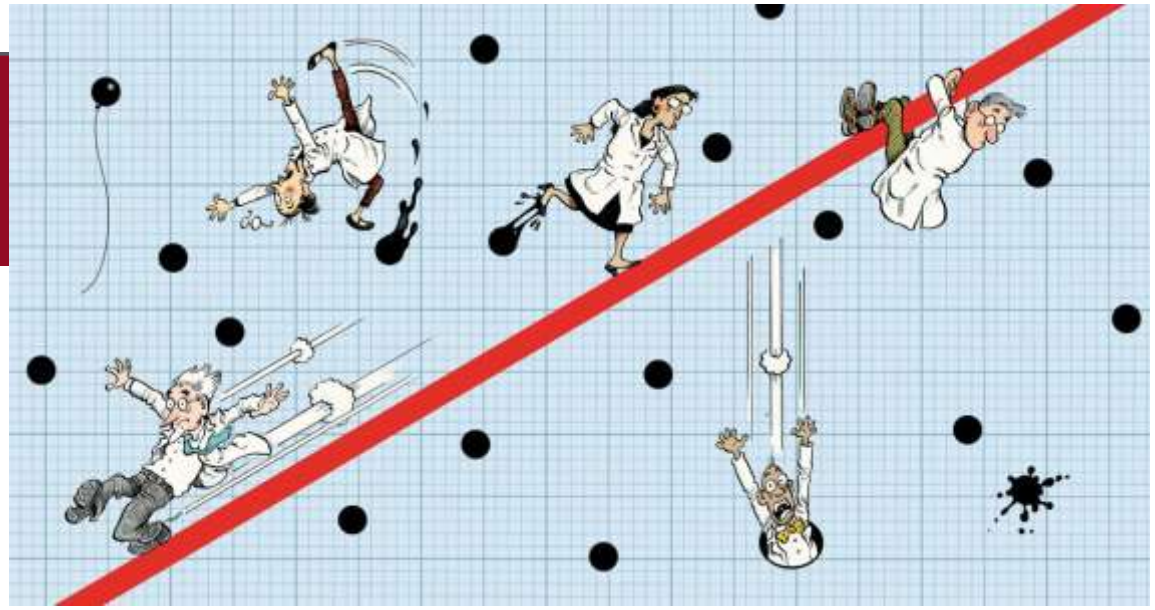
Statistics is experiencing a quality control crisis



Effect or no effect?

nature
International journal of science

nature
International journal of science

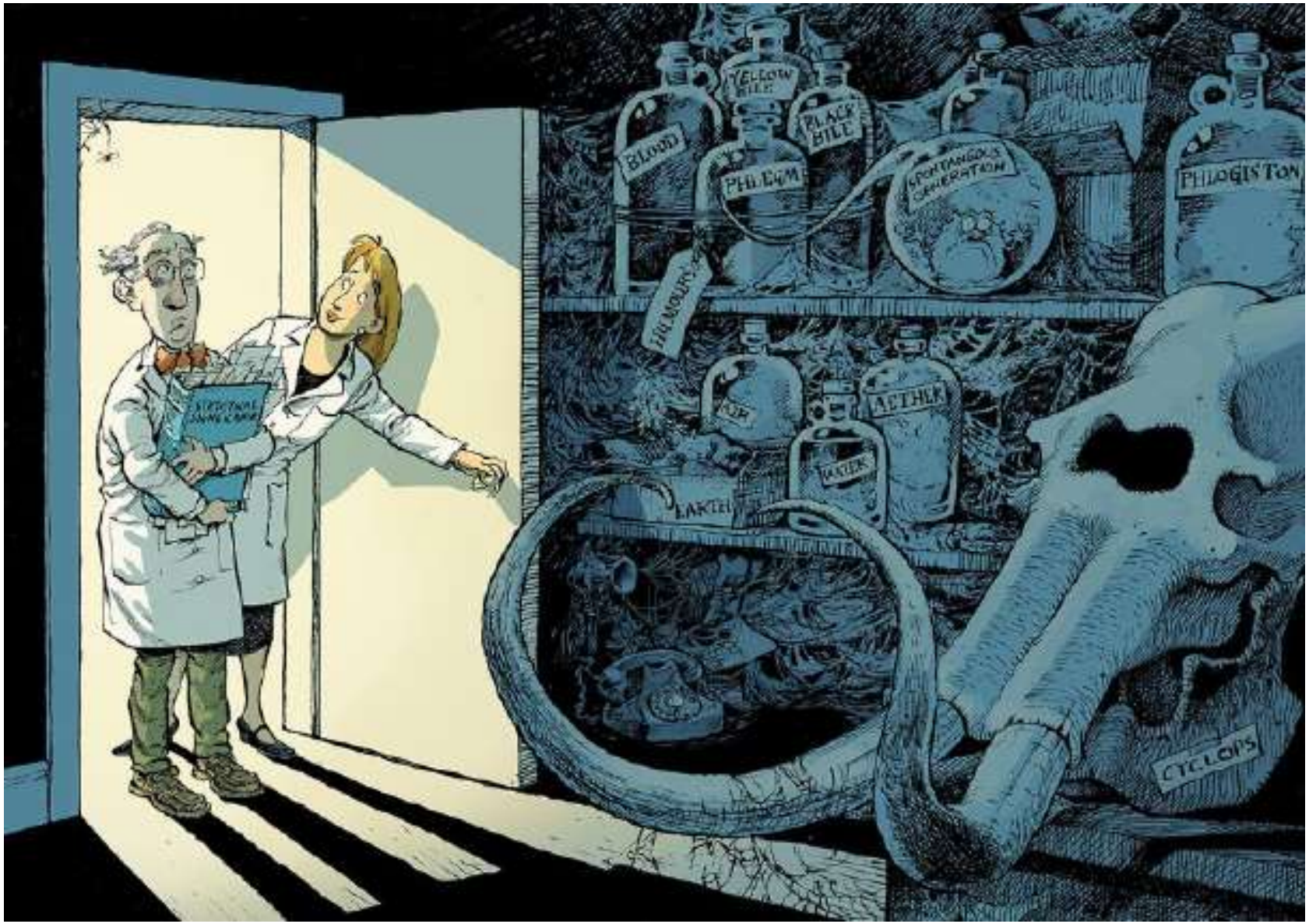


COMMENT · 28 NOVEMBER 2017

Five ways to fix statistics

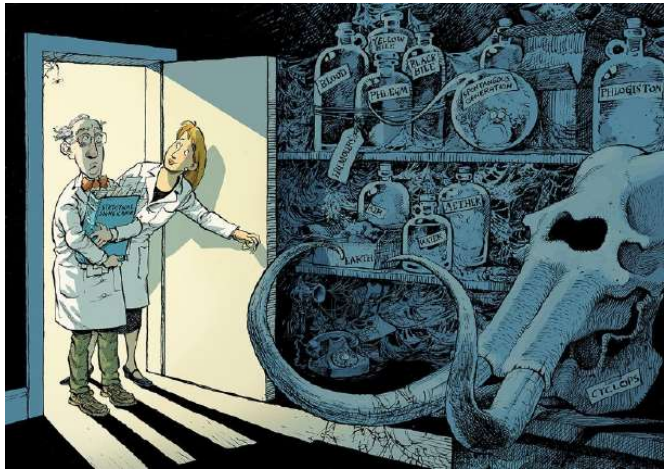
As debate rumbles on about how and how much poor statistics is to blame for poor reproducibility, Nature asked influential statisticians to recommend one change to improve science. The common theme? The problem is not our maths, but ourselves.

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



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/

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.

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**

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

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

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 ✉

…or a broader crisis?

Today, all that is controversial and relevant … operates simultaneously in science, technology, economics, law and policy…

COMMENT · 21 MAY 2019

Views from a continent in flux

nature
International journal of science

Nature asked nine leading Europeans to pick their top priority for science at this pivotal point. Love, money, and trust got most votes.

Social media gives this cocktail unprecedented reach and acceleration

COMMENT · 21 MAY 2019

Views from a continent in flux

Nature asked nine leading Europeans to pick their top priority for science at this pivotal point. Love, money, and trust got most votes.

The powerful agents of post-truth

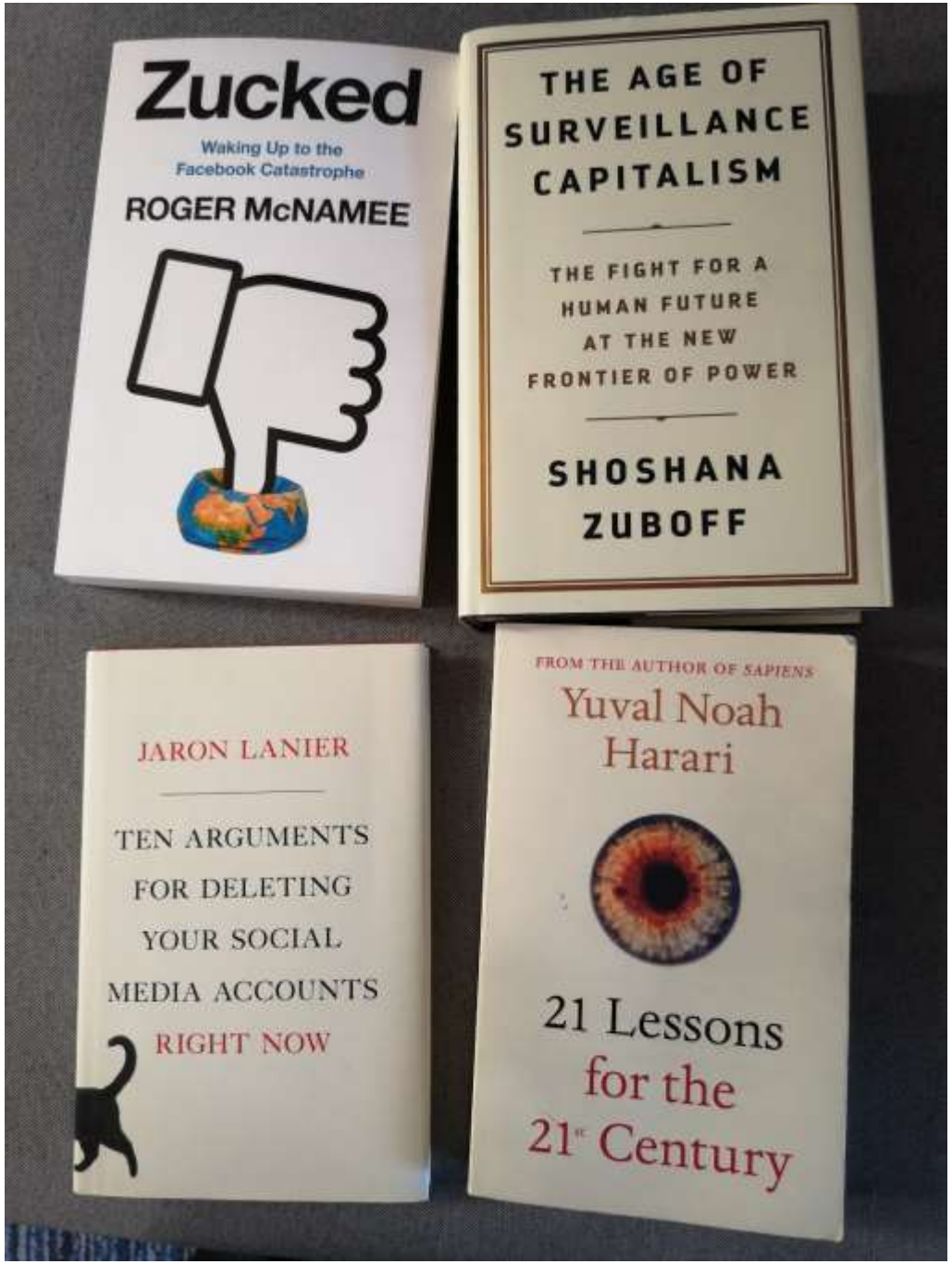
Jaron Lanier

Poisonous algorithms to stoke hatred and division

Yuval Noah Harari, Homo Deus 2015 & 21 Lessons for the 21st Century, 2018.

Jaron Lanier, 2018 Ten Arguments for Deleting Your Social Media Accounts Right Now

https://www.theguardian.com/society/2018/aug/23/russian-trolls-spread-vaccine-misinformation-on-twitter



How deep?

New or old?

Platform or surveillance?

Huxley or Orwell?

Expertise?

“People in this country have had enough of experts” (Michael Gove)

P. Stephens, Financial Times, June 23 2016, https://www.ft.com/content/bfb5f3d4-379d-11e6-a780-b48ed7b6126f

Andrea Saltelli, and Silvio Funtowicz, “Science cannot solve these problems alone because it helped to create them in the first place”, The Guardian, July 14, https://www.theguardian.com/science/political-science/2016/jul/14/six-leading-scientists-give-perspectives-on-uk-science-after-brexit?CMP=share_btn_tw

THE CONVERSATION

Academic rigour, journalistic flair

Arts + Culture Business + Economy Cities Education Environment + Energy FactCheck 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

Science as authoritative source of knowledge for policy & everyday life?

Major misdiagnoses in forensics, preclinical and clinical medicine, chemistry, psychology, economics…

Present zeitgeist = end of expertise? Or an older problem?

Issues tend to become “wicked” “where goal-formulation, problem-definition and equity issues meet”

Horst W. J. Rittel

Policy Sciences 4 (1973), 155–169

Dilemmas in a General Theory of Planning*

HORST W. J. RITTEL
Professor of the Science of Design, University of California, Berkeley

MELVIN M. WEBBER
Professor of City Planning, University of California, Berkeley

How do we appraise the work of experts when this feeds into policy? A complex matter for Clark and Majone

W. C. Clark and G. Majone, "The Critical Appraisal of Scientific Inquiries with Policy Implications," Sci. Technol. Hum. Values, vol. 10, no. 3, pp. 6–19, Jul. 1985.

Table 1. Critical criteria.

Critical Role	Critical Mode: Input	Critical Mode: Output	Critical Mode: Process
Scientist	Resource and time constraints; available theory; institutional support; assumptions; quality of available data; state of the art.	Validation; sensitivity analyses; technical sophistication; degree of acceptance of conclusions; impact on policy debate; imitation; professional recognition.	Choice of methodology (e.g., estimation procedures); communication; implementation; promotion; degree of formalization of analytic activities within the organization.
Peer Group	Quality of data; model and/or theory used; adequacy of tools; problem formulation. Input variables well chosen? Measure of success specified in advance?	Purpose of the study. Are conclusions supported by evidence? Does model offend common sense? Robustness of conclusions; adequate coverage of issues.	Standards of scientific and professional practice; documentation; review of validation techniques; style; interdisciplinarity.
Program Manager or Sponsor	Cost; institutional support within user organization; quality of analytic team; type of financing (e.g., grant vs. contract).	Rate of use; type of use (general education, program evaluation, decisionmaking, etc.); contribution to methodology and state of the art; prestige. Can results be generalized, applied elsewhere?	Dissemination; collaboration with users. Has study been reviewed?
Policymaker	Quality of analysts; cost of study; technical tools used (hardware and software). Does problem formulation make sense?	Is output familiar and intelligible? Did study generate new ideas? Are policy indications conclusive? Are they consonant with accepted ethical standards?	Ease of use; documentation. Are analysts helping with implementation? Did they interact with agency personnel? With interest groups?
Public Interest Groups	Competence and intellectual integrity of analysts. Are value systems compatible? Problem formulation acceptable? Normative implications of technical choices (e.g., choices of data).	Nature of conclusions; equity. Is analysis used as rationalization or to postpone decision? All viewpoints taken into consideration? Value issues.	Participation; communication of data and other information; adherence to strict rules of procedure.

Scientists

Public Interest Groups

Critical mode

	Input	Output	Process
Scientists	Resource and time constraints; available theory; institutional support; assumptions; quality of available data; state of the art.	Validation; sensitivity analyses; technical sophistication; degree of acceptance of conclusions; impact on policy debate; imitation; professional recognition.	Choice of methodology (e.g., estimation procedures); communication; implementation; promotion; degree of formalization of analytic activities within the organization.
Public	Competence and intellectual integrity of analysts. Are value systems compatible? Problem formulation acceptable? Normative implications of technical choices (e.g., choices of data).	Nature of conclusions; equity. Is analysis used as rationalization or to postpone decision? All viewpoints taken into consideration? Value issues.	Participation; communication of data and other information; adherence to strict rules of procedure.

Evidence based policy

PETRUCHIO: I say it is the moon.

KATHERINE: I know it is the moon.

PETRUCHIO: Nay, then you lie. It is the blessèd sun.

KATHERINE: Then God be blessed, it is the blessèd sun.

But sun it is not, when you say it is not,

And the moon changes even as your mind.

...

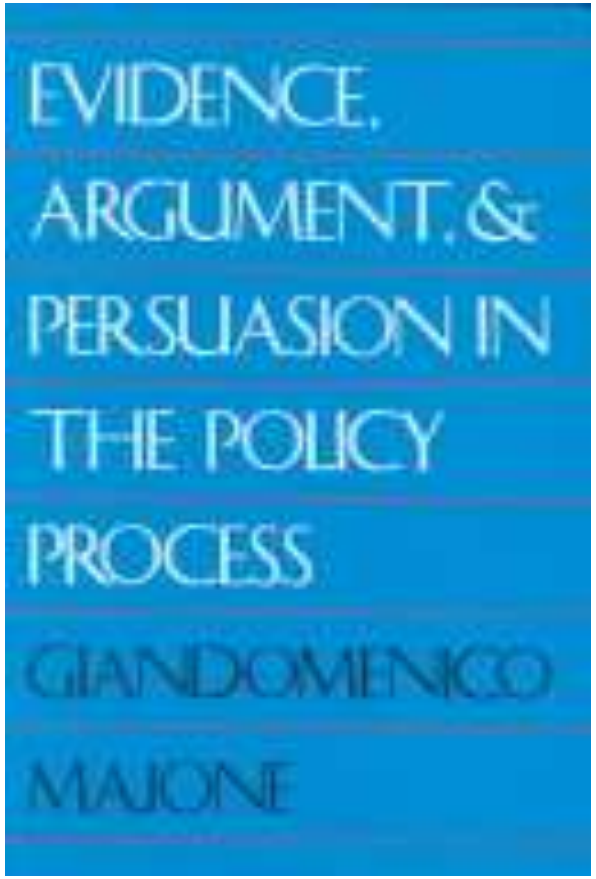
W. Shakespeare, the Taming of the Shrew, Act IV.

‘Policy based evidence’ has entered the public discourse

Warring parties accuse one another of the sin

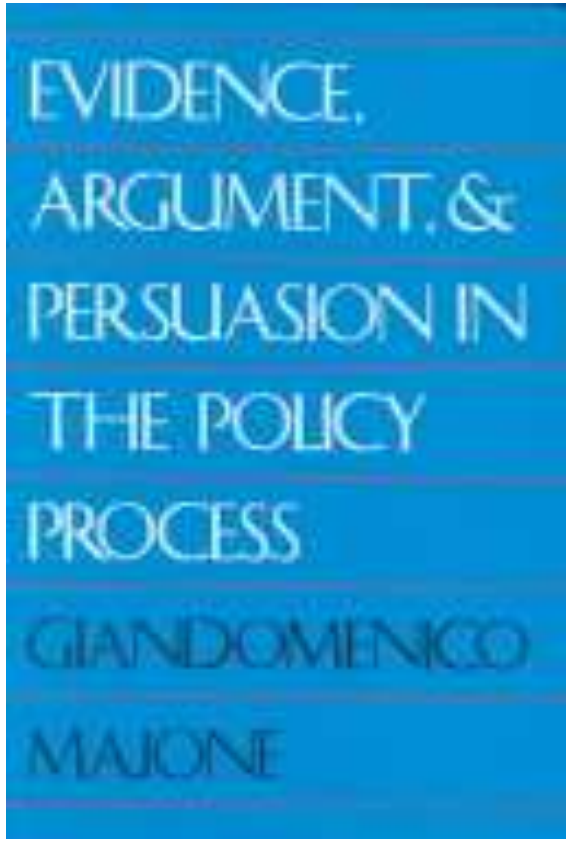
“Greenpeace […] wants is policy based evidence making not evidence based policy making” (Sanderson, 2015) …

Wilkes, G., 2015, Free Lunch: Policy-based evidence-making, Financial Times, July 3.
Sanderson, A.B., 3 Feb 2015, Breitbart, see http://www.breitbart.com/london/2015/02/03/academic-attacks-greenpeace-for-ignoring-the-evidence-on-gm-crops/; the politician is UKIP Energy Spokesman Roger Helmer MEP.

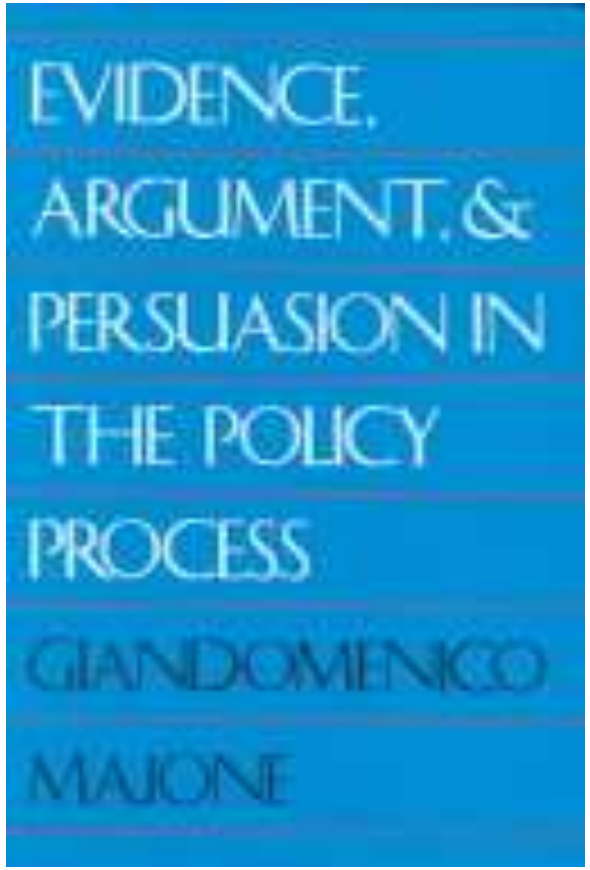


The pretended distinction between facts and value is used instrumentally

In the policy process fact and values cannot be separated in the making of an argument

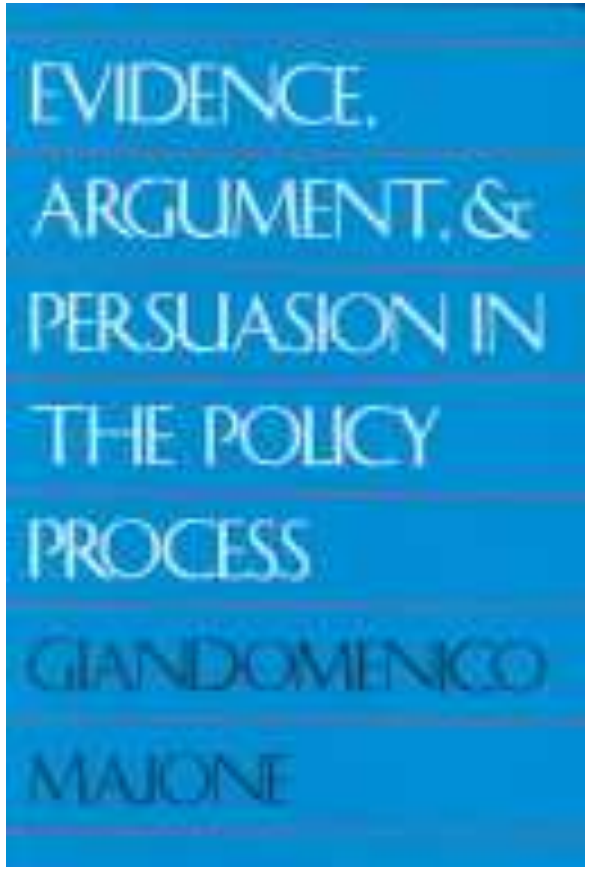


"When science, technology, and public policy intersect, different attitudes, perspectives, and rules of argument come into sharp conflict. Scientific criteria of truth clash with legal standards of evidence and with political notions of what constitutes sufficient ground for action"



Me: “the technique is never neutral”
https://arxiv.org/ftp/arxiv/papers/1712/1712.06457.pdf

Majone: “In any area of public policy the choice of instruments, far from being a technical exercise that can be safely delegated to the experts, reflects as in a microcosm all the political, moral, and cultural dimensions of policy-making”



"[my suggestion is to view a] policy analyst as a producer of arguments, capable of distinguishing between good and bad rhetoric, rather than as a "number cruncher" …

“A bewildering clamour of methods across wide areas of science, technology, the […]economy and society – complexities are routinely sidelined and expediently favourable numbers manufactured to suit the arguments of incumbent interests”

Andrew Stirling

https://steps-centre.org/blog/how-politics-closes-down-uncertainty/
https://www.prospectmagazine.co.uk/magazine/the-price-of-everything-what-people-get-wrong-about-cost-benefit-analysis

“‘tools’ like ‘externality assessment’, ‘impact analysis’ or ‘quantitative valuation’ help convince others which energy policy or health and safety standards or conservation strategy might be considered to be objectively ‘safest’, ‘safe enough’, ‘tolerable’ or even ‘best’”

Andrew Stirling

“Each technique routinely delivers its answers with formidable levels of precision. Yet the resulting impression of accuracy is deeply misplaced”

Andrew Stirling

[…] rhetoric clamour [surrounds] ‘expected utility’, ‘decision theory’, ‘life cycle assessment’, ‘ecosystem services’ ‘sound scientific decisions’ and ‘evidence-based policy’

Futures

Volume 91, August 2017, Pages 62-71

Original research article

What is wrong with evidence based policy, and how can it be improved?

Andrea Saltelli [a, b, c], Mario Giampietro [a, c, d]

Power asymmetries in the framing of issues: those who have the deepest pockets marshal the best evidence ➔ Instrumental use of quantification to obfuscate

A. Saltelli and M. Giampietro, “What is wrong with evidence based policy, and how can it be improved?,” Futures, vol. 91, pp. 62–71, Feb. 2017.

The JAMA Network Journals > Collections Store Physician Jobs About Mobile

JAMA Internal Medicine

Home Current Issue All Issues Online First Collections CME Multimedia

Online First >

Special Communication | September 12, 2016

Sugar Industry and Coronary Heart Disease Research

A Historical Analysis of Internal Industry Documents FREE

ONLINE FIRST

Cristin E. Kearns, DDS, MBA[1,2]; Laura A. Schmidt, PhD, MSW, MPH[1,3,4]; Stanton A. Glantz, PhD[1,5,6,7,8]

[+] Author Affiliations

JAMA Intern Med. Published online September 12, 2016. doi:10.1001/jamainternmed.2016.5394

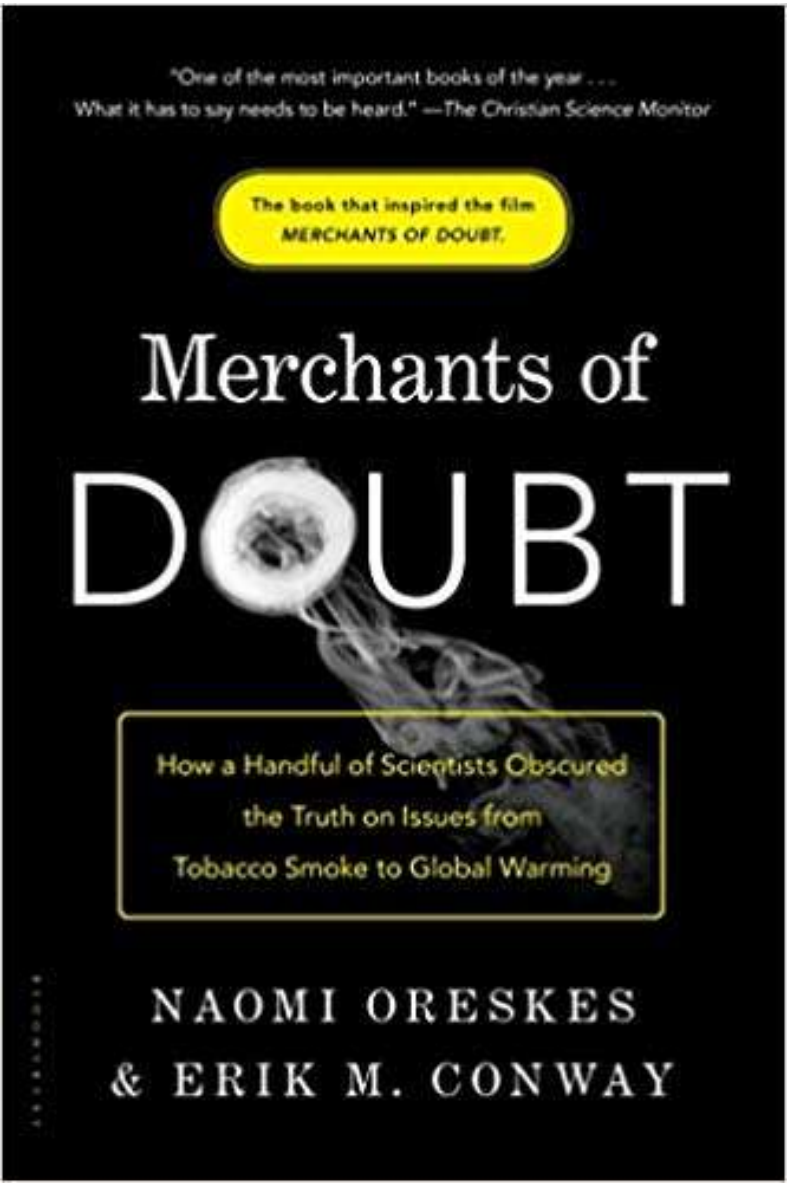
Text Size: A A A

September 12, 2016

See also https://www.theguardian.com/society/2016/apr/07/the-sugar-conspiracy-robert-lustig-john-yudkin, and the story of US President Dwight Eisenhower heart attack,…

“our findings suggest the industry sponsored a research program in the 1960s and 1970s that successfully cast doubt about the hazards of sucrose while promoting fat as the dietary culprit in CHD [coronary hearth disease]”

http://archinte.jamanetwork.com/article.aspx?articleid=2548255



Naomi Oreskes

nature
International journal of science

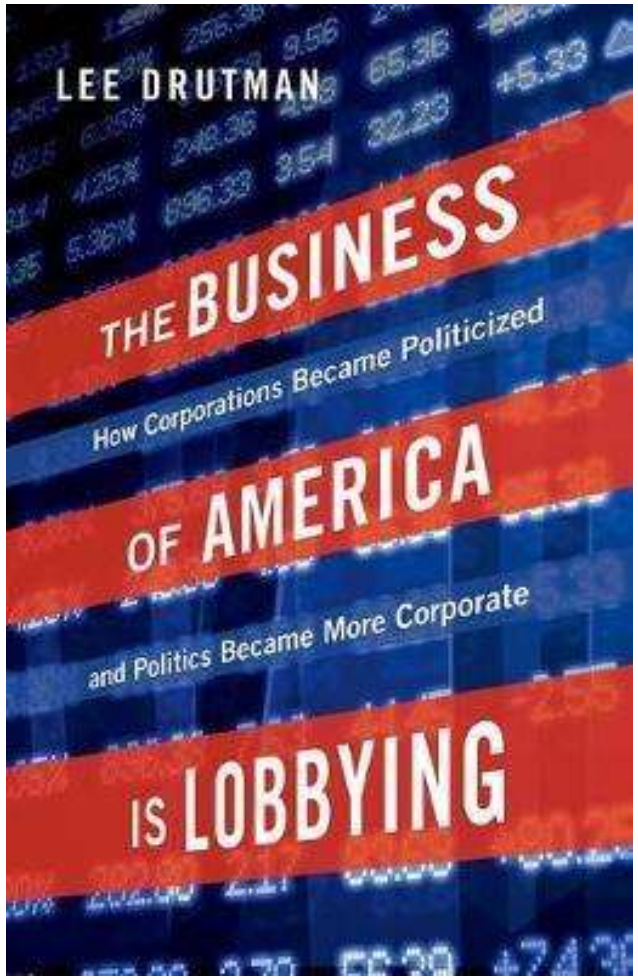
Beware: transparency rule is a Trojan Horse

Like tobacco lobbyists and climate-change deniers, the US Environmental Protection Agency is co-opting scientific trappings to sow doubt, warns Naomi Oreskes.

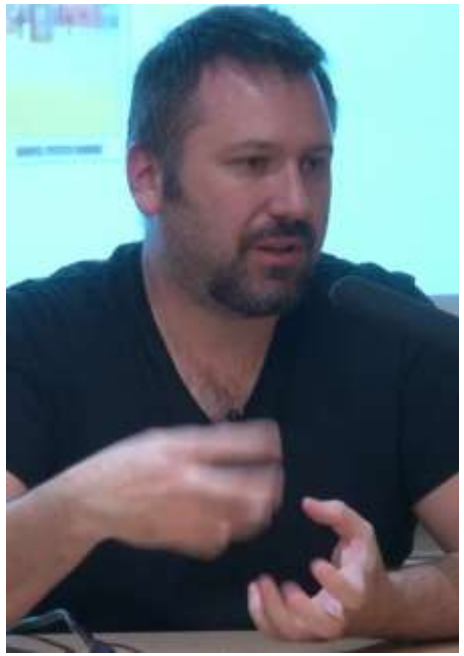
Science and lobbying

(US) corporate interest can spend on lobbying $34 for each dollar spent by diffuse interest and unions combined

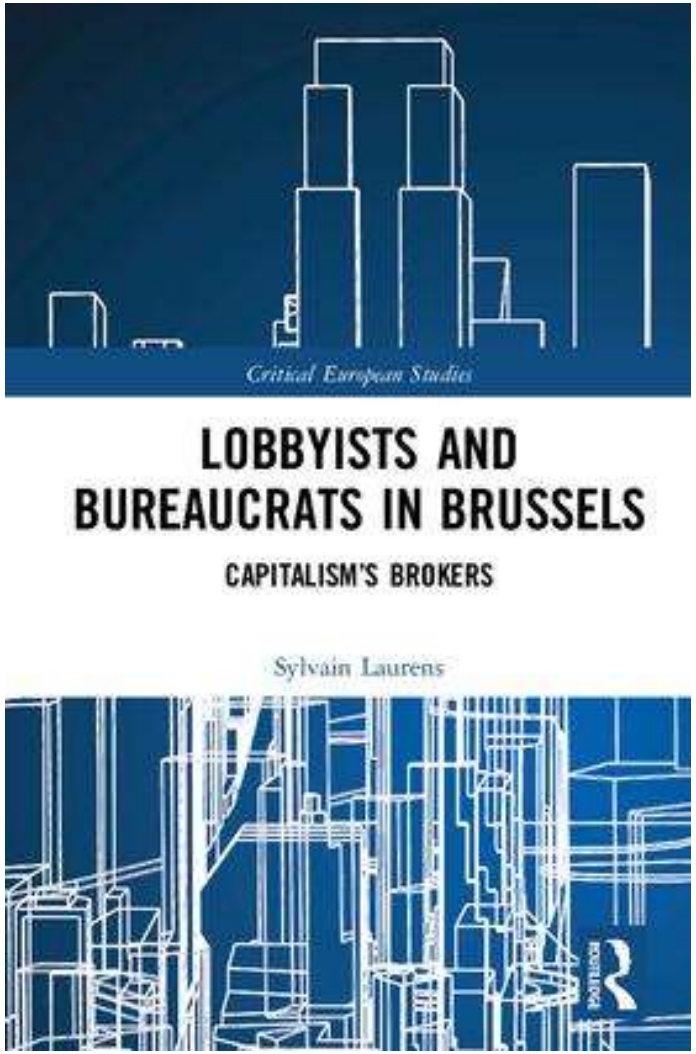
Lee Drutman



(EU) the Brussels concentration effect



Sylvain Laurens



For both scholars a salient aspect of this power is lobbyists' access to more and better disseminated science

➔ Urgent a remedial action to give citizens and political staffers some structured mechanism of access to independent scientific evidence (L. Drutman)

See discussion on OTA in Adam Keiper, 2004, Science and Congress, The New Atlantis, https://www.thenewatlantis.com/publications/science-and-congress


VICTOR
"HIS MASTER'S VOICE"
REG. U.S. PAT. OFF.

“Regulatory policy is increasingly made with the participation of experts, especially academics. A regulated firm or industry should be prepared whenever possible to co-opt these experts. This is most effectively done by identifying the leading expert in each relevant field and hiring them as consultants or advisors or giving them research grant or the like”

Owen, B. M., & Braeutigam, R., 1978 The regulation game, : Strategic Use of the Administrative Process, Ballinger Press

“This activity requires a modicum of finesse; it must not be too blatant, for the experts themselves must not recognize that they have lost their objectivity and freedom of action”

Thanks to Erik Millstone

Owen, B. M., & Braeutigam, R., 1978 The regulation game, : Strategic Use of the Administrative Process, Ballinger Press

US news

Science institute that advised EU and UN 'actually industry lobby group'

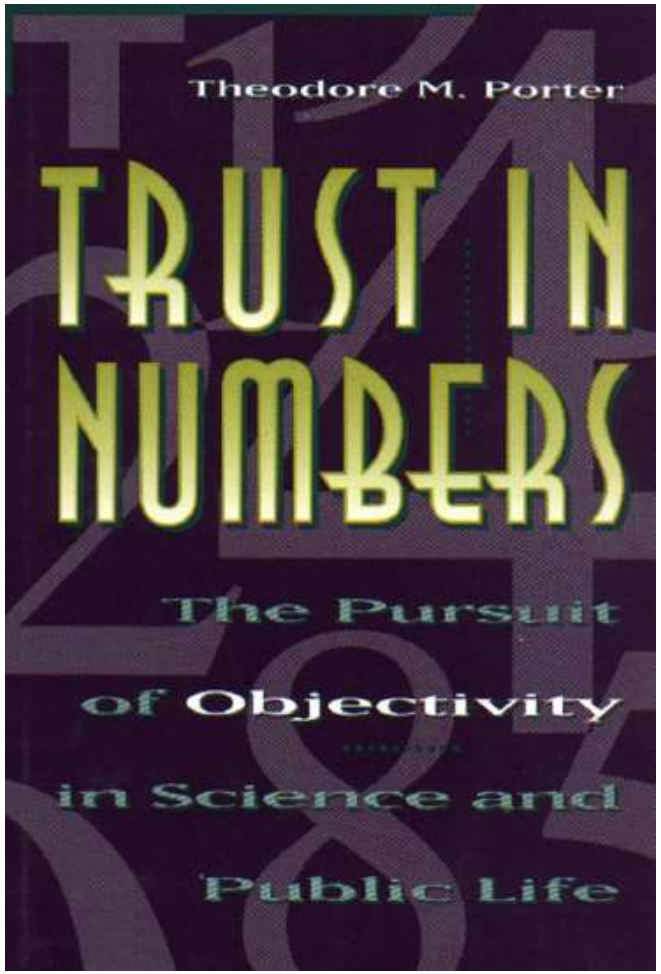
International Life Sciences Institute used by corporate backers to counter public health policies, says study

Arthur Neslen

Mon 3 Jun 2019 03.00 BST

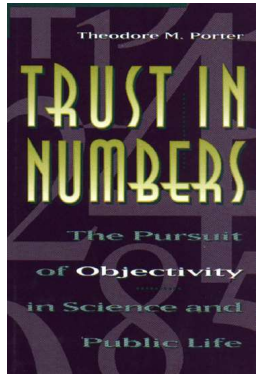
Numbers and trust

Theodor
M. Porter



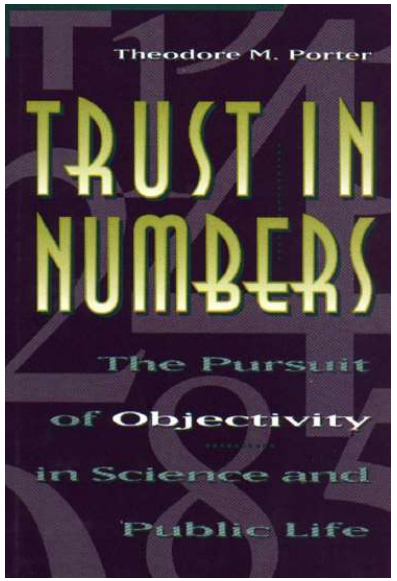
Theodore M. Porter, Trust in Numbers,
The Pursuit of Objectivity in Science and Public Life, Princeton 1995

p. 8: “The appeal of numbers is especially compelling to bureaucratic officials who lack the mandate of a popular election, or divine right.



Arbitrariness and bias are the most usual grounds upon which such officials are criticized.

A decision made by the numbers (or by explicit rules of some other sort) has at least the appearance of being fair and impersonal.”



p. 8: “Scientific objectivity thus provides an answer to a moral demand for impartiality and fairness.

Quantification is a way of making decisions without seeming to decide.

Objectivity lends authority to officials who have very little of their own.”

Trust, authority and styles of quantification: two different stories

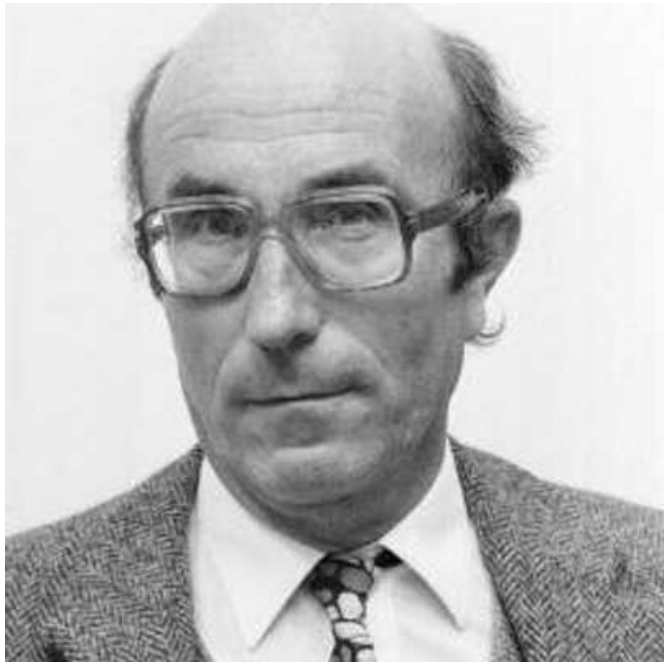
ECOLE NATIONALE
DES PONTS ET CHAUSSEES

US Army Corps of Engineers®

Porter's story: Quantification needs judgment which in turn needs trust …without trust quantification becomes mechanical, a system, and 'systems can be played'.

‘System trust’, is social system theory:

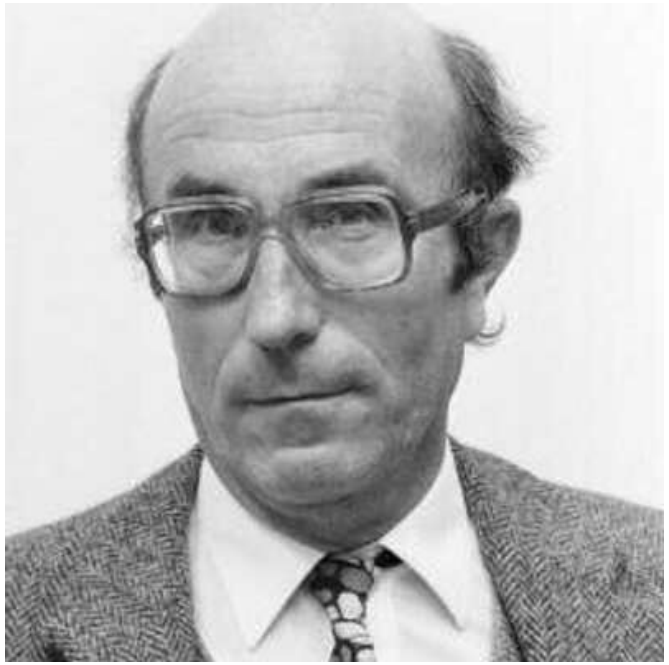
“The reduction of complexity [made possible by generalized media of communication as money, power and truth] **assumes trust** on the part of those who are expecting such reduction and of those who are supposed to accept it once it is accomplished”



Niklas Luhmann

N. Luhmann, Trust and Power. Polity Press, 2017.

"[System trust thus permits] the bank to lend more money than it possess, the state to issue more commands than it can enforce using the police, that more information is divulged in professional advice than could be backed up empirically or logically".



Niklas Luhmann

N. Luhmann, Trust and Power. Polity Press, 2017.

‘the essential fiduciary status’ of science= Trust in science is necessary for the general society to continue to support it, materially and with recruits. And mutual trust within science is necessary for its systems of quality assurance to function

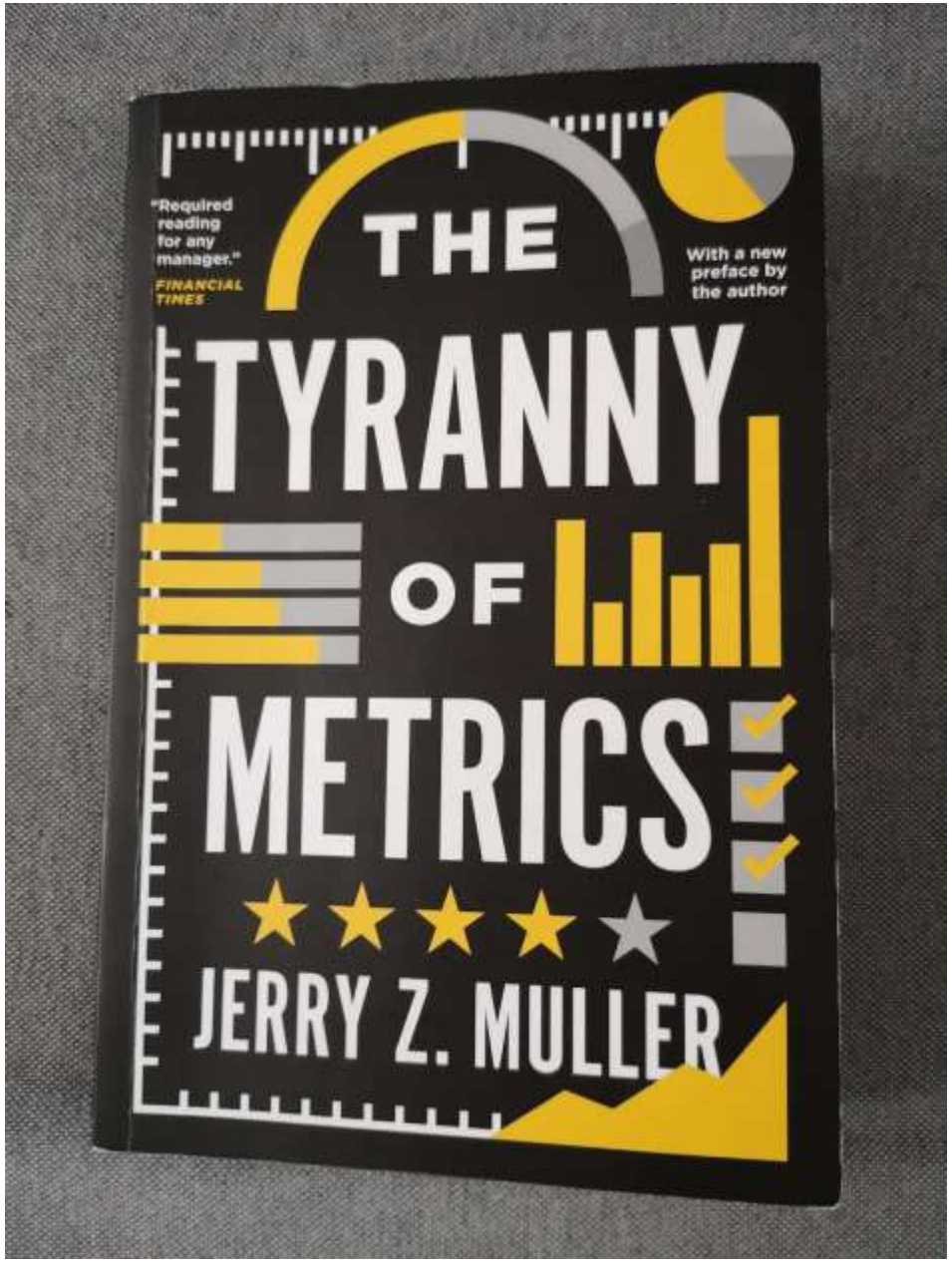
Jerome R. Ravetz

Charles Goodhart

p. 44 “Any … measures necessarily involve a loss of information … [and distorts behavior]” (Porter, 1995)

This is what we normally call Goodhart’s law, from Charles Goodhart. "When a measure becomes a target, it ceases to be a good measure."

http://cyberlibris.typepad.com/blog/files/Goodharts_Law.pdf



More reading

Frames

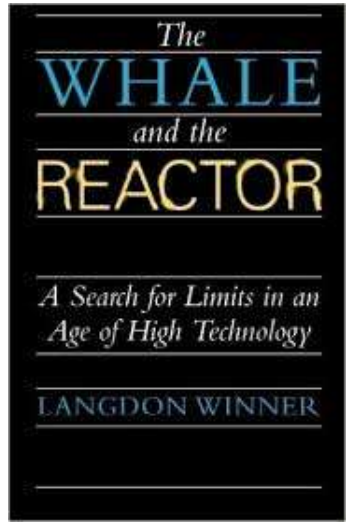
Frames

Most analyses offered as input to policy are framed as cost benefit analysis or risk analyses.

Langdon Winner

8

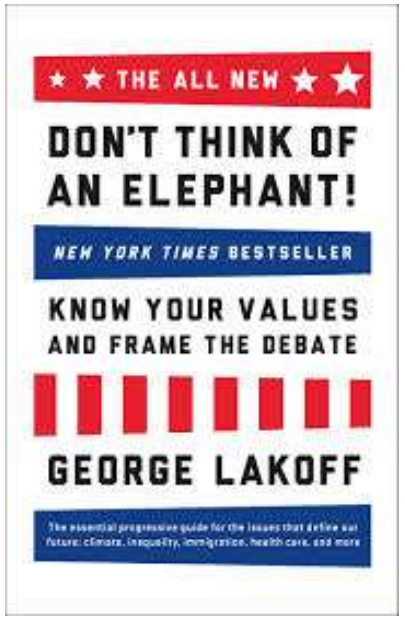
ON NOT HITTING THE TAR-BABY



Winner, L., 1986. The Whale and the Reactor: a Search for Limits in an Age of High Technology. The University of Chicago Press, 1989 edition.

Frames: The expression ‘tax relief’ is apparently innocuous but it suggests that tax is a burden, as opposed to what pays for road, hospitals, education and other infrastructures of modern life (Lakoff, 2004).

George Lakoff



Lakoff, G., 2010, Why it Matters How We Frame the Environment, Environmental Communication: A Journal of Nature and Culture, 4:1, 70-81.

Lakoff, G., 2004-2014, Don’t think of an elephant: know your values and frame the debate, Chelsea Green Publishing.

Frames as hypocognition & Socially constructed ignorance

For Rayner (2012) “Sense-making is possible only through processes of exclusion. Storytelling is possible only because of the mass of detail that we leave out. Knowledge is possible only through the systematic ‘social construction of ignorance’ (Ravetz, 1986)”

Steve Rayner

Jerry Ravetz

Ravetz, J., R., 1987, Usable Knowledge, Usable Ignorance, Incomplete Science with Policy Implications, Knowledge: Creation, Diffusion, Utilization, 9(1), 87-116. Rayner, S., 2012, Uncomfortable knowledge: the social construction of ignorance in science and environmental policy discourses, Economy and Society, 41:1, 107-125.

Rayner's (2012) strategies to deal with "uncomfortable knowledge".

1. Denial: "There isn't a problem"

2. Dismissal: "It's a minor problem"

Rayner, S., 2012, Uncomfortable knowledge: the social construction of ignorance in science and environmental policy discourses, Economy and Society, 41:1, 107-125.

Rayner's (2012) strategies to deal with "uncomfortable knowledge".

3. Diversion: "Yes I am working on it" (In fact I am working on something that is only apparently related to the problem)

Rayner, S., 2012, Uncomfortable knowledge: the social construction of ignorance in science and environmental policy discourses, Economy and Society, 41:1, 107-125.

Rayner's (2012) strategies to deal with "uncomfortable knowledge".

4. Displacement: "Yes and the model we have developed tells us that real progress is being achieved" (The focus in now the model not the problem).

Rayner, S., 2012, Uncomfortable knowledge: the social construction of ignorance in science and environmental policy discourses, Economy and Society, 41:1, 107-125.

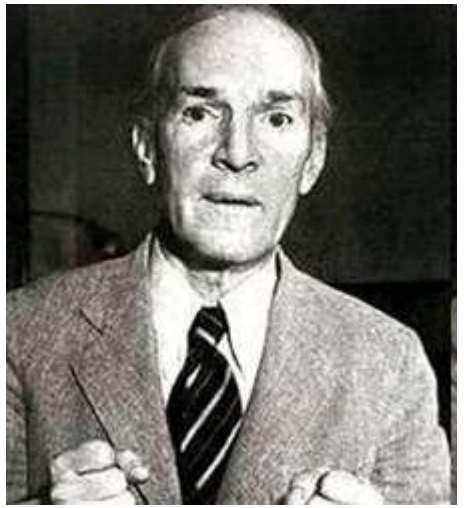
“Uncomfortable knowledge” can be used as a gauge of an institution’s health.

The larger the “uncomfortable knowledge” an institution needs to maintain, the closer it is to its ancient régime stage (Funtowicz and Ravetz, 1994).

Funtowicz, S.O. and Jerome R. Ravetz, 1994, Emergent complex systems, Futures, 26(6), 568-582.

Why frames ‘stick’

“If is difficult to get a man to understand something when his salary depends upon his not understanding it.”



Upton Sinclair

Second part: methods

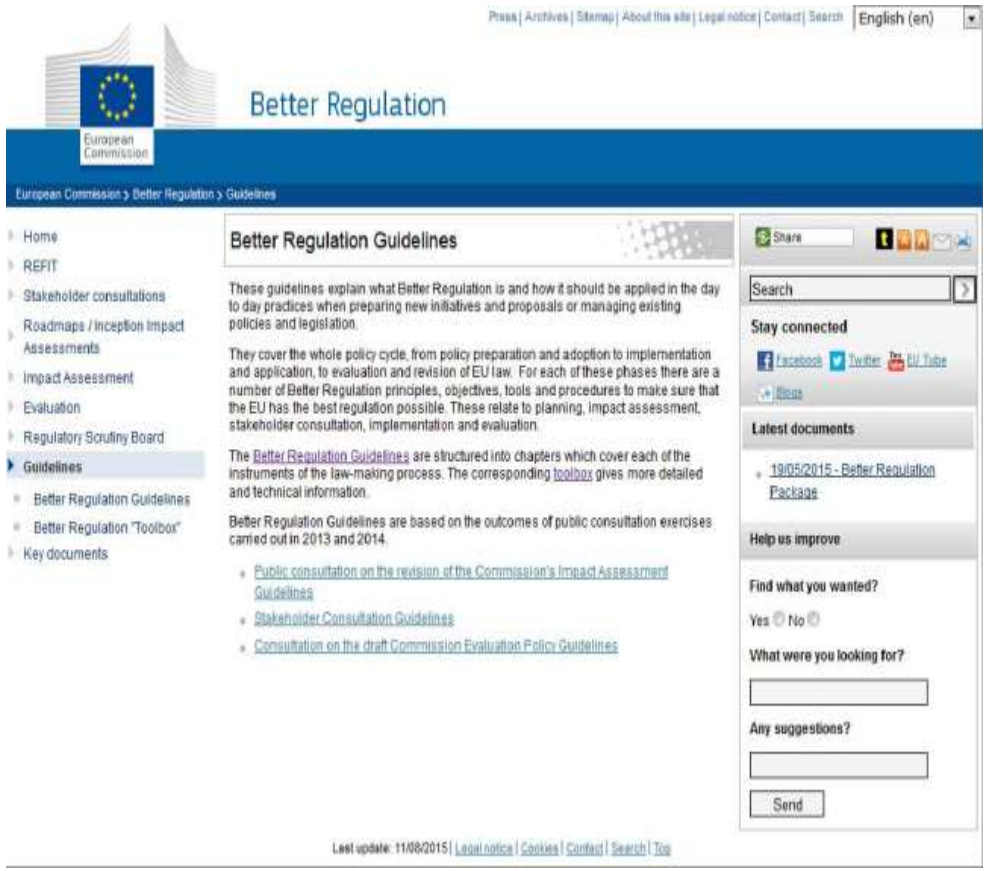
Sensitivity auditing
NUSAP
PNS
Indicators
Examples or practicum

Methods for responsible quantification

See slides of a recent course: 'Numbers for Policy'
http://www.andreasaltelli.eu/presentations/#Course

Sensitivity auditing

EC impact assessment guidelines: what do they say about sensitivity auditing ?



http://ec.europa.eu/smart-regulation/guidelines/docs/br_toolbox_en.pdf

p. 392

… where there is a major disagreement among stakeholders about the nature of the problem, … then sensitivity auditing is more suitable but sensitivity analysis is still advisable as one of the steps of sensitivity auditing.

Andrea Saltelli, Ksenia Aleksankina, William Becker, Pamela Fennell, Federico Ferretti, Niels Holst, Sushan Li, Qiongli Wu, Why so many published sensitivity analyses are false: a systematic review of sensitivity analysis practices, Environmental Modelling and Software, Volume 114, April 2019, Pages 29-39.

Sensitivity auditing, [...] is a wider consideration of the effect of all types of uncertainty, including structural assumptions embedded in the model, and subjective decisions taken in the framing of the problem.
[...]
The ultimate aim is to communicate openly and honestly the extent to which particular models can be used to support policy decisions and what their limitations are.

p. 393

“In general sensitivity auditing stresses the idea of honestly communicating the extent to which model results can be trusted, taking into account as much as possible all forms of potential uncertainty, and to anticipate criticism by third parties.”

The rules of sensitivity auditing

Rule 1: Check against rhetorical use of mathematical modelling;

Rule 2: Adopt an “assumption hunting” attitude; focus on unearthing possibly implicit assumptions;

Rule 3: Check if uncertainty been instrumentally inflated or deflated.

The rules of sensitivity auditing

Rule 4: Find sensitive assumptions before these find you; do your SA before publishing;

Rule 5: Aim for transparency; Show all the data;

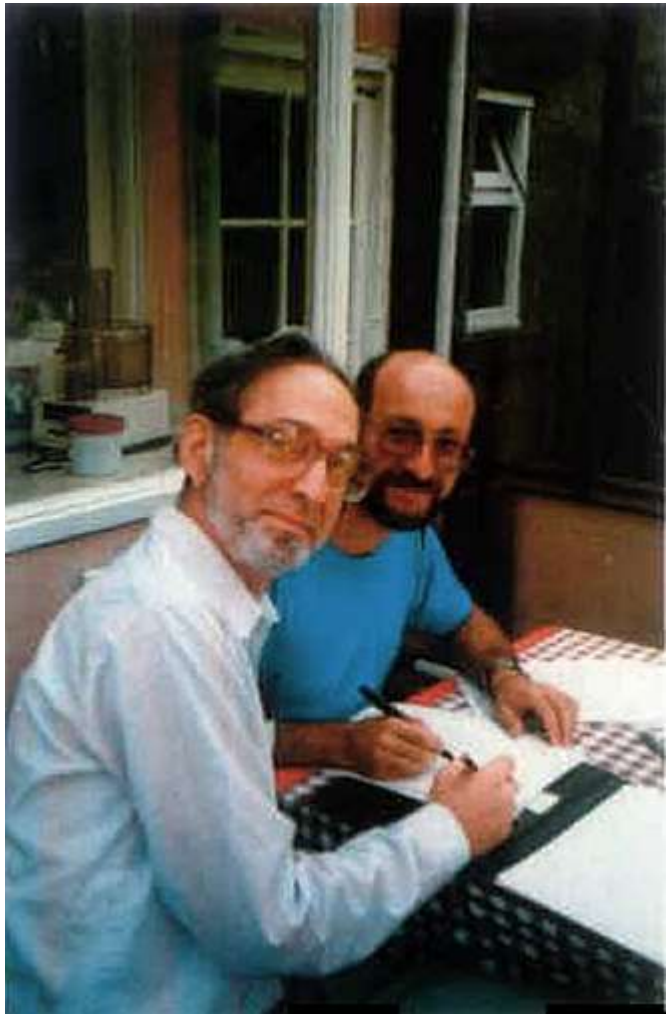
Rule 6: Do the right sums, not just the sums right; the analysis should not solve the wrong problem;

Rule 7: Perform a proper global sensitivity analysis.

NUSAP

NUSAP =

Numeral
Unit
Spread
Assessment
Pedigree



Jerome Ravetz and Silvio Funtowicz, circa 1988, at Sheffield

Numeral will usually be an ordinary number;

Unit refers to the units used in Numeral;

Spread is an assessment of the error in the value of the Numeral

Assessment is a summary of salient qualitative judgements about the information – this can be of statistical nature (a significance level) or more general, e.g. involving terms such as 'conservative' or 'optimistic'.

Pedigree is an evaluative description of the mode of production and of anticipated use of the information

Jeroen P. van der Sluijs, James S. Risbey and Jerry Ravetz, 2005, Uncertainty Assessment of VOC Emissions from Paint in the Netherlands Using the NUSAP System, Environmental Monitoring and Assessment (2005) 105: 229–259.

NUSAP pedigree matrix

Universiteit Utrecht

Example Pedigree matrix parameter strength

Code	Proxy	Empirical	Theoretical basis	Method	Validation
4	Exact measure	Large sample direct mmts	Well established theory	Best available practice	Compared with indep. mmts of same variable
3	Good fit or measure	Small sample direct mmts	Accepted theory partial in nature	Reliable method commonly accepted	Compared with indep. mmts of closely related variable
2	Well correlated	Modeled/derived data	Partial theory limited consensus on reliability	Acceptable method limited consensus on reliability	Compared with mmts not independent
1	Weak correlation	Educated guesses / rule of thumb est	Preliminary theory	Preliminary methods unknown reliability	Weak / indirect validation
0	Not clearly related	Crude speculation	Crude speculation	No discernible rigour	No validation

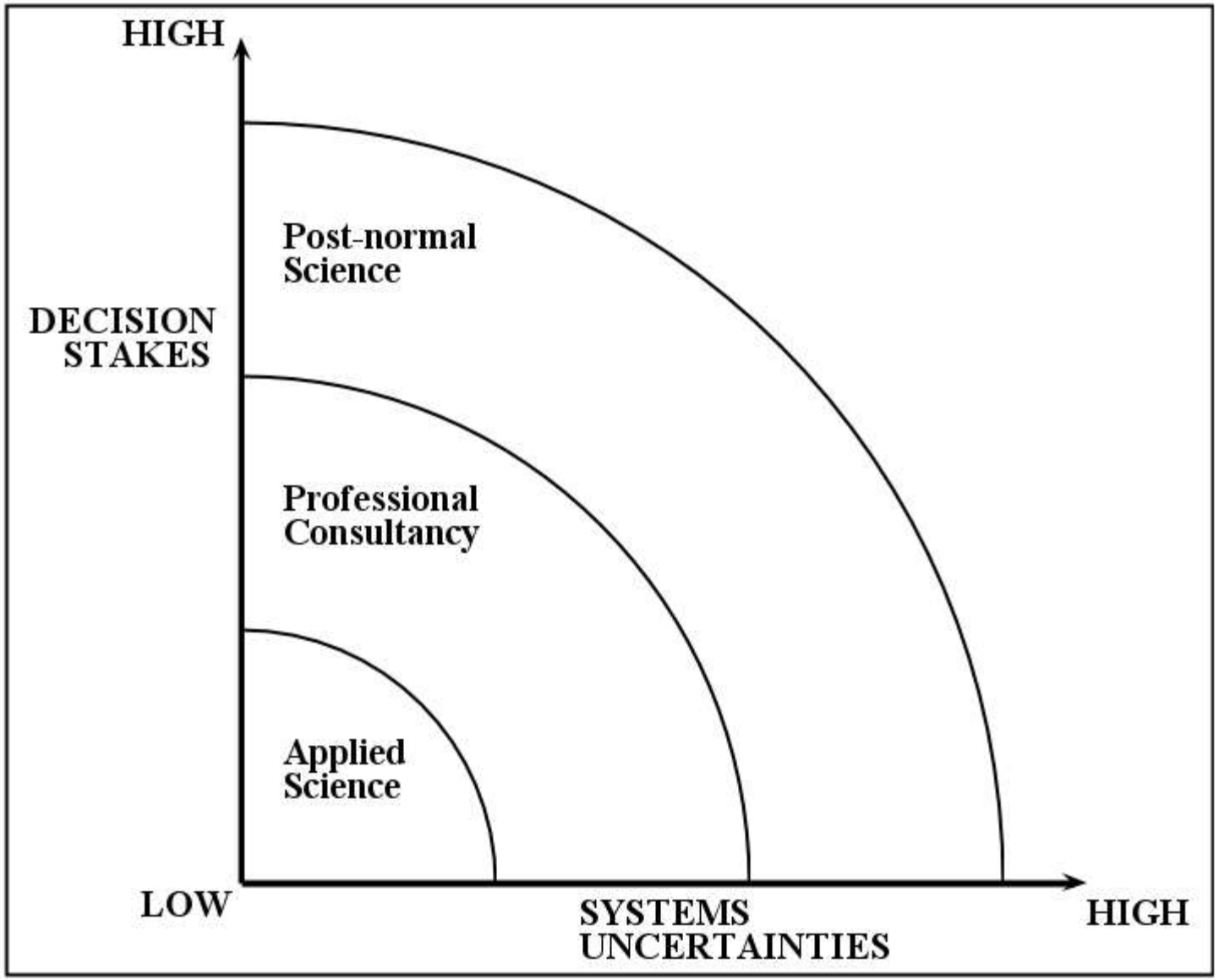
Jeroen van der Sluijs

http://www.nusap.net/

Example Pedigree matrix parameter strength

Code	Proxy	Empirical	Theoretical basis	Method	Validation
4	Exact measure	Large sample direct mmts	Well established theory	Best available practice	Compared with indep. mmts of same variable
3	Good fit or measure	Small sample direct mmts	Accepted theory partial in nature	Reliable method commonly accepted	Compared with indep. mmts of closely related variable
2	Well correlated	Modeled/derived data	Partial theory limited consensus on reliability	Acceptable method limited consensus on reliability	Compared with mmts not independent
1	Weak correlation	Educated guesses / rule of thumb est	Preliminary theory	Preliminary methods unknown reliability	Weak / indirect validation
0	Not clearly related	Crude speculation	Crude speculation	No discernible rigour	No validation

Post normal science



Funtowicz, S. and Ravetz, J., 1993. "Science for the post-normal age", Futures, 31(7): 735-755.

Funtowicz, S.O. and Ravetz, J.R. (1994). The worth of a songbird: Ecological economics as a post-normal science. Ecological Economics 10(3), 197-207.

… an approach for the use of science on issues where “facts are uncertain, values in dispute, stakes high and decisions urgent”

“the stage where we are today, where all the comfortable assumptions about science, its production and its use, are in question”

https://en.wikipedia.org/wiki/Post-normal_science

“… an inclusive set of robust insights more than as an exclusive fully structured theory or field of practice”

… a lens to see at the science-policy-technology interfaces with a hunch for context, purpose, assumptions, expectations, power relations, and for the non separability of facts and values

https://en.wikipedia.org/wiki/Post-normal_science

PNS's extended peer community

Participation: **PNS's extended peer community**

Extension to

1) more than one discipline

2) to lay actors, taken to be all those with stakes, or an interest (Why? Ask to Paul Feyerabend in Against Method) – including investigative journalism and whistle blowers.

Feyerabend, Paul (1975). Against method. Verso Publisher.

PNS's extended peer community

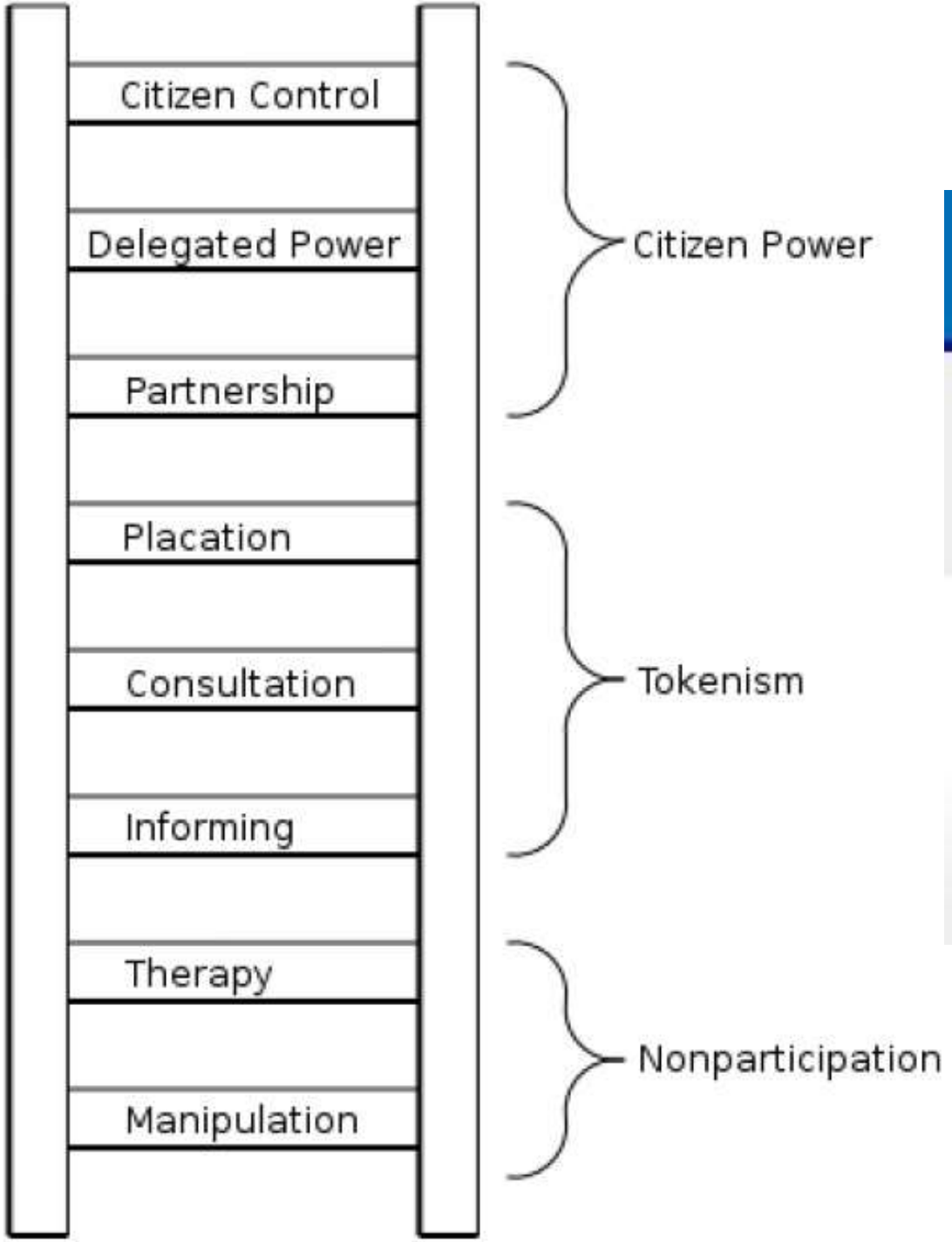
Inspiration: 'popular epidemiology', 'housewife epidemiology', early evidence-based medicine (the Cochrane collaboration), and the total quality management ideas of W. Edwards Deming, in particular quality circles.

Phil Brown, 1997, Popular Epidemiology Revisited, Current Sociology, Volume: 45 issue: 3, page(s): 137-156.

PNS's extended peer community

The extension of the peer community is not only ethically fair or politically correct – it enhances quality, see Brian Wynne & Cumbrian sheep farmers' against scientist and authorities in the relation to the Chernobyl radioactive fallout

Wynne, B. (1992). Misunderstood misunderstanding: social identities and public uptake of science. Public Understanding of Science, 1, 281–304.



Taylor & Francis Online

Journal
Journal of the American Institute of Planners ›
Volume 35, 1969 - Issue 4

Article

A Ladder Of Citizen Participation

Sherry R. Arnstein

Pages 216-224 | Published online: 26 Nov 2007

Download citation https://doi.org/10.1080/01944366908977225

Composite indicators

Composite indicators
What are they?

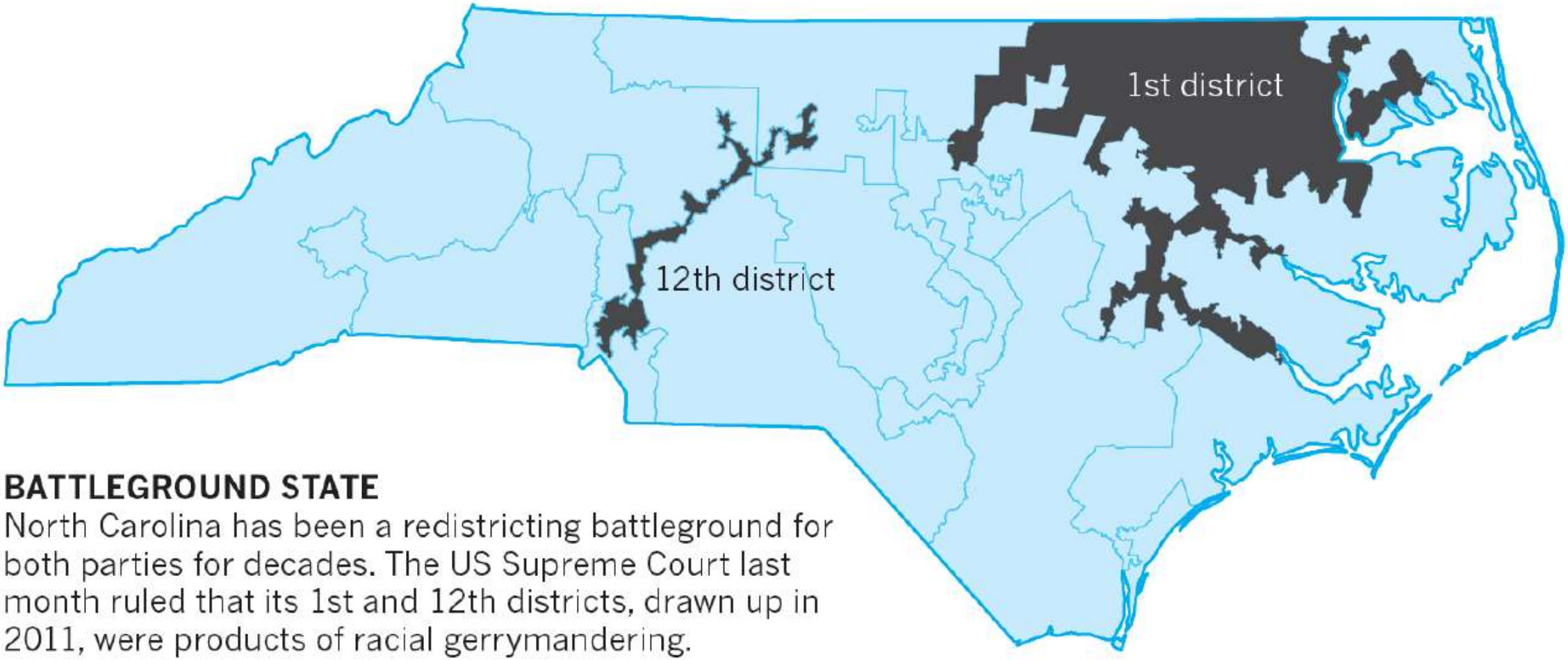
Eight factors further disaggregated into 44 sub-factors

1.1 Government powers are effectively limited by the legislature
1.2 Government powers are effectively limited by the judiciary
1.3 Government powers are effectively limited by independent auditing and review
1.4 Government officials are sanctioned for misconduct
1.5 Government powers are subject to non-governmental checks
1.6 Transition of power is subject to the law

One of the eight factors with its 6 sub factors …

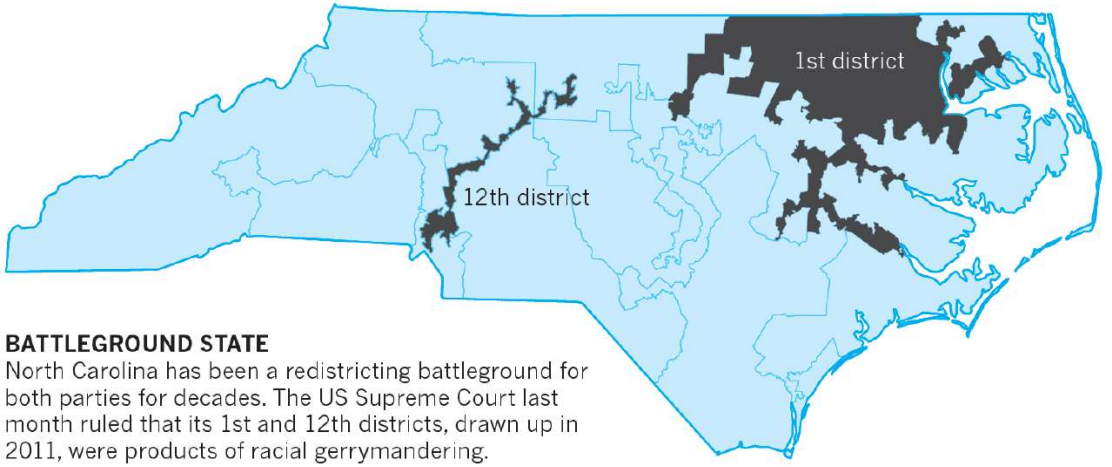
https://worldjusticeproject.org/sites/default/files/documents/WJP-ROLI-2019-Single%20Page%20View-Reduced_0.pdf

Making the case for gerrymandering?



BATTLEGROUND STATE
North Carolina has been a redistricting battleground for both parties for decades. The US Supreme Court last month ruled that its 1st and 12th districts, drawn up in 2011, were products of racial gerrymandering.

Nature June 2017 article on the mathematics of ‘nailing’ gerrymandering



BATTLEGROUND STATE
North Carolina has been a redistricting battleground for both parties for decades. The US Supreme Court last month ruled that its 1st and 12th districts, drawn up in 2011, were products of racial gerrymandering.

“[US] ranked 55th of 158 nations — last among Western democracies — in a 2017 index of voting fairness (Electoral Integrity Project)”

Carrie Arnold, 2017, The mathematicians who want to save democracy, 200, NATURE, VOL 546, 8 JUNE 2017.

Quality of composite indicators

Handbook on Constructing Composite Indicators

METHODOLOGY AND USER GUIDE

Is a theory for composite indicators possible?

CI as boundary objects, between analysis and advocacy, as:

- instruments of democratization of expertise;
- instruments of social discovery
- semiotic objects

Paul-Marie Boulanger, 2014, Elements for a comprehensive assessment of public indicators, Report EUR 26921 EN.
http://publications.jrc.ec.europa.eu/repository/bitstream/JRC92162/lbna26921enn.pdf

A triadic conception of the sign as structure connecting three elements: the sign properly said (S), an object (O) and an "interpretant"(I). But an example is needed …

Charles Sanders Peirce,
the father of semiotics
1839-1914

“This monkey possess a sophisticated repertory of vocal signs for signaling the presence of a predator [distinguishing a] terrestrial stalking one such as a leopard, an aerial raptor such as an eagle or a ground predator such as a snake.”

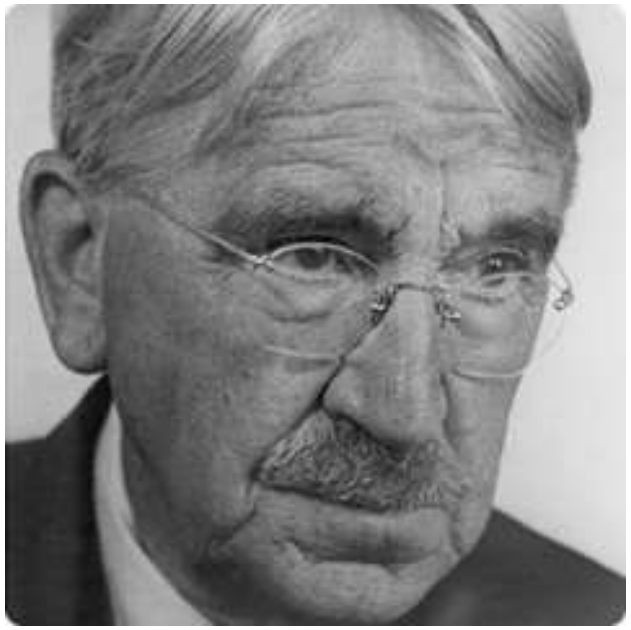
African vervet monkey
(Cercopithecus aethiops)

Sign ⟷ Cry

Object ⟷ Predator

Interpretant ⟷ Behaviour

Composite indicators as instrumental to the creation of a new public, through a process of social discovery (J. Dewey)



John Dewey
1859-1952

Dewey, J., 1938. The Public and its Problems, Read Book Ltd. Edition, 2013.

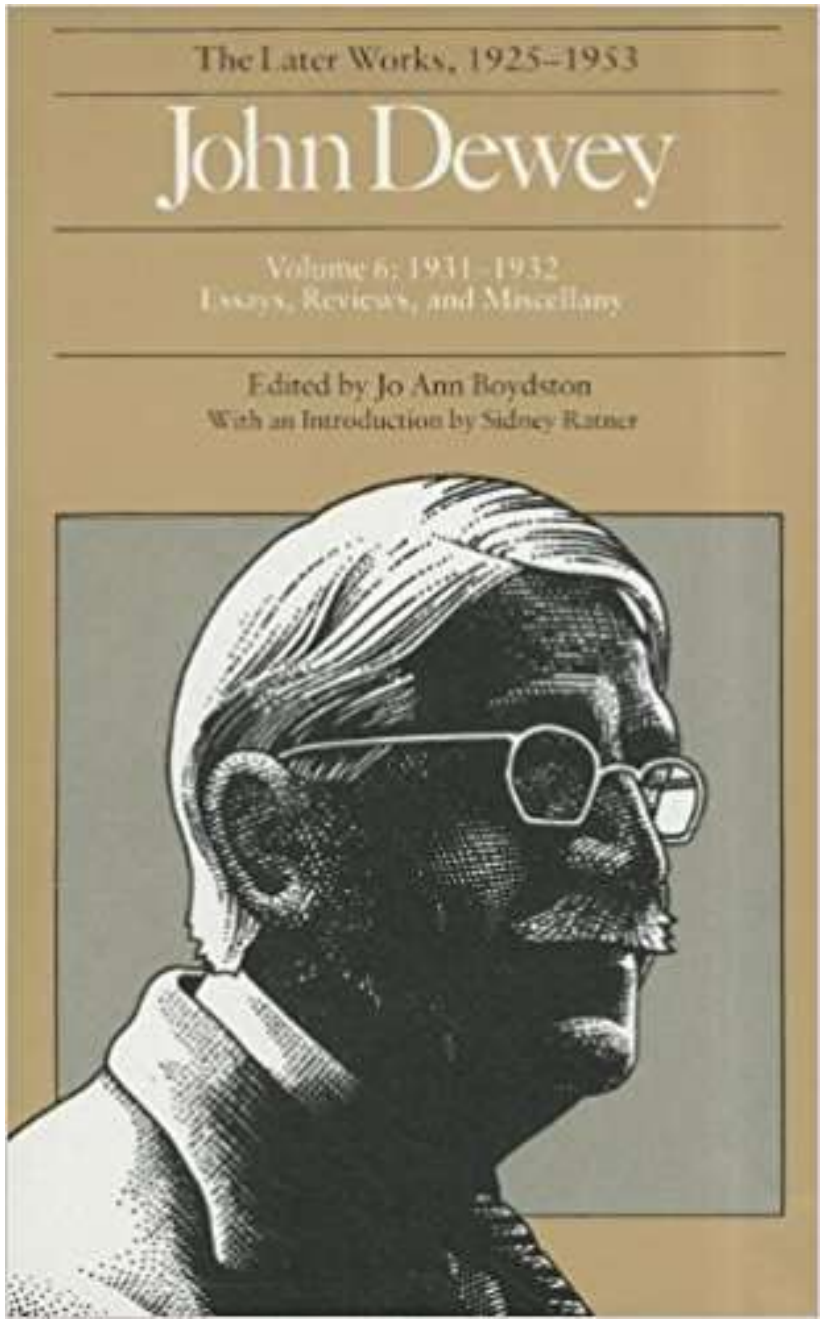
Why are ‘social discoveries’ needed?

Because there are ‘publics’ affected by transaction taking place somewhere else. “[…] machine age has so enormously expanded, multiplied, intensified and complicated the scope of the indirect consequences […] that the resultant public cannot identify and distinguish itself”

Dewey, J., 1938. The Public and its Problems, Read Book Ltd. Edition, 2013.

Social facts – unlike physical facts, are only meaningful in a context of desired ends

From J. Dewey ‘Social Science and Social Control’ in John Dewey: The Later Works, 1925-1953: 1931-1932, Vol. 6-ExLibrary,



Building a composite indicator can be seen as a process of social discovery for which a model of extended participation comes natural.
Frames and indicators are co-produced in the process which must be designed as to have a meaningful ‘interpretant’, or ‘end-in-sight’

Paul-Marie Boulanger, 2014, Elements for a comprehensive assessment of public indicators, Report EUR 26921 EN. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC92162/lbna26921enn.pdf

Critique of composite indicators: the Fitoussi-Stiglitz-Sen report

"a general criticism … frequently addressed at composite indicators, i.e. the arbitrary character of the procedures used to weight their various components … an aggregation procedure always means putting relative values on the items that are introduced in the index …

Jean-Paul Fitoussi,
Amartya Sen, Joseph Stiglitz

CMEPSP (2009). Commission on the Measurement of Economic Performance and Social Progress, URL: http://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+ Commission+ report, last accessed June 2017.

"The problem is not that these weighting procedures are hidden, non-transparent or non-replicable – they are often very explicitly presented by the authors of the indices, and this is one of the strengths of this literature. The problem is rather that their normative implications are seldom made explicit or justified"

Jean-Paul Fitoussi,
Amartya Sen, Joseph Stiglitz

CMEPSP (2009). Commission on the Measurement of Economic Performance and Social Progress, URL: http://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+ Commission+ report, last accessed June 2017.

Critique of composite indicators: Ravallion

There are types two indices: those built on economic theory / monetary aggregates / shadow prices and all others (=mashup indices)

Martin Ravallion

+ existing measures of e.g. development or poverty (Human Development Index, HDI, the Multidimensional Poverty Index, MPI) are bad at coping with tradeoffs

Martin Ravallion, 2010, Mashup indices of development, Policy Research Working Paper 5432 , The World Bank Development Research Group, http://documents.worldbank.org/curated/en/454791468329342000/pdf/WPS5432.pdf

To illustrate the distinction, consider two stylized examples of composite indices, both formed from the data on household assets and consumer durables found in the Demographic and Health Surveys (DHS). For index A the variables and their weights are set by the analyst, who has some concept of ―economic welfare‖ in mind, and thinks this is related to certain variables in the DHS, which are aggregated based on the analyst‘s judgments. For index B, the variables and weights are instead based on a regression model calibrated to another survey data set for which a comprehensive measure of consumption (though still containing measurement errors) could be derived. The model is calibrated to common variables in the expenditure survey and the DHS, and the regression model is used to predict wealth in the DHS.

A is a mashup index, B is not.

Martin Ravallion

Since composite indicators are here to stay how can we make them defensible?

Tools for evidence appraisal such sensitivity analysis and sensitivity auditing can be useful to gauge (and possibly deconstruct) measures

Sensitivity analysis

PDF

Info

Series A
Statistics in Society

Explore this journal >

Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators

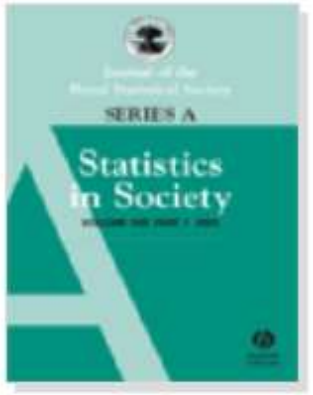
M. Saisana, A. Saltelli, S. Tarantola

First published: 3 March 2005 Full publication history

DOI: 10.1111/j.1467-985X.2005.00350.x View/save citation

Cited by (CrossRef): 181 articles Check for updates

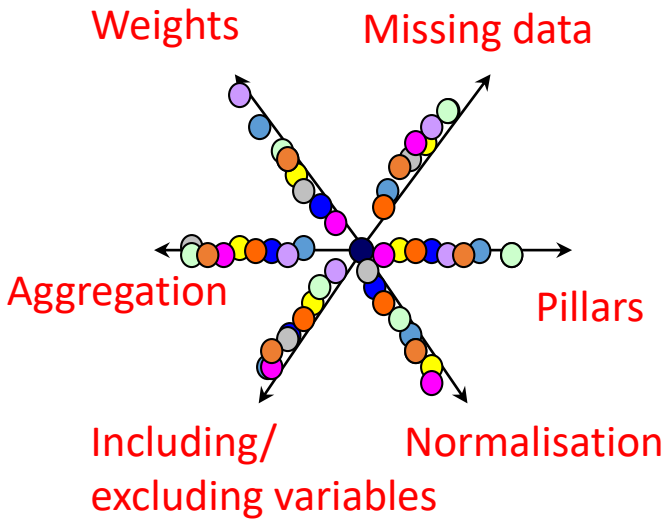
Citation tools

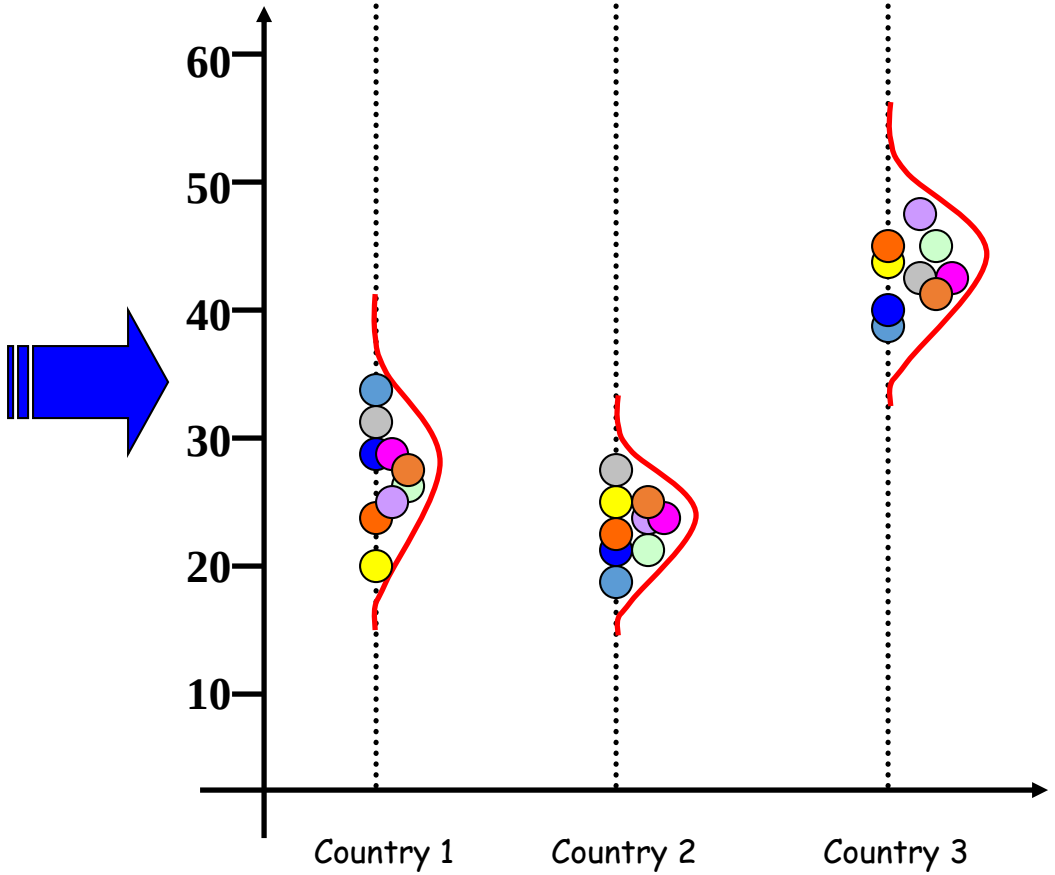


View issue TOC
Volume 168, Issue 2
March 2005
Pages 307–323

Assumption	**Alternatives**
Number of indicators	▪ all six indicators included or one-at-time excluded (6 options)
Weighting method	▪ original set of weights, ▪ factor analysis, ▪ equal weighting, ▪ data envelopment analysis
Aggregation rule	▪ additive, ▪ multiplicative, ▪ Borda multi-criterion

Space of alternatives





Using sensitivity analysis the volatility of country ranking can be exposed

Research Policy 40 (2011) 165–177

Contents lists available at ScienceDirect

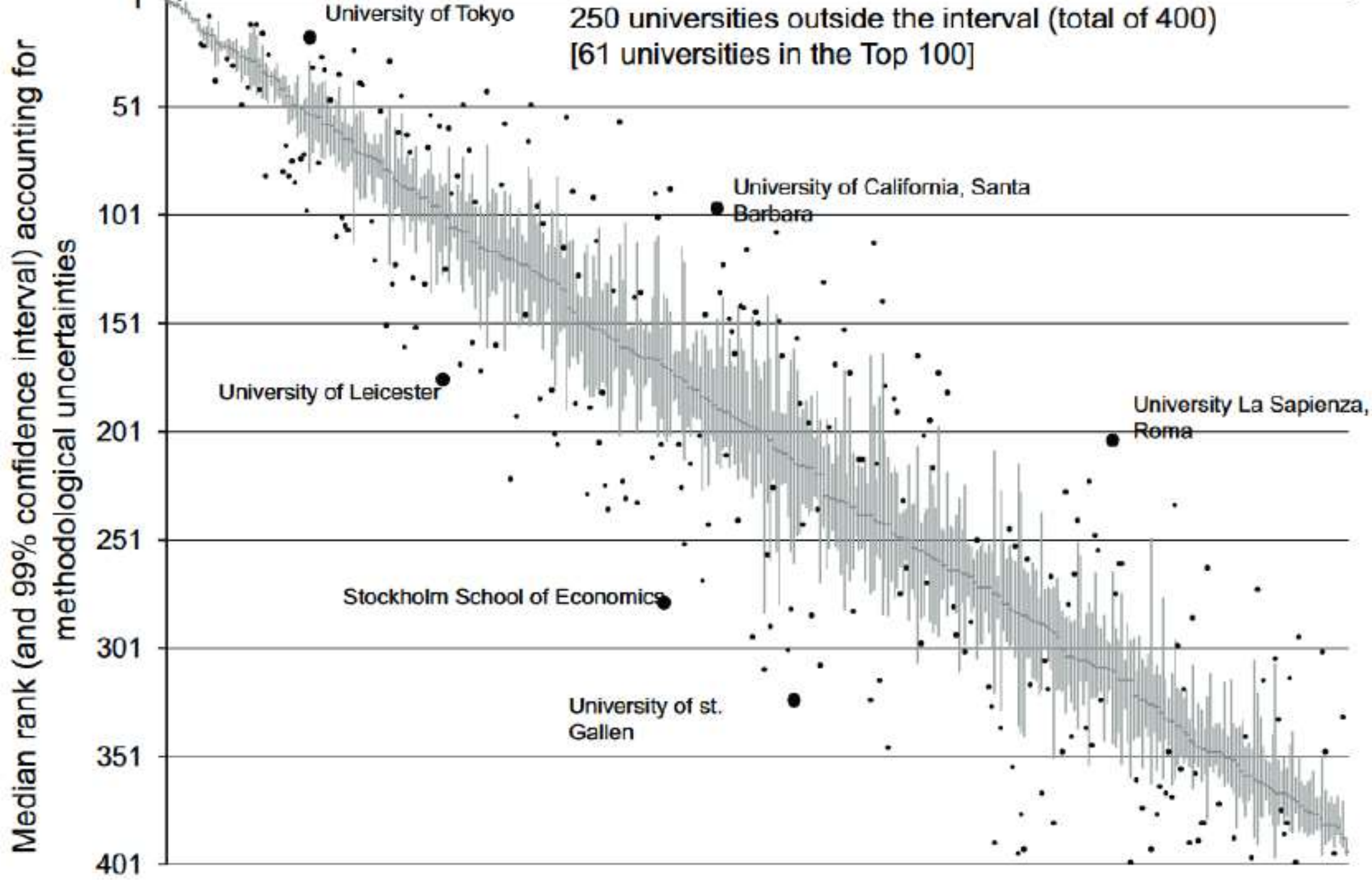
Research Policy

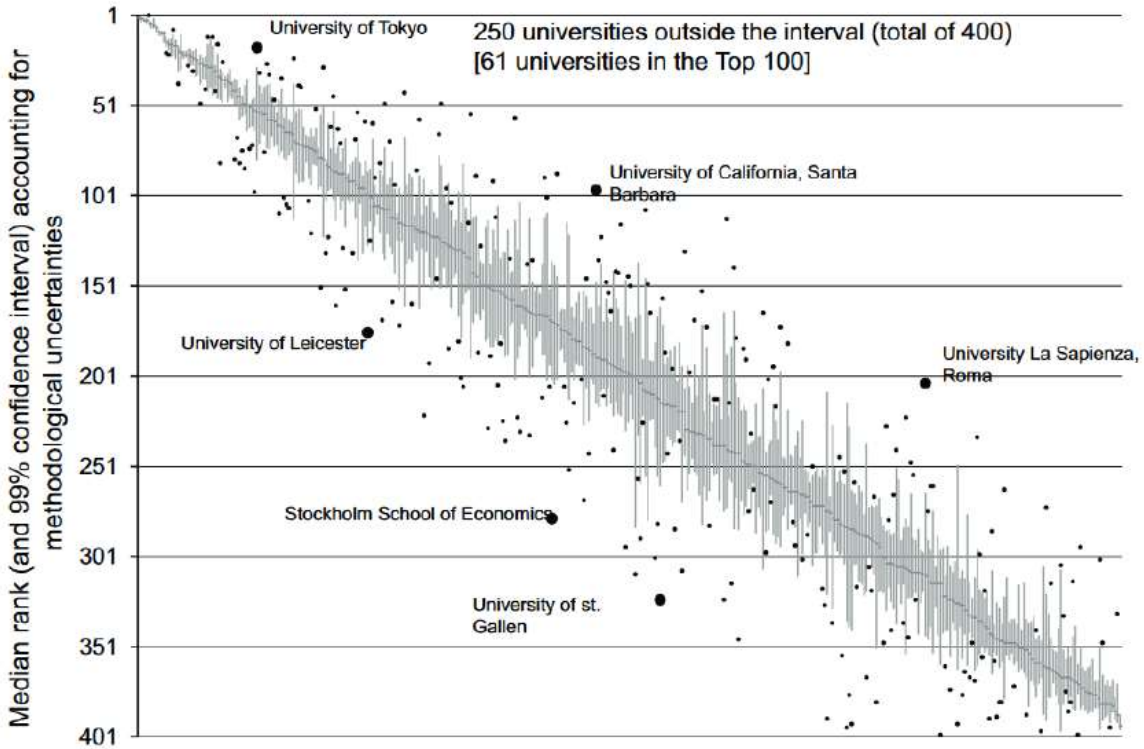
journal homepage: www.elsevier.com/locate/respol

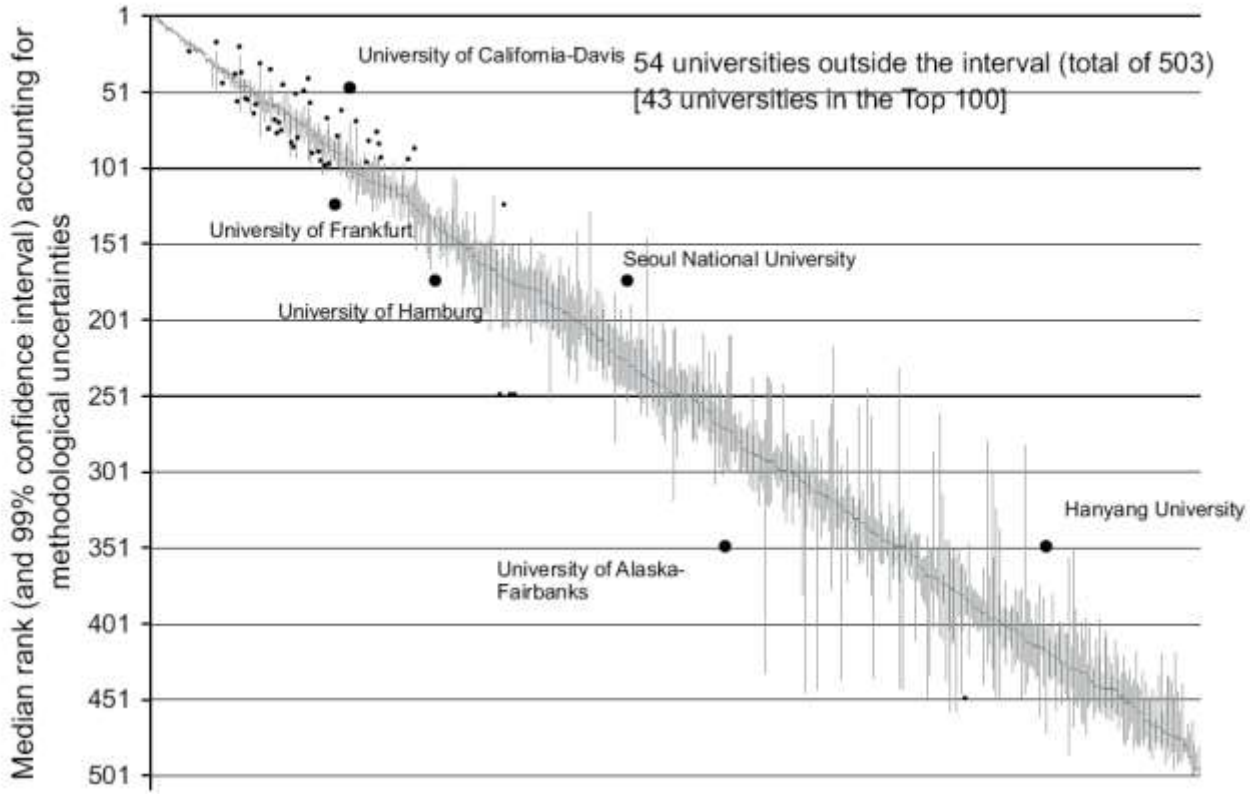
Rickety numbers: Volatility of university rankings and policy implications

Michaela Saisana*, Béatrice d'Hombres, Andrea Saltelli

Econometrics and Applied Statistics, Joint Research Centre, European Commission, Enrico Fermi 2749, 21027 Ispra, Italy


250 universities outside the interval (total of 400)
[61 universities in the Top 100]
University of Tokyo
University of California, Santa Barbara
University of Leicester
University La Sapienza, Roma
Stockholm School of Economics
University of st. Gallen
1
51
101
151
201
251
301
351
401
Median rank (and 99% confidence interval) accounting for methodological uncertainties


Median rank (and 99% confidence interval) accounting for methodological uncertainties
1
51
101
151
201
251
301
351
401
University of Tokyo
250 universities outside the interval (total of 400)
[61 universities in the Top 100]
University of California, Santa Barbara
University of Leicester
University La Sapienza, Roma
Stockholm School of Economics
University of st. Gallen


Median rank (and 99% confidence interval) accounting for methodological uncertainties
1
51
101
151
201
251
301
351
401
451
501
University of California-Davis
54 universities outside the interval (total of 503)
[43 universities in the Top 100]
University of Frankfurt
Seoul National University
University of Hamburg
Hanyang University
University of Alaska-Fairbanks

One can test whether assigned weights correspond to real importance

Journal of the Royal Statistical Society
SERIES A Statistics in Society

J. R. Statist. Soc. A (2013)
176, *Part* 3, *pp.* 609–634

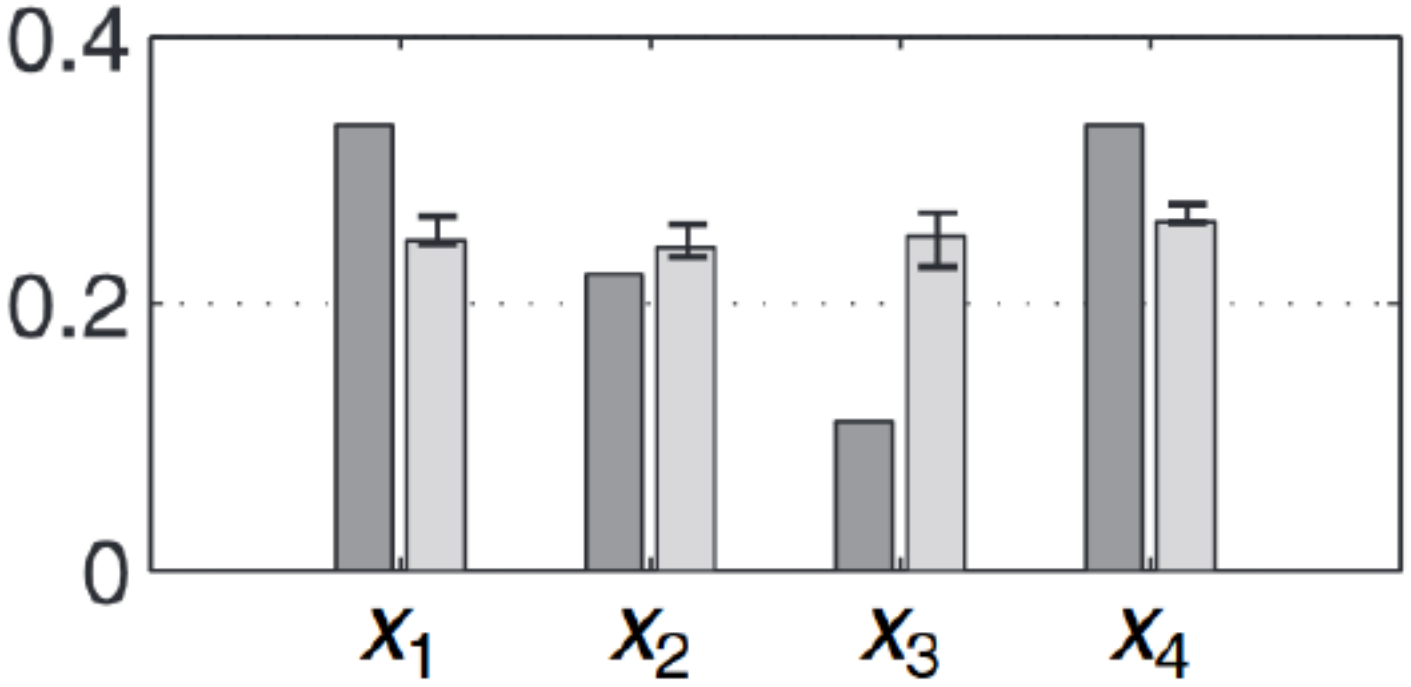
Ratings and rankings: voodoo or science?

Paolo Paruolo

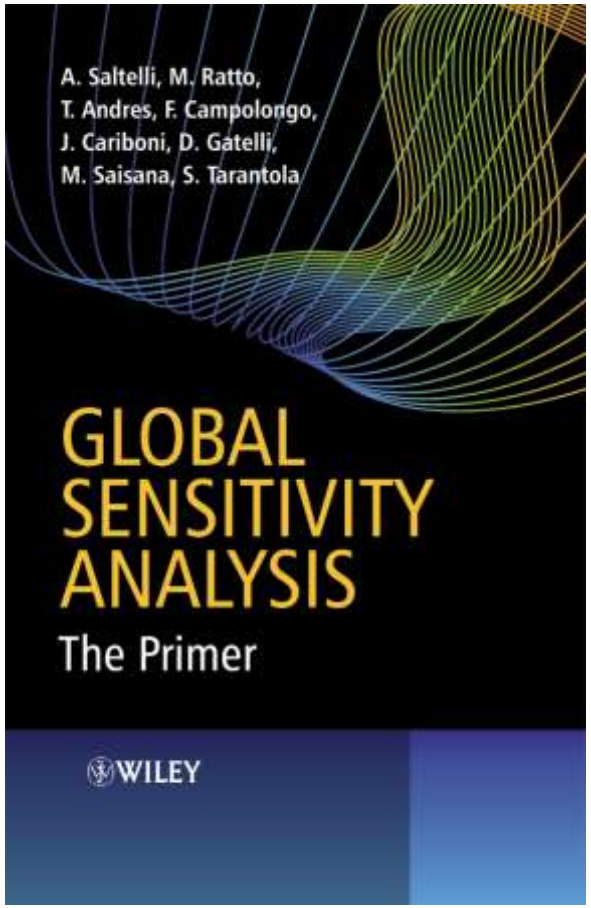
University of Insubria, Varese, Italy

and Michaela Saisana and Andrea Saltelli

European Commission, Ispra, Italy



Assigned weights (dark grey) versus measured importance for the 2010 HDI (Paruolo et al., 2011)



Sensitivity auditing

Int. J. Foresight and Innovation Policy, Vol. 9, Nos. 2/3/4, 2013 213

What do I make of your latinorum? Sensitivity auditing of mathematical modelling

Andrea Saltelli* and
Ângela Guimarães Pereira

European Commission,
Joint Research Centre,
Institute for the Protection and Security of the Citizen,
Via E. Fermi, 2749, 21027 Ispra (VA), Italy
E-mail: andrea.saltelli@jrc.ec.europa.eu
E-mail: angela.pereira@jrc.ec.europa.eu
*Corresponding author

Jeroen P. Van der Sluijs

Copernicus Institute of Sustainable Development,
Utrecht University (NL),
Heidelberglaan 2, 3584 CS Utrecht, The Netherlands
E-mail: j.p.vandersluijs@uu.nl

Silvio Funtowicz

Centre for the Study of the Sciences and the Humanities (SVT)
Allegaten, University of Bergen (NO),
34 – Postboks 7805 5020 Bergen, Norway
E-mail: silvio.funtowicz@svt.uib.no

Futures xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Futures

journal homepage: www.elsevier.com/locate/futures

Original research article

What is wrong with evidence based policy, and how can it be improved?

Andrea Saltelli[a,b,c,*], Mario Giampietro[a,c,d]

Conclusions: CI – instructions for use

Awareness of the imperfections and non-neutrality of measures

Investigate properties and assumptions (uncertainty and sensitivity analysis, sensitivity auditing)

Use for engaging the publics (social discovery), deliberative extended participation; quality as fitness for purpose (interpretant)

The End

@andreasaltelli

Some examples:

Sensitivity auditing: the OECD PISA study

Do PISA data justify PISA-based education policy?

International Journal of Comparative Education and Development
Vol. 19 No. 1, 2017
pp. 1-17

2396-7404
DOI 10.1108/IJCED-12-2016-0023

THE CONVERSATION

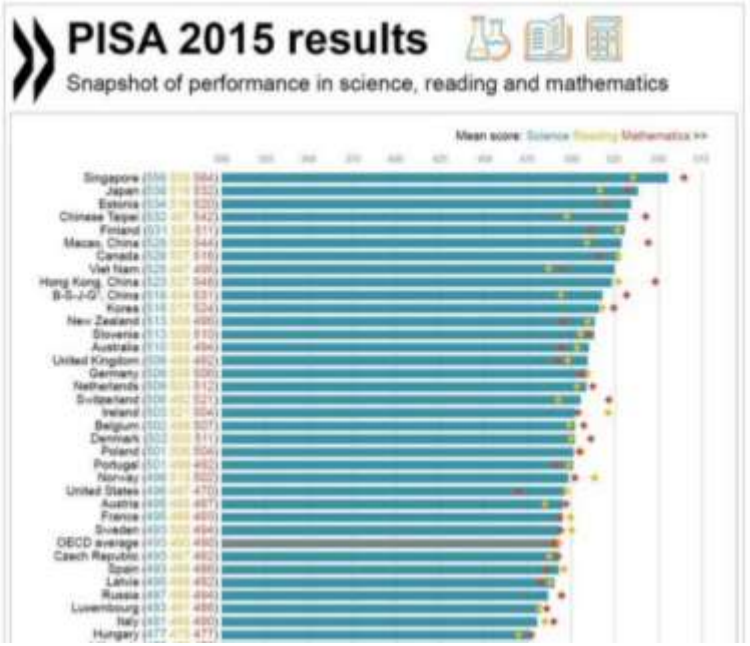
Academic rigour, journalistic flair

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

International PISA tests show how evidence-based policy can go wrong

June 12, 2017 3.55pm AEST

Chemistry class at the Dong Tien Secondary School, Thai Nguyen Province, Vietnam. Asian Development Bank/flickr, CC BY-SA



A condensed version of the article

With PISA the OECD gained the centre-stage in the international arena on education policies, which led to important controversies

http://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics

theguardian

OECD and Pisa tests are damaging education worldwide - academics

In this letter to Dr Andreas Schleicher, director of the OECD's Programme for International Student Assessment, academics from around the world express deep concern about the impact of Pisa tests and call for a halt to the next round of testing

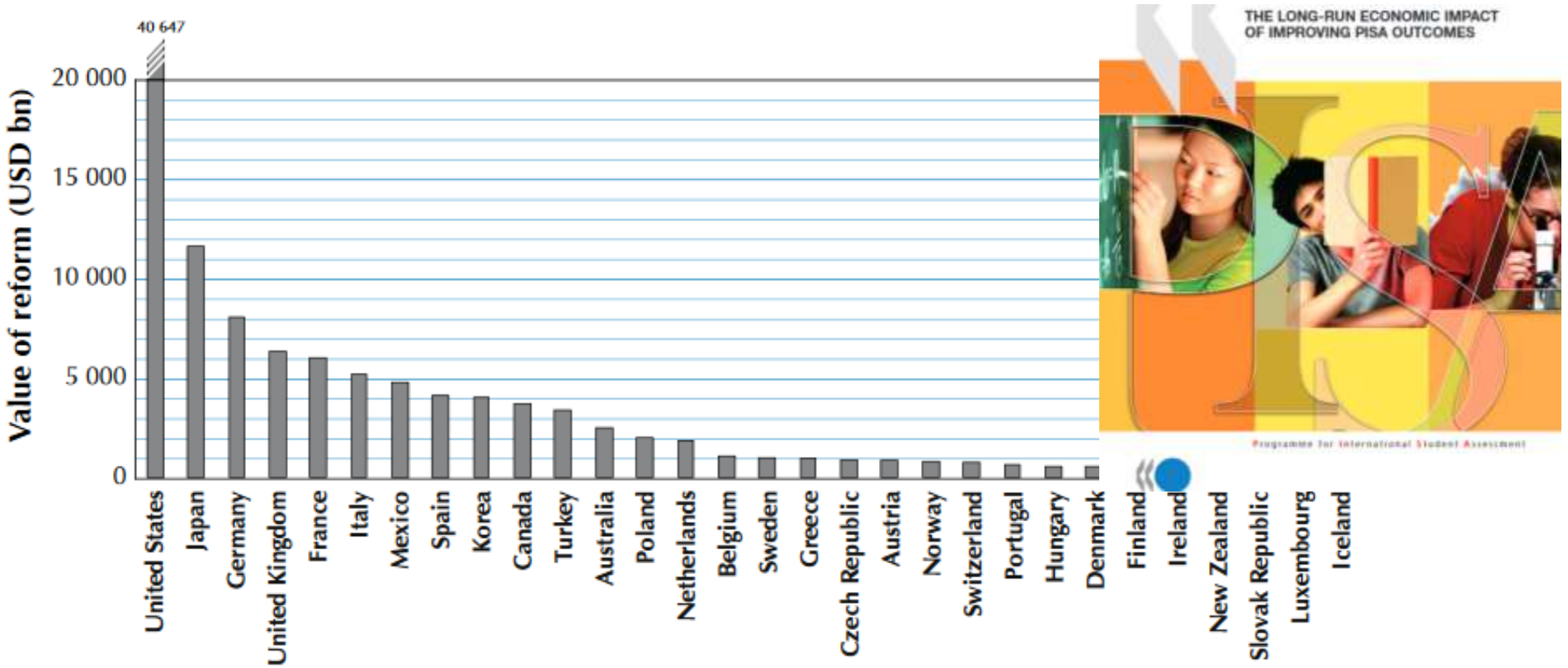
School children in Sichuan province in China. Academics say the OECD should develop alternatives to league tables and find more meaningful ways of reporting assessment, taking account of different cultures. Photograph: James Zeng Huang/Corbis Sygma

Critical remarks by the 80 signatories of the letter:

- Flattening of curricula (exclusion of subjects)
- Short-termism (teaching to the test)
- Promoting "life skills to function in knowledge societies"
- Stressing the student
- … ➔ Stop the test!
- A more participatory run of the study would be advisable

Figure 1

Present value of Scenario I (improve student performance in each country by 25 points on the PISA scale) in billion USD (PPP)



Note: Discounted value of future increases in GDP until 2090 due to reforms that improve student performance in each

http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/thehighcostofloweducationalperformance.htm

PISA's daring quantifications:

"If every EU Member State achieved an improvement of 25 points in its PISA score (which is what for example Germany and Poland achieved over the last decade), the GDP of the whole EU would increase by between 4% and 6% by 2090; such an 6% increase would correspond to 35 trillion Euro"

Woessmann, L. (2014), "The economic case for education", EENEE Analytical Report 20, European Expert Network on Economics of Education (EENEE), Institute and University of Munich.

Our study identifies both technical and normative issues:

1) Non response bias (what students are excluded; PISA non-response for England: the bias turned out to be twice the size of the OECD declared standard error in 2003.

2) Non open data, which makes SA impossible

Our study identifies both technical and normative issues:

3) Flattening curricula (do all countries wish to prosper by becoming knowledge societies?)
4) Power implications: power in the use of evidence. OECD (unelected officers and scholars) becoming a global super-ministry of education

Some examples:

Sensitivity analysis: the case of the Stern review

Global Environmental Change 20 (2010) 298–302

Contents lists available at ScienceDirect

Global Environmental Change

journal homepage: www.elsevier.com/locate/gloenvcha

Sensitivity analysis didn't help. A practitioner's critique of the Stern review

Andrea Saltelli *, Beatrice D'Hombres

Joint Research Centre, Institute for the Protection and Security of the Citizen, Ispra, Italy

The case of Stern's Review – Technical Annex to postscript

Nicholas Stern, London School of Economics

William Nordhaus, University of Yale

Stern, N., Stern Review on the Economics of Climate Change. UK Government Economic Service, London, www.sternreview.org.uk.

Nordhaus W., Critical Assumptions in the Stern Review on Climate Change, SCIENCE, 317, 201-202, (2007).

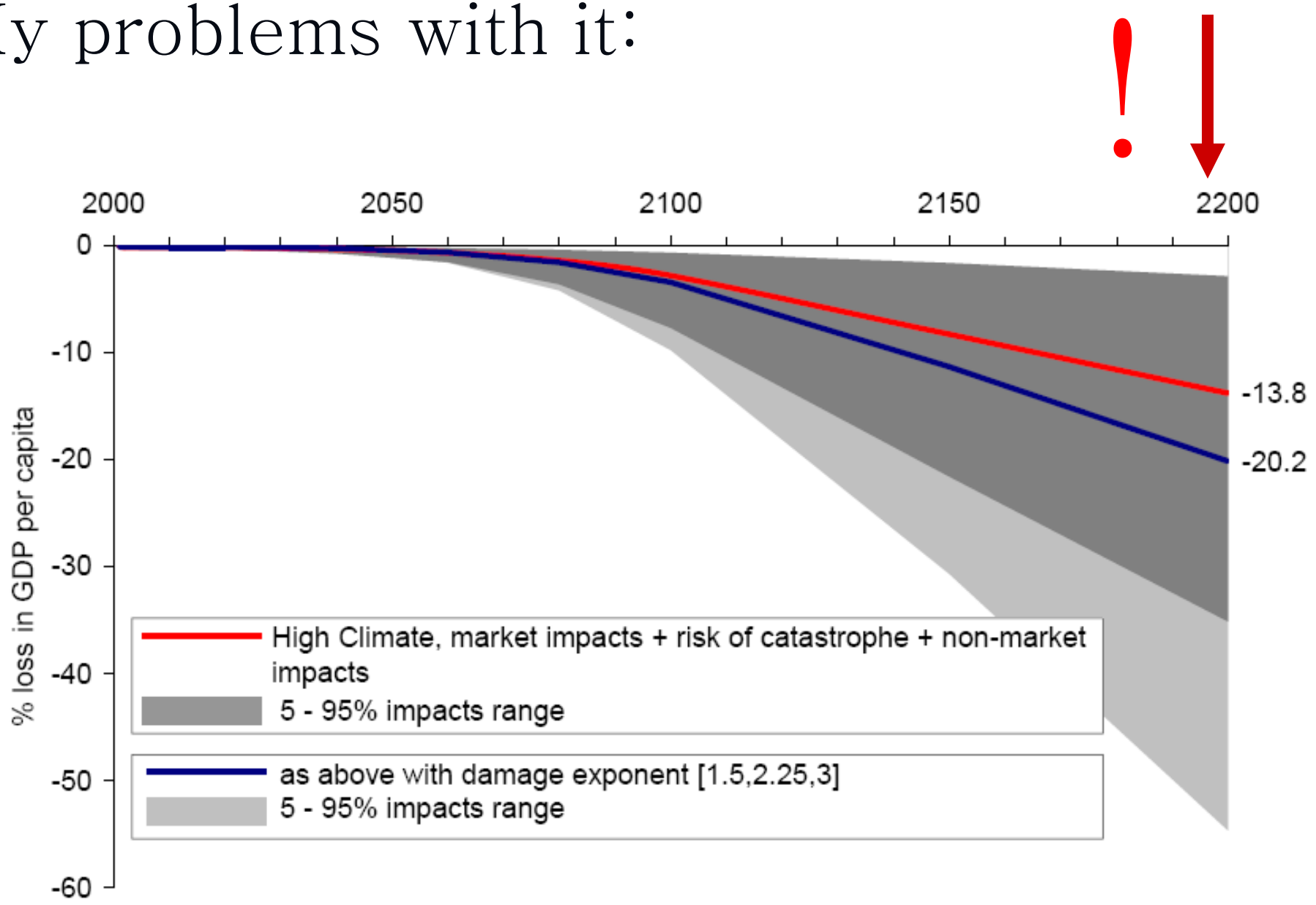
The Stern – Nordhaus exchange on *SCIENCE*

1) Nordhaus falsifies Stern based on 'wrong' range of discount rate

2) Stern's complements its review with a postscript: a sensitivity analysis of the cost benefit analysis

3) Stern infers: My analysis shows robustness'

My problems with it:

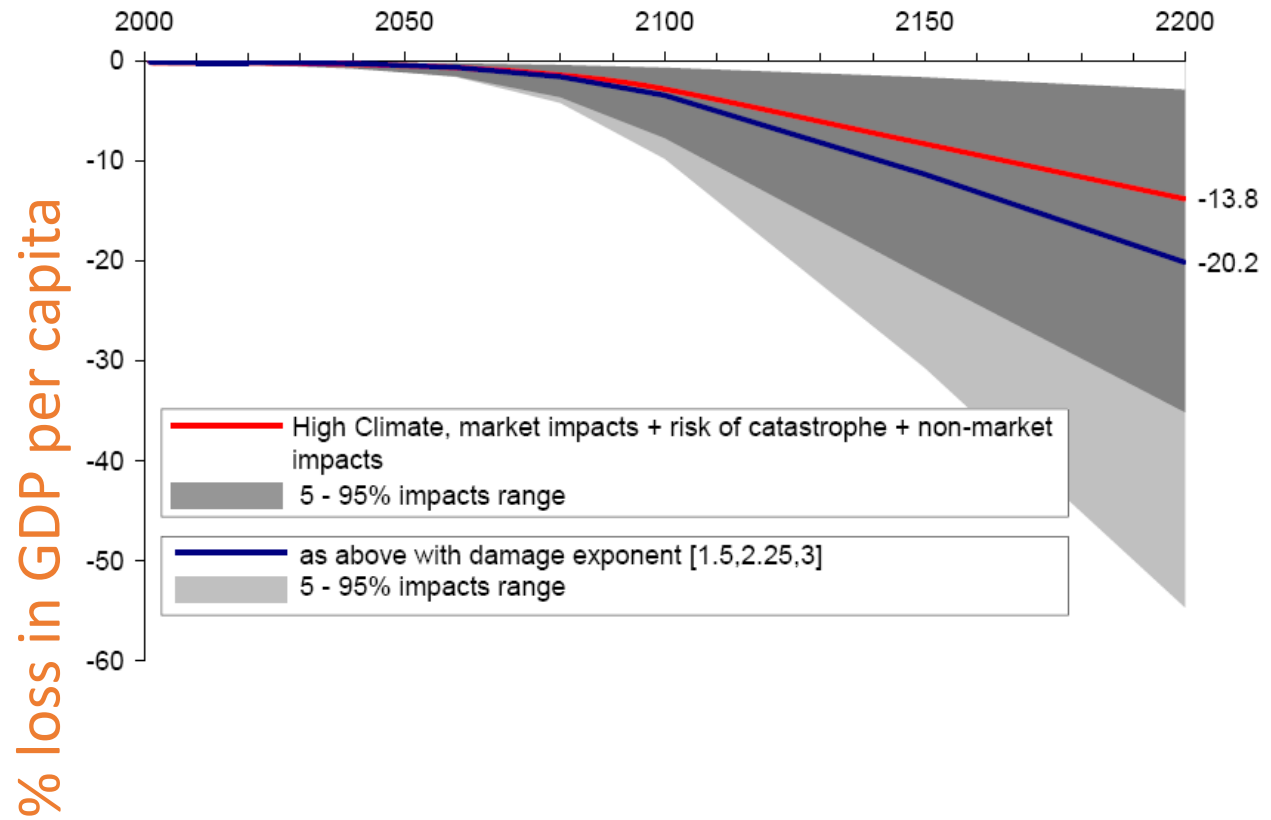


… but foremost Stern says:

changing assumptions → important effect

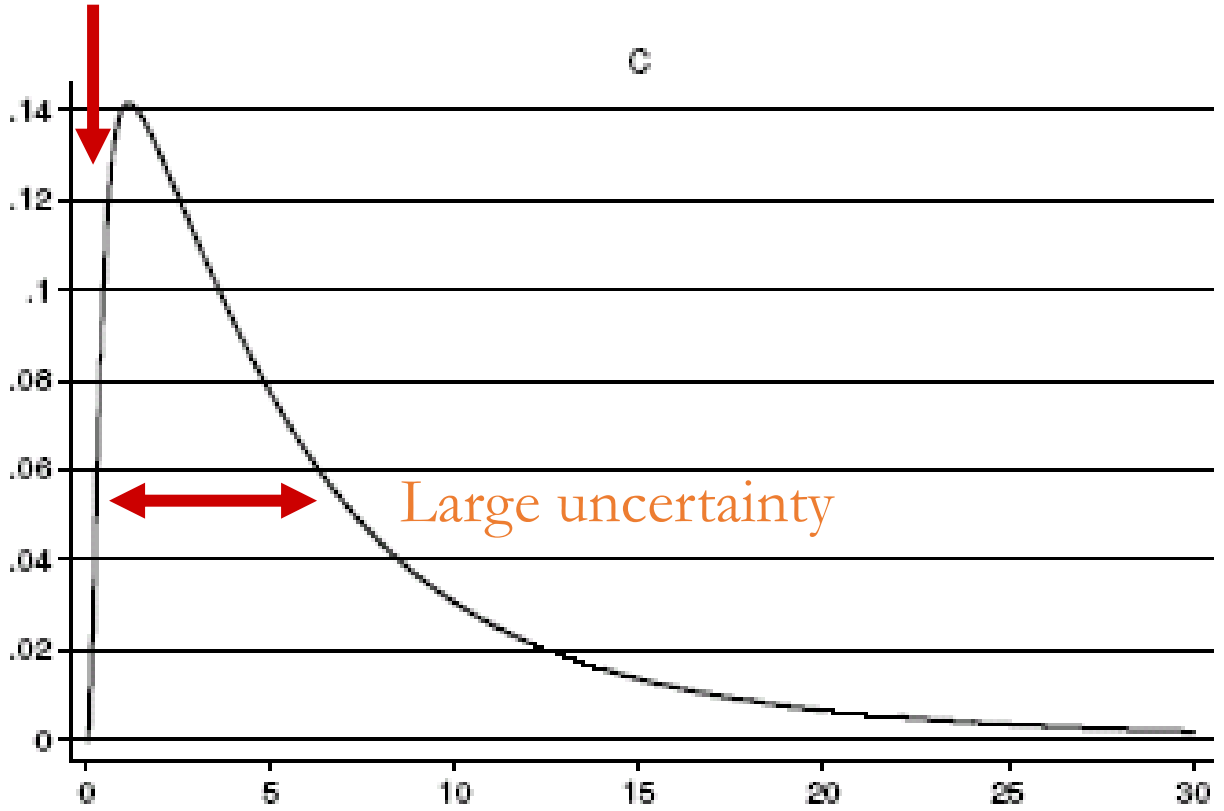
when instead he should admit that:

changing assumptions → all changes a lot

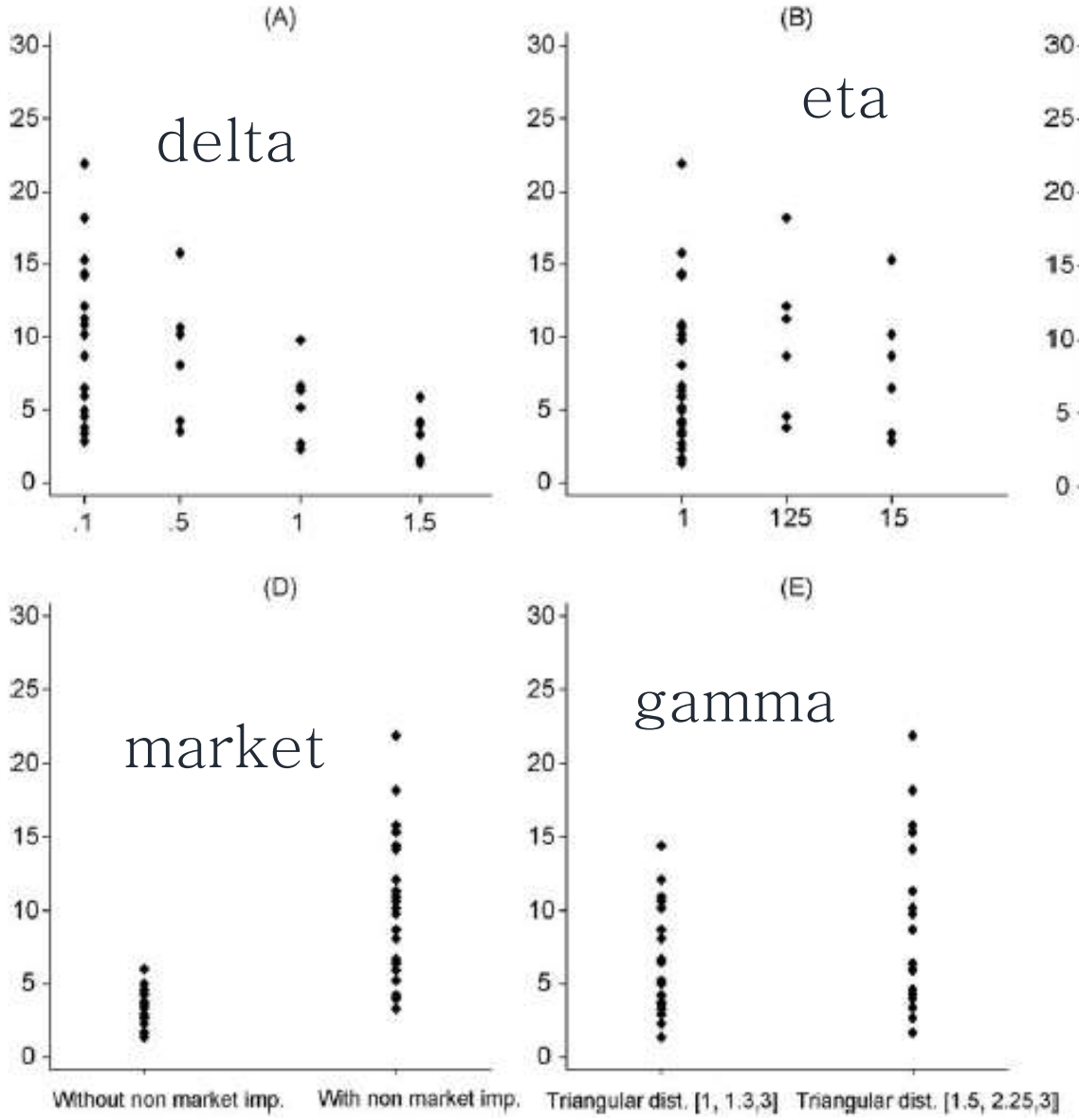


How was it done? A reverse engineering of the analysis

Missing points

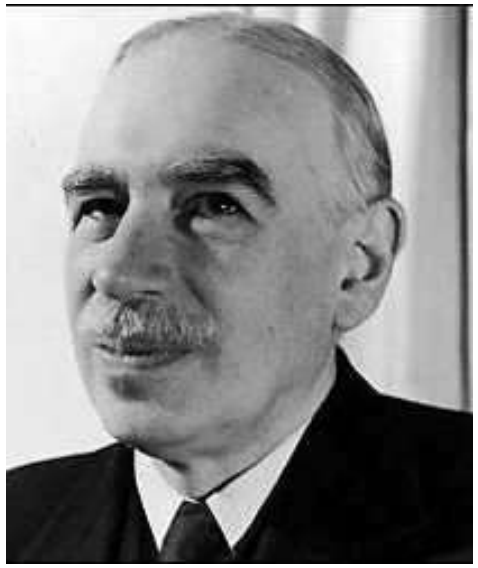


% loss in GDP per capita



Sensitivity analysis, also by reverse engineering

Same criticism applies to Nordhaus – both authors frame the debate around numbers which are …



… precisely wrong

From: Saltelli, A., D'Hombres, 2010, Sensitivity analysis didn't help. A practitioner's critique of the Stern review, *GLOBAL ENVIRONMENTAL CHANGE*, 20, 298-302.

The End

@andreasaltelli

Practicum

Grade a set of questions using a Likert scale

Likert scale

5. Strongly agree
4. Agree
3. Neutral
2. Disagree
1. Strongly disagree

A. Our duty is to provide objective numbers to policy makers. A cost benefit analysis is useful to make sure that taxpayer money is well spent.
B. Given proper statistical tools it is always possible to arrive at a number quantifying our present state of knowledge.
C. Numbers should be objective and not the result of 'stealth advocacy'.
D. Numbers can convey a misleading impression of accuracy and precision.
E. The analyst should strive to highlight the difference between risk and uncertainty.
F. The analyst should strive to identify different values underpinning different framing of the issue.

Practicum in sensitivity auditing

"What follows is a hypothetical executive summary from an imagined Food and Agriculture Organization (FAO) report on the state of the world's food systems, written from the perspective of the 2050s"

https://www.thesolutionsjournal.com/article/pathways-leading-sustainable-healthy-global-food-system/

Executive Summary: FAO State of World Agriculture in 2050 Draft Report

"[…]this FAO report presents evidence that the international food system of the second half of the 21st century is more sustainable than the food system of the late 20th or early 21st centuries.

[…] today more people are being fed on less land and agriculture is requiring fewer inputs"

Executive Summary: FAO State of World Agriculture in 2050 Draft Report

"[…] despite there being 10 billion people on the planet, today agriculture requires 438 million hectares* less land than it did in 2015, yet produces more adequate nutrition for all."

*Authors' estimate

This [438 Mha] figure was arrived at by assuming that:

- Agriculture shifts away from over production of cereals, oils, and sugars, but increases fruit and vegetables;
- Agricultural yields increase ~1%/y between now and 2050.
- Protein consumption shifts from 86% animals and 14% plants to 50% animal and 50% plant.

"Please contact the authors for references etc. pertaining to these calculations"

END

Our study:

- Gain in number of hectares: three significant digits (438 millions)?

- Balancing hectares growth and population growth (our computation) results in no change in food per capita at planetary scale.

Our study:

- Neglect of diminishing returns and ecosystem stress (fertilizers, pesticides)
- More adults (higher caloric intake) in 2050 population
- Can one educate citizens globally? The case of tobacco

In conclusion the

“mismatch between what the world needed for everyone to enjoy a nutritious diet and what the world was actually producing”

is the substitution of a political problem with a technical one

Food ethics (2017) 1:173–179
DOI 10.1007/s41055-017-0020-6

DISCUSSION PAPER

Problematic Quantifications: a Critical Appraisal of Scenario Making for a Global 'Sustainable' Food Production

Andrea Saltelli[1,2,3] · **Samuele Lo Piano**[1]

Accepted: 4 August 2017 / Published online: 15 August 2017

Reformation or resistance?