Uncertainty of science – is there a need for a new science policy?

Magnus Gulbrandsen

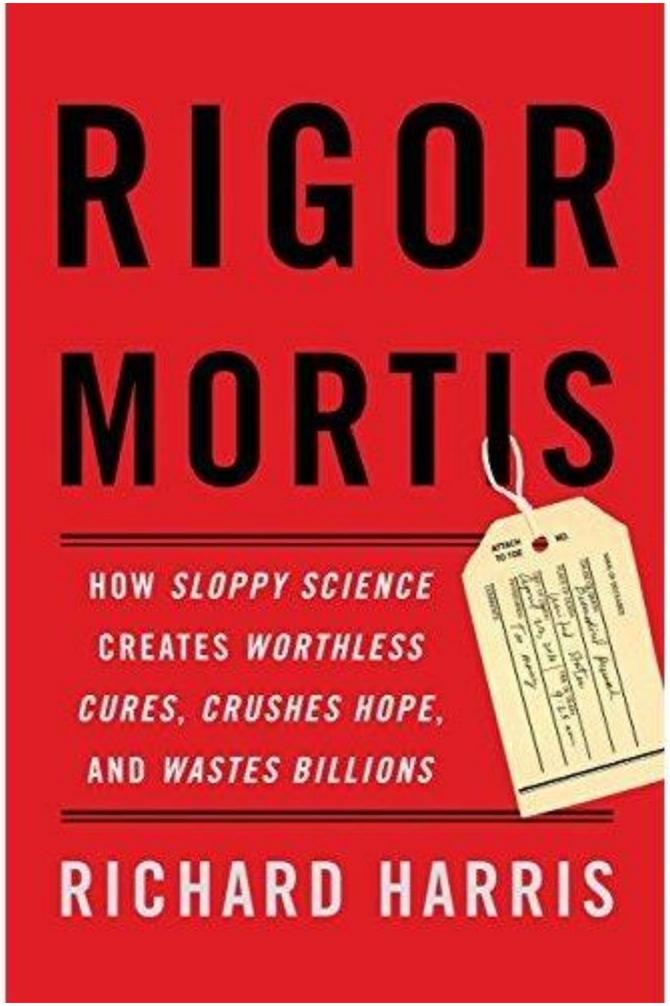
TIK Centre for Technology, Innovation and Culture, University of Oslo, Oslo
Institute for Research on the Impact of Science (OSIRIS)

Presentation at the 6th Research Policy Group meeting Oslo, 26 Oct 2017

Uncertainty of science?

- Science is by definition uncertain – oriented at new questions and puzzles
- Science is an elaborate mechanism of quality control – ensuring that our stock of knowledge is made more credible over time
- Recent claims: this mechanism *does not work* as well as it should; there are major problems with modern science making it more uncertain and less credible than it should be

Harsh critique



THE NEW ATLANTIS
A JOURNAL OF TECHNOLOGY & SOCIETY

CURRENT ISSUE | ARCHIVE | SUBSCRIBER SERVICES | A

Morgan Ray Schweitzer

Saving Science

Science isn't self-correcting, it's self-destructing. To save the enterprise, scientists must come out of the lab and into the real world.

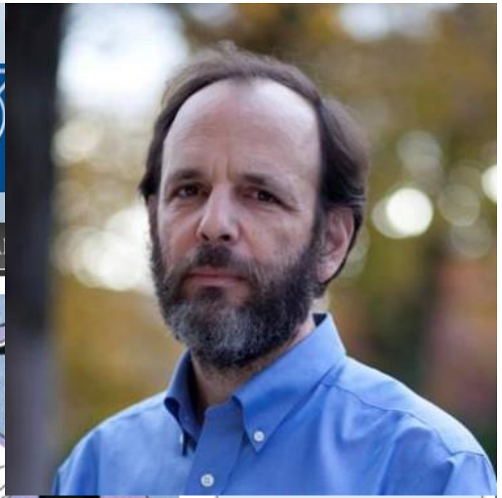
Daniel Sarewitz

Science, pride of modernity, our one source of objective knowledge, is in deep trouble. Stoked by fifty years of growing public investments, scientists are more productive than ever, pouring out millions of articles in thousands of journals covering an ever-expanding array of fields and phenomena. But much of this supposed knowledge is turning out to be contestable, unreliable, unusable, or flat-out wrong. From metastatic cancer to climate change to growth economics to dietary standards, science that is supposed to yield clarity and solutions is in many instances leading instead to contradiction, controversy, and confusion. Along the way it is also undermining the four-hundred-year-

- Editor's Note
- **Daniel Sarewitz** on why scientists must come out of the lab and into the real world
- **Ivan Oransky and Adam Marcus** praise the growing scrutiny of scientific publications
- **Barbara A. Spellman** on the role of technological and demographic changes

[Publication of this special section was made possible through the support of a grant to *The New Atlantis* from the John Templeton Foundation; the opinions expressed are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.]

PDF version



"Science isn't self-correcting, it's self-destructing"

Science seems to be ill

- These and other commentators argue that something is fundamentally wrong with modern science

- But what is wrong with it, exactly?
- What are the causes of this illness?
- What can be done about it? Do we need a new science policy – and what kind of policy?

- These are the main topics of my talk today

The diagnosis

Lack of replicability

nature International weekly journal of science

NATURE | COMMENT

Drug development: Raise standards for preclinical cancer research

C. Glenn Begley & Lee M. Ellis

Affiliations | Corresponding author

Nature **483**, 531–533 (29 March 2012) | doi:10.1038/483531a
Published online 28 March 2012
Clarification (May, 2012)

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

Subject terms: Cancer • Drug discovery • Publishing

Efforts over the past decade to characterize the genetic alterations in human cancers have led to a better understanding of molecular drivers of this complex set of diseases. Although we in the cancer field hoped that this would lead to more effective drugs, historically, our ability to translate cancer research to clinical success has been remarkably low[1]. Sadly, clinical trials in oncology have the highest failure rate compared with other therapeutic areas. Given the high unmet need in oncology, it is understandable that barriers to clinical development may be lower than for other disease areas, and a larger number of drugs with suboptimal preclinical validation will enter oncology trials. However, this low success rate is not sustainable or acceptable, and investigators

- A leading US biotechnology firm tried to repeat 53 “potentially groundbreaking” academic investigations
- Only in six of the cases did they get similar results
- Later studies have confirmed low replicability also outside of medical science
- Twisting the problem: failed rather than missing replicability

Statistical problems

- A number of investigations have shown that severe statistical problems are widespread
- False positives, small samples, use of wrong methods, misinterpretation or manipulation of p values (“p-hacking”) and much more

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific 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. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Published research findings are sometimes refuted by subsequent evidence, with ensuing confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims [6–8]. However, this should not be surprising. It can be proven that most claimed research findings are false. Here I will examine the key factors that influence this problem and some corollaries thereof.

The Essay section contains opinion pieces on topics of broad interest to a general medical audience.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a p-value less than 0.05. Research is not most appropriately represented and summarized by p-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

should be interpreted based only on p-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. “Negative” research is also very useful. “Negative” is actually a misnomer, and the misinterpretation is widespread. However, here we will target relationships that investigators claim exist, rather than null findings.

As has been shown previously, the probability that a research finding is indeed true depends on the prior probability of it being true (before doing the study), the statistical power of the study, and the level of statistical significance [10,11]. Consider a 2 × 2 table in which research findings are compared against the gold standard of true relationships in a scientific field. In a research field both true and false hypotheses can be made about the presence of relationships. Let R be the ratio of the number of “true relationships” to “no relationships” among those tested in the field. R

is characteristic of the field and vary a lot depending on whethe field targets highly likely relatio or searches for only one or a fe true relationships among thous and millions of hypotheses that be postulated. Let us also consi for computational simplicity, circumscribed fields where eith is only one true relationship (a many that can be hypothesized) the power is similar to find any several existing true relationshi pre-study probability of a relatio being true is $R/(R+1)$. The pr of a study finding a true relatio reflects the power $1 - \beta$ (one m the Type II error rate). The pro of claiming a relationship when truly exists reflects the Type I e rate, α. Assuming that c relatio are being probed in the field, t expected values of the 2 × 2 tab given in Table 1. After a researc finding has been claimed based achieving formal statistical signi the post-study probability that i is the positive predictive value, The PPV is also the complemen probability of what Wacholder have called the false positive re probability [10]. According to t × 2 table, one gets PPV = (1 – β – βR + α). A research finding is

Citation: Ioannidis JPA (2005) Why most p research findings are false. PLoS Med 2(8):

Abbreviation: PPV, positive predictive val

John P. A. Ioannidis is in the Department of and Epidemiology, University of Ioannina S Medicine, Ioannina, Greece, and Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States of America. E-mail: jioannid@cc.uoi.gr

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020124

JOURNAL OF ECONOMIC SURVEYS

doi: 10.1111/joes.12032

WHAT’S TO KNOW ABOUT THE CREDIBILITY OF EMPIRICAL ECONOMICS?

John Ioannidis
Stanford University

Chris Doucouliagos
Deakin University

Abstract. The scientific credibility of economics is itself a scientific question that can be addressed with both theoretical speculations and empirical data. In this review, we examine the major parameters that are expected to affect the credibility of empirical economics: sample size, magnitude of pursued effects, number and pre-selection of tested relationships, flexibility and lack of standardization in designs, definitions, outcomes and analyses, financial and other interests and prejudices, and the multiplicity and fragmentation of efforts. We summarize and discuss the empirical evidence on the lack of a robust reproducibility culture in economics and business research, the prevalence of potential publication and other selective reporting biases, and other failures and biases in the market of scientific information. Overall, the credibility of the economics literature is likely to be modest or even low.

Keywords. Bias; Credibility; Economics; Meta-research; Replication; Reproducibility

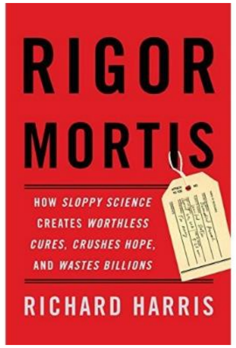
1. Introduction

Research is a public good produced by unavoidably self-interested researchers. The credibility of research and the market for evidence have been critically questioned (see Ioannidis, 2005, 2012a; Young *et al.*, 2008). Is evidence distorted? Is wrong evidence produced at a faster rate than correct evidence? Is research produced and replicated efficiently enough? Can we do better? These and other questions apply as much to economics as they do to other sciences. Over the years, investigators have identified several problems that affect the credibility of empirical economics research. These include but are not limited to: publication

“The credibility of the economics literature is likely to be modest or even low.”

Wider methodological challenges

- Problems with contamination of cell samples, labelling, transport, sharing, independent testing and authentication
- Problems with equipment, such as indications that a large share of fMRI based studies could be false
- More severe integrity problems: 10-fold increase in retractions in recent years
- Problems with citation patterns: studies that have been proven wrong keep getting cited



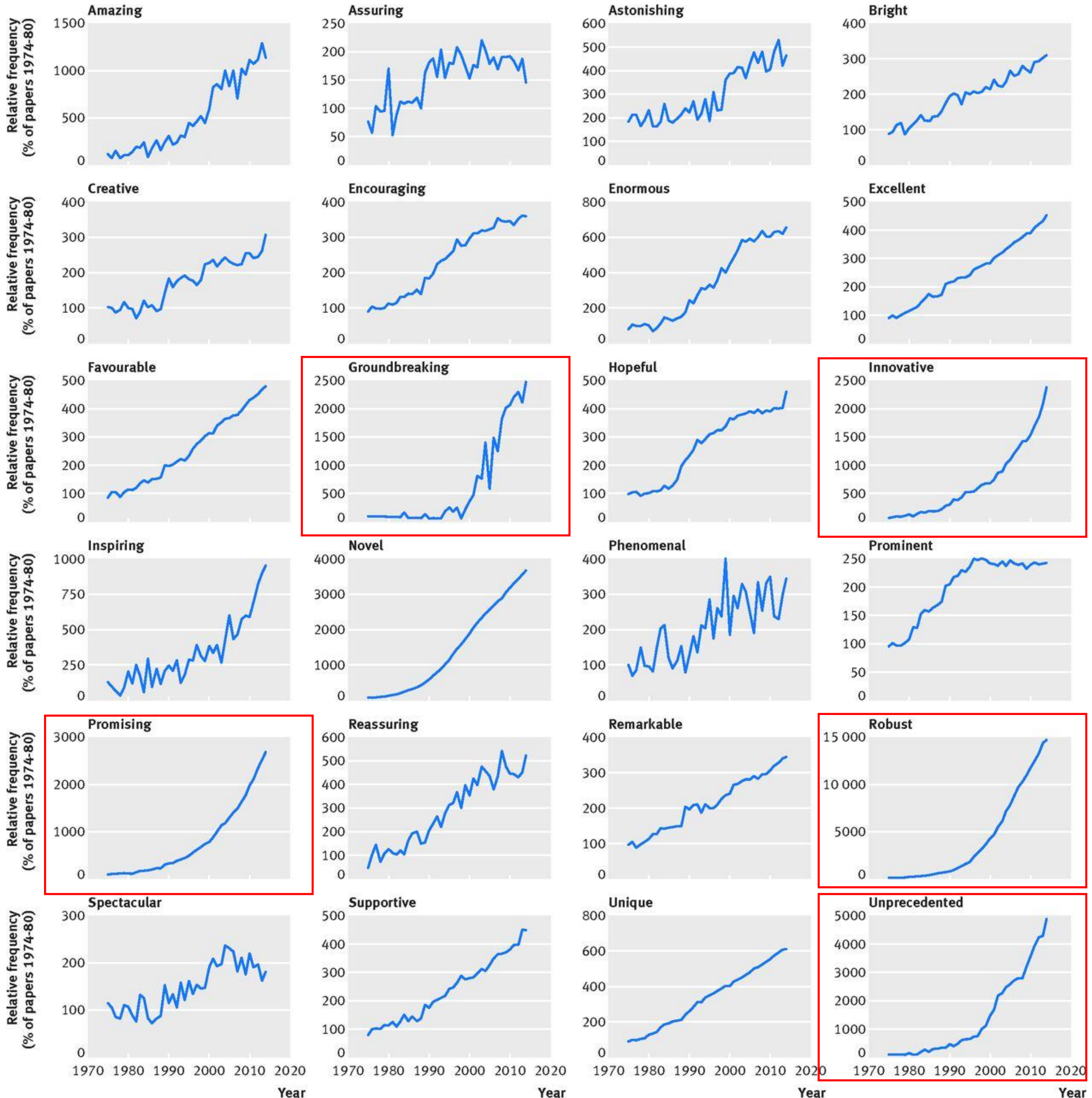
Paradigm problems

- Medical science: change from working closely with patients to hunting for the basic mechanisms of disease in the laboratory – few treatments from this (medical practice has also changed)
- Frequent problems with transferring results from animal tests to humans
- From context-sensitive data to “big, detached data” in many fields (“datageddon” rather than insight)

Language

- Inflation in the “voice of God” language in scientific publications and applications
- Harder to make realistic assessments of the contribution to the stock of knowledge

Source: Vinkers et al., BMJ 2014



Summary: the disease

- Various problems with methods, theories and the paradigms and practices of scientific work…
- … may signify widespread challenges of work standards in science, leading to major problems of credibility

- If we (for now) accept that this is a fair depiction, what are the causes of this disease?
- What are the cures?

The causes

Main explanations

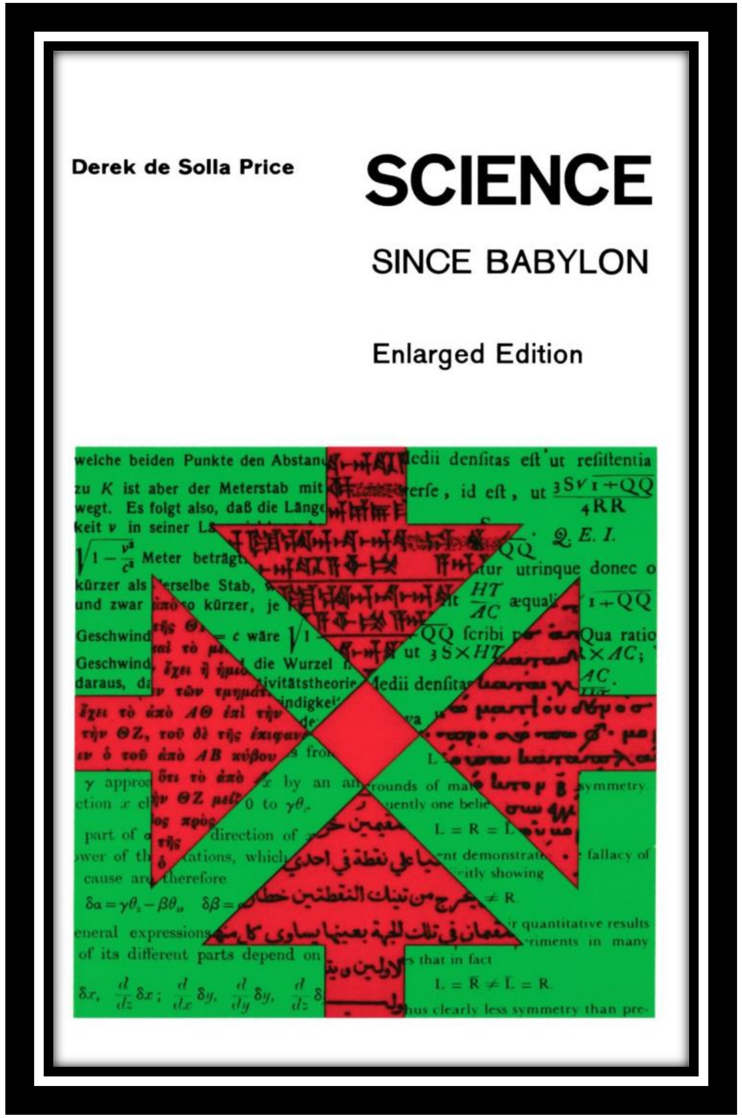
- **Cynical scientists** – researchers engage in dubious practices to gain personal advantages such as funding or scholarly recognition
- **Research system pressure** – academia is hyper-competitive and attractive positions are in short supply (“Publish or perish”)
- **Outside pressure** – increased competitive funding also means strong incentives to publish fast, early and frequently
- **Publication system** – the journals have had weak traditions for supporting openness, data sharing, retractions and so on
- **Peer review** – some signs that the peer review system, at the heart of competitive funding and publication, is under pressure

Sarewitz: detachment

- These problems become worse when science is detached from society
- "Scientific knowledge advances most rapidly (…) when it is steered to solve problems — especially those related to technological innovation"
- "When science is not steered to solve such problems, it tends to go off half-cocked in ways that can be highly detrimental to science itself"

Joint effects, size

- Researchers like Ioannidis view the causes of replicability and other problems as complex and interrelated
- Policy decisions and internal decisions in the research system may reinforce one another in a bad way
- The growth and size of the science system a wicked challenge in itself as it means data and publication overload – one of the “diseases of science” in de Solla Price (1961)



The cure(s)

Back to mission-oriented R&D

- Sarewitz: “Science will be made more reliable and more valuable for society today not by being protected from societal influences but instead by being brought, **carefully and appropriately**, into a direct, open, and intimate relationship with those influences”
- He prescribes a return to a type of research that is tied more directly to missions and to actors that have a clear responsibility for these missions
- His ideal: technological research funded by the military in the US (but could be organised through other actors in society, not least in a wider European context)

“But if your constituency … is society, not scientists, then the choice of what data and knowledge you need has to be informed by the real-world context of the problem to be solved. The questions you ask are likely to be very different if your end goal is to solve a concrete problem, rather than only to advance understanding. That’s why the symbiosis between science and technology is so powerful: the technology provides focus and discipline for the science.”

Many examples

- We see the same in Norway and elsewhere: some of the most important and radical impacts have come from researchers working in a mission-oriented or applied context

The GSM system for telecommunication

- Technology developed at the Telecom Research Institute and SINTEF (our largest technological research institute, an RTO or PRO)
- All-digital system well adapted to Norwegian geography which posed complex challenges
- Global utility, especially in neighbouring countries!

Aftenposten A-magasinet Osloby Sport Meninger Magnus Meny

Skapte revolusjon innen mobiltelefoni

GUNHILD M. HAUGNES

OPPDATERT: 30.OKT.2012 11:09 | PUBLISERT: 22.OKT.2012 12:02

Torleiv Maseng regnes for å være GSM-nettets far. Han ledet den lille forskergruppen som ved SINTEF på 80-tallet utviklet det som ble den europeiske standarden for radiodelen på det med tiden globale digitale mobilnettet GSM. Her er han fotografert hos Forsvarets Forskningsinstitutt i 2004, hvor han de siste årene har vært forskningssjef.

FOTO: Øhman, Rolf

De nordiske landene opparbeidet seg høy mobilkompetanse på feltet gjennom utviklingen av det analoge systemet NMT (nordisk

Multiphase flow

- A panel of scientists placed this as the most important research-based innovation in Norway the last 50 years
- R&D at the Institute for Energy Technology and SINTEF and related research and education at several universities
- Crucial technology for the Norwegian oil and gas industry based on huge practical challenges in the North Sea
- Benefits estimated at “hundreds of billions of NOK”

Rørene gjorde oljeeventyret mulig

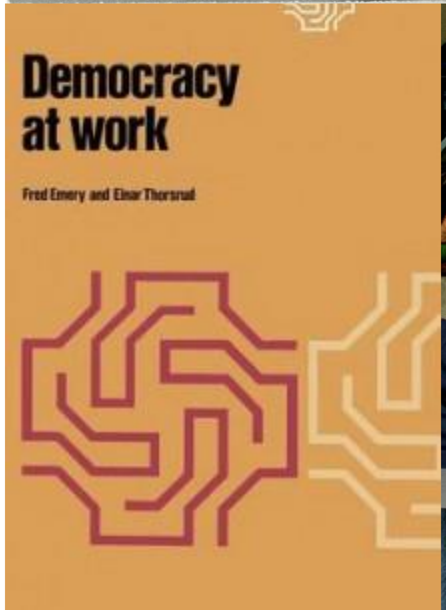
GRO STRØMSHEIM | GUNHILD M. HAUGNES
OPPDATERT: 29.OKT.2012 11:49 | PUBLISERT: 29.OKT.2012 07:45

Laboratoriet på Tiller utenfor Trondheim er enormt. Rørene er laget i samme størrelse som de som blir brukt på havbunnen ved oljeplattformene, for å sikre at resultatene av forsøkene stemmer med virkeligheten.
FOTO: SINTEF

Flerfaseteknologien har gjort det mulig å bygge ut flere og mindre felt enn det ellers ville vært mulig.

Democracy in the workplace

- The “collaboration experiments” between unions and business associations, led by social psychology professor and work researcher Einar Thorsrud, have probably had large effects since the start in 1962
- Related to the “Nordic Model” of “flat” organisational structures, high degrees of collaboration etc.
- New legislation, the “Basic Agreement”, employee rights to influence adaption of new technology
- R&D normative and based on challenges of bureaucracy, alienation and routinisation



Characteristics of this research

- Tied to a concrete societal challenge or problem – interaction with users in industry, healthcare and society in a wide sense – and often with a lead user with money, power and competence to put research results into use
- Long-term and with significant "core funding" and most of the time a relatively high degree of autonomy
- Driven by curiosity, high ambitions and often carried out in a combination of non-academic research organisations and universities
- Protected from short-term political priorities and the logic of the market – and often with other application areas than the original one

Recap: uncertainty of science

- Worries that – as a whole – science is becoming less credible than it used to or should be, even the “elite” or “excellent” research published in the leading journals
- Different dynamics within and outside of the science system can help explain why these problems are on the rise
- Solutions favour policies that bring research organisations into somewhat closer and committed relations with society
- Sarewitz’ and others’ overall recommendation seems valuable, but close science-society relations are fairly common in Europe and in broader ways than what the US commentators prescribe

Pitfalls

- Worries of elite US life science not always transferrable to other settings
- The framing may be overly negative – science is still a major success story and a strong institution with self-correcting mechanisms
- Current policies may have unintended consequences that increase the problems, for example "open access" can solve some problems but increase other forms of uncertainty
- We know much less about mission-oriented and applied research, and large projects in these categories have also had major problems of uncertainty and waste of money
- Such research may require strong/specific users

Some points for discussion

- The argument is that science works best in *combination* with something else
- Do we then need a broader development of and more powerful ideas about the different hybrid natures of research organisations?
  - Research and teaching
  - Research and (public) missions
  - Research and contracts/markets
  - Research and stakeholder engagement, including RRI
  - Hybrid nature: the organisation of such research including overall funding pattern, career structure, competences and more
- Do we need to develop more and stronger independent institutions that synthesise and assess the credibility of science to help the “combination” effort?

Final challenge

- The argument is that good research is often a result of its usefulness rather than the cause of usefulness – partly because science is a great testbed for new ideas rather than the main source of ideas
- Useful/applied/mission-oriented research is often long-term and curiosity-driven, and many researchers carry out and are motivated by this form of research
- Challenge: this research may suffer under a dichotomous or polarised science policy that either supports an introspective notion of excellence or short-term practical benefits and impact – this is perhaps where a new science policy direction is needed

More information

Email: magnus.gulbrandsen@tik.uio.no

Web: http://www.sv.uio.no/tik/english/research/projects/osiris/

Twitter: @osiris_tik

Pictures from Wikipedia and public domain image databases