

“A formidable share of the research we pay for is simply wrong.”

By JON KÅRE TIME, appeared in MORGENBLADET, No. 22/8–14 June 2018

Translated from the Norwegian by Heidi Sommerstad Ødegård, AMESTO/Semantix Translations Norway AS,

Torture of data, perverse reward systems, declining morals and false findings: Science is in crisis argues statistician Andrea Saltelli.

“There’s been a lot of data torturing with this cool dataset,” wrote Cornell University food scientist Brian Wansink in a message to a colleague.

Wansink was famous, and his research was popular. It showed, among other things, how to lose weight without much effort. All anyone needed to do was make small changes to their habits and environment. We eat less if we put the cereal boxes in the cabinet instead of having it out on the counter. But after an in-depth investigation by BuzzFeed earlier this year, Wansink has fallen from grace as a social science star to a textbook example of what respectable scientists do not do.

“Data torturing” is also referred to as “p-hacking” – the slicing and dicing of a dataset until you find a statistical context that is strong and exciting enough to publish. As Wansink himself formulated it in one of the uncovered emails, you can “tweak” the dataset to find what you want. “Think about all the different ways you can cut the data,” he wrote to a new grad student about to join his lab.

It’s a crisis! “Is science in crisis? Well, it was perhaps debatable a few years ago, but now?”

Andrea Saltelli seems almost ready to throw up his hands in despair at the question. He’s an Italian chemist and statistician with a background as the head of the European Commission’s “Econometrics and Applied Statistics Unit.” Today, he’s a guest researcher at the Centre for the Study of the Sciences and the Humanities at the University of Bergen. Lately, he specialises in what he believes are serious problems with the reliability of science, especially how the statistics and research used to justify political decisions, are, on closer scrutiny, often more uncertain and less impartial than the impression we are given.

The issue is recognisable in Norwegian social debates. Should we trust Fisheries Minister Per Sandberg when he, backed by the Institute of Marine Research and the Norwegian Food Safety Authority, says we can safely eat farmed salmon? Is Statistics Norway’s “immigration

account” a suitable basis for policy? And is it really as simple as the research shows that more teachers don’t have an “effect”?

Issues of this nature are what Saltelli and his colleagues discuss in the book *Science on the Verge* (2017). In the book, Saltelli draws a picture of a serious and comprehensive system crisis.

This statistician from Rome has levelled withering criticism, claiming in a lecture that “the pathologies” exhibited by science can be likened to the traffic in indulgences that fomented Martin Luther’s rage against the Roman Catholic Church in the 1500s.

The trouble with marshmallows. But hey, wait a minute! How can we talk about a “crisis” at the dawn of the gene-editing age? When scientists create enzymes that eat plastic and finally begin to get the hang of robots. When children can be born with three parents. Just recently (last year!), astrophysics gained measurable evidence that Albert Einstein’s theory about gravitational waves was right. Oh yes, Saltelli answers.

“Science is still delivering miracles big-time. My point is that a formidable share of the research we pay for is quite simply wrong. Science has lost the ability to self-police its production, and it doesn’t seem as if society has any idea what it should do,” says Saltelli.

Among other things, it’s about the uncertainty that has spread in recent years through parts of the world of science, which both the media and research literature have dubbed the “reproducibility crisis” or “the replication crisis”: It turns out that a lot of research is difficult or impossible for other scientists to review accurately. And when independent scientists try to reproduce previous experiments, many have been frequently amazed that the findings can’t be verified.

At the heart of the current debate about the crisis is psychology, where the discussion has to some degree been harrowing. The final chapter was written last week when the famous Marshmallow experiment got a shot across the bow: In this influential experiment, the self-control of four-year-olds was tested to see if they wait a short while to eat a treat. If they managed they would receive a second treat a little later.

A major attempt at replicating the study nevertheless indicates that the popular idea that there is a correlation between the ability of small children to exercise self-restraint and their success later in life may be false (see sidebar). But the discussion on reproducibility is also ongoing in everything from cancer research and medicine to economics and sports science.

THE MARSHMALLOW TEST FALLS

The so-called Marshmallow experiment, which tested the ability of 4- to 6-year-olds to delay gratification by seeing if they could manage to wait a short period to eat a treat in exchange for being rewarded with a second a little later, has become famous and influential. Not the least, this is because the test could effectively predict how the children would fare later in life. For many, attempting to instill self-restraint in young children seemed like a very good idea.

An attempt has now been made by other researchers to recreate the experiment in extensive tests with many more study subjects. Like before, the researchers also found an effect on adolescent math and reading skills, but it was only half the size, and disappeared when they controlled for circumstances such as family background. They also found that the test could not predict other types of behaviour or how personalities developed. Published in the journal *Psychological Science*, the new results may mean that efforts intended to make children better at delaying gratification are of little value.

Two classic experiments in research on social priming, which deals with how we are affected by subtle signals in our surroundings, also recently “failed.” One showed how we more easily interpret acts by others as unfriendly after being prepared (primed) with stories about a hostile environment, who in the

This is “a significant crisis.” Fifty-two per cent of scientists agreed that this was the case when the journal Nature conducted a (controversial) interdisciplinary survey on the topic in 2016, which is often referred to in the debate. Fully 90 per cent agreed that the crisis “exists.”

Studies with few participants, secrecy of data, poorly designed experiments and misunderstood use of statistical analysis are some of the explanations for the phenomenon. Not least, many people point to target-oriented management and the pressure to publish in academia as the root cause: Do the reward systems create overproduction of articles and an abundance of wasted or bad research? Is it too tempting to take shortcuts to get more or more sensational articles in print?

Democratic problem. In Norway, biologist Dag O. Hesse expresses concern about the conditions for free research in his most recent book, *Sannhet til salg* (Truth for Sale). He warns against increased political control, market and utilitarian thinking and harmful

publication points. Not to mention the explosion of rogue journals and fake science.



39.2%

The share of researchers who state that they have been pressured by research heads or partners to produce “positive” results, according to a study in Clinical Cancer Research.

According to Saltelli, everything correlates – the reproducibility crisis is just the tip of a massive iceberg.

“Science itself is compromised,” he says.

The poor quality control will ultimately undermine democracy, Saltelli claims.

Modern society rests on a well-functioning marriage of science and political power, where research legitimises power by standing as a guarantor of truth.

“Dark forces are entering because a void has been created. Society is vulnerable. It’s like a kind of ecosystem, where predators take over because other species are weakened.”

The inexplicit message of numbers. The news went around the world in 2015: 7.9 per cent of all the species on Earth will die out due to climate change.

“That’s ridiculous. For how can you predict this, right down to the decimal, when we don’t even know how many species we have today?” asks Saltelli.

Excessive trust placed in numbers and models is part of the reason for people’s distrust of science, he believes. Mathematical modelling is being misused to create all kinds of amazing constructions: Actually, it is often the case that series of uncertain assumptions are given a numerical value and put into a model. The result the computer grinds out is taken to be fact, but the fact hides several layers of scientific uncertainty and valuations, which thus become invisible in the political debate.

GLOSSARY

Replication:

When scientists repeat other scientists' experiments to see if a similar result can be obtained.

Reproducibility:

Can mean that the way in which research is conducted is documented in such a way that it is possible for others to examine the analysis or to repeat the experiment.

P-hacking:

When more data is added, or the dataset is analysed in different ways until a correlation is found that is statistically significant ($p < 0.05$). This is usually the key to getting a finding published, even though the p-value

itself proves little. Pre-registration of experiments may counteract p-hacking.

H-index:

Popular metric of a researcher's influence, based on how much the researcher has published and how many times they have been cited by other researchers.

Post-normal science:

An approach to how we can use research and new technologies by emphasising scientific uncertainty, especially when a lot is at stake and the values that are to steer the decisions are unclear. The approach has been embraced by the Centre for the Study of the Sciences and the Humanities at the University of Bergen.

“You read sentences in scientific articles that read ‘with a sea temperature that is predicted to increase by 2.5 degrees by 2100...’ and it feels like it’s a matter of fact. But who knows? It can be 5 degrees or 1 degree or whatever. The system is so complex that any attempt at predictions over such a time span is hubris. Yet such numbers have become part of the public debate,” says Saltelli.

Statistics Norway’s now discredited immigration account, the report that deals with the impact of immigration on the state budget up to the year 2100, may perhaps be another example: Not only is the calculation based on highly uncertain population projections, it also does not take into account that the labour market adapts to wages and prices. And the comparative basis on which the entire report is based is a very hypothetical Norway where the boundaries are closed and the population actually declines without it having any negative financial ramifications.

The problem is that there is uncertainty on top of uncertainty on top of uncertainty?

“Any sensible scientist would be very careful to make such claims. But unfortunately, scientists now use this strategically to defend their positions, to promote their agenda in public,” says Saltelli.

Denial. There is an “attitude of denial” to the crisis in scientific institutions and in research policy, Saltelli believes. He claims the reproducibility issue is not even mentioned in key research policy papers and documents that deal with how research should influence politics.

“If you want to use research to give advice on policy, can you ignore being concerned about the fact that science is experiencing this crisis in quality control? It makes no sense, so this behaviour is strategically defensive. The silence speaks volumes.”

Saltelli then lets loose with: The whole notion of knowledge-based (evidence-based) policy is insane. All too often, research becomes, in practice, an instrument that legitimises predetermined political purposes. Knowledge from research becomes a currency that can be used to buy political clout, he has written.

“People with deep pockets can gain control over more research-based knowledge. They can promote it more aggressively than others and use it as a framework that everyone else must deal with.”

Never clean. Saltelli is one of several researchers at the University of Bergen who market an approach to science, technology and the use of research in society called “post-normal science.” It entails being open about uncertainty, complexity and disagreement, especially when values are at stake and the risks are great.



People with deep pockets can gain control over more research-based knowledge.

ANDREA SALTELLI

Saltelli believes that far too many researchers still cling to the idea that research is neutral and objective truth – ideas that also live on in popular culture. But he believes that more scientists now experience a conflict between this romantic vision and what he believes is the reality.

“For example, it could be because they work for commercial companies or have to defend their skin or career or whatever. This conflict is one of the reasons for the collapse in trust in science and its reputation.”

Can it be fixed? After looking at how scientists use statistics, John Ioannidis at Stanford University claimed as early as 2005 that most published research findings must be incorrect. Ioannidis is a leading figure in the field of meta-research, in other words research about research. He has also estimated that 85 per cent of the money that goes to biomedical research is “wasted.” Since then, he has done similar critical studies on several disciplines, most recently in *Nature* a few weeks ago where he was one of the co-authors of an article on brain research on gender differences, where they point out systematic weaknesses in the research and conclude that gender differences may in reality be less than current scientific literature tells us.

Ioannidis is one of those who have now taken the lead in the effort to find good solutions to the problems with unreliable results. They include tightening methodologies and facilitating more open data sharing, multiple replications and pre-registration of studies. But such

measures are not enough to deal with the “crisis,” Saltelli believes. It is too extensive to be fixed by giving doctoral students courses in statistics or by making slight adjustments to publication points.

“I have great respect and admiration for my colleagues who are working on this. There aren’t many of them and they need all the encouragement they can get. But I still believe that the rabbit hole goes deeper,” says the statistician.

How deep?

The monster in the rabbit hole. “If we dig deep enough, this is about the fact that we have a quality control system in science that was appropriate when research was carried out in smaller communities of experts. Today, however, we can talk about mega science that is largely aimed at promoting innovation and growth.” Saltelli says that, on average, two million articles are published in 30,000 journals each year. This “monster” cannot obviously not be controlled by the same gentlemen’s agreements, and with the same ethos as when everyone knew everyone, he believes.

“When there are few players and they know each other, you can have behaviour that is more constructive than in the virtually Hobbesian situation we are now in where there are millions of us, and everyone needs to make a living and promote their careers.”

Do not read this! “Where’s the crisis? There is no crisis. Science is a never-ending process.”

Daniele Fanelli is a researcher at the London School of Economics and himself a prominent figure within “meta-science,” i.e. research on research.

“It’s ironic that part of the core evidence basis for that there is indeed a reproduction crisis, if you take a dispassionate look at it, does not show this to be the case at all,” says Fanelli via Skype.

He would like to properly lay out his views on the so-called crisis: “How much time do you have?” he asks.

Fanelli has written one of several articles in the journal PNAS which now warns against newspaper articles such as this one. After the difficulties with reproducing research results in psychology and medicine became known, articles about science that is in crisis, or even “broken,” have appeared multiples times both in scientific literature and specialist science reporting. Stories about fraud and immorality, erroneous, stupid and unreliable research are captivating reading material. But critics believe the crisis-maximisation undermines confidence in science and makes it easier for those who, for ideological reasons, would like to discredit research areas, for example in the case of vaccines, genetic modification or climate change.

Fanelli reviews the knowledge base for the crisis claims in an article. When you look closely, you can’t find support for the dramatic conclusion, he believes. To the contrary. Referring to several new and successful attempts to repeat earlier research, Fanelli points out studies that

indicate that unethical or dubious research practices are not more common than before. For example, the spike in the number of articles withdrawn is probably mainly due to the fact that the journals have become more diligent. While it is true that you do see some false-positive findings and p-hacking, Fanelli believes this has little impact when you compile and summarise the research in an area.

Fanelli has made a career of studying cheating in research, questionable practices and reliability in research. He says that one of the reasons he “enthusiastically” entered this discipline was that he had “a feeling” that something had to be dysfunctional.

“If I had seen a knowledge base that supports that viewpoint, I would have supported it. It’s not that I have an agenda. I just think it’s wrong.”

So you have changed your mind?

“Absolutely, although I don’t think I ever talked about a ‘crisis.’”

Positive psychology. Much psychological research is simply “psychobabble,” concluded the newspaper *The Independent* following the hitherto biggest attempt to review psychological research in “The Reproducibility Project” in 2015. “Only” 40 of the 100 experiments were successfully recreated on a large scale. Fanelli looks at the study as a turning point, but with opposite signs. Before this, you could easily get the impression that psychology did not have any standards for presenting evidence.

– “I assumed that practically no research in psychology could be reproduced. Not because it’s that easy to be a fraudster in this area, but because humans are so complex that you can’t expect it to work the same way as in particle physics. If there’s one field that’s really in trouble, it’s psychology. But for me, 40 per cent was pretty good; you can’t logically expect 100 per cent. And this was supposed to be the discipline in the worst shape! And then the statisticians started looking at the numbers, and all the reanalyses suggest that the estimates really are better,” says Fanelli.

Do you fear that the narrative is being hijacked by people with a special agenda in the debate on research policy and the use of research in politics?

“I saw people use my own studies to support scepticism against climate science, or people who said: ‘Why should we vaccinate our children if biomedical research is corrupt and broken,’” says Fanelli.

“This is a zeitgeist that questions all types of expertise. This is therefore a challenge that must be met.

Fanelli is an avid proponent of better training in statistics and other initiatives that can make research more robust. However, he has little belief in the notion that bad or unreliable research is due to the pressure to publish. People were complaining about the pressure to publish as early as the 1950s and certainly before then too, he points out. And he refers to his own research, which shows that scientists simply do not publish more than they did before if you take into account the fact that articles today frequently have several co-authors. Some surveys also suggest that published studies are getting longer, more sophisticated, and richer in data content.

“In many ways it runs counter to what the classic pressure to publish tale preaches.” So why are people so concerned about the dangers of publishing pressure?

“I suspect it’s a narrative that serves the academics themselves. No one wants to be pressured,” says Fanelli.

Defence. “Science can’t be self-correcting as long as the incentive schemes produce poor science,” says Saltelli, who has read Fanelli’s article in PNAS.

He says the amount of bad research far exceeds the number of withdrawn articles, and points out that the journals still do not encourage enough replication studies. It’s still the case that these studies mainly appear in print when scientists do not manage to confirm the research. This proves, he believes, that the system is still not equipped to handle the “coming crisis.”

In a currently unpublished response to the crisis critics in PNAS, Saltelli, together with UiB



Where’s the crisis? There is no crisis.

DANIELE FANELLI

colleague Silvio Funtowicz, one of the founders of the “post-normal” way of thinking, write that they think Fanelli’s criticism resembles a “religious view,” complete with the idea that “doubters are corrupting young people.” But what about his own approach. What does he

think about it potentially being used, for example, to justify opposition to vaccines or deny climate change?

Congressional battle. It was apparently an ordinary presentation of a report on Capitol Hill in Washington, D.C. Called “The Irreproducibility Crisis of Modern Science,” the report was received with open arms on 17 April this year by Texas Republican Lamar Smith, chair of the House Committee on Science, Space and Technology.

This was “an important study,” said Smith, according to Undark Magazine. The report could give the impression of being a serious and neutral overview of the “crisis” in science, with several well-known proposals for how research could be improved.

It was published by the National Association of Scholars, a conservative research policy group best known for its support of climate change sceptics. In an article in the Wall Street Journal, the authors of the report elaborated on their views: “The whole climate science discipline is a jumble of unreliable statistics, random research methods and politicised groupthink,” they wrote.

The debate about reproducibility is already being exploited by political activists, declared Naomi Oreskes, historian of science at Harvard and known for her book *Merchants of Doubt*, to *Undark*: “Climate sceptics and other activists are having a feeding frenzy because this is exactly what they want. And what they want to do is use this now to try to discredit all science.”

Among other things, the report supports a bill that would bar the US Environmental Protection Agency (EPA) from using research that is not “substantially reproducible.”

“Even the best arguments can be expropriated, exploited and abused,” says Saltelli.

This could prevent the environmental authorities from taking advantage of the best knowledge in the regulation of, for example, environmental toxins because of lingering scientific uncertainty or because data is kept secret, for example, for privacy reasons. Scandal-ridden EPA Administrator Scott Pruitt, who was appointed by Donald Trump, is now trying to introduce this rule, over the strong protests of American research organisations. His proposal, which is now being circulated for comment, points directly to the “replication crisis” as justification.

They are indeed the “enemies of science,” admits Saltelli about the authors behind the report submitted to Congress.

“They are turning the knowledge base into a weapon, there is no easier way to put it,” he writes in an e-mail.

The Italian fears that events like this will cause scientists to not speak openly about issues in science. However, as he has also written, he believes that the combination of corruption, anger and new digital technology can mobilise change: Because science needs a reformation.

jkt@morgenbladet.no

ONE REASON WHY THINGS ARE GOING WRONG

Many warnings have been issued over the years about the different ways experiments can produce errors through poor design or misunderstood analyses of the data material. Nevertheless, scientists make the same mistakes over and over again. This paradox is the starting point for the study *The Natural Selection of Bad Science* (Royal Society, 2016) in which British scientists attempt to explain why they believe there are so many false-positive findings in research despite all the warnings. The mechanism, which the authors call “the natural selection of bad research,” is simple: As long as publishing is the most important factor for an academic who wants a career, it is reasonable to assume that the researchers choose methodologies with the idea of getting published rather than with a view to discovering something new.