ST.GALLEN SYMPOSIUM

#NEWGENERATIONALCONTRACT

THOUGHT PROVOKING IDEAS OF THE GLOBAL ESSAY COMPETITION 2023

The problem with the jury of one's peers assessing their bad and not good deeds: the proposal for replacement of scholarly peer review.

Marcel Mordarski is one of the top 25 contributors to this year's Global Essay Competition Award. He studies at the University of Cambridge and attended the 52nd St. Gallen Symposium as a Leader of Tomorrow.

One of the age of information's biggest fights has historically been against fake news. Journalists, security professionals, and scientists trying to limit the spreading of Donald Trump's election fake news in 2020 are in fact no different from those complaining about Otto v. Bismarck's war-mongering Ems Dispatch in 1870. The scientific community, particularly concerned about what is truthful and what is not, tried to reduce fake news in science.

A threshold

The peer review process, i.e., challenging new pieces of knowledge by a competent specialist before publishing in a journal, was established as a common practice sixty years ago. We lack certainty, however, that its mechanisms are advancing science and improving its public image. "`Disruptive' science has declined. [...] The proportion of publications that send a field in a new direction has plummeted over the past half-century." says a recent Nature article[1] self-critically suggesting a correlation to peer review. On the other hand, along with the decline of Americans' trust in scientists during the pandemic, the importance of improving the clarity of the peer review keeps being mentioned[2].

The undesirable effects are because peer review aims at scrutinising if new knowledge exceeds an arbitrary threshold of quality rather than if it

influences research. It is as if one made sure that ordinary citizens are punished for repeating election/war lies without appreciating the leading role of Trump and Bismarck. Correspondingly in the scientific world, it is state-of-the-art research that leads to scientific new technologies, illness treatments, and discoveries of new materials.

The Internet age requires a well-suited metric to raise well-informed societies and bring about scientific advancements. To avoid propagating one of the worst legacies my generation has inherited from those before, this essay will describe how a metric like this could be constructed and implemented.

Information is everything

What is customarily called a "scientific paper" used to bear the name of a "letter", (hence, e.g., a journal name "Physical Review Letters"). The essence of scientific discourse from the pre-Internet times when scientists exchanged thoughts through letters (and later through journals) remains an underpinning of the nowadays social construct of the Western academia. The function of a valuable paper is to move a conversation within a particular scientific community forwards. To change its beliefs. Overwhelmingly, it is not to add to an infinite stockpile of interesting facts (a notion implied for the majority of society, even some scientists themselves, by the rigid high-school curriculum)[3]. Peer review was introduced to check if a paper does meet its function.

An uncontrolled, mass-scale experiment began[4]. The majority of journals started asking scientists to challenge soon-to-be-published letters on the behalf of the entire community. The work of the few was expected to improve research quality and reduce the time scientists needed to spend on finding valuable papers. However, under the non-scientific assumption of peer review's "common sense", no system of verifying if it was verily beneficial for scientific progress was put in place. It might have been because peer review had been around since 1731[5] and scientists became familiar with it.

Today, peer review is the basis of both funding and planning of scientific work. Only highly-cited writers, i.e., those going through the peer review process successfully more often than others, are given grants. Only those can get tenured positions at a university and attract bright spark students (in turn frantically consulting supervisors' impact factors on Google Scholar). Only those scientists can decide on strategic directions of research locally, nationally and internationally.

Crucially, through the flocculation of these consequences, peer review decides what is knowledge and what is not.

Based on this, governments approve research funds, military development, and public investment to name a few – the US government decided to multi-billion-fund the Human Genome Research project based on peer-reviewed evidence[6]. Companies equally decide which technologies should be part of their commercial portfolio. Finally, people form their beliefs about the world through knowledge that is accepted as valuable by the scientific communities.

As peer review looks to "appeal to common sense", it is desirable to ask why science does progress but not so disruptively.

Problems

“How can you say one thing when your data shows something else? [...] the literature maintains an attitude somewhat like the approach of lawyers. If the jury buys it, it doesn't matter whether or not it's true. In scientific publishing, the juries are the reviewers and the editors. If they are already convinced of the conclusion, if there is no voir dire, you will surely win the case.” said Richard D. Feinman[7] addressing the peer review’s root of evil. It aims to please the journals and not the scientific community worldwide.

Problematically, peer review in essence is about converging new research with the existing one. Motivations for such modus operandi of the reviewers are several.

Firstly, reviewers struggle to prioritise someone else’s research over their own. The review process is usually not paid for (even though the journals profit millions from it) and consumes time one could devote to scholarly work. Wei Li from Harvard says that an average scientist is asked to peer review five times a year which may take almost a full day of work every time. This does not include subsequent correspondence with the journal and the authors[8]. Insightful and extensive remarks are thus less likely to be provided unless they relate to the specialist’s domain.

Secondly, even if a researcher is fascinated by the work being reviewed, it might be challenging to provide remarks beneficial for the entire community and not the reviewer’s own research. Since a re-viewer has the ultimate authority in the “publish or perish” scientific world of today, the alignment of topics is frequently exerted upon the authors.

Thirdly, the commentaries provided are often high-level. They focus on the language and presentation of the paper. Data giving rise to the research communication is provided along, but according to Adam Mastroianni, a sociologist investigating peer review, it is rarely looked upon[4]. Its interpretation may vary depending on the technique (often unknown to the reviewer), and in critical cases, only the repetition of the experiment may say if it is factual (costly, time-consuming). Furthermore, studies showed that reviewers are bad at detecting fraudulent data and conclusions[9,10]. On average, only 25% of purposefully inserted errors were spotted.

Fourthly, the notion that a paper needs to exceed a threshold of quality encourages authors to report only working frameworks. It does progress science, but it displays just half of the medal. Information that a particular method works can save valuable researchers’ time, but equal time can be saved with the knowledge that a particular method does not work. Richard Smith, an editor of the British Medical Jour-nal, says that “authors often do not even bother to write up such studies.”[11].

Fifthly, having likely contributed to the field themselves, reviewers may be trying to avoid losing their status and sense of intellectual superiority. Also on the personal side of matters, run-ins between the authors and reviewers are not a rarity, particularly when there is an imbalance of power within the scientific community between them. The latter may try to exert their thinking on the former.

Finally, the arbitrariness of reviewers’ demands for the paper results in equally arbitrary rejections. Scientists tend to resubmit their work then to a different

journal, hoping to be more successful. This is a clear waste of both parties' time and effort. While the supporters of the peer review would call it an opportunity to improve research, why not let the community judge for itself whether the presented results are useful? This is particularly important with the studies that may become relevant only in some time, like a paper by Karl Popper[12] which was reluctantly published due to his fame, but described a problem in medicine that the community found important only 20 years later.

Anecdotally discovered by the author of this essay underthings sadly go further and corrupt the system in many other ways than just discussed problems.

Obsessed with increasing citation indices, scientists ask each other for citations just to bias the statistics. This is often done not to stimulate science but to trick it.

Another corruption is the dry and impersonal writing style of papers. They used to provide intuition, value, and passion behind the paper, which have been scrapped to minimise the risk of overstepping the reviewer's personal preferences. Hence, the concepts stopped being intuitively explained and any-body only starting to join a particular field struggles with jargon and synthesising previous work. The passion is gone, and many papers start with impersonal "X has recently become a promising intermediate for Y that is used in Z".

Finally, peer review struggles with the new forms of sharing research. It does not cover podcasts and Twitter discussions. Neither does it scrutinise iPython/Mathematica notebooks which increasingly be-come the preferred publishing standards due to the ubiquitous coding studies in any field of science.

Solution

What should the scientific community do? Doing nothing, in a sense of removing the review process at all, would be a promising start. Notably, websites like arXiv already offer a platform for scientists to share their preprints unreviewed. Effect? Good science defends itself. Bad science is a starter for counterattack papers that point out flaws in scientists' reasoning. Resultantly, scientists, the community of ever-lasting doubt, save money and time on publishing.

It is also becoming more and more common to address the clue of scientific discourse – the impact that new knowledge has on others – in a more quantitative way. A mere count of citations, that does not outline the influence of the cited work on the one being published, is replaced by more holistic metrics like Altmetrics. It measures the impact of a paper based on diverse online research output, such as social media, online news media, online reference managers and so on[13]. Such a metric stimulates publications for a wide range of specialists and laymen alike.

I am thus proposing to merge these two with a data-based approach. The widely-used natural language processing algorithms like BERT and ELMO are learning to compare two pieces of text not only based on their words, but also the meaning of the words. Whenever a paper is cited, it should be checked against the newly presented knowledge to get the overlap score. The semantics analysis should indicate whether the cited paper's contribution is negative or positive,

valuable or not. Such summation of impacts would do justice to the quality of the scholarly work. It would remove paradoxes where Einstein is not the most impactful scientist of the 20th century by current metrics[14].

The solution would remove the disadvantages of peer review. It would not diverge scientists from their own research. It may be impartial and accepting of disruptive knowledge as computers have a larger database of scientific facts than any human brain. In many cases, powerful computers could scrutinise the source data against the presented statements in the text. Another advantage of such semantics comparisons is the ability to unscrupulously challenge whatever a scientist claims disregarding their status and affiliation. The ability to analyse Twitter threads, pieces of code, patents, and other sources is implied.

Centrally, this metric would prioritise groundbreaking papers of high quality in place of low-impact pa-pers, and effortless collaborations which are the pandemics of today's scientific discourse. Under the new order, it would be better to have one thought that shows plenty of scientists, entrepreneurs, politicians, and regulators a more insightful way of thinking instead of diluted series of articles which take time to process.

The solution would also retain the benefits of peer review.

In the first place, it would verify the published facts. The difference is that they would do it reactively once we know if the idea has the potential to change the world. The fabrication of data would also be harder as the algorithm would check it.

Additionally, it would retain the anonymity of the review process. The suggestions would be made public as opposed to many of today's journal review processes.

Eventually, it could promote unknown scientists from lesser-status institutions. Authority would be built through quality research regardless of race and geographic region.

The perfect is the enemy of good. It has not escaped the author of this essay that the implementation of his solution has to face challenges relating to mindsets rather than technicalities. Lobbying may kill the effort. But where if not in the scientific world would it be possible to humbly admit the wrongdoing and construct a brave new world on the rejection of ignorance?

References

1. M. Kozlov, " 'Disruptive' science has declined — and no one knows why," Nature, vol. 613, no. 7943, pp. 225–225, 2023. 'Disruptive' science has declined — and no one knows why (nature.com).

2. T. V. Karin Wulf, Alice Meadows, "Does Trust in Research Begin with Trust in Peer Review? - The Scholarly Kitchen — scholarlykitchen.sspnet.org," Does Trust in Research Begin with Trust in Peer Review? - The Scholarly Kitchen (sspnet.org).

3. L. McEnerney, "LEADERSHIP LAB: The Craft of Writing Effectively — youtube.com," LEADERSHIP LAB: The Craft of Writing Effectively - YouTube.

4. A. Mastroianni, "The rise and fall of peer review — experimentalhistory.substack.com," The rise and fall of peer review - by Adam Mastroianni (experimental-history.com).

5. H. S. Hadas Shema, "The Birth of Modern Peer Review — blogs.scientificamerican.com," The Birth of Modern Peer Review - Scientific American Blog Network.

6. "Report on the Human Genome Initiative for the Office of Health and Environmental Research — web.ornl.gov," Report on the Human Genome Initiative for the Office of Health and Environmental Research (ornl.gov).

7. R. Feinman, The world turned upside down. Nms Press - Duck in a Boat, 2014.

8. W. Li, "Peer Review in Science: the pains and problems - Science in the News — sitn.hms.har-vard.edu," Peer Review in Science: the pains and problems - Science in the News (harvard.edu).

9. F. Godlee, C. R. Gale, and C. N. Martyn, "Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial," Jama, vol. 280, no. 3, pp. 237–240, 1998.

10. S. Schroter, N. Black, S. Evans, J. Carpenter, F. Godlee, and R. Smith, "Effects of training on quality of peer review: randomised controlled trial," Bmj, vol. 328, no. 7441, p. 673, 2004.

11. R. Smith, "Peer review: A flawed process at the heart of science and journals," Journal of the Royal Society of Medicine, vol. 99, no. 4, pp. 178–182, 2006. [Online]. Available: Peer Review: A Flawed Process at the Heart of Science and Journals - Richard Smith, 2006 (sagepub.com).

12. N. McIntyre and K. Popper, "The critical attitude in medicine: the need for a new ethics." BMJ, vol. 287, no. 6409, pp. 1919–1923, 1983. The critical attitude in medicine: the need for a new ethics. | The BMJ.

13. “Altmetrics - Wikipedia – en.wikipedia.org,” Altmetrics - Wikipedia.

14. “Scholar indices and impact - Wikipedia – en.wikipedia.org,” Scholar indices and impact - Wikipedia.