

# Written Evidence Submitted by Professor Les Hatton and Professor Gregory Warr (RRE0031)

## Introduction

With respect, the committee is ten years behind, but hopefully some useful pointers will emerge.

There is a systemic problem in lack of reproducibility in science. It spans disciplines and countries and with the increasing use of large quantities of opaque buggy software in all scientific disciplines, and in the continuing absence of any systematic method of eliminating software bugs, it is getting worse.

## Previous initiatives

It was obvious over 25 years ago that we had a problem and this has been reinforced repeatedly, [CK92, HR94, Hat97, +SSD09, Don10]. Ten years ago, in 2011, the National Institute of Science and Technology in Boulder Colorado held an IFIP / SIAM / NIST Working Conference on Uncertainty Quantification in Scientific Computing discussing various aspects of this problem [DB11].

In the biomedical sciences, in 2015, it was reported the US wasted 28 billion dollars on irreproducible biomedical research with the startling conclusion that "Currently, many published research findings are false or exaggerated, and an estimated 85% of research resources are wasted", [FCS15].

## What is it we are trying to achieve?

The scientific method demands that we have the means to falsify a theory. To falsify a theory requires peer review of the methodology including any statistical arguments, access to the original data, and access to any and all software used to produce the results. Anything less than this is not scientific. Theories which are not falsifiable are not scientific.

This is deceptively simple but very difficult to bring about in practice.

## What is the problem?

This is a multi-faceted problem.

### Declining standards in Peer-review

Peer-review is a volunteer activity. Journals make money and the reviewers do it for free. Like many others, the authors of this submission freely gave their time as reviewers and editors for years. 20 years ago, it was not uncommon to have five reviewers for a paper. Now it is not uncommon to have one. Yes, that's one reviewer to judge a detailed scientific paper which could take months of work.

This is compounded by the fact that academics are under such pressure to produce papers, generate funding, run research students and all the rest, that there is little time left over for the essential activity of peer-reviewing.

Not surprisingly, glaring problems have been reported because of the declining standards. For example a significant segment of the biomedical literature has been tainted by misidentified [Smi06] and contaminated [Nei15] cell lines.

The lack of suitable reviewers puts an intolerable pressure on editors who then have to make important decisions on little evidence. The explosion in journal numbers further degrades the system.

Currently the peer-review system amounts to little more than rather incompetent censorship, [HW17], and is routinely ignored by the media. For example, during the Covid pandemic, most of the science reported by the BBC was **not** peer-reviewed, pushed forward by a university's over-active PR department. The BBC carefully points this out at the end, when nobody is listening by then having swallowed whatever was in the leader. This is an abysmal practice.

Peer-review is very simple. Either you strike down the theory, or you strike down the methodology or you show the data processing is in error or in disagreement with the theory. If you can't do any of these, you must however reluctantly, accept it. No other options are available, particularly "not liking it".

## **Naivety about computation**

Nobody, especially computer scientists, knows how to write a correct program but today, many if not most avenues of science depend heavily on successful computation. In spite of this, computer programs are often thrown together with little idea of how they could be tested, if at all.

This has been pointed out on numerous occasions, [IHGC12]. Things have got nominally better in that the complete means to produce computational results is supposed to be included in a journal submission and there are guidelines on how to do this in bioinformatics [HW16]. However, who is supposed to check them? The reviewers are under enough pressure as it is and the authors of this submission have never had a submitted computational reproducibility package even downloaded let alone checked.

In other words it is a box-ticking exercise with no tangible benefit.

## **Naivety about the scientific method**

The scientific method does not appear to be taught to young scientists who do not generally appreciate that the role of science is to attempt to falsify new theories. That's how theories survive to become dependable. This is down to education.

Even worse, reviewers have forgotten what this means also, often rejecting papers based on opinion.

## **Naivety about the appropriate use of statistics**

We will only record this but the average quality of statistical analyses is poor in our experience, both in authors and in the reviewers. This too is down to education.

## **Naivety about the meaning of evidence**

The meaning of evidence in quantitative science is very clear and it is entirely data-related. There is no room for opinion as in some parts of the Civil Service and elsewhere where "documented stakeholder opinion" is now regarded as evidence, a truly ridiculous notion.

## **What might be done?**

This has taken a long time to deteriorate and it will take a long time to put right. However there are some worthy innovations in peer review<sup>1</sup>. These include dispensing with its traditional forms and are to be encouraged. It may be that open publication through servers such as arXiv and bioRxiv, along with public and signed post publication comment<sup>2</sup> are the solution to the problems noted above.

However, for any innovations in scientific publication to succeed two conditions would need to be met. The first, as repeated numerous times above, is the provision with a publication of all the information necessary for independent reproduction and replication of the research. Furthermore, there needs to be some way of recording whether anybody has actually bothered to replicate it. The second will take longer. There must be an improvement in the culture of science such that less than rigorous work and deceptive publication practices are no longer tolerated.

With the scientific method itself at risk, the stakes could not be higher or the 21st century will go down as the bullshit century.

## **About the authors**

Les Hatton Ph.D trained as a mathematician at King's College Cambridge and is an Emeritus Professor of Forensic Software Engineering at Kingston University. He was prior to this a Professor of Computer Science at the University of Kent and of Geophysics at T.U. Delft in the Netherlands and was the recipient of the 1987 Conrad Schlumberger prize as a geophysicist. In later years, he has written extensively on the impact of software failure in commercial systems and in scientific research.

Greg Warr Ph.D is a biochemist who also trained at King's College Cambridge and is an Emeritus Professor of Biochemistry and Molecular Biology at the Medical University of South Carolina. He retired as

---

<sup>1</sup><https://www.elsevier.com/reviewers-update/story/innovation-in-publishing/is-open-peer-review-the-way-forward>, accessed 28-Sep-2021.

<sup>2</sup><http://blog.scienceopen.com/2016/09/disambiguating-post-publication-peer-review/>, accessed 28-Sep-2021.

a Program Director in the Molecular and Cellular Biosciences Division of the National Science Foundation in Washington DC. He has published over 200 papers in peer-reviewed journals and has an h-index of 52.

## References

- [CK92] Jon F. Claerbout and Martin Karrenbach. Electronic documents give reproducibility a new meaning. In *Proc. 62nd Ann. Int. Meeting*, pages 601–604. Soc. of Exploration Geophysics, 1992.
- [DB11] A.M. Diensfey and R.F. Boisvert. In *10th IFIP WG 2.5 Working Conference, WoCoUQ 2011, Boulder, CO, USA, August 1-4, 2011, Revised Selected Papers*, Berlin, Heidelberg, 2011. Springer.
- [Don10] D.L. Donoho. An invitation to reproducible computational research. *Biostatistics*, 11(3):385–388, 2010.
- [FCS15] L.P. Freedman, I.M. Cockburn, and T.S. Simcoe. The Economics of Reproducibility in Preclinical Research. *PLoS Biol*, 2015. doi:10.1371/journal.pbio.1002165.
- [Hat97] Les Hatton. The T experiments: Errors in scientific software. *IEEE Computational Science and Engineering*, 4(2):27–38, April 1997.
- [HR94] Les Hatton and Andy Roberts. How accurate is scientific software ? *IEEE Transactions on Software Engineering*, 20(10):785–797, 1994.
- [HW16] Les Hatton and Greg Warr. Full Computational Reproducibility in Biological Science: Methods, Software and a Case Study in Protein Biology. *ArXiv*, 2016.
- [HW17] Les Hatton and Greg Warr. Scientific peer review: an ineffective and unworthy institution. *Times Higher Education*, 9<sup>th</sup> December, 2017.
- [IHGC12] Darrell C. Ince, Leslie Hatton, and John Graham-Cumming. The case for open program code. *Nature*, 482:485–488, February 2012. doi:10.1038/nature10836.
- [Nei15] Jill Neimark. Line of Attack. *Science*, 347, 2015.
- [Smi06] Reginald Smith. Peer review: a flawed process at the heart of science and journals. *J. R. Soc. Med.*, 99, 2006.
- [SSD+09] M. Shahram, V. Stodden, D. L. Donoho, A. Maleki, and I. Rahman. Reproducible research in computational harmonic analysis. *Computing in Science & Engineering*, 11(01):8–18, jan 2009.

**(September 2021)**