

Open Research Online

The Open University's repository of research publications and other research outputs

Workings of science: Trust in science and mathematics

Journal Item

How to cite:

Johnson, Jeffrey and Odlyzko, Andrew (2022). Workings of science: Trust in science and mathematics. Ubiquity, 2022(June), article no. 1.

For guidance on citations see [FAQs](#).

© 2022 Jeffrey Johnson; 2022 Andrew Odlyzko



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:
<http://dx.doi.org/doi:10.1145/3512337>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's [data policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Ubiquity

Volume 2022, Number June (2022), Pages 1-9

Ubiquity Symposium: Workings of science: Trust in science and mathematics

Jeffrey Johnson, Andrew Odlyzko

DOI: [10.1145/3512337](https://doi.org/10.1145/3512337)

Concerns about the trustworthiness of science are not confined to fringe groups that reject science entirely. There is substantial unease even among researchers about the reliability of peer review and the reproducibility crisis—where scientific results are not or cannot be tested by replication. Here the authors point out that such worries even apply to mathematics—for many the language of science. This is largely caused by the growing complexity of our knowledge base, where results are more complicated and investigators sometimes have to rely on the results of others that they do not fully understand. This means, as with the sciences, mathematical discoveries increasingly have to be treated as not absolutely reliable, but as part of a process of searching for the truth, and even what it means for something to true.

The importance of science in people's lives is growing. Yet trust in science seems to be decreasing. The global pandemic has brought into focus the question of trustworthiness of the science as it is used to support policy. Most of the general public have been bewildered at the lack of consensus among scientists on the likely outcomes of policies costing billions of dollars and millions of lives.

Much of this concern can be dismissed as due largely to a lack of understanding of what science is and how it works. And of course, much is politically motivated, based on an unwillingness to accept inconvenient findings. However, there are serious questions about the trustworthiness of science that arise from internal trends, primarily growing complexity, and those should be faced. In this article, we point out that even mathematics, the language of science, and so a discipline on which the whole scientific edifice rests, suffers from growing questions about its reliability.

Our conclusion is not that science and mathematics are unworthy of trust. Rather, what needs to be emphasized is that scientific and mathematical research is a process and does not generally produce absolute truths. There are self-correcting elements at work, and humanity will surely be able to benefit from further research advances. However, it should never be forgotten that one needs to retain a certain level of skepticism—likely a growing level of skepticism—about all results, even those in the widely accepted peer-reviewed literature.

SCIENCE AND SOCIETY

Although there are skeptics such as anti-vaxxers, climate change deniers, and many others, most people to a large extent do trust science and its application in social policy. For example, almost 90% of adults in the U.K. had at least one Covid-19 vaccine injection as of September 2021 [1]. However, applied science is not infallible and occasionally results in terrible mistakes, such as the thalidomide tragedy of the 1950 and '60s, when thousands of babies died within a few months or were born with missing or malformed limbs. Lay people are not trained to judge the evidence of medical trials and trust regulators to protect them. In the U.S. the Food and Drug Administration did not approve thalidomide, but in 46 other countries regulators failed their citizens [2]. This illustrates that worries about applications of science by policy makers are not wholly misguided.

The scientific community has protocols for deciding the validity of its theories. All scientific theories are contingent—they cannot be proved to be correct, but if not consistent with observation they must be modified or rejected. The gold standard is to use theory to make predictions, conduct

replicable experiments, and show that the data support the prediction each time the experiment is conducted. However, in a complex world these protocols may be hard or impossible to apply, and there are loud voices from within science itself questioning the validity of much of what is announced and even published in the peer-reviewed literature.

Outright fraud has always been a problem in science, and there are claims it is increasing with the growing pressure to publish, get grants, and so on. But it is a much less serious problem than the "replication crisis." This refers to some results that have been in the literature for years or decades before finally being refuted. This problem arises partially because there are few incentives for researchers to engage in the non-original work of checking the work of others. To an increasing extent it also arises from the practical impossibility of duplicating the original research. How do you verify the discovery of an elementary particle that was found by a massive computation on gigantic data sets coming from the world's only accelerator capable of producing that data? And how do you verify a network science result that came from a sophisticated analysis of giant data sets that are not publicly available for proprietary or privacy reasons? Thus, even the peer-reviewed scientific literature increasingly seems to rest on shaky foundations. For example a study of a large collection of prominent psychology results found a large portion of replications produced weaker evidence for the original findings despite using materials provided by the original authors [3]. Similar issues are increasingly visible in mathematics.

MATHEMATICS AND ITS RELATION TO SCIENCE

Mathematics is not an empirical science, although much of it is inspired by empirical facts, whether from various sciences or from the exploration of mathematical concepts. It serves as the language of science, but it is largely a deductive discipline. The majority of mathematicians are Platonists—meaning they do not claim they are constructing something new in their research. Instead, they believe they are discovering underlying mathematical structures that exist independent of them. In this they resemble most scientists.

Mathematics is proud of the degree of rigor that it has attained. This rigor was never absolute, in spite of the care exercised. History offers many cases of incorrect results that were widely accepted before defects in their proofs were found, as well as proofs regarded as faulty before being demonstrated to be correct. Yet, overall, professional mathematicians have generally felt confident in the reliability of their peer-reviewed literature. But this rigor appears increasingly questionable. This is distinct from the basic foundational questions about mathematical knowledge.

MATHEMATICAL FOUNDATIONS

There are some very fundamental issues in mathematics that have, to some extent, shaken the faith of its practitioners in the underlying beauty and simplicity of their field. Gödel's incompleteness theorems of the 1930s, for example, showed in any axiomatic system satisfying some natural properties, there are "true" statements that cannot be proved to be true in that system. This means we cannot hope to prove everything; every non-trivial logical system will contain propositions that are "undecidable" propositions that cannot be proved to be true or false within the system. This means Hilbert's great wish—expressed in 1930 as "we must know, we will know"—cannot be fulfilled.

Decades before the seismic shock of Gödel's 1930s incompleteness theorems, there were already some schisms that were appearing in mathematics, e.g. with the emergence of the intuitionists who wanted a more constructive approach to mathematics. In particular, they wished to avoid the law of excluded middle, one of the basic principles of traditional logic, which states for every proposition

either that proposition or its negation is true. However, the intuitionist approach was not widely accepted, and it was not until Gödel's results appeared that most mathematicians had to acknowledge that there were serious foundational issues in their field.

Some other results, such as the independence of the Axiom of Choice, or general issues of constructivity, also served to upset mathematicians' hopes for an intellectually clean edifice. But while these phenomena were disappointing, they did not seriously undermine the belief in the reliability of the published literature. There are established and universally accepted standards for rigor in proofs, and there is a system of peer review before publication. Further, unlike most fields, mathematics has specialized journals in which recognized experts review recently published papers. This provides an extra layer of scrutiny and thus greater reassurance to mathematicians as well as experts in other fields that what is published is trustworthy.

MATHEMATICS AND 'THE BOOK'

The main problem facing mathematics is that its literature is moving away from the ideal, in which individuals can be certain they can trust the results they read or publish. This can be illustrated by reference to the concept of the great prolific, eccentric, and itinerant mathematician Paul Erdős (1913–1966) called "The Book." Erdős would show up, often unannounced, on the doorstep of a fellow mathematician, declare "My brain is open!", and stay as long as his colleague served up interesting mathematical challenges. The Book was a hypothetical collection, maintained by God (referred to by Erdős as the "Supreme Fascist"), of proofs of mathematical theorems with the greatest possible simplicity and elegance. The highest compliment Erdős could pay to a colleague's work was to say "That's straight from The Book" [4]. The idea of The Book has been embraced by many mathematicians, and there are debates as to which proofs do or do not qualify for it. A selection by two eminent researchers, with input from many others, has already gone through six editions [5].

What kinds of proofs qualify for The Book? There are differences of opinion, but it seems everyone agrees on the necessity of including Euclid's proof that the sequence of primes does not end. The proof in Aigner and Ziegler's *Proofs from THE BOOK* goes roughly as follows:

For any finite set $\{p_1, \dots, p_r\}$ of primes, consider the number

$$n = p_1 p_2 \dots p_r + 1.$$

Then either n has a prime divisor p smaller than n , or n is a prime p . But p is not one of the p_i : otherwise p would be a divisor of 1, the difference of n and the product $p_1 p_2 \dots p_r$, which is impossible. So a finite set $\{p_1, \dots, p_r\}$ cannot be the collection of all prime numbers [5].

This also shows there must be an infinite number of primes. The argument is simple, does not require any deep background knowledge of mathematics, and leaves no room for doubt of its correctness.

In his 1940 book *A Mathematician's Apology*, G. H. Hardy points out that Euclid's proof is by reductio ad absurdum. This goes as follows. Suppose we want to prove p is true, and we already know q is true. Then suppose, contrary to what we want to prove, we assume p is not true, and follow this by proving that $not-p$ implies $not-q$. Since, by hypothesis, q is true, this is an absurd contradiction from which it is concluded p must be true. Hardy continues "reductio ad absurdum, which Euclid loved so much, is one of a mathematician's finest weapons. It is a far finer Gambit than any chess gambit: a chess player may offer the sacrifice of a pawn or even a piece, but a mathematician offers the game" [6].

Hardy notes: "[Euclid's] proof can be arranged so as to avoid reduction ad-absurdum, and logicians of some schools would prefer that it should be." This would include the intuitionist mathematicians following the work of L. E. J. Brouwer from 1907. Intuitionistic logic can be succinctly described as classical logic without the Aristotelian law of excluded middle, p or $\text{not-}p$; or the classical law of double negation elimination, $\text{not-}(\text{not-}p)$ implies p [7]. Intuitionists require that mathematical objects be constructed and are not satisfied by an argument that says a mathematical object exists because if it did not it would lead to a contradiction.

MATHEMATICS AND ITS GROWING COMPLEXITY

Hardy's argument shows sometimes it is possible to reconcile the simplicity demanded by The Book and special limitations on the allowed range of arguments that philosophies such as those of the intuitionists demand. But that is rare.

Modern mathematical research seldom produces results that qualify for The Book. Esoteric knowledge is increasingly required to not just do the research and write the papers, but to understand the results. The growing sophistication, and need for expertise in several topics, has led to a phenomenon also observed in other fields, namely more collaborations. Whereas a century ago more than 95% of mathematical papers were by single authors, today fewer than half are. This of course means often authors do not fully understand everything they put their names to. And it is harder for peer review to provide a trustworthy evaluation of a paper. In some cases, claimed solutions for famous results float around for years, without a definitive verdict as to their correctness [8].

Such factors affect much of modern mathematical literature. We do not have quantitative data on the degree to which published papers have become less correct, but there is anecdotal evidence, including comments from editors that it is harder to obtain detailed referee reports.

Our arguments have a lot in common with those of Bordg [9]. However, we do not regard what is occurring as a crisis. It appears to be more of a continuation and intensification of a trend that has been developing for a long time.

What we can do is illustrate some of the key issues of trustworthiness through a discussion of some famous mathematical problems whose solutions have aroused controversy about their validity. One prominent example is the four color conjecture [10]. It was proved in the mid-1970s by Appel and Haken with the crucial assistance of a computer. The basic processes of the proof were not very complicated, but there were many special cases to consider. Those could be framed in simple algorithmic ways and done by a computer. In principle, they could have been done by hand, but the number of cases was impracticably large. This led to considerable debate as to whether the proof could be accepted as valid.

Since that time, the proof of Four Color Theorem has been improved. The basic analysis is simpler and easier for people to understand, and the number of cases to be checked by computer is smaller. Hence it is generally accepted as valid, both because there are now separate proofs, and because the use of computers has in general become more widespread and pervasive. But there is still some uneasiness in mathematics, as the currently known proofs of the Four Color Theorem are not regarded as coming close to qualifying for The Book. They all require dealing with many special cases.

The early history of the four color problem illustrates the social nature of mathematical truth. In 1879 Alfred Kempe gave a proof that was widely accepted. For 11 years the consensus amongst

mathematics was that the conjecture had been proved true. However, Kempe's proof was shown to be incorrect by Percy Heawood in 1890 [10]. Thus, for 11 years the conjecture was believed to be true by the mathematics community, but then reverted to being "not-proven." For some mathematicians it flip-flopped back to be true after Appel and Haken's 1976 computer proof, but for others it did not. The proof of Fermat's last theorem by Andrew Wiles caused much excitement as it flipped from "unproved" to "proved" in June 1993. Unfortunately it flopped back to unproved in September 1993 after the proof was found to have an error. In 1994 Wiles and Taylor had an inspiration that allowed Wiles to publish a proof in 1995 that satisfied the mathematical community, and the conjecture flipped back to proved where it remains today. In science a theory is accepted until there is evidence to reject it; in mathematics it seems that a proof is valid or invalid until the community of mathematicians votes otherwise. The idea that correct proofs are social constructs is far from the Platonic ideal.

The issue of trustworthiness of the proof of the Four Color Theorem rested, and still rests in some minds, on the use of computers. But similar concerns arise even in some cases where computers play a minor role. The classification of finite simple groups was announced in the early 1980s, not long after the Appel and Haken proof of the Four Color Theorem. It was a major achievement on an important fundamental problem in mathematics. But in some ways, it was even more controversial—not because of reliance on computers, but because the result relied on the work of hundreds of mathematicians, working in a loosely coordinated way. It was a landmark in mathematical collaboration, started in the days before the internet was widely available. But the output of that effort was a collection of what some estimated as more than 15,000 pages of papers, quite a few of which were known or suspected to have non-trivial gaps. (There was also a very large gap that had been overlooked in the triumphal announcement made in 1983, the lack of treatment of a particular class of groups, but let's ignore that.) So was this really a proof? It has always been treated with some suspicion, and there are several papers, for example in the analysis of algorithms, which explicitly state they are valid only if the classification of finite simple groups is correct.

There is an ongoing effort to simplify the original proof, and to make it more nearly correct and trustworthy. But even that enterprise has already produced more than 5,000 pages of books and papers. Can that possibly be completely valid? Many experts are more comfortable with a relatively simple computer program checking billions of cases than with having to rely on 5,000 pages of text generated by people.

The controversies associated with both the Four Color Theorem and the classification of finite simple groups are fundamentally the same, namely, how can we trust arguments that are extremely complicated and beyond the full comprehension of any single person, no matter how gifted? There is, of course, the hope that much simpler proofs will be found, but that hope is slim. At the moment we have to rely on either long computer calculations or the labor of human researchers who do possess human frailties.

One way to increase our trust in complicated arguments is shown by recent work on the four centuries old Kepler's conjecture about the densest packing of equal spheres in three dimensions [11]. After some earlier unsuccessful attempts, in 1998 Hales produced an argument that was generally accepted by experts. As with the Four Color Theorem, it relied on extensive computer computations, and there continued to be doubts about its validity. This led Hales, together with collaborators, to obtain formal proofs of the validity of the proof (or to be more precise, a version of the proof that had been re-written so it could be checked by the proof assistant). But of course that did not eliminate all doubts, as one then has to accept that the automated theorem provers are correct, that the software and hardware they run on are operating according to specifications, that

no cosmic rays flip bits in memory, and so on. In addition there are some serious doubts, raised, for example in De Millo et al [12], in the context of formal verification of programs, but applicable more widely, as to how far effective such methods might be.

CONCLUSIONS

Going forward, computer hardware and software are getting better, so we can hope formal method approaches such as those used for Kepler's Conjecture can help deal with other complicated mathematical proofs, perhaps even with the classification of finite simple groups. But there will surely always be questions. First, can those formal systems be trusted? And second, will they be applied in enough cases? In the case of software, the complexity of actual deployed systems has outpaced the ability of formal verification.

Thus there are promising ways to provide increasing trustworthiness of the mathematical literature. But how effective they will be remains to be seen, and in the meantime mathematics, like most areas, is suffering from growing complexity, and less and less of it qualifies for 'The Boo'.

Mathematics continues to develop and remains a vibrant and imaginative source of discovery for new structures and relationships. It continues to find valuable applications in a variety of fields. Even very abstract results have been used in areas ranging from bioinformatics to robotics. But we need to realize, as with the sciences, one has to treat even mathematics published in the most prestigious journals not as the definitive truth, but as part of an ongoing process of searching for the truth. We should be prepared to encounter some missteps along the way.

Acknowledgements

The authors thank Robert Akscyn, Joe Buhler, Peter Denning, Dennis Hejhal, Jeffrey Lagarias, Ted Lewis, Peter Olver, and Philip Yaffe for their comments.

References

- [1] The Visual and Data Journalism Team. [Covid vaccine: How many people in the UK have been vaccinated so far?](#) BBC News. Sept. 30, 2021.
- [2] The Science Museum. [Thalidomide](#). Dec. 11, 2019.
- [3] Open Science Collaboration. [Estimating the reproducibility of psychological science](#). *Science* 349. 6251 (2015). DOI: 10.1126/science.aac4716.
- [4] Encyclopaedia Britannica. [Paul Erdős: Hungarian mathematician](#).
- [5] Aigner, M. and Ziegler, G. M. *Proofs from THE BOOK*, 6th edition. Springer Verlag, 2018.
- [6] Hardy, G. H., *A Mathematician's Apology*. Cambridge University Press, 1990. (First published 1940).
- [7] Stanford Encyclopedia of Philosophy. [Intuitionistic Logic](#). First published Wed Sep 1, 1999; substantive revision Tue Sep 4, 2018; <https://plato.stanford.edu/entries/logic-intuitionistic/>
- [8] Hartnett, K. [New math book rescues landmark topology proof](#). *Quanta Magazine*. Sept. 9, 2021.
- [9] Bordg, A. [A replication crisis in mathematics?](#) *Mathematical Intelligencer* 43 (2021) 48-52.
- [10] Wilson, R. *Four Colors Suffice: How the map problem was solved*, 2nd ed. Princeton University Press, Princeton, 2014.

[11] Lagarias, J. C., ed. *The Kepler Conjecture. The Hales-Ferguson Proof*. Springer New York, 2011.

[12] De Millo, R. A., Lipton, R. J., and Perlis, A. J. [Social processes and proofs of theorems and programs](#). Communications of the ACM 22, 5 (1979), 271-280.

Authors

Jeff Johnson, Professor of Complexity Science and Design at Open University, joined the university in 1980 after three years as a senior research associate in the Geography Department of Cambridge University, and six years as a research fellow in the Mathematics Department of Essex University. He is head of the Department of Design, Development, Environment, and Materials and is President of the Complex Systems Society.

Andrew Odlyzko has had a long career in research and research management at Bell Labs, AT&T Labs, and most recently at the University of Minnesota, where he built an interdisciplinary research center, and is now a professor in the School of Mathematics. He has worked in a number of areas of mathematics, computing, and communications, and is concentrating on the diffusion of technological innovation. His recent works are available through his web page.

2022 Copyright held by the Owner/Author.