Génova, G., Astudillo, H. & Fraga, A. The Scientometric Bubble Considered Harmful. Sci Eng Ethics 22, 227–235 (2016).

DOI: https://doi.org/10.1007/s11948-015-9632-6

The scientometric bubble considered harmful

Gonzalo Génova *(corresponding author)*[*]
Departamento de Informática
Universidad Carlos III de Madrid, España
ggenova@inf.uc3m.es
Centro de Innovación en Tecnologías de la Información para Aplicaciones Sociales
Universidad de Santiago de Chile, Chile
gonzalo.genova@usach.cl

Hernán Astudillo
Departamento de Informática
Universidad Técnica Federico Santa María, Santiago, Chile
hernan@inf.utfsm.cl

Anabel Fraga
Departamento de Informática
Universidad Carlos III de Madrid, España
afraga@inf.uc3m.es

Published online in *Science and Engineering Ethics*, 2015
The final publication is available at Springer via http://doi.org/10.1007/s11948-015-9632-6

======

Abstract

This article deals with a modern disease of academic science that consists of an enormous increase in the number of scientific publications without a corresponding advance of knowledge. Findings are sliced as thin as salami and submitted to different journals to produce more papers. If we consider academic papers as a kind of scientific 'currency' that is backed by gold bullion in the central bank of 'true' science, then we are witnessing an article-inflation phenomenon, a scientometric bubble that is most harmful for science and promotes an unethical and antiscientific culture among researchers.

The main problem behind the scenes is that the impact factor is used as a proxy for quality. Therefore, not only for convenience, but also based on ethical principles of scientific research, we adhere to the San Francisco Declaration on Research Assessment (DORA) when it emphasizes "the need to eliminate the use of journal-based metrics in funding, appointment and promotion considerations; and the need to assess research on its own merits rather on the journal in which the research is published".

Our message is mainly addressed to the funding agencies and universities that award tenures or grants and manage research programmes, especially in developing countries. The message is also addressed to well-established scientists who have the power to change things when they participate in committees for grants and jobs.

[*] Partially funded by project PMI USA1204, Centro de Innovación en Tecnologías de la Información para Aplicaciones Sociales, Universidad de Santiago de Chile, Chile.

Keywords

Ethics in Scientific Publications, Careers in Academia, Research Assessment, Scientometrics, Impact Factor

Introduction

We wonder how many readers may have been attracted to start reading this article by our marketing-wise title. The fact that we have felt compelled to use this gimmick to attract your attention is a symptom of something, something very bad... but the matter is serious.

This article deals with a modern disease of academic science that has received several names: the Numbers Game (Parnas 2007), the Salami Science (Nature Materials 2005), etc. Basically, it consists of an enormous increase in the number of scientific publications without a corresponding advance of knowledge. Findings are sliced as thin as salami and submitted to different journals to produce more papers. These spurious achievements of Academia, represented by mountains of unloved and unread publications, are indeed a waste of *write-only papers*. In the Physical Review journals, for example, around 30% of the papers published from 1893 through 2003 got one or no citations at all (Redner 2005). And these are reference journals in the field! What about papers published –but never read– in less prestigious journals and conferences? It is a *publish-and-perish* process in which most of the papers are lost (Moro 2009).

If we consider academic papers as a kind of scientific 'currency' that is backed by gold bullion in the central bank of 'true' science, then we are witnessing an article-inflation phenomenon, a Lehman Brothers bubble of words that could explode at any time, dragging down the shareholders of science. It could be even worse if, not reaching this level of drama, the disease does not receive due attention and gets untreated, leading to scientific paralysis. The situation was described as early as 1981 in *Science* journal (Broad 1981), with a critic to the shrinking length of papers and the abuse of the so-termed Least Publishable Units (LPU); but things have gone worse since then.

Why scientific publications are necessary

We do not question the necessity of publishing scientific results. Science is a public affair that has to be discussed in the marketplace, i.e. in scientific workshops, conferences and journals. Besides, nowadays anyone can publish anything in whatever corner of the global network. Therefore, a previous filtering by a responsible program committee or editorial board is beneficial. This filtering adds value as long as the core of science (the gold bullion) is made more *accessible*… because it is kept *small*. The bigger the bubble gets, the less accessible the core becomes.

Scientific publications should be a remedy for 'information overload' (a term popularized by Alvin Toffler in his 1970 book *Future Shock*). Instead, the Academia has created an artificial necessity of publishing, not for the advance of knowledge, but for the advance of professional careers. Academia has succumbed to 'infoxication'. We all know the consequences: science is published in smaller and smaller bits (LPUs) so as to maximize the number of publications one can make out of his or her daily work, which is very questionable from the viewpoint of research ethics. Besides, this publication race runs against the first and natural goal of scientific publishing, because the information overload makes the gold bullion of science less accessible.

Publications, like currency, are devaluated when their number increases without a real necessity.

Why evaluating scientific productivity is necessary

Science is expensive. Governments and private investors rightfully expect that supporting the salaries of scientists will pay out. Therefore, it is desirable to promote good scientists and research centers, while discouraging bad ones. How are we tempted to achieve this in our modern industrial society? By *measuring* productivity. But scientific productivity is not alike industrial productivity. Ideas cannot be measured like bricks.

We can say, metaphorically speaking, that the House of Science is made out of bricks. But this is only a metaphor. Ideas are not bricks, ideas do not weigh, ideas do not have volume. Assessing the value of new ideas is necessary, but reducing their value to numbers is pernicious, most pernicious.

Why current metrics of scientific productivity are pernicious

Current metrics of scientific productivity are aimed at measuring the quality of publications. The quality of a publication is basically measured according to the impact factor of the venue (journal or conference) where it gets published. The impact factor is measured according to the number of citations other papers in the venue have received in recent years. Implicit assumptions underlying this measurement procedure are: (i) a publication is good if it gets published in a good venue; (ii) a venue is good if it has deserved sufficient attention from scientists. In other words, it is assumed that there is a positive correlation between impact factor and scientific quality.

With this kind of metrics, in a certain sense, the judgment on the quality of a publication is outsourced onto a huge and anonymous mass of judges, achieving a kind of 'blind justice' that is supposed to avoid corruptions. The idea is interesting, but we all can observe the effective corruptions of the system (excluding the cases of outright frauds):

- The publishing journal is given more importance than the scientific message. Scientists may become desperate to publish in certain journals, instead of concentrating on the scientific quality of their results (Lawrence 2003).
- Popularity is favored over quality. Results that are of interest for a small population are discouraged. Papers with a more general and divulgative character have more probability to get cited than 'hard science' papers (Mattern 2008).
- All citations are counted the same way, regardless the fact that there are multiple reasons to cite, reasons that are not always positive (Mattern 2008).
- Fast science is favored over slow science. But science needs time to think, time to read, and time to fail (The Slow Science Academy 2010). Notable or surprising results are preferred to more modest and solid ones (Brembs et al. 2013). Short term interest and applicability are preferred to long term research projects that pursue difficult, uncertain results. As Wolfgang Pauli reportedly told a colleague, "I don't mind your thinking slowly; I mind your publishing faster than you think" (quoted in Mackay 1977).
- Papers that find little resistance because in accordance with tradition are favored over revolutionary or heterodox ideas, which will find few heroes willing to propose them against the establishment of science. It is a good thing that new ideas find a certain opposition, so that they have to prove their value. But the proponents themselves should not see their careers at risk for being audacious.

- The more cited papers, scientists and journals get even more cited, the less cited ones get less cited: 'the rich get richer and the poor get poorer', a phenomenon long ago known as the Mathew Effect (Merton 1968) or, more recently, Preferential Attachment (Barabási & Albert 1999). Modest, 'middle class' scientists die of starvation, unable to compete with the most powerful and consolidated ones.
- Also, the business of big scientific publishing and scientometric companies is favored. Finding a place in the publishing market is more and more difficult, since new venues have to compete with numbers that tend to consolidate themselves. This promotes the concentration of editorial power, instead of its democratization.
- Local and regional venues are destroyed in favor of a multitude of 'international' conferences and journals. In spite of the indubitable interest of local venues, scientists find little academic profit in discussing new or preliminary ideas with near colleagues. This effect is more pernicious for second-line countries in the scientific arena.
- Indirectly, this increasingly widespread tendency to hire faculty based mainly or solely on scientific productivity leads to poor teaching, or at least to less investment in teaching resources.

The main problem behind all these is that the impact factor is used as a proxy for quality. But the Journal Impact Factor, as the San Francisco Declaration on Research Assessment (DORA 2012) recalls, was originally created by Thomson Reuters as "a tool to help librarians identify journals to purchase, not as a measure of the scientific quality of research in an article"; besides, "data used to calculate the Journal Impact Factors are neither transparent nor openly available to the public". Thomson Reuters acknowledges that the Journal Impact Factor does not measure the quality of an individual article in the journal, but the reputation of the journal in its field (Thomson Reuters 2013). Thomson Reuters claims, too, that the problem does not lie on how the Impact Factor is calculated, but on how it is used by funding agencies, publishers and universities.

However, the Journal Impact Factor (IF) and related indices have received severe criticism. Larsen & Ins (2010) indicate problems of Thomson's Science Citation Index (SCI) and Social Sciences Citation Index (SSCI), derived of being a monopoly with declining coverage and unable to cope with new and growing publication channels. Brembs et al. (2013) show three other fundamental problems: (a) the index is *not calculated* but negotiated (adjusting the denominator of the IF quotient, i.e. which published articles are counted; reducing this number can dramatically affect the result); (b) the index is *not reproducible* (i.e. the same calculations performed on the same public data do not yield Thomson's results); and (c) the index is *mathematically unsound* (because citation distribution is strongly left-skewed, i.e. a small number of publications receive most of citations, thus the use of the arithmetic mean is inappropriate). In other words, Thomson Reuters Journal Impact Factor violates even the most basic scientific standards; it generates "an illusion of exclusivity and prestige based on an assumption that it predicts scientific quality, which is not supported by empirical data".

Supporters of scientometrics will argue that, in spite of all its deficiencies, it is the best system we can have, because it is based on *objective measurements*. This reminds us of the drunkard looking for the keys under the lamppost because it was the only place where there was light, though he had actually lost them several feet away. True, the system measures objectively. But maybe it measures something different from the purported quality. Or maybe scientific quality is simply something that cannot be measured. Scientometrics presents the inevitable tendency of every performance indicator to measure what can be measured, and to set aside that which cannot, so that the measurable assumes unwarranted importance (Tipple 1990). Numbers achieve objectivity, but they miss quality.

Why whatever metrics of scientific productivity is chosen, it will be pernicious

Scientometrics can probably avoid some of its worst effects by improving the measurement systems. But, in the end, whenever we design a mechanical feedback system, we obtain a mechanical feedback system (obvious!). The problem itself is the conception of Academia as a mechanical feedback system, because it makes Academia behave according to Darwinian principles.

If we design a mechanical ecosystem to assess scientific productivity, scientists and publishing venues will adapt to assure their own survival, by developing strategies like Salami Science, self-citation or friend-citation, and so on (Reinach 2013). These strategies all combine to create an *unethical and antiscientific culture* where political skills are rewarded too much, and imaginative approaches, high-quality results and logical argument, too little (Lawrence 2003).

Darwinism assures survival of those who are able to evolve and adapt to survive. They are the best because they are survivors, they are survivors because they are the best. Unless we can connect survival with some other aspect of being 'good', apart from the ability to survive, Darwinism is tautological, and assures nothing more than survival. Natural selection is 'mechanical' in this sense: it is the undesigned effect of blind, unintelligent forces. It is highly questionable that it can achieve an intelligent result such as a scientific activity of good quality. Adler et al. (2009b) share this concern for the effects of measurements and ranking systems on the behavior of researchers via natural feedback mechanisms; they also indicate the danger that automated quantitative assessment, even if effective in the short term, may be more easily fooled in the long term than qualitative peer reviewing. Adopting a system, for short term gains, that is so easily open to abuse is a risk to long term research standards.

Therefore, any mechanical feedback system to measure scientific productivity will be pernicious for science in the long term… unless we consider that the goal of science is the survival of adaptable scientists in the Academia ecosystem. An ecosystem where the ethical point of view in research is completely lost. This is the pernicious effect of measuring the quality of science with numbers given by a ranking of journals.

We cannot dispense with human judgment

There is only one way to escape from this vicious circle: recognizing that 'quality', 'goodness', is something that essentially *cannot be measured*, something that is beyond numbers and algorithms, something that can be judged only by humans, despite the fallible character of human judgment. The postulate that there is a positive correlation between impact factor and scientific quality is far from having been demonstrated (Brembs et al. 2013). According to a report from the International Mathematical Union (IMU), the belief that citation statistics are inherently more accurate than human judgment, and hence overcome the possible subjectivity of peer review, is *unfounded*: "using the impact factor alone is like using weight alone to judge a person's health" (Adler et al. 2009a). No doubt, objective measures can aid human judgment. But we deceive ourselves if we think that we can avoid corruption and achieve blind justice by using mathematical formulas.

Moreover, by relying on a formulaic approach to measuring quality and performance, we discourage the creative young men and women whose research careers we want to support and promote. As one of the supporters of the IMU report says, "If they enjoyed being narrowly sized and measured by bean-counters, they'd most likely have chosen a different profession" (Hall 2009).

Therefore, we do not think there is an algorithmic solution to the problem of measuring scientific quality, and we are not proposing it. Instead, not only for convenience, but also based on ethical principles of scientific research, we adhere to the San Francisco Declaration on Research Assessment when it emphasizes "the need to eliminate the use of journal-based metrics, such as Journal Impact Factors, in funding, appointment and promotion considerations; the need to assess research on its own merits rather than on the basis of the journal in which the research is published" (DORA 2012). We do not intend to convert human judgment in a sort of metric (in a certain sense, this is what the impact factor already does, by averaging many individual judgments). We claim that scientific quality is simply beyond what any metrics and algorithms can guarantee. A more intelligent way to collect and interpret citation data is a must, but it will fail again to solve the problem if it is used as a mechanical procedure to assess quality.

It is much easier to collect some figures than to think seriously about what a researcher has achieved. As Lindsay Waters puts it, "there are certain advantages to this way of doing things. One does not need to look directly at colleagues and say that the group of us read your work and found it wanting in the following ways, so please rebut us or you must go, despite the fact that you are a wonderful person" (Waters 2005). It is easier to rely on anonymous numbers to fire or promote someone, to approve or discard a research project.

Conclusion

Our main concern is to raise awareness of the problem. Thousands of scientists have already signed DORA (more than 12.000 at the time of writing this manuscript), and the DORA site collects also dozens of supporting articles in scientific journals, but we think the message deserves to be spread far and wide: *the scientometric bubble is unethical and harmful for science*. We are not against journals or peer-reviewing, and we are not against a proper use of impact factors for the ends they were devised. But we are against the overwhelming value that numbers and formulas are gaining in Academia, at the loss of true quality assessment of individual works. We do not deny the existence of a certain loose correlation between impact factor and scientific quality, but we claim that too much stressing its value corrupts academic life and, what is worse, goes against the dissemination of knowledge.

Ours is not a purely utilitarian argument of convenience. On the contrary, we think there are important issues of fairness and ethical values at stake. In the first place, *values that affect the way a professional career is assessed*. One of the commenters of the IMU report emphatically writes: "reducing an assessment of an individual to a single number is both morally and professionally repugnant" (Silverman 2009). We do not think using impact factors is an absolute evil that should not be done at all. The ethical problem is not in using, but in abusing. However, assessing through impact factors and journal ranking is *so cheap*, compared with other, more qualitative methods, that the evaluator will be strongly tempted to be abusive. Indeed, the true beneficiaries of numerical assessment are neither researchers nor science itself, but evaluation agencies, who can replace scientists (capable of peer reviewing) with mere bureaucrats (capable of counting citations). Using numbers should not be forbidden, but numbers must be relativized, especially avoiding the intention to reduce everything to a one-dimensional scale: "research quality is not something that ought to be regarded as well-ordered", "research quality is an inherently multidimensional object and should be treated as such" (Adler et al. 2009b). Besides, any professional assessment should recognize that success in publishing depends on many factors, some of them purely fortuitous and having little to do with personal scientific merit. We suggest that the contribution of impact factors and journal rankings in an individual's assessment should count less than peer review evaluation.

In the second place, there is also a threat to *ethical values that affect how a researcher approaches his or her scientific activity*. The perversion in the way scientific productivity is assessed perverts the scientist, who becomes worried about publishing for not perishing, instead of being focused on obtaining truer and more reliable knowledge. The researcher, urged to survive within this perverted feedback system, will prefer popularity to intrinsic value, will regard 'where' to publish better than 'what' to publish.

Of course, the utilitarian part is not negligible, either. We have already mentioned some bad effects of rating-by-counting: salami science, friend citation, poor teaching… not the least the very bubble of unnecessary papers that gives title to this essay, because it makes knowledge *less* accessible.

Our message is mainly addressed to the funding agencies and universities that award tenures or grants and manage research programmes, especially in developing countries, more tempted to use the 'cheap' method of counting papers. The message is also addressed to well-established scientists who have the power to change things when they participate in committees for grants and jobs. In the words of David Parnas (in relation to computer science but easily generalizable to other realms of scientific research), "Those who want to see computer science progress and contribute to the society that pays for it must object to rating-by-counting schemes every time they see one being applied" (Parnas 2007).

Albert Einstein is often quoted as having said: "Not everything that can be counted counts, and not everything that counts can be counted". In fact the words must be credited to sociologist William Bruce Cameron (1963). Even though quantitative measurement constitutes one of the foundations of modern science, numerical measures must be used with care and wisdom. Assessing the quality of scientific publications requires human judgment. This judgment can be assisted, but not replaced, by objective measurements.

The obsession to find quantitative and algorithmic methods to evaluate scientific productivity hides an intellectual cowardice, i.e. the evaluator's abdication of his or her responsibility to give a personal judgment on the scientific quality of the evaluated work, so as to become an obedient but absurd bureaucrat that simply applies the maths. Replacing the human factor by 'objective' metrics in the evaluation of science will not avoid corruption.

Human judgments are fallible, but at least they do not promote this scientometric bubble that threatens to paralyze the advance of knowledge by hiding the gold bullion of 'true' science under an enormous overload of publications. In our opinion, allowing the growth of this scientometric bubble is as irresponsible as allowing the growth of financial bubbles. Should we wait until blind Darwinian selection excludes this mechanism?

References

Adler, R., Ewing, J., & Taylor, P. (2009a). Citation Statistics. A Report from the International Mathematical Union (IMU) in Cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). *Statistical Science* 24(1): 1-14.

Adler, R., Ewing, J., & Taylor, P. (2009b). Rejoinder: Citation Statistics. *Statistical Science* 24(1): 27-28.
Barabási, A.-L. & Albert, R. (1999). Emergence of scaling in random networks. *Science* 286 (5439): 509–512.
Brembs, B., Button, K. & Munafò, M. (2013). Deep impact: unintended consequences of journal rank. *Frontiers in Human Neuroscience* 7(291): 1-12.
Broad, W.J. (1981). The publishing game: getting more for less. *Science* 211 (4487): 1137–1139.
Cameron, W.B. (1963). *Informal Sociology: A Casual Introduction to Sociological Thinking*. New York: Random House.
DORA. (2012). San Francisco Declaration on Research Assessment. http://am.ascb.org/dora/. Accessed 23 October 2014.
Hall, P.G. (2009). Comment on Citation Statistics. *Statistical Science* 24(1): 25-26.
Lawrence P.A. (2003). The politics of publication. *Nature* 422 (6929): 259-261.
Larsen, P.O., Ins, M.v. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics* 84(3): 575–603.
Mackay, A.L. (1977). *The Harvest of a Quiet Eye: A Selection of Scientific Quotations*. Bristol: Institute of Physics.
Mattern, F. (2008). Bibliometric Evaluation of Computer Science - Problems and Pitfalls. *European Computer Science Summit - ECSS 2008.* http://www.vs.inf.ethz.ch/publ/slides/Bibliometry-ECSS-Summit-08.pdf. Accessed 23 October 2014.
Merton R.K. (1968). The Matthew Effect in Science. *Science* 159 (3810), 56–63.
Moro, E. (2009). Publish and… perish. http://estebanmoro.org/2009/01/publish-and-perish/. Accessed 23 October 2014.
Nature Materials (2005). Editorial. The cost of salami slicing. *Nature Materials* 4, 1, 2005.
Parnas, D.L. (2007). Stop the Numbers Game. Counting papers slows the rate of scientific progress. *Communications of the ACM* 50(11): 19-21.
Redner, S. (2005). Citation Statistics from 110 Years of Physical Review. *Physics Today*, 58(6): 49-54.
Reinach, F. (2013). Darwin e a prática da 'Salami Science'. http://sao-paulo.estadao.com.br/noticias/geral,darwin-e-a-pratica-da-salami-science-imp-,1026037. Accessed 23 October 2014.
Silverman, B.W. (2009). Bibliometrics in the Context of the UK Research Assessment Exercise. *Statistical Science* 24(1): 15-16.
The Slow Science Academy. (2010). The Slow Science Manifesto. http://www.slow-science.org/. Accessed 23 October 2014.
Thomson Reuters. (2013). Thomson Reuters Statement Regarding the San Francisco Declaration on Research Assessment. http://researchanalytics.thomsonreuters.com/statement_re_sfdra/. Accessed 23 October 2014.
Tipple, C. (1990). Reactions from a CEO. In Fitz-Gibbon, C.T. (Ed.), *Performance Indicators*, BERA Dialogues 2, 1990.
Toffler A. (1970). *Future Shock*. New York: Random House.
Waters, L. (2005). *Enemies of Promise: Publishing, Perishing, and the Eclipse of Scholarship*. Chicago: Prickly Paradigm Press, Second printing.