[Preprint accepted on 20/02/2023; waiting peer rewiev]

Citation metrics and strategic mutations of scientific research: narratives and evidence

Luca Guerra
[orcid: 0000-0002-8903-7744]

Again to her,
Eleonora

Abstract
Following the diffusion of the management model promoted by the New Public Management even within universities, sectors until then endowed with particular protection, such as the academic one, were progressively subjected to new controls and constraints, which gradually took the form of quantitative surveys, with a growing role of citation metrics. These evaluation processes have given rise to various important critical positions at an international level.
If it is true that, according to Goodhart's law, "when a measure becomes an objective it ceases to be a good measure", it is even more significant to note that when a measure becomes an objective, what is measured ceases to be what it was before. The quantitative measurement of academic performance has in fact triggered forms of gaming such as to alter the very game of scientific research, its purposes and the forms of its sharing. In the article we will carry out a comparison between the forms of mutation and the narratives that accompany them to see to what extent today we can legitimately talk about the transformation of scientific research into s*trategic scientific research*.

Upheavals
For several years now we have been witnessing global upheavals against the forms of measurement of research, of researchers and the related policies for the allocation of funds, promotion in academic careers and scientific qualifications. These mobilizations have grown, in particular, following the affirmation of the New Public Management which, since the beginning of the 1990s, after a period of economic crisis, has begun to invest the managerial practices of public administrations and considered the issue of accountability of the foremost importance. When this need for evaluation and measurement entered the univesities, unexpected phenomena, counter-purposes and conflicts arose. The forms of unease have grown a lot in recent years within the community of researchers and have led precisely to organized forms of criticism and opposition.
One of the most important and influential positions taken, in particular, against the evaluation centered on bibliometric indices is the **San Francisco Declaration on research assessment (DORA)**, proposed during the annual conference of the American Society of Cell Biology in December 2012. This declaration specifically targets the Impact factor (IF) as a source of pernicious wide-ranging consequences in the world of scientific research. This declaration does not dispute, on a general level, the use of quantitative methods[1] in the evaluation of research, but specifically the distorted use of the IF, the first and best-known indicator developed by the Institute for Scientific Information (ISI) in 1961 and used on a systematic basis in the Journal of Citation Report (JCR) since 1974. This is indeed an index for which – Biagioli notes - an article paradoxically is already evaluated when coming off the press, already endowed with impact, even before anyone has read and admitted that anyone will ever read it. Eugene Garfield himself, founder of ISI, had decided to take a critical stand[2] with respect to how badly his creature is used. That index - he explains - had been invented in order to help librarians in the selection of journals to subscribe. In general terms - we could also consider - it is plausible to think of highly cited periodicals as high-quality scientific sources even if, on the other hand, this does not give us the certainty that, for example, journals oriented towards the development of new promising lines of research and therefore with a low IF contain articles of poor quality.

[1] For example, instead of the simple IF, it is suggested to adopt a set of numerical indicators such as: impact factor after five years, EigenFactor, SCImago, h-index, (combined with the evaluation of a vague "influence on policies and practices").
[2] In 1998 for example, while defending the overall work of ISI against the criticisms of two German professors, Garfield wrote "The source of much anxiety about Journal Impact Factors comes from their misuse in evaluating individuals, e.g. during the habilitation process".
Here is the full text of the letter to the editor:
http://www.garfield.library.upenn.edu/papers/derunfallchirurg_v101(6)p413y1998english.html

Therefore, the drafters of DORA, after having specified some reasons for opposing the way the IF is used, turn to five interlocutors: Funding agencies, Institutions, Publishers, Organizations that supply metrics, Researchers, with the aim they do not rely on their respective activities on the IF and invite everyone to explore new quantitative indicators, to measure the impact of the single article (rather than its container) and to assign greater importance to qualitative evaluation. From that manifesto, the operational project called Tools to advance research metrics[3] (TARA) came to be, carried out in collaboration with the Center for Science and Technology Studies (CWTS) in Leiden, one of the best-known index developing agencies.

As Figà-Talamanca (2000) aptly pointed out in a harsh critique of the IF (qualified as a "virus"), not only does it make very little sense to use the citations of an abstract entity such as the "average article"[4] that no one has ever written, but also the transformative effect that the diffusion of the IF has produced and produces in the publishing and distribution system must be taken into account. The preferential and competitive orientation to publish in the cluster of higher IF journals determined and in fact determines the concentration of the most interesting articles in that cluster, making the prediction come true by means of the prediction itself. This shift has thus led to an ever greater focus on journals; it has enormously amplified the importance of the "container" and has in fact led to a relatively reduced focus on the article. In a short time, this event generated a rigid hierarchy of journals and laid the foundations for an increasing growth of power in the hands of publishers. Furthermore, what Robert Merton described as the "coin" of science, that is to say above all the citations that are used to pay intellectual tribute, enters a hyperinflationary trend and loses value. Not only does the number of currency-citations explode in parallel with the increasingly widespread use of the IF, but false currency[5] begins to circulate: authors, journals, groups agree to cite each other in abundance. If we recognize that in the very act of citing there are already strategic attitudes (to support certain authors and currents or, on the contrary, to try to marginalize them, even sometimes for personal rivalry), we must recognize that currently the strategic component of citations has enormously increased, where they no longer aim only at intellectual recognition but at influencing assignments, careers and funding[6].

Goodhart's law is widely cited in the literature on research evaluation. It states that “when a measure becomes a goal, it ceases to be a good measure”. However, the most important point to consider is that when a measure becomes a goal, what is measured ceases to be what it was before. The IF has in fact generated and spread adaptive behaviors that the Anglo-Saxons call “gaming behaviors”. According to the classic distinction between play and game, the latter is a self-transforming game, in which the rules themselves become object of the game. The rules that guide the games of scientific research and scholarly communication have thus entered a phase of mutation, the forms of which should begin to be outlined.

A few years later, following the International Conference on Science and Technology Indicators, held in Leiden in September 2014, the **Leiden Manifesto**[7] for research metrics was published in *Nature* in 2015, subdivided into ten principles. The introduction refers to the practice – which has by now become prevalent and "obsessive" – of evaluating research by means of data, often poorly collected and misinterpreted. The principles exhort to: use metrics only as an auxiliary function with respect to the qualitative evaluation carried out by experts; correlate the evaluated performances to the specific goals of the various research projects (basic, frontier, development); not penalize research of local interest and social and humanistic research; make indices and their applications transparent; allow assessees to verify data and analyses; use a suite of indicators diversified by disciplinary field and use normalizations[8] for the placement of journals by sector; evaluate researchers according to several qualitative dimensions of the research work;

[3] https://www.cwts.nl/projects/current-projects/tools-to-advance-research-assessment-tara

[4] Especially in a context of asymmetry. Garfield, in support of the significance of his citation indices, albeit based on a small nucleus of journals, could cite Bradford and his "multiplier" (of sectoral application) for which it resulted that the most "relevant" articles within a discipline are concentrated in few journals. Lotka similarly noted that, within science fields, few authors account for most of the "relevant" papers in those fields. Garfileld too has created his own "concentration law", which is also an economic blessing for him because it saved him the work of having to index mountains of journals. (We spare ourselves Zipf's law here).

[5] Marco Biagioli, relating to this trading of citations, refers to a “post-production misconduct” (Biagioli, 2016).

[6] See, for example, the considerations of Paul Wouters (2020).

[7] https://www.nature.com/articles/520429a

[8] In reality indices normalized by quartiles in the various disciplinary sectors for the purpose of their comparability already exist both in Web of science and in Scopus (naturally each index is a "construction" and therefore the tracing of the boundaries of the disciplines between journals and the disciplinary grouping of the articles is an option exercised by compilers).

not expect numerical precisions that would only be boasted; use a suite of indicators instead of just one, which risks leading to "systemic" behaviour; and regularly review and update indices that need to relate to changes in research practices. In conclusion, the drafters acknowledge that "research metrics can provide crucial information that would be difficult to gather or understand by means of individual expertise" and invite the evaluators to a wise mix of quantitative data and qualitative evaluation.

In mid-2015 **The metric tide**: report of the independent review of the role of metrics in research assessment and management (Wilsdon et al. 2015) was published. It is a study that develops considerations from multiple perspectives starting from the results of a "call of evidence" in which the UK academic community was called to express its views, in particular in relation to the Research excellence framework (the process of evaluation of academic research aimed at the efficient allocation of research fundings). The working group promotes DORA's recommendations, takes note of the growing "tide" of the increasingly widespread and systematic use of metrics ("indices", as they prefer to call them) and, after having evoked the risk of a "new barbary", advocates the adoption of a pragmatic approach that leads to the creation of "responsible metrics"[9]. "Proceeding cautiously, in an exploratory way, seems an appropriate approach" (Wilsdon et al. 2015, 29) - the authors put it. Since the relationship between citation metrics and judgments of quality is problematic as it is highly variable within the various disciplines[10], the report also presents the "informed peer review", "the idea that the judicious application of specific bibliometric data and indicators may inform the process of peer review, depending on the exact goal and context of the assessment" (Wilsdon et al. 2015, 64). But even this evaluation proposal must be taken with caution since - the report suggests - there are various interpretations regarding its concrete implementation and furthermore it is difficult to exclude that the peer review is not in turn mainly influenced by the metrics relating to scientific products being evaluated.

Instead of opposing the tide, almost considered as if it were a natural occurrence, in order to avoid or minimize "strategic behaviors and gaming" and, even worse, forms of "goal displacement" (Mertonian formula upon which the purpose of scientific research to produce new knowledge is replaced by the aim of producing bibliometric impact), the report proposes five principles that should inform these somewhat mysterious entities called "responsible metrics". The proposed principles are divided into: "robustness" (the metrics must be based on the best possible collection of data in terms of accuracy and scope); "humility" (it must be recognized that the quantitative evaluation should support the qualitative evaluation carried out by the experts, and not a substitute); "transparency" (data collection and analysis processes should be kept open and transparent, so that the assessees can check and verify the results); "diversity" (the diversity between fields must be taken into account and a range of indicators that reflect and support a plurality of research and career paths should be used); "reflexivity" (potential and systemic effects of indicators must be recognized and anticipated, and updated in response to those effects).

Since then, manifestos and public positions have multiplied on behalf of disciplinary scientific communities, funding organizations, publishers and various other subjects interested in the production and circulation of scientific knowledge which, with more or less converging accents, envisage the risks of using citation metrics, underline the dangers of increasing pressure on impact and productivity and advocate the adoption of different criteria for evaluating scientific products and researchers' performance in general in order to preserve the integrity of science.

An important step from the level of complaint to that of the proposal was recently taken by the international Coalition for advancing research assessment (CoARA) with the **Agreement on reforming research assessment**[11] (also signed by Italian Anvur at the end of 2022) through which it finally opens an operational site which commits members to sharing their solutions within a year of signing up and to set the end of 2027 as a date by which to verify their suitability and effectiveness. Articulated by principles and commitments, the agreement invites starting from the ongoing projects of TARA for STEMM and from Humetrics[12] (a project financed by the Mellon Foundation) for HSS and in which the affirmation of general principles is combined with the invitation to the rule of thumb (to a pragmatic approach). The transition to an operative plan is certainly to be judged positively and almost certainly such a large deployment of men and means will lead to new proposals. However we can at least note that bibliometrics[13] already makes available a vast

[9] A blog is available to encourage discussion at www.ResponsibleMetrics.org

[10] "The correlation between bibliometrics and peer review is weaker in most fields in the humanities, the applied fields, the technical sciences, and the social sciences. This is partly caused by lower coverage in the citation databases, but also by varying citation and publication cultures" *(Wilsdon* et al. 2015, 64)

[11] https://coara.eu/agreement/the-agreement-full-text

[12] https://humetricshss.org/about

[13] For an introduction in Italian see Nicola De Bellis (2014) and Simona Turbanti (2018)

panoply of indices (variously "responsible"), that the reference to the SCOPE framework[14] (extremely demanding and laborious for the institutes that wish to adopt it) risks creating a babel of further indices, and that peer reviewing appears there as a lifeline, without recognizing its limits and costs[15].

Mutations and teratological cases

In the Leiden Manifesto, reference is made to strong pressure on PhD students to publish in a hurry and in journals with a high IF in order to attract funding; the case of Spanish sociologists who focus on abstract models or on US data in order to obtain higher IF scores is also cited.

How much in general does the objective of bibliometric impact guide and shape scientific research itself, in its contents and styles? To what extent does the process of production and sharing of knowledge change, in its practices, objects and purposes? To what extent can we speak of a transformation of scientific research into *strategic* scientific research? Let's see some of these mutations or at least apparent ones. Retraction watch, which maintains a database of retracted articles (and even corrected) after publication, informs us on an increasing number of scientific publications that have been withdrawn. Oransky, co-founder of Retraction watch, refers to a "retraction boom" (Oransky 2022) recorded in recent years. The author tells us that in 2010 there were an average of 45 retracted articles per month and in 2021 it was almost 300. The overall rate of retractions is a little less than 0.1% but Oransky adds that, based on some reports and investigations conducted by "hunters" of erroneous or fraudulent publications, the rate of articles that should be retracted would be approximately 2%. What is this growth in retractions due to? To a progressive worsening of the qualitative level of the research? It would not seem so. For Oranski, the research community has become increasingly attentive to falsified research and has therefore been able to activate a more efficient system of self-control. Therefore, if the number of retracted publications is growing, it is not so much because the fake ones are growing but above all because they are more effectively intercepted. However, the disturbing fact remains that a not negligible number of researchers knowingly and unscientifically falsify data in order to publish and gain impact. Furthermore, what is worrying is that often “researches continue to cite retracted papers” and that the retraction process “remains comically clumsy, slow and opaque — often taking years, if it ever happens at all” (Oransky 2022, 9). It may be of interest to know that there is a positive correlation (Fang and Casadevall, 2011) between retraction frequency and IF, which can be interpreted by the fact that in those journals it is easier to publish proposing highly relevant discoveries and therefore the authors who want to publish there feel the incentive to unduly or even fraudulently inflate the scope of their results) and also because there is a much higher rate of control and surveillance of articles after publication than in other periodicals. A study (Gaudino et al. 2021) specific to the biomedical field relating to the period 1971-2020 appeared on the JAMA which showed that the rate of retractions had been increasing, especially from 1980 to 2014 (particularly after 2004) and began to decrease starting from 2015, where the absolute number was constantly growing at least until 2018. In the context of the oncological studies Pantziarka and Meheus (2019), in addition to confirming the growing trend of retractions for this filed, highlight how the articles examined had a great potential impact also on learned patients looking for alternative treatments to conventional ones, with a connected increase in the risk of mortality. The good news is that the time between the publication of an article in the oncology field and its identification as erroneous or fraudulent is constantly decreasing, with a reduction in the risk that these articles enter the circle of citations. Particularly disquieting is, as shown by various notes on the Retraction watch blog, that in more than one case the pharmaceutical industries had taken into consideration for the production of drugs some researches which were later unmasked as fraudulent.

The imagination of researchers engaged in goal displacement is very prolific. Among the methods that have become more popular, for example, there is "p-hacking" (also known as "dredging" or "snooping") which consists of at least two variants: performing many statistical tests on the data and reporting only those that give significant correlation results or operate statistical tests on data presented as random ones and instead collected on the basis of some undeclared criteria. Another incorrect way of producing results is HARKing (hypothesizing after the results are known) which

[14] https://inorms.net/scope-framework-for-research-evaluation

[15] See for example: Brezis and Birukou (2020); Dondio et al. (2019); Richard Smith (2006). Biagioli speaks about “fake peer-review rings" to indicate the permeability of the model through fraudulent conduct aimed at the "collaborative" exchange of citations.

The considerations of Figà-Talamanca are still very instructive, as he reconstructs the peer reieving / citation metrics conflict for the Italian biomedical field as a conflict of intergenerational power within the academy.

consists in presenting as a correlation hypothesis following the collection and testing of the data what is instead the previous hypothesis, thus supporting the "publication bias", where it would be less profitable to publish results that give no statistical significance. Another worrying phenomenon within the world of research is what is defined as the "crisis of reproducibility" which precisely put at risk one of the principles of scientificity itself and it is due to many possible factors and it is difficult to say how many are intentional and malevolent and which ones not. The lack or partiality of the data made available, the use of licenses that inhibit or hinder the reuse, the application of proprietary software or with undeclared versioning, all this negatively affects the possibility of control and validation of the results by other research groups. Other pernicious forms for the integrity of science derive from the diffusion of "predatory journals"[16], those journals that boast scientific credits, illustrious editorial committees (whose members often ignore that they belong to them), accuracy in peer reviewing, improbable IF levels or other phantasmagorical indices invented from scratch, international awards and so on, where in reality these journals in many cases do not even carry out the work of checking and selecting publications but simply aim to collect as much revenue from the article processing charge (APC) according to the pay-to-play model. Naturally one can legitimately ask: who are these preys that are so helpless and clueless? Perhaps it is more likely to imagine that these journals are often opportunistically used by researchers not so much to publish articles intended for reading, but for the mere increase of the bibliometric score. The extraordinarily long Bealls' list[17] of "potential predatory journals and editors" is regularly updated on the web. Then there is that emerging multifaceted array of "post-production misconducts" (hypothetically without prejudice to the intrinsic quality of the research) which ranges from outright fraud to adaptive behavior and which varies as the metrics vary, and to which the already mentioned citation explosion has to be connected. This is the circle of "never ending tuning" (of the indices) and "hacking" (of the same) on which Biagioli (2016) insists a lot, however representing it almost as a self-generating process, without - in our opinion - adequately recalling the fact that it comes full circle each time because a researcher fails to respect the founding principles of research integrity and the public sharing of its results (because, as is known, there is virtually nothing that cannot be manipulated).

Again, there were frequent complaints of cases for which the author had suggested some reviewers as evaluators and communicated email addresses he himself owned.

Then there is the case, bordering on the incredible, of the sw SCIgen, invented by some goliardic MIT students to create pseudo computer science articles that appear credible at first sight, and which were used in particular by some Chinese academics to publish articles in a sort of continuous stream (also published by renowned publishers). The most recent ChatGpt bot (and who knows for its brothers to come), based on artificial intelligence, promises new pseudoscientific goodies being able, in addition to answer questions, to produce texts, sw codes, contracts, images, etc. That the system of controls and metrics can in fact be widely played is also testified by Mr. Ike Antkare (I can't care) endowed with an h-index higher than that of Einstein, even though he never existed, except in the imagination of the impertinent Cyril Labbé.

Less worrying but nonetheless significant is the denounced practice of "salami-slicing" consisting of splitting the communication of the results into several articles in order to increase productivity but at the cost of making the reader's reconstruction work more arduous and cumbersome. However, it should be highlighted that there is statistically a long-term trend in the average growth of the length of publications, so that from the salami-slicing we can infer a lengthening of the content (and not a reduction in the length of the articles). Furthermore, the growing practice of co-authorship[18], whereby "the average number of authors of publications in the scholarly literature continues to increase" (Wilsdon et al. 2015, 32) is even more ambiguous to interpret because it can constitute a response to the incentive to show high individual levels of scientific prolificacy by strategically increasing the research groups with which one collaborates, or it can indicate the choice of a purely figurative and negotiated collaboration as a response to pressure to publish or it can indicate an increase in the level of (multidisciplinary and transdisciplinary) collaboration typical of current research.

[16] In reality there are few cases in which there are final judgements for forgery against publishers of scientific journals. There is a very large gray area within which publishers can adopt predatory or spamming behaviors and also publish validated scientific articles. However, it would perhaps be easier to identify which are the *non-predatory publishers*, given the outrageous pricing policies (and now, following the affirmation of the "transformative agreements", APC policies) by renowned publishers who legally undermine the process of sharing scientific knowledge.

[17] https://beallslist.net

[18] The co-authorship record probably belongs to an article published in 2014 by the Physical review letters signed by 5,154 researchers

Some authors believe that the overall changes that have occurred in the world of research have profoundly altered the physiognomy of the scholarly communication process in the space of a few decades, more than what has happened since the mid-seventeenth century. Mark Edwards and Siddharta Rey, in an article with the striking title “Science is broken”, after having considered that “Competition among researchers for funding has never been more intense, entering an era with the worst funding environment in half a century”, write (prudently referring however to a hypothetical future): “We believe that this system presents a real threat to the future of science. If immediate action is not taken, we risk creating a corrupt professional culture akin to that revealed in professional cycling (i.e. 20 out of 21 Tour de France podium finishers during 1999-2005 were conclusively tied to doping), where an uncontrolled perverse-incentive system created an environment in which athletes felt that they had to cheat to compete” (Edwards and Rey, 2017). The results of a questionnaire submitted to a few thousand researchers in the United States led the authors to hypothesize that "certain features of the research working environment may have unexpected and potentially detrimental effects on the ethical dimensions of scientists' work" (Martinson 2005). Other authors write “some violations are considered to be related to the increased pressures to publish” (Hayer 2013).
Parallel to these alarming complaints, alongside consolidated institutions such as the US Office of Research Integrity, often informal initiatives have arisen managed by groups of researchers (sometimes anonymous and with a taste for the sneer) such as Retraction watch, PubPeer and other blogs as “watchdogs” precisely to intercept these unprecedented forms of gaming. Consequently, the issue of the ethics of science, of the responsibility and integrity of those engaged in it has also returned to great relevance. A particularly significant recent document is the Hong Kong Five Principles (Moher et al. 2020), a guide for researchers and evaluators to preserve the “integrity” of science. The "Allea European code of conduct for research integrity" proposes principles and good practices for the preservation of research freedom and the promotion of forms of self-regulation and self-correction[19]. The "Slow science manifesto"[20] denounces the risks for the quality of research and refers to knowledge as a common good, today considered at risk due to recent forms of science management. The guidelines of the Committee on publication ethics[21] (COPE) establish some ethical principles by which scholarly communication must abide.
The theme of science ethics is probably more radical today than is generally believed and - in our opinion - constitutes the most important and still largely opaque nucleus within what in Italy today is defined as the "third mission" of universities. Away from the rhetoric and cloying paternalism that often accompanies reflections on the third mission and also on citizen science[22], scientists are asked a radical question on their own exercise, on the consequences it produces in society, on the planet and in general on all objects of meaning, to the point of questioning the scientific method itself, in its age-old and joint epistemological structures of disinterest and observation.

Data production and publish or perish culture

At the beginning we said that the application of the new public management principles and in particular accountability had a substantial role in reshaping the structures and practices of public administrations. However, at least one other aspect must be taken into account in our reconstruction that is the ever-increasing use of quantitative methods in the collection, processing and evaluation of data. Like never before we are submerged in a gigantic sea of data, like never before we have advanced tertiary societies produced such an industrial quantity of data. This is not the place to deepen the notion of *datum*, which is mostly assumed objectively, but it is good at least to be aware that the *datum* is never an object, but rather a *relation*. It is located at the precipitation point of a complex interweaving of practices, operations and tracings that put it and keep it at work. To remember a famous statement by Foucault, datum is not waiting in a limbo for an act of our attention and therefore the well-known rigmarole data-information-knowledge does not help us much. This exponential growth of data is mainly due to the fact that automatic electronic writings have been connected

[19] https://allea.org/code-of-conduct
[20] https://slowscience.be/the-slow-science-manifesto-2
[21] https://publicationethics.org
[22] In the declarations of principles of the European Citizen Science Association and the American Citizen Science Association, there are exhortations to consider citizen research as equal to the professional one. However, if one reads the statutes of some initiatives they will find almost systematically (and - in our opinion - inevitably) a clear distinction between researchers and practitioners. A slice of spicy chorizo salami, recently passed off as an image of Proxima Centauri, was enough to cool the enthusiasm for crowd science.

to our gestures and therefore in explicating they produce increasingly dense and "granular" recordings, very often without either intention or awareness. In addition to being automatic, these writings take place at very low cost and therefore expand very easily producing the "audit society" (Power 1997) or the "surveillance society". Again, to refer to the Foucauldian perspective, this move brings about an overlapping of the disciplinary rationality (based on the recording and control of detail and which makes use of epistemological, organisational, spatial, architectural, electronic "grids" (*quadrillages*)) and the biopolitical rationality which identifies and shapes normal and statistical trends of social, ergonomic and biological order). The new technologies make the detail and the massive possible at the same time and therefore give computation (algorithms) an unprecedented and growing range of social action. Therefore - it is good to make this clear - the issue of evaluation cannot be resolved in a discussion that is limited to citation metrics and researchers'performance only because it affects mostly privileged and more protected social groups such as those of the academy. For the purpose of its understanding, it cannot remain on a corporative level but requires consideration within a wider horizon.

In the world of research, this capillary and invasive process, has begun to trigger elusive conducts, resistances, and games.

It has long been argued that a hyper-competitive culture has been introduced into academia, legitimized by the assumption that the more intense the competition between researchers, the greater the results in terms of productivity and improvement in the quality of research. A great deal of literature - conversely - maintains that what it defines as a culture of *publish or perish* (and also *impact or perish*) produces opposite effects to those expected and generates perverse incentives which, instead of improving the scientific enterprise, undermine its foundations, trigger mutations and teratologies. The risk - as has been said - is that a *broken science* emerges from publish or perish. Converging evidences are also gathered on the level of experience. For example, the results of a questionnaire on the topic submitted to Dutch biomedical researchers lead the authors of the survey to consider that "substantial proportion of the responding medical professors judge publication pressure as having become excessive, and a substantial part believes that this affects the validity and credibility of medical science" (Tijdink et al. 2013). According to the statements, the problem also moves to the level of one's own well-being and health, which in turn has consequences in terms of scientific, clinical and teaching efficiency. In fact, the authors write "Apart from general effects on medical science and practice, our study suggests that excessive publication pressure has detrimental effects on personal well-being. […] publication pressure causes them to develop a cynical view on medical science, and may be associated with increased risk of developing burn-out" (Tijdink et al. 2013).

There is no reason to doubt that those statements about a growing personal and collective unease are not authentic and consequent to effective transformations. Yet it is not at all easy to recognize a certain and direct link between the use of citation metrics, POP culture and questionable or fraudulent scientific conduct. Already in 2011 in a very complex (and very abstract) behavioral model, the researchers Lacetera and Zirulia questioned the causal link between POP culture and manipulative or fraudulent conduct. On the contrary, the pressures on impact and productivity and the high level of competition can have - in the authors'view - a dissuasive effect on incorrect conduct. We read: "This pressure can actually serve as a powerful mechanism to deter fraud as it increases the incentives of peers to scrutinize each other's work. This is due both to the increase in the reward from discovering fraud and to the anticipation of the author's higher temptation to cheat" (Lacetra and Zirulia 2011, 588). More recently, Fanelli on the basis of a large statistical study[23] that focuses on the first 15 years of career of a large number of researchers tends to exclude some effects of POP and concludes "In sum, according to our findings, scientists today are not publishing, on an individual basis, at higher rates than their colleagues in the 1950s" (Fanelli 2020, 117). It is also difficult to discriminate between behaviors induced by the institutional pressure of the POP (for which one either has a high impact and high productivity or is left out) and those instead subjectively motivated by careerism and the desire to collect bonuses (option left to the freedom of the individual). Fanelli writes "The pressure to publish hypothesis is not supported because the only significant patterns we detected in these and previous analyses support a role for cash incentive policies, which do not constitute real 'pressures' on researchers, but rather an incentive that is misaligned with the scientific norms of disinterestedness" (Fanelli 2022, 15). While aknowledging that Fanelli's observations concern a limited sample of journals, concentrate on specific misconduct cases, consider circumscribed disciplinary fields, select (necessarily) a limited number of variables, hypothesize "intentions" and therefore are neither entirely generalizable nor conclusive, they however warn us not to access easy simplifications on the phenomena taking place in the field of scientific research and communication. Let us

[23] In relation to these studies the pompous name of metascience has been used, where more prosaically they consist in the application of statistical methods to *scholarly communication*

add that what we have defined as upheavals concerns small minorities of researchers; for example, the DORA declaration, one of the best-known documents, to date is signed by about 20,000 individual researchers and 2,700 organizations, numbers that are by no means negligible, where however the global number of researchers is about nine million[24]. Letting aside the naiveties of some champions of peer reviewing fiercely hostile to metrics, it is good to know that this is also a game *internal* to scientific community, not just a battle of pure principles. The measure of the reciprocal weight between the two methods of evaluation is polarized within a field of forces and can determine changes precisely in terms of careers, distribution of funds, assignment of tasks, etc.
In short, we are not allowed to simplistically say that competition, performance measurement, the use of citation metrics have destructive effects by themselves. From the worrying signals of some mutations, from the multiple upheavals and from the evidence of unease, what we can hypothesize is that we are probably approaching a *threshold* (perhaps already exceeded in many areas) which can induce the transformation of scientific research into *strategic* scientific research. Hence now is the time for careful and critical study of scholarly production and scholarly communication and not for caricatural statements about evil bureaucrats in conspiracy against science.

References

Biagioli, Marco. 2016. "Watch out for cheats in citation game". *Nature* 535: 201 https://doi.org/10.1038/535201a.

Biagioli, Mario e Lippman, Alexandra (edited), ed. 2020. *Gaming the metrics: misconduct and manipulation in academic research*. Cambridge: The Mit press.

https://mitpress.mit.edu/9780262537933/gaming-the-metrics

Brezis, Elise S., and Aliaksandr Birukou. 2020. "Arbitrariness in the Peer Review Process". *Scientometrics*, 2123: 393–411. https://doi.org/10.1007/s11192-020-03348-1;

Cari-Ann, Hayer (et al.). 2013. "Pressures to publish: catalysts for the loss of scientific writing integrity?". *Papers in Natural Resources* 689. Doi: 10.1080/03632415.2013.813845
https://digitalcommons.unl.edu/natrespapers/689

De Bellis, Nicola. 2014. *Introduzione alla bibliometria*. Roma: AIB.

Dondio, Pierpaolo (et al.) 2019. "The 'Invisible hand' of peer review: the implications of author-referee networks on peer review in a scholarly journal". *Journal of Informetrics* 13: 708–716. https://doi.org/10.1016/J.JOI.2019.03.018.

Edwards, Mark A. and Roy Siddhartha. 2017. "Science is broken. Perverse incentives and the misuse of quantitative metrics have undermined the integrity of scientific research". Aeon.
https://aeon.co/essays/science-is-a-public-good-in-peril-heres-how-to-fix-it

Fanelli, Daniele. 2020. "Pressures to publish: what effects do we see? In *Gaming the metrics: misconduct and manipulation in academic research*, a cura di Biagioli, Mario e Lippman, Alexandra, 111-122. Cambridge: The Mit press. Doi: https://doi.org/10.7551/mitpress/11087.003.0011

Fanelli, Daniele (et al.) 2022. "Do individual and institutional predictors of
misconduct vary by country? Results of a matched-control analysis of problematic image duplications". *PLoS One* 17 (3) e0255334. https://doi.org/10.1371/journal.pone.0255334

Fang C., Ferric and Casadevall, Arturo. 2011. "Retracted science and the retraction index". *Infective Immunology* 79 (10): 3855-9. Doi: 10.1128/IAI.05661-11

Figà-Talamanca, Alessandro. 2000 "L'impact factor nella valutazione della ricerca" [Draft of the speech, Bologna 27 June 2000].

[24] https://sciencebusiness.net/news/number-scientists-worldwide-reaches-88m-global-research-spending-grows-faster-economy

http://siba.unipv.it/biblioteche/info/SISSCOWEB_A_Fig_Talamanca_L%27impact_factor_nella_valutazione_della_ricerca_SISSCOWEB.htm

Garfield, Eugene. 1998. "The impact factor and using it correctly". *De Unfallchirurg*, 48 (2): 424. http://garfield.library.upenn.edu/papers/derunfallchirurg_v101%286%29p413y1998english.html

Gaudino, Mario (et al.). 2021. "Trends and characteristics of retracted articles in the biomedical literature, 1971 to 2020". *JAMA international medicine* 181 (8): 1118-1121
Doi: 10.1001/jamainternmed.2021.1807

Lacetera, Nicola and Zirulia, Lorenzo. 2011. "The economics of scientific misconduct". *The Journal of Law, Economics, and Organization*. 27: 568–603. https://doi.org/10.1093/jleo/ewp031

Martinson, Brian C. (et al.). 2005. "Scientists behaving badly". *Nature*. 435: 737-738

Moher, David, et al. (2020). "The Hong Kong Principles for assessing researchers: fostering research integrity". *PLoS Biology* 18 (7). https://doi.org/10.1371/journal.pbio.3000737.

Naujokaitytė, Goda. 2021. "Number of scientists worldwide reaches 8.8M, as global research spending grows faster than the economy" in *Science business*
https://sciencebusiness.net/news/number-scientists-worldwide-reaches-88m-global-research-spending-grows-faster-economy

Oranski, Ivan. 2022. "Retractions are increasing, but not enough". *Nature* 608: 9. https://doi.org/10.1038/d41586-022-02071-6

Pantziarka, Pan and Meheus, Lydie. 2019. "Journal retractions in oncology: a bibliometric study". *Future oncology* 15 (31): 3597-3608. https://doi.org/10.2217/fon-2019-0233

Power, Michael. 1997. *The Audit Society*. Oxford: Oxford University Press

Smith, Richard. 2006. "Peer review: a flawed process at the heart of science and Journals". *Journal of the Royal society of medicine*, l99: 178-182.

Tijdink, Joeri K (et al.). 2013. "Publication pressure and burn out among Dutch medical professors: a nationwide survey". *PLoS One* 8 (9):e7338. Doi: 10.1371/journal.pone.0073381.

Turbanti, Simona. 2018. *Strumenti di misurazione della ricerca*. Milano: Editrice Bibliografica.

Wilsdon, James et al., ed. 2015. *The metric tide: report of the Independent review of the role of metrics in research assessment and management*. Thousand Oaks: SAGE publications.
doi: 10.13140/RG.2.1.4929.1363. https://sk.sagepub.com/books/the-metric-tide

Wouters, Paul. 2020. "The mismeasurement of quality and impact". In *Gaming the metrics: misconduct and manipulation in academic research*, edited by Mario Biagioli and Alexandra Lippman, 66-75. Cambridge, The Mit press. https://doi.org/10.7551/mitpress/11087.003.0006