International Journal of Communication 9(2015), Feature 1880–1887 1932–8036/2015FEA0002

The Index and the Moon: Mortgaging Scientific Evaluation

FLORENCIO CABELLO FERNÁNDEZ-DELGADO[1]
MARÍA TERESA RASCÓN GÓMEZ
University of Malaga, Spain

Keywords: scientific evaluation, impact factor, journal ranking, scholarly communication, open access, DORA

Warning: This article represents an explicit invitation to foolishness. If the old Chinese proverb "When the wise man points at the *moon*, the fool looks at his *finger"* is right, we confess we are fools unable to take our eyes off a particular kind of finger that has become increasingly important in today's society: *indexes*—whether they are linked to pension evaluations, mortgage rates, the budget deficit, sovereign debt ratings, or any other focus of measurement.

Our work at the university has led us to focus on one of these indexes in particular. We are referring to the finger that points toward the heaven of academic papers, research careers, and funding of scientific projects: *the impact factor*. This is an annual metric of the relevance of a scientific journal based on the average number of citations received by the articles published in it in the previous two years. The task of calculating and publishing this metric in the *Journal Citation Reports* (JCR) has been performed since 1975 by the Institute for Scientific Information. This organization is part of the publicly traded company Thomson Reuters, self-defined as "the *world's* largest international multimedia *news agency.*" The impact factor is vested with official recognition through its transposition to the scales with which evaluation agencies and academic institutions calibrate the relevance of scientific contributions and, consequently, decide which projects and researchers deserve credit (in the form of funding, hiring, promotion, and tenure) and which do not.

In our defense, we must admit that we would have never lost sight of the moon of academic impact if we had not met Juan Moreno Yagüe and Francisco Jurado. Back in 2012, these two Spanish jurists were foolish enough to look at the Euribor (the European Interbank Offered Rate) while all the wise men of banking prestidigitation pointed at the ether of the financial markets. To better understand this foolishness, we should begin with the explication of Euribor included on the site of its Operation Euribor campaign:

[1] We want to express our gratitude to Professors Langdon Winner and Antonio Lafuente for their inestimable comments and support. We are also indebted to Ashley Ashkenazy for her revision of the text. Of course, any remaining mistakes are alone our responsibility.

> The Euribor . . . is a daily published reference rate indicating the average interest rate at which Eurozone banks offer to lend unsecured funds to other banks in the euro wholesale money market (or interbank market). It is calculated by using data from the main 44 banks operating in Europe, some of which are international. Its monthly value is routinely used as reference for bank loans. (#OpEuribor, n.d., para. 1)

The task of calculating and publishing this metric in the Telerate (a sort of financial teletext) has been performed since 1999 by Thomson Reuters. This metric is vested with official recognition by its monthly publication in the official state gazettes by the national central banks in the European Union, as a reference value with which financial institutions set the interest rates of their variable-rate loans (not to mention its influence in public procurement and other official procedures).

Juan Moreno Yagüe and Francisco Jurado's foolishness led them to a shocking conclusion: The Euribor is "the biggest financial scam in European history" (Jurado, 2013, para. 15). The main hypothesis of their investigation can be read in detail at http://opeuribor.es/our-research/, but it can be summarized as follows: Since 2007 and the collapse of subprime mortgages, the data offered by the panel banks about interest rates are merely declarations of intent without the backing of real loan operations (Enrich, 2012, para. 11–13; Jurado, 2014, pp. 74–76).

The record fines the European Commission imposed in 2013 to six banks in the Libor/Euribor panel "for colluding to fix key interest rate benchmarks" (Treanor, 2013, para. 1) seems to endorse the #OpEuribor hypothesis, even if the commission ended up creating a sort of smoke screen about the nature and quantity of this manipulation.

At this point, these questions are inevitable: How was it that large, private banks were able to set interest rates arbitrarily and for their own benefit? Where were the national central banks, the European Central Bank, and the European Banking Authority when they were most needed? We can affirm that they had their heads in the clouds. In fact, all the regulatory bodies consulted by #OpEuribor made themselves scarce and ended up referring to Thomson Reuters.

One might argue that if the European regulators transferred their responsibilities of financial control to a "news agency" that is publicly traded, and if this news agency only received declarations of intent without comparing them with the real banks' operations, then the Euribor calculation would be little more than an opinion poll. And Thomson Reuters would not be to blame if the official state gazettes then recorded its results on their pages. Indeed, this is the same objection the rating agencies Standard & Poor's, Moody's and Fitch raised when they testified before the U.S. Congress Committee on Financial Services in 2011. According to their testimony, their credit risk assessments (including the triple-A awarded to Lehman Brothers shortly before its failure) were just opinions and therefore fell under the umbrella of First Amendment protection.

At this point, some readers might be asking what the Euribor has to do with the impact factor in scientific evaluation. Are we suggesting that Thomson Reuters's JCR also lacks a solid basis? Surely it would be reckless to consider that the criticism of Thomson Reuters's calculation of Euribor might sow

doubts on the other activities of this global information company. However, this does not preclude pointing out three disturbing shortcomings in the calculation process of the impact factor: It is negotiated, methodologically flawed, and irreproducible.

Regarding negotiation, *PloS Medicine* revealed some of the controversial editorial strategies it discovered while speaking with Thomson Reuters to understand the rules of the "impact factor game" (Editorial, 2006). Especially shocking among those strategies is the attempt by journal editors to have Thomson Reuters remove certain types of (less citable) articles from the denominator of the impact factor equation. This may lead to dramatic variations in the resulting number, "given that all citations count toward the numerator whether they result from publications included in the denominator or not" (Brembs, Button, & Munafò, 2013, p. 6). A good illustration of this manipulation is the "retrospective change in the denominator" (ibid.) whereby *Current Biology* increased its impact factor from 7.007 in 2002 to 11.910 in 2003.

A second shortcoming concerns the biased statistical approach employed to quantify a journal's impact, taking into account that citation data are strongly left-skewed (with a large number of citations concentrated in a few publications): "The use of an arithmetic mean as a measure of central tendency on such data (rather than, say, the median) is clearly inappropriate, but is exactly what is used in the IF [impact factor] calculation" (Brembs et al., p. 6). According to Moustafa (2015), another "disastrous effect" (p. 141) of the statistical approach is that the same unsound methodology is applied to rank universities at a global level: "Universities' ranking systems are based on subjective and biased criteria that compare heterogeneous institutions, which are different in almost everything, including staff, number of students and professors, number and rhythm of workday hours, infrastructure, specialization and equipment" (Moustafa, 2015, p. 141).

With respect to the lack of transparency of the original data handled by Thomson Reuters, Rossner, Van Epps, and Hill (2008) criticized Thomson Reuters's estimates as "irreproducible results" (p. 183) that would be unprintable in any JCR-indexed journal: "Just as scientists would not accept the findings in a scientific paper without seeing the primary data, so should they not rely on Thomson Scientific's impact factor, which is based on hidden data" (Rossner, Van Epps, & Hill, 2007, p. 1092).

Any of these concerns alone is serious enough to suggest eradicating the impact factor from science metrics, but something worries us even more: Nobody doubts that the information monopoly is a precondition of Thomson Reuters's global leadership, and that publishing negotiated and irreproducible results obtained via an unsound methodology does not deprive news agencies of their right to free speech. Our greatest concern is whether our scientific assessment agencies are giving official recognition to an index appropriate for what they are supposed to evaluate. We fear that both the original design of the impact factor and the review of empirical data lead to a negative answer.

With respect to the latter, the current empirical literature suggests a significant correlation between journal rank and scientific unreliability. Thus, in Fang, Steen, and Casadevall's (2012) analysis of the "the ongoing retraction epidemic" (p. 17028), the authors find "a highly significant correlation . . . between the journal impact factor and the number of retractions for fraud or suspected fraud and error"

(p. 17030). Moreover, the comprehensive literature review carried out by Brembs, Button, and Munafò (2013) takes these disturbing (albeit counterintuitive) findings even further: "Publications in high ranking journals are not only more likely to be fraudulent than articles in lower ranking journals, but also more likely to present discoveries which are less reliable (i.e., are inflated, or cannot subsequently be replicated)" (p. 4).

With respect to the original design of the impact factor, we should not overlook the fact that the impact factor was originally devised to measure the relevance of journals (thus helping librarians select their acquisitions), but it is actually used for measuring the relevance of articles and individual authors. Are we discovering anything new? Of course not; the co-creator of the impact factor himself (and now Thomson Scientific shareholder) recognizes that this was its origin. Nonetheless, he still defends it as "a good technique for scientific evaluation," because "there is nothing better" (Garfield, 2006, p. 92).

There Is Nothing Better?

More and more international scientific researchers, associations, and journals believe that if that is the best argument available to defend the impact factor, then the time has come to transform scientific evaluation from its very foundations. This idea is formulated in the bluntest possible terms by Brembs, Button, and Munafò (2013): "Given the data we surveyed above, almost anything appears superior to the status quo" (p. 8). As a result of this growing conviction, on December 16, 2012, the American Society for Cell Biology (ASCB) released the *San Francisco Declaration on Research* Assessment (DORA). As of this writing, this declaration is supported by 12,377 researchers and 572 scientific societies, including the addition of the prestigious League of European Research Universities (2015) on March 16, 2015. The lucid recommendations contained in the three pages of the DORA deserve careful reading, but here we highlight some of its key proposals:

1. "The need to eliminate the use of journal-based metrics, such as Journal Impact Factors, in funding, appointment, and promotion considerations" (ASCB, 2012, para. 7). Even though the DORA envisages transitory mechanisms, its diagnosis is unavoidable: It is not a matter of patching up the impact factor of "abuses" (Van Noorden, 2013, p. 7453) or "distortions" (Alberts, 2013, p. 787), but rather of abandoning it altogether. One early warning along these lines was issued by the British House of Commons Science and Technology Select Committee (2004): "We urge HEFCE [Higher Education Funding Council for England] to remind RAE [Research Assessment Exercise] panels that they are obliged to assess the quality of the content of individual articles, not the reputation of the journal in which they are published" (para. 210). The committee's concern was that the widespread (and supposedly wrong) assumption among UK researchers that the impact factor was used to assess the quality of their articles was leading the RAE to "indirectly support[s] a hierarchy of journals, making it difficult for new and little-known journals . . . to compete" (para. 210). This point was recently taken up by Nobel Laureate in Physiology or Medicine Randy Schekman (2013), who defended that "funders and universities . . . must tell the committees that decide on grants and positions not to judge papers by where they are published" (para. 8) with a view to

"break the tyranny of the luxury journals" (para. 9). Kansa (2013) takes the criticism of the impact factor one step further, denouncing that even if its origins were "relatively benign," then it has become "a tool for Taylorism and the (coercive) monitoring of research outputs by university bureaucracies" (para. 11).

2. The DORA also recommends that journal publishers "remove all reuse limitations on reference lists in research articles and make them available under the Creative Commons Public Domain Dedication" (ASCB, 2012, para. 20). What is the purpose of this recommendation? To transfer the commitment to the Semantic Web and Open Data (Berners-Lee, 2009) to the field of scientific metrics.

   In this context, the promotion of open access (Suber, 2012) also appears to be peremptory in order to "bring[ing] scholarly communication back to the research institutions in an archival publication system in which both software, raw data and their text descriptions are archived and made accessible" (Brembs et al., 2013, p. 8). However, even if journals were not open access, the freedom to access and reuse the bibliographic metadata would suffice to generate a database of citations that could open-source the code of impact ratings. This would mean the end of irreproducible results, because anyone (scientific or not) could verify them!

3. "The need to assess research on its own merits rather than on the basis of the journal in which the research is published" (ASCB, 2012, para. 8). Three aspects seem especially noteworthy in this truism/heresy (delete where not applicable): First, the importance of transferring this change of focus to early-stage researchers (who are the most sensitive to these indexes); second, the expansion in the definition of research output to encompass data sets and software (anyone who has watched Hans Rosling's TED Talks can easily understand why); and, third, the promotion of qualitative indicators of research impact, such as the influence on policy and practice (Editorial, 2006, para. 6). Obviously, evaluating these facets is complex (undoubtedly far more complex than the mere transcription of a precooked index), but it is worth doing if it encourages an enriching dialogue between science and society.

All these proposals make the DORA an inspiring document to help tackle the issues of research assessment in a sensible and accurate manner. Nonetheless, we consider that, in order to offer a more powerful interpretation of this declaration (and, by extension, of the open-access movement), we should situate it in a broader framework that allows us to discuss the "political challenges of open science" (Rendueles & Aristegui, 2014, p. 45).

In our opinion, this approach is relevant to avoid the "procedural inertia" (Rendueles & Aristegui, 2014, p. 56) of open science, by which it "accepts discussing within a field where institutional and political conflicts tend to be minimized by conceiving them as formal problems that can be solved technically" (ibid.). In fact, it is worth noting the growing awareness that behind the above-mentioned "aberrations" (Moustafa, 2015, pp. 140–141) are institutional factors linked to the "publish or perish" mandate (Wesel,

2015, p. 3) and to the social pressure generated by "the general increase in competitiveness, and the precariousness of scientific careers" (Shapin, 2008, as cited in Brembs et al., 2013, p. 4). Rendueles and Aristegui (2014) extend this criticism to question the "amazing level of unanimity" peer review has reached among open-access journals despite its notable "limits" (pp. 53–55). For them, both journal ranks and peer review can be seen as devices aimed at "hiding, but not eliminating, the frontiers between actual scientific knowledge and the contextual and institutional factors whose contingency is always uncomfortable" (p. 55).

In other words, we do not think we should read the DORA as an instruction book to replace *wrong* metrics with *right* ones, but as a lucid account of concerns and principles regarding the social value of research assessment. On this issue, Brembs, Button, and Munafò (2013) make their point very clear by stating that "any journal rank (not only the currently favored Impact Factor) would have this negative impact" (p. 1). This leads us directly to the skepticism about the role of metrics in the advocacy of open science and collaboration expressed by Kansa (2013): "But are more metrics (even Alt-metrics) really the solution to the perverse incentives embodied by our existing metrics?" (para. 13). However valuable they might be, the fact that they are used in the neoliberal academia as a coercive tool to decide (for instance) whether we are hired or fired makes it difficult to consider the question as a mere procedural one. Instead, paraphrasing Langdon Winner (1989), one might reasonably wonder whether the quest for sophisticated metrics means just a commitment to "building a better mousetrap" (pp. 78–79):

> With employment increasingly precarious, professional pressures balloon in ways that make risk taking and going outside of established norms unthinkable. Adding more or better metrics without addressing the underlying job security issues just adds to the ways people will be ejected from the research community. (Kansa, 2013, para. 16)

Without losing sight of the urgent need of reform in the funding and organization of research, we can conclude that the DORA offers a solid alternative to mainstream scientific indexes that finds resonance in many other social realms. After all, attending to indexes means attending to what a society believes—that is, to the way it gives *credit* (be it financial, political, journalistic, or scientific). Blind faith is not just the antithesis of the attitude scientists are supposed to adopt, but also an antidemocratic threat that no engaged citizen can afford to ignore. Hence, the ultimate sense of our initial invitation to foolishness: If we aim to gain a clear (in)sight of the moon, we must prevent that, by looking away, our indexes get out of our hands until they end up eclipsing it.

References

Alberts, B. (2013, May 17). Impact factor distortions [Editorial]. *Science*, *340*(6134), 787. doi:*10.1126/science.1240319*

ASCB (American Society for Cell Biology). (2012, December 16). *San Francisco declaration on research assessment*. Retrieved from http://am.ascb.org/dora/

Berners-Lee, T. (2009, February 4). The next web [Video file]. Retrieved from http://www.ted.com/talks/lang/en/tim_berners_lee_on_the_next_web.html

Brembs, B., Button, K. & Munafò, M. (2013). Deep impact: Unintended consequences of journal rank. *Frontiers in Human Neuroscience, 7*(291)*,* 1–12. doi:10.3389/fnhum.2013.00291

Editorial: The impact factor game: It is time to find a better way to assess the scientific literature [Editorial]. (2006). *PLoS Medicine, 3*(6), e291. doi:10.1371/journal.pmed.0030291

Enrich, D. (2012, December 9). Banking industry squirms over European rate probe. *The Wall Street Journal*. Retrieved from http://www.wsj.com/articles/SB10001424052970203880704578087211853601302

Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America, 109*(42), 17028–17033. doi:10.1073/pnas.1212247109

Garfield, E. (2006, January 4). The history and meaning of the journal impact factor. *Journal of the American Medical Association, 295*(1), 90–93.

House of Commons Science and Technology Select Committee. (2004, July 7). *Science and Technology. Tenth Report*. Retrieved from http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39912.htm

Jurado, F. (2013, November 19). The biggest financial scam in European history. *eldiario.es/guerrillatranslation*. Retrieved from http://guerrillatranslation.com/2013/11/19/the-biggest-financial-scam-in-european-history/

Jurado, F. (2014). *Nueva gramática política. De la revolución en las comunicaciones al cambio de paradigma* [New political grammar. From the communication revolution to the paradigm shift]. Barcelona, Spain: Icaria.

Kansa, E. (2013, December 11). It's the neoliberalism, stupid: Why instrumentalist arguments for open access, open data, and open science are not enough. *Digging Digitally*. Retrieved from http://www.alexandriaarchive.org/blog/?p=931

League of European Research Universities. (2015, March 16). Not everything that can be counted counts . . . [Press release]. Retrieved from http://www.leru.org/index.php/public/news/not-everything-that-can-be-counted-counts-

Moustafa, K. (2015). The disaster of the impact factor. *Science and Engineering Ethics, 21*(1), 139–142. doi:10.1007/s11948-014-9517-0

#opEuribor (n.d.). What Is Euribor? Retrieved from http://opeuribor.es/what-is-euribor/

Rendueles, C., & Aristegui, D. G. (2014). Abierto, libre . . . y público. Los desafíos políticos de la ciencia abierta [Open, free. . . and public. The political challenges of open science]. *Argumentos de Razón Técnica*, *17*, 45–64. Retrieved from http://institucional.us.es/revistas/argumentos/17/art_4.pdf

Rossner, M., Van Epps, H., & Hill, E. (2007). Show me the data. *Journal of Cell Biology, 179*(6), 1091–1092. doi:10.1083/jcb.200711140

Rossner, M., Van Epps, H., & Hill, E. (2008). Irreproducible results—A response to Thomson Scientific. *Journal of General Physiology, 131*(2), 183–184. doi:10.1085/jgp.200809957

Schekman, R. (2013, December 9). How journals like *Nature*, *Cell* and *Science* are damaging science. *The Guardian*. Retrieved from http://www.theguardian.com/commentisfree/2013/dec/09/how-journals-nature-science-cell-damage-science

Suber, P. (2012). *Open access*. Cambridge, MA: MIT Press.

Treanor, J. (2013, December 4). Banks fined record €1.7bn over benchmark interest rate rigging cartel. *The Guardian*. Retrieved from http://www.theguardian.com/business/2013/dec/04/banks-rate-rigging-libor-euribor-rbs-citigroup-jpmorgan

Van Noorden, R. (2013, May 16). Scientists join journal editors to fight impact-factor abuse [Editorial]. *Nature, 498*, 7453. Retrieved from http://blogs.nature.com/news/2013/05/scientists-join-journal-editors-to-fight-impact-factor-abuse.html

Wesel, M. V. (2015). Evaluation by citation: Trends in publication behavior, evaluation criteria, and the strive for high impact publications. *Science and Engineering Ethics,* 1–27. doi: 10.1007/s11948-015-9638-0

Winner, L. (1989). *The whale and the reactor. A search for limits in an age of high technology*. Chicago, IL: University of Chicago Press.