

Andrzej Kiepas

Silesian University of Technology (Poland)

ORCID: 0000-0001-5279-5288

e-mail: andrzej.kiepas@us.edu.pl

Science in the Age of Big Data

Book Review: Terje Tüür-Fröhlich, *Non-trivial Effects of Trivial Errors in Scientific Communication and Evaluation. Schriften zur Informationswissenschaft Band 69. Herausgegeben vom Hochschulverband für Informationswissenschaft (HI) e. V., Glückstadt 2016.*

For many years, the question of how to evaluate scientific research results, both concerning individual scientists and scientific institutions, has aroused understandable interest on the part of these distinguished entities. Such an evaluation determines as much the course of individual scientific careers as the condition of institutions involved in obtaining and disseminating the results of scientific research. Terje Tüür-Fröhlich's book deals with this important area and highlights many serious, though ostensibly trivial, errors that can be found in the databases that are currently used, endorsed, and favored in the mentioned assessment process. The authoress comes from Estonia and works in information science and related issues of using different kinds of data. The book discussed here constitutes her doctoral dissertation; it is also the result of the authoress's collaboration with the Institute of Philosophy and Scientific Method at Johannes Kepler University, Linz (Austria). Scientometrics and issues related to them constitute the main core of considerations in the discussed book; however, what arises from the more detailed discussion and serves as an important context for the analysis and research conducted by the author are more general questions about the rationality of science. Reflections on the rationality of science indicate, inter alia, that the measures of this rationality are subject to various historical transformations concerning two interrelated areas, namely:

- a) the internal area, whose basis is formed by the values and internal measures (cognitive, methodological) that are pertinent to the activities carried out within the framework of science and technology;

- b) the external area, which refers to the model that takes into account how science and technology function in society (in this regard, current discussions indicate, *inter alia*, the need to link the processes of creating and disseminating the results brought about by science and technology with society, e.g., in the form of the RRI concept, i.e., Responsible Research and Innovations) (Kiepas, 2016, p. 132).

In the modern tradition, science and technology were located within dominant instrumental rationality subject to criticism, which, given the emerging postmodern social and cultural order, was associated with various expectations of change in this respect. The development of new media, information technology, and the Internet, as well as the current prospect of Industry 4.0, i.e., the rise of artificial intelligence, robotization, the Internet of Things, and big data, have opened up new opportunities for instrumentalization and algorithmizing of various fields, including science and technology. Here, science and technology have produced instruments that, as a consequence of their use in scientometric (parameterization) processes, have become tools for development and enslavement. In this respect, Terje Tüür-Fröhlich's work can contribute to the discussion on the role that databases (Big Data) can play when applied as tools for evaluating the performance of scientific institutions and individuals scientists. The very necessity of seeking such evaluation tools seems to be quite widely recognized and unquestioned, at least when they are used to control and consequently eliminate the negative effects of disseminating scientific and technological findings and innovations (as mentioned regarding the external area). Doubts and perceived shortcomings arise when specific quantitative information tools are applied to the processes of evaluation and management of science and the activity of its representatives.

The discussed work analyzes errors, gaps, and ambiguities that we encounter when using various tools and databases, e.g., Web of Science, Journal Impact Factor, Hirsch Index, Social Sciences Citation Index (SSCI) or Times Higher Education World University Ranking, are presented from a broader context of the ethos of science (Merton, 1973) or, if one may put it this way, of a specific scientific culture related to K.R. Popper's concept of falsifiability (Banse & Kiepas, 2002; Perek, 2002).

Terje Tüür-Fröhlich's monograph consists of three main parts:

1. *Trivial errors & commercial science evaluation* constitutes an introduction to the issues under research and an indication of their position in a broader perspective of modern science's functioning, its rationality, and ethos; this part also specifies the objectives and methods of the research, its limitations, and general characteristics of the subject under research with examples of selected databases;
2. *Errors in Thomson Reuters' SSCI* – qualitative and quantitative case studies; case studies of various errors in specific databases are carried out not only in the form of their immanent research but also by taking into account the discussions and critical positions around them;
3. *Historical roots–discussion–conclusions* – this part is devoted to presenting conclusions resulting from the earlier conducted analyses, which is also intended

to formulate recommendations relating to the problems of evaluation and parameterization of scientific activity.

As a result of a systematic overview and critical reflection on the contents of various databases, the authoress formulates her conclusions focusing on three main theses confirmed by the conducted case studies. These conclusions are in general as follows:

“(T1) Trivial errors are of high relevance in the evaluation context. Under today’s evaluation pressure, the not detected, not publicly eliminated, or retracted errors can be important for the “sake” of the careers of the scientists and their institutions, too” (Tüür-Fröhlich, 2016, p. 133).

“(T2) Trivial errors are associated with biases by the power structure and symbolic capital (prominence, reputation, ‘impact’). These Matthew and Matilda Effects – the rich get richer, the poor get poorer – impinge on authors, journals, institutions, scientific disciplines and fields, countries. They are linked with language biases and gender inequality in the sciences. These errors and biases tend to persist, to interact with each other and to exaggerate” (Tüür-Fröhlich, 2016, pp. 133-134).

“(T3) The difficulties to handle different formats of references and footnotes, non-Anglo-American names, and publications in non-English languages were known to the pioneers of commercial citation indexing. The blunt ignorance of lingual, disciplinary and cultural differences has led to errors and the underestimation of errors, in other words: “The tomato evaluation has adverse effects on scientific community (i.e., the first citation index SCI) was rotten from the beginning” (Tüür-Fröhlich, 2016, p. 136).

In relevant chapters of her work, Terje Tüür-Fröhlich provides detailed justifications for the mentioned conclusions. While pointing out various errors, often trivial, and their non-trivial consequences, she does not completely reject or disqualify the possibility of using certain databases to evaluate the effects of scientific activity. One must also agree with her conclusion, in which she states: “Closing one important remark: We should not forget that quantitative citation per se, with errors and without errors. To do only formal citation analysis is an error per se. Semantic citation analyses would be a valuable method to study scientific/scholarly communication and a useful way to assess research” (Tüür-Fröhlich, 2016, p. 139). Also, e.g., M. Kokowski similarly describes the issue when he refers to the social sciences and humanities and emphasizes the necessity to eliminate “...numerous errors of technocratically developed scientometrics” (Kokowski, 2015, p. 178). It is an important but at the same time complicated and not easy task; all the more so as consequently, it encounters various barriers not only of scientific or legal nature but also of a social and cultural one. The progressing algorithmizing of various fields of human activity and various areas of our lives reveals its various consequences in evaluating and parameterizing scientific activity. By revealing these consequences, the work by Terje Tüür-Fröhlich discussed here may raise awareness of what implications the use of such tools may bring so that they do not become “weapons of math

destruction” (O’Neil, 2017), for it is primarily the social sciences and humanities that may fall victim to these weapons.

References:

- Banse, G., & Kiepas, A. (2002). *Rationalität heute. Vorstellungen, Wandlungen, Herausforderungen*. LIT Verlag.
- Kiepas, A. (2017). *Filozofia techniki w dobie nowych mediów (Philosophy of Technology in the Age of New Media)*. Wydawnictwo Uniwersytetu Śląskiego.
- Kokowski, M. (2015). Jakiej naukometrii i bibliometrii potrzebujemy w Polsce? (What kind of scientometrics and bibliometrics do we need in Poland?). *Prace Komisji Historii Nauki PAU*, XIV, 135-184.
- Merton, R.K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.
- O’Neil, C. (2017). *Broń matematycznej zagłady. Jak algorytmy zwiększają nierówności i zagrażają demokracji (Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy)*. PWN.
- Perek, M. (2002). Grenzen der Rationalität in der Wissenschaft. Schweigegebiete intellektueller Traditionen. In G. Banse, & A. Kiepas, (Eds.), *Rationalität heute. Vorstellungen, Wandlungen, Herausforderungen* (pp. 287-298). LIT Verlag.
- Tüür-Fröhlich, T. (2016). *Non-trivial Effects of Trivial Errors in Scientific Communication and Evaluation*. *Schriften zur Informationswissenschaft Band 69*. Herausgegeben vom Hochschulverband für Informationswissenschaft (HI) e. V. Verlag Werner Hülsbusch.