



## Science evaluation and future-proof science

**Peter Vickers: Identifying future-proof science. Oxford: Oxford University Press, 2023, 288 pp, £72 HB**

**Nicola Mößner<sup>1,2</sup>**

Accepted: 9 October 2023 / Published online: 7 November 2023

© The Author(s) 2023

In his book, Peter Vickers defends the idea of future-proof science. By this, he means scientific results that will not lose their scientific status over the course of time, i.e., that will remain a proper part of the scientific body of expertise as long as there are human beings. In his concluding remarks, Vickers states that “what’s driven this entire project is a desire to identify *facts*” (238). Identifying future-proof science thus means identifying scientific facts.

Vickers’s project comprises two interesting questions: (a) what is future-proof science, i.e., what are the criteria that scientific statements have to meet to be called “facts”? And (b), how can non-experts identify future-proof science? The author tries to answer these questions by examining case studies from a variety of academic disciplines. The spectrum of examples ranges from fundamental physics to the recent Covid pandemic.

In this context, Vickers analyses several proposals from the philosophy of science regarding criteria for determining the scientific status of hypotheses. For instance, he takes a closer look at the assumed connection between the capacity of hypotheses to make successful novel predictions and their truth value (Chapters 3 and 4). He also examines the hypothesis that, in order to identify future-proof science, the whole body of evidence has to be taken into account. The latter is a difficult task for non-experts. However, it is primarily laypeople who need tools to distinguish between facts and mere chimaeras, as the flourishing of conspiracy theories during the Covid pandemic has shown. Vickers is aware of this and, as a consequence, suggests an alternative for laypeople, which is to consider claims about the scientific consensus (Chapters 4 and 5). Moreover, Vickers discusses (a) whether such a consensus can be reached, (b) what distinguishes a mere majority agreement from a solid scientific consensus, (c) how laypeople can know about a solid scientific consensus, i.e.,

---

✉ Nicola Mößner  
nicola.moessner@rwth-aachen.de

<sup>1</sup> Department of Philosophy, Leibniz University Hannover, Hannover, Germany

<sup>2</sup> Research and Teaching Unit Philosophy of the Cultural World, RWTH Aachen University, Aachen, Germany

which criteria have to be met by the latter, (d) how the respective conditions can be identified by the public, and (e) what role the course of time might play regarding the establishment of a solid consensus (Chapters 6 to 8).

The upshot of this discussion is the proposal of the following two conditions that future-proof science must meet: “(1) At least 95 per cent of relevant scientists are willing to state the claim unambiguously and without caveats or hedging. If prompted, they would be willing to call it an ‘established fact’. (2) The relevant scientific community must incorporate a substantial diversity of perspectives” (111). Hence, the author suggests that laypeople have to look for a “solid scientific consensus” in order to decide whether a particular hypothesis can be regarded as a scientific fact. In this, he follows Naomi Oreskes’s claims about what laypeople can regard as a reason to still trust scientific experts, despite the fact that scientific reasoning can be as fallible as everyday cognitive activities. The merit of this indirect route to assess the status of scientific hypotheses is that the focus is on laypeople for whom it is particularly relevant to find out whether a given claim belongs to future-proof science.

However, it is also the source of some difficulties. The second of the two conditions mentioned by Vickers entails a certain degree of vagueness by calling for “a substantial diversity of perspectives”. When exactly is this condition fulfilled? This problem is addressed by the author in a footnote (see 111, fn. 46). However, in his subsequent discussion, he constantly repeats the relevance of an international and gendered scientific community, which might lead readers to think that these are the only aspects to take into consideration regarding diversity. Moreover, it raises the question of whether agreement on facts in science could even be possible prior to the time when women were permitted to enter universities, around the end of the nineteenth and the beginning of the twentieth century—which seems odd.

Beyond that, there is also a more substantial problem entailed in the background assumption, namely that the diversity condition is meant to make sure that a relevant *plurality of perspectives* is involved. More often than not, such a plurality is the result of bringing together people with different mindsets developed through individual experiences or different modes of education. Differences in ethnic background and gender can be *helpful markers* to indicate whether a plurality of perspectives can be expected within a certain group of people, but such characteristics are *no guarantee*. One point that has to be addressed—not only by Vickers, but by all philosophers who insist on such a diversity condition—is the connection between diversity and plurality. They have to spell out the kind of diversity that is needed in different contexts of science to make sure that the required plurality of perspectives is obtained. It can be assumed that this task is much more complex than a mere scoring of gender and ethnic backgrounds within particular groups. It is also a question of degree: how much plurality is required? Is there a threshold in the sense that, as a result of adding more perspectives to a given set, the entire strategy becomes counterproductive because a consensus can no longer be reached? These are inconvenient questions, but addressing them will do more for plurality in science than merely gesturing towards adding more women, etc. to the community.

Regarding Vickers’s first condition—the solid consensus—one question is how laypeople can identify such a consensus. Again, the author focuses on practicability,

i.e., he discusses concrete examples as to how laypeople can come to know about opinion-building processes and their results in the scientific community. However, by presenting these examples, and apparently, unbeknownst to him, Vickers also points out some crucial current problems in science.

He claims that “one good rule of thumb when trying to ascertain whether opinion has reached 95 per cent is this: in most cases where it has *not*, evidence of substantial debate in the community will be relatively easy to find, and in most cases where it *has*, any serious opposition (within the relevant scientific community) will be extremely difficult to find” (222). Vickers then explains how a layperson can find out whether such debates are happening in a particular community: by (1) taking a look at relevant conferences and/or (2) scientific journals. However, the author does not clarify how a layperson could identify what the “relevant” conferences within a certain academic field might be. Presumably, this will be difficult for an outsider, because in many academic fields there is more than one expert association holding conferences, and not all relevant conferences take place on an annual basis, etc. With regard to identifying “relevant journals”, Vickers does suggest how to pick them out. For instance, using Wikipedia references would be one possible route to these resources (see 222f.).

However, this reference to the journal system is also problematic: Vickers obviously thinks that scientific debates that laypeople should take seriously, and where expert and consensual opinions can be found, are published in what is commonly regarded as “reputable” journals. The criterion for the latter is a *properly working* peer-review system (see 95). So the author trusts in the proper functioning of the current academic journal system. This also becomes apparent when Vickers argues—again in accordance with Oreskes’s approach—that what makes scientific practices reliable is the inherently critical attitude of researchers. Scientific claims are constantly tested. Consequently, the more scientists are involved in these vetting processes, the more reliable the respective hypotheses and data will be. And—so the argument goes—the more researchers are involved, the more publications are produced within the scientific community. Vickers optimistically announces scientific progress as a fact by pointing out the exponential growth of scientific publications (see 39f.). He adds that, as a consequence of this growth, “any contemporary theoretical idea will be subject to far more scrutiny—in a relatively short period of time—than were theoretical ideas of the past” (40).

His considerations are based on quantitative assumptions inferred from our current academic journal system: the number of articles published on a regular basis is used to determine scientific growth and thus likely progress; the number of quotations of an individual scientist is used to infer her status of expertise; and the fact that a particular thesis has been published in a highly ranked journal—i.e., a journal that is often cited (for details on “journal impact factor”, “h-index”, etc., and related problems see Andersen 2020)—is used to discern its impact status within the scientific community. To make this relation between numbers and quality plausible, Vickers emphasizes that our current publication system is based on “rigour and general professionalism” (38)—but is this actually the case?

It does not come as a surprise that the author highlights these quantitative markers as indicators of quality. However, this traditional approach has already, and for

some time, been the target of legitimate critique (see, for example, Holzer 2022; Retzlaff 2022). For example, Hanne Andersen summarizes some of the main worries concerning these metrical means, starting with their basic assumption that citation numbers are a reliable basis for inferring scientific success or impact (see Andersen 2020, 149). Eric Retzlaff asks provocatively whether, due to their metrical differences, the physicist and recent noble prize winner Peter Higgs (h-index of 9) is less relevant to the scientific community than Stephen Hawking (h-index of 76, see Retzlaff 2022, 149). The h-index or “Hirsch index” is a bibliometric means to indicate the individual scientist’s impact within the scientific community. It squares her number of publications with the number of their citations in a determined period of time and is meant to provide an average number of an individual’s citations, not the peaks.

As long as such bibliometric means are used as indicators of scientific quality and impact, most researchers will try to increase their own numbers. Andersen explains that one prominent strategy for this is to publish findings in a series of small units in order to increase the number of publications (“salami publication”, Andersen 2020, 150). A corollary of this is that Vickers’s assumption that a mere numerical growth of scientific publications implies scientific progress does not hold. Andersen explains the mistaken background inference: “But ideally, the output of research is new knowledge, or new ideas. Publications are merely a dissemination channel for his knowledge, and how much new knowledge individual publications present varies considerably” (Andersen 2020, 150).

It has become a more general point of concern in scientific practice that quantitative benchmarks are problematic when it comes to science assessment. For instance, the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG)—one of the biggest research funding organizations in Germany—notes in its “Guidelines for Safeguarding Good Research Practice” that “to assess the performance of researchers, a multidimensional approach is called for; in addition to academic and scientific achievements, other aspects may be taken into consideration. Performance is assessed primarily on the basis of qualitative measures, while quantitative indicators may be incorporated into the overall assessment only with appropriate differentiation and reflection” (DFG 2022a, 11). Therefore, Vickers’s and others’ hypothesis that a numerical increase, e.g., of science publications, citations, etc., indicates a qualitative development in science has to be handled with great care.

Another point of concern, based on similar considerations, is related to what constitutes the “reputation” of a publication. Vickers suggests that laypeople should use scientific journals as a source of information concerning the question of whether a consensus regarding a particular hypothesis has been reached (see 95). Actually, the author argues that journal articles are much more trustworthy in this respect than, for example, books, because “it is relatively easy to publish a book making any claim whatsoever if one is willing to pay and/or one doesn’t care who publishes it” (110, fn. 47). Obviously, the author thinks that peer-review processes of academic journals will prevent or reduce the publication of nonsense. Journals are therefore preferable as a source of information. However, it can be questioned whether these assumptions are correct (see DFG 2022b, Chapter 2.4 for a discussion of problems related to peer-review processes). Vickers assumes that the scientific quality of ideas

can be determined by merely taking a look at their places of publication, but is this actually the case?

Again, the German Research Foundation calls for a more cautious stance, stating that “the scientific/academic quality of a contribution does not depend on the medium in which it is published” (DFG 2022a, 19). It would be too hasty a decision to rely on “big names”, journals as well as publishers, alone when evaluating scientific findings. Hypotheses which do not make their way into scientific journals should not automatically be regarded as inferior to those that do. Björn Brembs et al., for example, discuss serious problems related to the current academic journal system from the perspective of the scientific community and make suggestions on alternative ways to distribute data and hypotheses (see Brembs et al. 2021).

Although it is Vickers’s intention to offer a handy strategy that laypeople can use, it has to be kept in mind that, by giving such advice, already problematic ways of science evaluation are consolidated even further. Therefore, the proposal put forward here is to make this insight a part of Vickers’s own suggestion of improvement, namely to add such considerations to the amendment of science education (see 234ff.). In this context, Vickers claims “that a significant intervention in science education programmes around the world is called for, on the grounds that our children really do need to leave school with richer conceptions of expertise, consensus, and scientific community dynamics” (237). These modifications are necessary due to “the shift from ‘internal’ evidence to ‘external’ evidence” (22). As pointed out convincingly in Vickers’s approach, laypeople can use external evidence to find out what constitutes future-proof science and whether a certain hypothesis belongs to this category. Hence, the author’s proposal to enhance students’ capacities in this regard is indeed more than recommendable. However, such a training should not stop at school level, but continue at university level. It is here that students should be taught about science evaluation processes. They should be made aware of problems regarding current science metrics and publication processes. Eventually, as future researchers and scholars, it is up to them to develop science. So the question is not only what the characteristics of future-proof science are, but also what we want good science to be like?

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Andersen, Hanne. 2020. Can scientific knowledge be measured by numbers? In *What is scientific knowledge? An introduction to contemporary epistemology of science*, ed. K. McCain and K. Kampourakis, 144–159. New York and London: Routledge.
- Brembs, Björn; Philippe Huneman, Felix Schönbrodt, Gustav Nilsson, Toma Susi, Renke Siems, Pandelis Perakakis, Varvara Trachana, Lai Ma, & Sara Rodriguez-Cuadrado. 2021. Replacing academic journals. <https://doi.org/10.5281/zenodo.5793611>.
- Deutsche Forschungsgemeinschaft (DFG). 2022a. *Guidelines for safeguarding good research practice. Code of conduct*. <https://doi.org/10.5281/zenodo.6472827>.
- Deutsche Forschungsgemeinschaft (DFG). 2022b. *Academic publishing as a foundation and area of leverage for research assessment. Challenges and fields of action (Wissenschaftliches Publizieren als Grundlage und Gestaltungsfeld der Wissenschaftsbewertung)*. <https://doi.org/10.5281/zenodo.6538163>.
- Holzer, Angela. 2022. “Die Vermessbarkeit der Wissenschaft. Digitalisierung, wissenschaftliches Publizieren, Verhaltenstracking und Wissenschaftsbewertung”, In Mößner, N., and K. Erlach (eds.): *Kalibrierung der Wissenschaft. Auswirkungen der Digitalisierung auf die wissenschaftliche Erkenntnis*, Bielefeld: transcript, pp. 163–181, <https://doi.org/10.1515/9783839462102>.
- Retzlaff, Erik. 2022. “Wer bewertet mit welchen Interessen wissenschaftliche Publikationen? Eine Skizzierung des Einflusses kommerzieller Interessen auf die Forschungsoutput-Bewertung”, In Mößner, N. and K. Erlach (eds.): *Kalibrierung der Wissenschaft. Auswirkungen der Digitalisierung auf die wissenschaftliche Erkenntnis*, Bielefeld: transcript, pp. 139–162, <https://doi.org/10.1515/9783839462102>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.