1

Gaming Metrics Before the Game: Citation and the Bureaucratic Virtuoso

Alex Csiszar

Henceforth, academic interviewees, it won't be: How many papers have you published? but: Where? … Publish at the *top* or be damned.
—Correspondence in *New Scientist*, November 23, 1972.

On June 13, 1974, in the foothills of Stanford University, a group of social scientists, information entrepreneurs, and government officials gathered to deliberate what had recently become pressing questions: Can science be measured? How? And what might be the consequences of doing so? The conference on "science indicators" at the Center for Advanced Studies in the Behavioral Sciences (CASBS) was among the first ever held on the subject. But the invitations sent to participants had already raised the question how measuring scientific productivity might alter the very object under study: "the creation and official use of indicators may lead to shifts in the range and kind of scientific activity itself."[1]

Proposals to establish objective measures of individual and national scientific activity have been around for nearly two centuries.[2] But during the late 1960s in the United States, renewed pressure on scientific funding bodies, sociological interest in studying what had become known as "the scientific community," and the increasingly wide availability of citation data put a spotlight on these questions.

It is commonly supposed that rising incidence of gaming is an unintended consequence of these developments, one that an era more innocent than our own simply did not have much reason—or enough foresight—to confront. We have even adapted a term from economics—Goodhart's law—to capture the idea that when measures become standards, they cease to be good measures. By unearthing a set of episodes from the early history of citation analysis, I want to argue two things. First, this supposition is historically mistaken: the idea that attempts to measure scientific output might lead to changes in researcher behavior has been a part of the conversation about citation metrics from the beginning. Second, the very framing of the problem in terms of unintended consequences not only has a long history, but also has caused observers to ask a limited set of questions about gaming and misconduct and thus to produce an equally limited set of responses to it.[3]

Many circumstances had come together to bring about the conference on science indicators. The catalyst was the publication by the National Science Foundation (NSF) of the first edition of *Science Indicators* (hereafter SI-72), a collection of statistics on the relative performance of science in the United States during 1972. In the late 1960s, at the instigation of Congress, the NSF had undergone its first major restructuring since it was established in 1950. The resulting 1968 amendment to the NSF act brought the foundation more closely under the control of Congress and the executive branch (Committee on Labor and Welfare, 1968). In particular, there would now be annual hearings to approve expenditures, and approval would depend on an annual report from the foundation on the "status and health of science and its various disciplines." The legislation did not specify the form of these reports but there were strong hints that numbers were what was wanted. Henceforth, the NSF owed the people hard data about the outputs of science.

The new mandate led NSF officers to the notion that they might construct a comprehensive and standardized set of indicators of the health of science. Congress received the first such report in 1973. The NSF then reached out to the Social Science Research Council—already studying the construction of *social* indicators—about how they had done in this first effort. The latter thought this an ideal job for sociologists of science. Robert K. Merton, the towering figure in the field, got the call.

The timing could not have been better. That academic year, Merton and his collaborator and partner Harriet Zuckerman were on leave at CASBS in Palo Alto, where they had assembled an interdisciplinary working group to pursue projects in the historical sociology of science. The core group included Nobel Prize–winning biologist Joshua Lederberg and the historians Arnold Thackray and Yehuda Elkana. The project on indicators could prove the relevance of the new program in sociology of science, Merton thought. It would be "a test for the guild."[4]

"Fun and Games with Citations"

In a small room uncomfortably packed with scientists and sociologists (Merton explained that "enforced intimacy is more conducive to straight-out serious talk"), one of the most consequential participants had no academic affiliation at all. Eugene Garfield was the proprietor of the Institute for Scientific Information (ISI) and the inventor of the Science Citation Index (SCI). Much of the debate about measuring science that had arisen in the past decade was focused on analyzing citation data obtained (at a price) from his company. Garfield was asked to precirculate a paper on the ways in which citation data could be used to track the health of science, its cognitive development, and its goals. Nearly all the other papers were dedicated to more or less savage critiques of SI-72 from a variety of perspectives, including statistics (William Kruskal), economics (Zvi Griliches), history and philosophy of science (Gerald Holton), and political philosophy (Yaron Ezrahi). The only other constructive paper—by the brothers Jonathan and Stephen Cole, students of Merton—also relied on Garfield's work; it explored how citation analysis could be used to identify scientific consensus.[5] Garfield's invention was under the microscope.

In his role as convener, Merton often steered the conversation. He was especially keen to direct attention to one key problem: "I'd like to get on the record the problem of goal displacement." He explained that "when certain indicators get officially, or quasi-officially, established as measures of this, that, or the other, there will be, one should look for, efforts to manipulate the numbers, by one's behavior":

> In the case of citations, I make a small self-fulfilling prophecy, that has already been fulfilled in one area I know, namely, as more and more citations are used both officially and unofficially as measures of contribution of work of relative standing and the like, there will be diverse patterns of adaptations to those uses, in a feedback way, and citation behavior will change in part in certain sub-sets of the community.

Merton noted, cryptically, that he knew already of a "highly organized effort … to operate on citations for political purposes" and he worried that the "purity of the community of science is itself, maybe, put in jeopardy" by this new behavior.[6]

So important did Merton believe this problem to be that he continued to impress it on Garfield in the months following the conference. "I enjoyed your fun and games with citations," he wrote, perhaps with a hint of mockery. And he warned Garfield to exercise care:

> Watch out for goal displacement: Whenever an indicator comes to be used in the reward system of an organization or institutional domain, there develop tendencies to manipulate the indicator so that it no longer indicates what it once did. Now that more and more scientists and scholars are becoming *acutely* conscious of citations as a form of behavior, some will begin, in semi-organized fashion, to work out citation-arrangements in their own little specialties.

This was as true for science, Merton said, as it was for any bureaucratic system involving counting and rewards: censuses, employee performance measures, or government agencies.[7]

Merton knew all this well, for he had introduced the concept of goal displacement to sociology himself several decades earlier. In 1940, he had written an essay, "Bureaucratic Structure and Personality," which focused on the dysfunctional consequences of bureaucratic organizations. "Goal displacement" was the process by which "an instrumental value becomes a terminal value." This was an important case of the unintended consequences of purposive action; it occurred whenever rules or performance measures were set up, which gradually came to be adhered to by actors as ends in themselves. Over the next decades, many of Merton's students helped found the US field of organizational sociology, studying government agencies, industrial firms, and other organizations to determine how rules and performance metrics can lead to deviant behavior and unintended outcomes—goal displacement, in short.

At the end of the 1950s, after major contributions to several areas of sociology, including a major intervention in structural-functionalism sociological theory, Merton shifted his research focus back to the sociology of science (the topic that he had written for his PhD). In this field, two of his papers had already become classics (Merton, 1938, 1942). These papers outlined a set of behavioral norms that governed the social structure of science. Insofar as these norms—universalism, communism, organized skepticism, and disinterestedness—were adhered to by practitioners, he had argued, science progressed toward greater and more certain knowledge. But this had always been a rather theoretical idea. As he returned to the issue (Merton, 1957), he realized that to observe these norms in action, he needed to get more concrete: just how were they instantiated in actual practice? Publishing practices seemed to be the nerve center where issues of reward, responsibility, and status merged on an everyday basis.

But the value placed on publishing papers, which made "open disclosure" the rule in science, also contained within it the possibility of pathology. As this value became institutionalized, Merton (1969) explained, it increasingly encouraged "through the displacement of goals, a spurious emphasis on publication for its own sake, almost irrespective of the merit of what is published."[8] Scientists published more and more papers, but each of them said less and less.

The same went for other measures of achievement. Merton began to enroll graduate students and to run seminars on the sociology of science in the 1960s, and he was on the lookout for sources of data for research projects. The group explored bibliometric data, biographical data, and survey methods. Through Garfield's efforts to market his products to sociologists and historians, Merton learned of the SCI in 1962. By 1965, Merton's students were mining the citation index for insights on the development of schools and fields, the growth of scientific consensus, and reward systems. By 1970, he was tentatively employing citation analysis in his own work, but already warning of its dangers. In a paper on age stratification in science (Zuckerman and Merton, 1972), a long footnote warned that the growing use of citation analyses "as aids in deciding upon the appointment and promotion of scientists" could lead to changes in citation practices that would invalidate them as measures altogether.

By the time of the CASBS conference in 1974, Merton was far from alone in expressing concern over the role that paper counting and citation counting might do to science. The conference had come at a critical moment for Garfield's enterprise. During the 1960s, he had worked diligently to market the SCI to scientists as an aid to information retrieval (Wouters, 1999). Looking to explore other markets for his expensive tools, he had also reached out tentatively to sociologists and others concerned with studying science, asking politely whether they might find some use in it. But he tended, in public, to play down suggestions that the citation index might be used as a means of evaluating scientists or research programs.

A decade later, however, Garfield had become bolder, and he mounted a public campaign to argue for new uses of citation analysis. In 1970, he wrote in *Nature* that "increasingly scarce intellectual and financial resources for supporting research could be managed more efficiently" using citation counting to identify the most promising and creative researchers. In 1972, he published in *Science* his first public ranking of journals according to the frequency with which they came up in citations—both absolutely and relative to the number of articles published. That year he also began marketing *Journal Citation Reports*, which became the permanent home of the "Impact Factor."

Garfield's early steps toward legitimating the evaluative use of citation analysis—along with other investigations by Derek J. de Solla Price (1969), Joel Margolis (1967), and the Cole brothers (1972)—made it a topic of controversy in scientific circles. Many observers were appalled, arguing that such surveillance and the encouragement of hierarchy were inimical to the spirit of creativity and equality central to the scientific community. Garfield's ranked list of journals was "a pedantic dunghill," suggested one letter in *New Scientist*, a "highly invidious pecking order" calculated to appeal to "connoisseurs of hierarchies."[9] Others, such as Samuel Goudsmit (1974), legendary editor of the *Physical Review*, argued that the implicit theory of citation of these studies was mistaken, for acknowledgement was only a minor use of footnotes. He dismissed citation analysis and quoted approvingly a warning that equating "high frequency of citation with worth or excellence will end in disaster."

Many critiques were explicit about the potential for gaming. Journal rankings were the equivalent of "pop charts for science," suggested Alex Comfort (1970) in *Nature*. The corrupt music industry gave scientists a taste of "the abuses we might expect (and even … indulge in): Commercial concerns dedicated to lobbying, so that our rating might be raised? Promotional quotation data on the relative one-upmanship of particular journals? In California nothing is impossible." And the mathematician and historian Kenneth O. May (1967) noted that these new uses of citations would lead authors to cite "their friends' papers more (a friend is someone who cites in return)" and also to increase their citations to marginally relevant papers "so as to attract people who might (and perhaps should) miss the paper."

Thus, while Garfield went into the conference with optimism, he also knew that these new uses for citation analysis were coming under attack. But that Merton himself should challenge their legitimacy must have been jarring. Merton had become a great supporter of Garfield and the ISI; he sat on its board, proselytized the uses of the citation index to others, and gave Garfield strategic advice when problems arose. Garfield nevertheless took Merton's warning in stride. "If it should turn out that over the next 10 to 20 years goal displacement does occur with the Science Citation Index," Garfield admitted, "then we will simply have to abandon its uses as an indicator of quality, etc." But he did not really think that this was likely. Maybe editors could be taught "to look out for self-serving attempts to distort citation data?"[10] Merton felt able to reassure Garfield that while localized citation misconduct was sure to happen, it was doubtful that things would get so bad that science itself would really be transformed "for a long, long time, if at all. It would follow upon a massive use of citation counts as a part of the reward system of science: appointments, promotions, research grants, elections to office in professional societies, awards, etc. And that is not in the cards."[11]

Merton was wrong. In the next decades, citation analysis and the Journal Impact Factor (JIF), which became enshrined in *Journal Citation Reports*, spread far and wide. The field of scientometrics blossomed. By 1980, Garfield was considering ways to capitalize on the growing use of citation data in evaluations of individuals and institutions. After several years of publishing Impact Factor data, in 1981, the ISI formulated a plan for a new suite of products marketed for the "deliberate use of citation counts for evaluative purposes." The proposed Science Citation Analysis System (SCAS) would be an online tool for "administrators of research faculty/staff" to measure the productivity of individuals and institutions without relying on such crude markers as the JIF.[12] Again, Garfield and Merton engaged in intense conversation about the potential consequences of such a move for science. Were Garfield's tools simply making more efficient a set of evaluative practices that were already flourishing? Or would the availability of such tools encourage these practices and thus lead to their misuse by administrators and deviant behaviour by those subject to the measures? Merton assured Garfield that the latter would be the case, and that there were no moral grounds on which he could disavow his role in any such an outcome: "You're helping to build the practice which will include a lot of pollution, this misuse which you can't control, but you've been instrumental in making possible."[13]

Even as Merton continued to warn Garfield about goal displacement behind the scenes, he remained a strong supporter of the ISI in public. The term that has come most often to be used to describe the susceptibility to manipulation of metrics of science was not "goal displacement" but rather Goodhart's law.[14] Although the latter originated in monetary economics, the common formulation in this context comes from Marilyn Strathern (1996): "When a measure becomes a target, it ceases to be a good measure." While "goal displacement" describes a similar phenomenon, the emphasis is inverted: where Goodhart's law highlights the consequences for the validity of the metric, Merton's concept highlighted the consequences for the object being measured. This emphasis was especially clear in the studies of bureaucratic organizations pursued by Merton's students. In many cases, measures and rewards that were instituted to encourage behavior that would further the goals of an organization led not simply to dysfunctional behavior, but also to reshaping the perceived goals of the organizations themselves. In the big picture, the phenomenon of gaming matters not simply—or even principally—because it invalidates any given metric, but because it can profoundly shape the everyday conditions under which research is carried out, not simply in terms of how and when people publish their findings, but of what kinds of projects they choose to pursue and how they go about them. Ironically, Strathern's now-famous formulation of Goodhart's law was derived from a paper by Keith Hoskin (1996), which was itself a critique of this formulation along these lines. Hoskin pointed out that the law seems to imply that the target is the problem, and that measures are not in themselves potentially problematic, thus leading to a search for "either better targets, or better ways of anticipating and handling a presumed natural-born predilection for beating the target system." (This is a reasonable characterization, incidentally, of a key argument made by advocates of altmetrics.[15])

More generally, however, the ideas behind goal displacement and Goodhart's law have both constrained the kinds of questions we ask about metrics and their problems. By adapting the theory of bureaucracy to the study of science, Merton encouraged the idea that the social study of science could be pursued along broadly similar lines to the social study of corporations and firms, and in relative isolation from the rest of society. But this obscured a fundamental disanalogy. In a democratic state, there is little chance of there being widespread consensus about what the goals of science ought to be in the first place. Broad public views about the legitimate ends of scientific activity cannot in general be separated from public representations of scientific activity and its productivity. For that reason, public indicators of science affect not only the practice of research, but also the public legitimacy of science.[16] To put it another way, understanding metrics of science is not just about the internal sociology of research, but also about the politics of scientific knowledge (Ezrahi, 1978; 1990).

My sense is that many of the most common critiques of the implementation and use of scientometrics essentially ignore this consideration. They tend to suppose that there is some pure and efficient form of scientific practice—and of scientific publishing—that was once directed wholly to the production of knowledge, but which has recently been turned upside down (for good or ill) by the implementation of accounting practices.[17] The late-Mertonian mode in the sociology of science was at its peak in the 1960s and 1970s, but it remained alive among those experts producing increasingly sophisticated tools with which to measure scientific productivity, even as they grew increasingly frustrated that these tools were subject to misuse and abuse.

When we debate the efficacy of a given measure of scientific achievement, we are also debating (perhaps implicitly) the grounds of the public legitimacy of science itself.[18] This perspective implies that the analysis of metrics in use ought to attend to the political values and modalities of power within which they are constructed, deployed, and spread (Merry, 2016). Metrics have become mediating devices in the interaction between the groups and institutions involved in science policy, its administration, and research. Scientific publishing—whose ascendancy has often been linked to public representations of the legitimacy of modern science—is now front and center in public debate about inefficiency and corruption (Sarewitz, 2016; Oransky, this volume, chapter 10).

Previous histories of measuring science (Godin, 2005; Bellis, 2009; Gingras, 2016) have not generally paid much attention to the question of misconduct and gaming, but we need a new history that puts these matters at the heart of the story. This is not simply because I think we need to get the history right, but because producing a better historical narrative will help to break the spell of Goodhart's law and its several variations. It is not yet clear what the history of scientific metrics *in use* might look like. But I suspect that we will need to formulate Goodhart's counter-law to describe the more common historical sequence: "*only when a measure becomes a target is it widely taken up as a measure worth using*."

Notes

1. See the rough transcript of R. K. Merton's remarks, and the generic conference invitation template dated March 21, 1974, in Robert K. Merton Papers, Columbia University Rare Book & Manuscript Library [hereafter RKMP], 328.4.

2. For an early example of measuring scientific output, see Granville (1830).

3. I am inspired here in part by James Griesemer's essay in this volume.

4. Quoted in Robert Parke letter to R. K. Merton, December 21, 1973, RKMP 328.7.

5. Conference agenda, RKMP 328.4.

6. Rough transcript of R. K. Merton's remarks, RKMP 328.4.

7. R. K. Merton letter to E. Garfield, August 3, 1974, RKMP 168.5.

8. See also, earlier, Merton (1957) and, later, Zuckerman and Merton (1971), in which the authors also make the fascinating hypothesis that the spread of referee systems was a further unintended consequence of the impulse to publish as an end in itself.

9. "At the Heart of the Paper Blizzard," *New Scientist*, November 23, 1972, p. 464. On resistance to "the moral equivalence of the scientist," see Shapin (2008).

10. E. Garfield letter to R. K. Merton, August 19, 1974, and R. K. Merton's reply, September 2, 1974, RKMP 168.5.

11. R. K. Merton to E. Garfield, September 2, 1974, RKMP 168.5.

12. SCAS is described in L. Simon, "Preliminary Specification for Science Citation Analysis System (SCAS)," January 29, 1981, in RKMP 168.4.

13. Transcript of a recorded conversation between E. Garfield and R. K. Merton on May 23, 1981, Papers of Eugene Garfield, Science History Institute, Philadelphia, 44.5.

14. This is not to say that "goal displacement" has wholly disappeared from the conversation. See, for example, Hicks et al. (2015).

15. See Jennifer Lin in this volume, chapter 16.

16. These points were in fact made at the 1974 conference by the political theorist Yaron Ezrahi (1978). But they seem largely to have been ignored.

17. See, for example, Yves Gingras in this volume, chapter 2. For the historical counterargument, see Csiszar (2018).

18. A similar point is made at the end of Paul Wouters's contribution to this volume, chapter 4.

References

Bellis, Nicola De. 2009. *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*. Lanham, MD: The Scarecrow Press.

Cole, Jonathan R. 2000. "A Short History of the Use of Citations as a Measure of the Impact of Scientific and Scholarly Work." In: Cronin, B., Atkins, H. B. (Eds.), *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*. Information Today, Medford, NJ, pp. 281–300.

Cole, Jonathan R., and Stephen Cole. 1972. "The Ortega Hypothesis." *Science* 178:368–375.

Comfort, Alex. 1970. "Pop Charts for Science." *Nature* 227:1069.

Committee on Labor and Welfare. 1968. National Science Foundation Act Amendments of 1968: Hearings Before the United States Senate Committee on Labor and Public Welfare, Special Subcommittee on Science, 90th Congress, First Session, on Nov. 15 and 16, 1967. Washington, DC: US GPO.

Csiszar, Alex. 2018. *The Scientific Journal: Authorship and the Politics of Knowledge in the Nineteenth Century*. Chicago: University of Chicago Press.

Elkana, Yehuda, et al., ed. 1978. *Toward a Metric of Science: The Advent of Science Indicators*. New York: Wiley.

Ezrahi, Yaron. 1978. "Political Contexts of Science Indicators." In: Elkana, Y., Lederberg, J., Merton, R. K., Thackray, A., Zuckerman, H. (Eds.), *Toward a Metric of Science: The Advent of Science Indicators*. Wiley, New York, pp. 284–327.

Ezrahi, Yaron. 1990. *The Descent of Icarus: Science and the Transformation of Contemporary Democracy*. Cambridge, MA: Harvard University Press.

Garfield, Eugene. 1970. “Citation Indexing for Studying Science.” *Nature* 227: 669–671.

Garfield, Eugene. 1972. “Citation Analysis as a Tool in Journal Evaluation.” *Science* 178:471–479.

Garfield, Eugene. 1973. “Citation and Distinction.” *Nature* 242:485.

Gingras, Yves. 2016. *Bibliometrics and Research Evaluation*. Cambridge, MA: MIT Press.

Godin, Benoît. 2005. *Measurement and Statistics on Science and Technology: 1920 to the Present*. London: Routledge.

Goudsmit, Samuel A. 1974. “Citation Analysis.” *Science* 183(4120):28–33.

Granville, Augustus Bozzi. 1830. *Science Without a Head*. London: T. Ridgway.

Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah de Rijcke, and Ismael Rafols. 2015. “The Leiden Manifesto for Research Metrics.” *Nature* 520:429–431.

Hoskin, Keith. 1996. “The ‘Awful Idea of Accountability’: Inscribing People into the Measurement of Objects.” In: Munro, R., Mouritsen, J. (Eds.), *Accountability, Power, Ethos and the Technologies of Managing*. International Thomson Business Press, London, pp. 265–282.

Margolis, Joel. 1967. “Citation Indexing and Evaluation of Scientific Papers.” *Science* 155:1213–1219.

May, Kenneth O. 1967. “Abuses of Citation Indexing.” *Science* 156 (3777): 890–892.

Merry, Sally Engle. 2016. *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking*. Chicago: University of Chicago Press.

Merton, Robert K. 1938. “Science and the Social Order.” *Philosophy of Science* 5:321–337.

Merton, Robert K. 1940. “Bureaucratic Structure and Personality.” *Social Forces* 18(4):560–568.

Merton, Robert K. 1942. “Science and Technology in a Democratic Order.” *Journal of Legal and Political Sociology* 1:115–26.

Merton, Robert K. 1957. “Priorities in Scientific Discovery.” *American Sociological Review* 22:635–659.

Merton, Robert K. 1969. “Behavior Patterns of Scientists.” *American Scientist* 57(1):1–23.

National Science Board, National Science Foundation. 1973. *Science Indicators 1972*. Washington, DC: US GPO.

Price, Derek J. de Solla. 1969. “Measuring the Size of Science.” *Proceedings of the Israel Academy of Sciences and Humanities* 4(6):98–111.

Sarewitz, Daniel. 2016. “Saving Science.” *New Atlantis* 49:4–40.

Shapin, Steven. 2008. *The Scientific Life: A Moral History of a Late Modern Vocation*. Chicago: University of Chicago Press.

Strathern, Marilyn. 1996. “From Improvement to Enhancement: An Anthropological Comment on the Audit Culture.” *Cambridge Anthropology* 19(3):1–21.

Wade, Nicholas. 1975. “Citation Analysis: A New Tool for Science Administrators.” *Science* 188:429–432.

Wouters, Paul. 1999. “The Creation of the Science Citation Index.” In: Bowden, M. E., Hahn, T. B. (Eds.), *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*. Information Today, Medford, NJ, pp. 127–136.

Zuckerman, Harriet, and Robert. K. Merton. 1971. “Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System.” *Minerva* 9(1):66–100.

Zuckerman, Harriet, and Robert K. Merton. 1972. “Age, Aging, and Age Structure in Science.” In: *Aging and Society*. Russell Sage Foundation, New York, pp. 292–356.