

A hundred million acts of whimsy?

Blaise Cronin

School of Information Science, Indiana University, Bloomington, IN 47405, USA

Keywords: Citation analysis, citation index, publishing ecosystem.

I know it would be acceptable to most readers, if I were less punctual and scrupulous in my quotations; it being by many accounted a more genteel and masterly way of writing, to cite others but seldom, and then to name only the authors, or mention what they say in the words of him that cites, not theirs, that are cited.

Robert Boyle (1665)
quoted in Shapin¹ (p. 118)

In the words of Paul Wouters², the advent of computerized citation indexes ‘created a completely novel symbolic universe, based on the semiosis of the citation. Garfield turned lead (the reference) into gold (the citation) in his search for the philosopher’s stone’. As I write, it is almost 50 years to the day since the semiotic gold rush began, a fitting moment for reflection. Robert Merton (eminent sociologist) and Joshua Lederberg (Nobel prize-winning geneticist) were remarkably quick to grasp the wider sociological and science policy implications of Eugene Garfield’s pioneering, though still embryonic work in citation indexing³. Alongside Lederberg and Merton, Garfield’s lifelong friends and doughty supporters, bibliometric grandees such as Derek de Solla Price and Belver Griffith helped fashion the field now most commonly known as scientometrics – a field whose conceptual foundations were lucidly sketched in Eugene Garfield’s seminal *Science* paper⁴. Here, at Arunachalam’s suggestion, I summarize briefly some of the theoretical and behavioral strands of the still fervid debate on the pros and cons of citation analysis that I laid out originally in *The Citation Process*⁵ and have since elaborated on in *The Scholar’s Courtesy*⁶ and *The Hand of Science*⁷.

Although the scholarly publishing ecosystem has experienced significant changes in the last couple of decades, the principal themes of the debate on the validity, legitimacy and utility of citation analysis have not altered fundamentally. With rapid developments in online communication, web-based publishing, digital libraries, institutional repositories and, more generally, open access publishing there has been a corresponding growth of interest in both established bibliometric and emerging cybermetric indicators. It is clear that we will soon have access to a critical mass

of web-based digital objects and usage statistics with which to develop multi-dimensional models of scholars’ communication behaviors – publishing, posting, blogging, scanning, reading, downloading, glossing, linking, citing, recommending, acknowledging – and with which to track scholarly salience, influence and impact, broadly conceived and broadly felt, over time^{8–10}.

In their landmark book, *Laboratory Life*, Bruno Latour and Steve Woolgar¹¹ describe how ‘rats had been bled and beheaded, frogs had been flayed’ in the service of science. They portray laboratory activity at the Salk Institute ‘as the organization of persuasion through literary inscription’. The end result of the spilling of blood and guts is the journal article, a sanitized and depersonalized artifact far removed from the messiness of what went before: ‘A laboratory,’ they aver, ‘is constantly performing operations on statements; adding modalities, citing, enhancing, diminishing, borrowing, and proposing new combinations¹².’ (Ref. 12, pp. 86–87). Ultimately, these literary operations mask the literal operations carried out on experimental rats by scientists in wet labs. We should not, of course, be surprised by any of this: ‘In scientific articles, reason proceeds down a high road that leads from darkness to light with not the slightest error, nor a hint of a bad decision, no confusion, nothing but perfect reasoning’ (ref. 13, p. 315). Finely crafted articles also draw a convenient veil over ‘research scribbling’, a set of integral activities ranging from preliminary sketching to trying out calculations that lies in the space between ‘the materialities of the experimental systems and the various written communications that are eventually released to the scientific community’ (ref. 13, p. 314).

It occurs to me, and I say this without positing any form of strategic silencing of the social on the part of scientometricians, that the messiness of professional and human relations is elided in similar fashion from most standard bibliometric and informetric accounts of knowledge growth, diffusion and sharing: ‘Never mind the underlying social reality, feel the data,’ seems to be the unspoken credo. Sociological relativists, for their part, are characterized as being antipathetic to grand narratives and large-scale number crunching. A case in point is citation, an integral part of scientific writing, yet one that is sometimes dismissed as inherently subjective and thus an unreliable basis for visualizing the intellectual structure of disciplines or for evaluating scientific performance. But, although I want to steer a judicious course between the Scylla of uncritical normativism and the Charybdis of naive

e-mail: bcronin@indiana.edu

constructivism (see Cronin⁵, for an early review of the dominant epistemological positions and Jeppe Nicolaisen¹⁴, for a more recent critique), I confess that I find it hard not to identify with Howard White's pragmatism (ref. 15, p. 9) in the matter of citations and what they tell us, once aggregated, about scientific communication: '[W]hen one sees that scores, hundreds, and even thousands of citations have accrued to a work, an author, a set of coauthors, it is difficult to believe that all of them are suspect. Why not believe there is a norm in citing... and that the great majority of citations conform to it?' Why not, indeed?

Critics of the normative position, with its emphasis on objectivity, rationalism and collectively (albeit tacitly) understood rules, note that the material reality of science is very different from the 'storybook' account with its emphasis on Great Men, selfless action and standpoint neutrality. Social constructivists stress the following: scientific knowledge is contingent, socially negotiated; knowledge claims are often fiercely contested; for every Mertonian norm there is a counter-norm; subjectivity and self-interest are inescapable facts of scientific life. In terms of citation this means, and here as elsewhere I am necessarily simplifying often complex arguments, that there is a battery of social and psychological reasons for citing, which may have as much to do with, for instance, rhetorical gamesmanship (persuading the reader of one's viewpoint through selective under- or over-citation) or strategic coat-tailing (citing friends, immediate colleagues or celebrity authors) as with the topical appropriateness or semantic suitability of the citations themselves. On top of which there is the matter of erroneous citation and the inescapable implication that authors may not always have read the materials they themselves cite^{16,17}. These and related concerns have provided considerable ammunition for the anti-citationist camp. In a series of articles in the eighties and nineties, Michael MacRoberts and Barbara MacRoberts¹⁸ argued that authors' citation behavior was prone to systematic bias. As such, they and others conclude, the approach should not be used in institutional, programmatic or individual evaluation exercises. But perhaps we don't have to throw out the bibliometric baby with the presumptively biased bathwater. Instead, why not inject a little countervailing evidence, some critical self-reflection and a soupcon of common sense into the occasionally testy debate?

Support for the so-called normative position comes from two noteworthy studies of recent origin. Michael Kurtz and colleagues¹⁹ at the Harvard-Smithsonian Center for Astrophysics have used the NASA (National Aeronautics and Space Administration) Astrophysics Data System to compare the obsolescence function as measured by 'reads' of records in the system with the obsolescence function as measured by citations. Their statistical analyses show that reads and cites 'fundamentally measure the same thing, the usefulness of an article'. The authors conclude that this 'proves that the normative theory of citation is true in

the mean'. Pertinently, Kurtz *et al.*¹⁹ go on to say that 'the private act of reading an article entails none of the various sociological influences that the public act of citing an article does [which] suggests that in the mean these factors do not influence the citation rate'. Citation, in other words is about what one knows, not whom one knows. These results seem to confirm the earlier findings of Steven Baldi's multivariate analysis of citations in the literature of celestial masers²⁰, a sub-field (coincidentally) of astrophysics: 'Authors are most likely to cite articles that are relevant to their work in terms of subject, recency of knowledge, theoretical orientation, and seem to have little concern with the characteristics of the authors who write them. This finding suggests that, at least in the research area examined, one's position in the stratification structure of science is likely to be the result of the worth and usefulness of one's scientific contributions rather than the reverse, as social constructivists would have us believe'. Content counts for more than connections.

Truth be told, we are invariably challenged to cite the most precise and most relevant work on a given subject, for the simple reason that few, if any, of us are wholly and authoritatively familiar with the scattered literature of our specialties, let alone the wider scientific literature. Even if we were familiar with the entire corpus of relevant literature, we would still have to make difficult choices. As Anthony Grafton²¹, a professor of English, noted matter-of-factly in his engaging history of the footnote, 'a historical work and its notes can never, in the nature of things, reproduce or cite the full range of evidence they rely on'. This scarcely contentious view was also held and elaborated on by Merton²²: '*By itself*, citation analysis cannot trace all the complex sources of cognitive influences upon a particular work since explicit citations, which are ordinarily the only kind entered into quantitative citation-analyses, do not adequately reflect the story ... A fine-grained analysis would have to be supplemented by focused interviews with scientists reporting on contexts of what they have set in print'. One could, and in future likely will, move beyond reliance on citation counting alone to the parallel analysis of acknowledgments, those complementary, ubiquitous, and heretofore largely ignored, paratextual ledgers of informal collaboration and influence in science^{7,23}. By way of an aside, I am inclined to think that Merton would have welcomed the communicative transparency of web-based and open access publishing and the opportunities they afford for developing novel forms of sociometric analysis and influence mapping.

Garfield's 1955 paper⁴ includes 16 citations. I looked at each one in context and concluded, impressionistically I concede, that each seemed to serve a serious rather than frivolous purpose. Nor is 16, in absolute terms, a massive number of citations for a paper of such length. It is certainly not as if Garfield was loading the text with scholarly citations to lend greater credence to his central thesis. In this vein, Latour¹² has observed that citations can serve a but-

trussing function: 'Attacking a paper heavy with footnotes means that the dissenter has to weaken each of the other papers, or will at least be threatened with having to do so'. The idea of citations as 'threat signals' has been explored recently and in some detail by Nicolaisen¹⁴. Parenthetically, the same point has been made in respect of acknowledgments: Graham Harman (ref. 24, p. vii) feels that a long list of names at the front of a book is 'like some Praetorian guard [that] often serves to intimidate readers, to make them feel outclassed by a competent network of college professors, research institutions, and fellowship foundations'. But back to Garfield's path-breaking paper⁴, specifically, citations 3–7 which contextualize the idea of an 'impact factor' and provide the reader with pointers to related work, both pro and contra. Two of the 16 citations were self-citations (one to an earlier *Science* paper), hardly surprising or excessive for a paper presenting a novel proposal. It would require painstaking detective work to reconstruct the entire pool of candidate citations that Garfield might have drawn upon in the course of writing his paper. Perhaps one or more of the 16 selected citations could have been replaced with another that was somehow either more topically focused or less ideologically one-sided, but who, ultimately, is to act as Solomon in such matters? What one can say with confidence is that the 16 citations are ostensibly tightly coupled with the central ideas of the paper and furnish the interested reader with meaningful connections to the prior literature.

A simple, self-referential illustration may help further. In the course of writing *The Hand of Science*⁷, to which I alluded earlier, I cited, amongst many others, Eugene Garfield (EG) and Elisabeth Davenport (ED), both of whom I have known personally for two decades. The former is best described as a professional friend (for example, I co-edited a *Festschrift* in his honor), the latter as a frequent collaborator, co-author and close friend. I did not cite them *because* of our social ties, but because their ideas were relevant to the work in hand. At the same time, the odds on their being cited by me are increased as a result of the pre-existing personal connections: I know them and their publications well; I interact with them, exchanging thoughts and materials. In the case of the latter we have been active, and occasionally co-located, collaborators for many years. A consequence of my citing EG and ED (and others with whom I am personally acquainted) is that it reduces, potentially, the likelihood of others in the citable author pool from being selected. All other things (the citable work's topicality, relevance, currency, etc.) being equal, strong social ties will presumably trump weak or non-existent ties. Granovetter^{25,26} has discussed weak and strong ties. Call it preferential attachment, a statistical fact of not only scientific but also social life. I shall return to the subject of social citation later.

In sum, to equate quotidian citation behavior with systematic bias, as MacRoberts and MacRoberts¹⁸ do, is to lose sight of what Bertram Brookes²⁷ (p. 118) called 'in-

dividuality and social interactions' in statistically based social scientific research. I cannot speak for Garfield, but I am willing to believe that he cited two of his own and 14 other papers because in some way each had influenced the evolution of his thinking and/or – amongst many other possible motivations – because each could provide the reader with a sense of the larger intellectual milieu in which the citing paper was located. I am also willing to believe that Garfield actually read the papers he cited. And I, for my part, would have you believe that I have likewise read and been influenced, to some extent, by the ideas embedded in the citations dangling at the end of this short essay. There are, to continue this line of reasoning, many other topically pertinent works that I could have cited; some perhaps more germane than those I have actually employed. To be sure, a few citations are included to buttress my case, but what author, or defense attorney for that matter, would not make the most of the evidence to hand, especially if the evidence is inherently credible? Undoubtedly, I have my personal preferences and biases, consciously held or not, but I have also, I think it is fair to say, provided a sufficiently balanced mix of recent and not-so-recent citations to reflect equitably the various dimensions of the construct validity debate, such that the interested, but undecided or neophyte reader can chain easily through the related literature in search of further theoretical and empirical arguments. I do not cite Merton because he is famous; I cite him because his arguments are compelling: his fame is due to the quality of his scholarship, not the reverse – which is not to say that the Great Man does not instantiate the Matthew effect that he popularized so effectively. I cite MacRoberts and MacRoberts (ornithology is their primary discipline, by the way) because they mounted a sustained, and, as it turned out, much cited, attack on the normative position. I do not cite them because their citation-related publications are frequently invoked in the literature, but because the suite of arguments marshaled in those papers constitutes one of the most cogent rejoinders to the normative position.

One of the many measures used by Baldi²⁰ in his study of citation behaviors was the existence of social ties, which he operationalized as potentially citing or potentially cited articles' authors having 'ever worked at the same institution or received their Ph D from the same graduate department' (ref. 20, p. 837). To his credit Baldi does not think that this is the end of the story. First, caution is needed in generalizing from this single-domain study to other scientific disciplines, not to mention social scientific or humanistic fields where there may be considerably less consensus as to what constitutes high quality work. Second, as he concedes (ref. 20, p. 844), more work is needed 'to elaborate the kinds of social ties that matter' and that '[f]uture studies should assess the effects of various social relationships among authors on citation occurrence by collecting more detailed information on type of ties'. I want to take this one step further.

It is important to state that a commitment to the normative view of citation does not necessarily preclude or somehow invalidate one's having an interest in the social and biographical minutiae of scholarly communication. Paula Mählck and Olle Persson²⁸ introduced the term 'socio-bibliometric mapping' and showed that 'if you collaborate with a certain author, you will tend to cite him or her'. Ian Rowlands²⁹ has also produced evidence of 'convergence between recognition, social ties and collaborative forms of activity'. Being together in a shared space, or meeting face-to-face, in turn shapes subtly the intellectual spaces in which a scholar's work is located. The central issue has been formulated thus: 'Is it primarily *who* citers know (social structure) or *what* they know (intellectual structure)?'³⁰ Of course, things are not quite that simple, as Howard White *et al.* (ref. 30, p. 125) readily acknowledge: 'Who you know pays off only if the people you know have something worth knowing – something plainly relevant to your own claims.' But unraveling the social dimensions of citation links is still not the whole story. Citations are the most visible chunk of the influence iceberg; acknowledgments are the part largely hidden below the water line⁶. Scholastic debts are routinely recorded in a variety of ways; citations just happen to be relatively easy to capture and display. In short – and Eugene Garfield has long recognized this – citation-based maps and networks are necessarily incomplete representations of scholars' interactions and interdependencies.

Informational imperfections are a feature of everyday life, as every stockbroker, intelligence analyst and gambler knows. The fact that there are gaps in our knowledge of the motivational substrate of citation does not require us to conclude that citations are dispensed randomly. The weight of empirical evidence seems to suggest that scientists typically cite the works of their peers in a normatively guided manner, and that these signs (citations) perform a mutually intelligible communicative function. As Robert van Braam³¹ has demonstrated, giving 'operational information' is the most important reason for citing because of the interaction which takes place in the reviewing process among authors, editors and referees. Either we have to assume that authors are engaged in repeated acts of whimsy, which just happen to be overlooked or go undetected by those responsible for quality control in the primary communication system – a curiously baroque and singularly unparsimonious hypothesis – or we have to assume that the reasons for, and ways in, which authors invoke the works of other authors have become, in Merton's words, 'normatively operative in modern science' (ref. 22, p. 48). Elsewhere he states that there exists 'a distinctive pattern of institutional control of a wide range of motives which characterizes the behavior of scientists' and goes on to say that 'once the institution of science enjoins disinterested activity, it is to the interest of scientists to conform on pain of sanctions and, insofar as the norm has been internalized, on pain of psychological conflict'³². This

is a fine encapsulation of the normative position, but one unlikely to pass muster with social constructivists.

'There are,' as Scott Montgomery³³ observed succinctly, 'no boundaries, no walls, between the doing of science and the communication of it; communicating *is* the doing of science.' New genres of academic writing and forms of authorship will continue to emerge, co-evolving with both the material practices of scientists and advances in information and communication technologies. The ways in, and means by, which scientists and scholars create new knowledge, publish their ideas and subject them to the scrutiny of their peers are undergoing continuous and occasionally profound change. Behavioral diversity and heterogeneity of norms, rather than standardization and uniformity, are defining features of the contemporary scholarly communication system: the publishing practices of high-energy physicists are very different from those of chemists; those of economists from those of historians, and so on. Different epistemic cultures may beget different modes of knowledge construction, validation and communication³⁴, but certain supra-disciplinary norms (e.g., bestowing credit where credit is due) remain constant – which, of course, is not quite the same as saying that infractions do not occur with greater or lesser frequency in particular domains.

Social constructivists may decry the lack of attention given to individual practice, materiality, grounded observation and situated action in sweeping functionalist interpretations of science, but I am happy to live with the 'noble fiction' – Plato's idea of the useful lie – as far as citation behavior is concerned if only because the alternative is repugnant to common sense and at variance with the observed conduct of most scientists. Eugene Garfield's legacy is not just a scientifically valuable concept, but a foundational tool (ISI's citation indexes), one that reveals 'the extent to which scientific progress depends upon the pervasive and persistent constitution of trust among a truly heterogeneous population of social actors' (ref. 35, p. 522). Trust, though constructivists may not like to hear this, is the ghost in the scientific machine. But that was true long before Garfield conceived of the *Science Citation Index*. In his superb study of science in seventeenth-century England, Steven Shapin (ref. 1, p. xxv) writes as follows: 'Knowledge is a collective good. In securing our knowledge we rely upon others, and we cannot dispense with that reliance. That means that the relations in which we have and hold our knowledge have a moral character, and the word I use to indicate that moral relation is *trust*.' It is the mutual reliance of which Shapin speaks that confers credibility upon authors' citation behaviors. A hundred million acts of whimsy? I trust not.

1. Shapin, S., *A Social History of Truth: Civility and Science in Seventeenth-century England*, University of Chicago Press, Chicago, 1994.

2. Wouters, P., Garfield as alchemist. In *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield* (eds Cronin, B. and Atkins, H. B.), Information Today, Inc., Medford, NJ, 2000, pp. 65–71.
3. Cronin, B. and Atkins, H. B. (eds). *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*, Information Today, Inc., Medford, NJ, USA, 2000.
4. Garfield, E., Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, 1955, **122**, 108–111.
5. Cronin, B., *The Citation Process: The Role and Significance of Citations in Scientific Communication*, Taylor Graham, London, 1984.
6. Cronin, B., *The Scholar's Courtesy: The Role and Significance of Acknowledgement in the Primary Communication Process*, Taylor Graham, London, 1995.
7. Cronin, B., *The Hand of Science: Academic Writing and its Rewards*, Lanham, Scarecrow Press, MD, 2005.
8. Brody, T., Stamerjohanns, H., Harnad, S., Gingras, Y., Vallieres, F. and Oppenheim, C., The effect of open access on citation impact. Paper presented at National Policies on Open Access (OA) Provision for University Research Output: an International Meeting, Southampton University, 2004; <<http://opcit.eprints.org/feb19prog.html>>
9. Harnad, S. and Brody, T., Comparing the impact of open access (OA) vs non-OA articles in the same journals. *D-Lib Mag.*, 2004, **10**; <<http://www.dlib.org/dlib/june04/harnad/06harnad.html>>
10. Cronin, B., Bibliometrics and beyond: some thoughts on web-based citation analysis. *J. Inf. Sci.*, 2001, **27**, 1–7.
11. Latour, B. and Woolgar, S., *Laboratory Life: the Social Construction of Scientific Facts*, Sage Beverly Hills, CA, 1979, p. 88.
12. Latour, B., *Science in Action*, Harvard University Press, Cambridge, MA, 1987, pp. 86–87.
13. Rheinberger, H.-J., 'Discourses of circumstance': A note on the author in science. In *Scientific Authorship: Credit and Intellectual Property in Science* (eds Biagioli, M. and Galison, P.), Routledge, New York, 2003, pp. 309–323.
14. Nicolaisen, J., Social behavior and scientific practice – Missing pieces of the citation puzzle, Royal School of Library and Information Science. Ph D thesis, Copenhagen, 2004.
15. White, H. D., Author co-citation analysis: overview and defense. In *Scholarly Communication and Bibliometrics* (eds Borgman, C. L.), Sage, Newbury Park, CA, 1990, pp. 84–106.
16. Cardona, M. and Marx, W., Confused, forgotten, rediscovered: References – the deficient memory of science. *Phys. J.*, 2004, **3**, 27–29; <<http://www.prophysik.de/Phy/External/PhyH/1.7058.2-3-0-0-1-phy.toc.content.page-0-21221.00.html>>
17. Simkin, M. V. and Roychowdhury, V. P., Read before you cite! *Complex Syst.*, 2003, **14**, 269–274.
18. MacRoberts, M. H. and MacRoberts, B. R., Quantitative measures of communication in science: a study of the formal level. *Social Stud. Sci.*, 1986, **16**, 151–172.
19. Kurtz, M. J., Eichorn, G., Accomazzi, A., Grant, C., Demleitner, M., Murray, S. S., Martimbeau, N. and Elwell, B., The bibliometric properties of article readership information. *J. Am. Soc. Inf. Sci. Technol.*, 2004, **56**, 111–128.
20. Baldi, S., Normative versus social constructivist processes in the allocation of citations: A network-analytic model. *Am. Sociol. Rev.*, 1998, **63**, 829–846.
21. Grafton, A., *The Footnote: A Curious History*, Harvard University Press, Cambridge, MA, 1997.
22. Merton, R. K., The sociology of science: an episodic memoir. In *Merton, The Sociology of Science in Europe* (eds R. K. and Gaston, J), Southern Illinois University Press, Carbondale, IL, 1977, pp. 3–141.
23. Giles, C. L. and Councill, I. G., Who gets acknowledged? Measuring scientific contributions through automatic acknowledgement indexing. *Proc. Nat. Acad. Sci. USA*, 2004, **101**, 17599–17604.
24. Harman, G., *Tool-Being: Heidegger and the Metaphysics of Objects*, Open Court, Chicago, 2002.
25. Granovetter, M., The strength of weak ties. *Am. J. Sociol.*, 2001, **78**, 1360–1380.
26. Granovetter, M., The strength of weak ties: A network theory revisited. *Sociol. Theory*, 1983, **1**, 201–233.
27. Brookes, B. C., People versus particles. In *Theory and Application of Information Research* (eds Harbo, O. and Kajberg, L.), Mansell, London, 1980, pp. 106–118.
28. Mählck, P. and Persson, O., Socio-bibliometric mapping of intra-departmental networks. *Scientometrics*, 2000, **49**, 81–91.
29. Rowlands, I., Patterns of author cocitation in information policy: Evidence of social, collaborative and cognitive structure. *Scientometrics*, 1999, **44**, 533–546.
30. White, H. D., Wellman, B. and Nazer, N., Does citation reflect social structure?: Longitudinal evidence from the 'GloboNet' interdisciplinary research group. *J. Am. Soc. Inf. Sci. Technol.*, 2004, **55**, 111–126.
31. van Braam, R. R., *Mapping of Science: Foci of Intellectual Interest in Scientific Literature*, DSWO Press, University of Leiden, 1991.
32. Merton, R. K., The normative structure of science. In *The Sociology of Science: Theoretical and Empirical Investigations* (ed Merton, R. K.), University of Chicago Press, Chicago, 1973, pp. 267–278.
33. Montgomery, S. L., *The Chicago Guide to Communication Science*, Chicago University Press, Chicago, 2003.
34. Knorr Cetina, K., *Epistemic Cultures: How the Sciences Make Knowledge Work*, Harvard University Press, Cambridge, MA, 1999.
35. Davenport, E. and Cronin, B., The citation network as a prototype for representing trust in virtual environments. In *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield* (eds Cronin, B. and Atkins, H. B.), Information Today, Inc., Medford, NJ, 2000, pp. 517–534.