



HAL
open science

What is the primordial reference for ...?—Redux

Guillaume Cabanac

► **To cite this version:**

Guillaume Cabanac. What is the primordial reference for ...?—Redux. *Scientometrics*, 2018, 114 (2), pp.481-488. <10.1007/s11192-017-2595-4>. <hal-02348282>

HAL Id: hal-02348282

<https://hal.science/hal-02348282v1>

Submitted on 5 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in:

<http://oatao.univ-toulouse.fr/22347>

Official URL

DOI : <https://doi.org/10.1007/s11192-017-2595-4>

To cite this version: Cabanac, Guillaume *What is the primordial reference for ...?---Redux*. (2018) *Scientometrics*, 114 (2). 481-488. ISSN 0138-9130

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

What is the primordial reference for ...? — Redux

Guillaume Cabanac

Abstract Eugene Garfield’s quest of the primordial reference for the familiar and ubiquitous phrase ‘Publish or Perish’ led him to a 1942 monograph (*The Scientist*, 1996, 10(12)). This quest is resumed two decades later here. Text mining applied to a sample of the mainstream and academic literature ever published, as well as crowdsourcing, yielded earlier references dating from 1934 and 1927. This search experiment suggests that ‘primordial reference chasing’ in full-text corpora remains an open problem for the community intersecting bibliometrics and information retrieval. Addressing it has the potential to rejuvenate Garfield’s work on historio-bibliography to improve our understanding of the genesis and diffusion of ideas, concepts, and associated metaphors.

Keywords Primordial reference chasing, Publish or Perish, Eugene Garfield

Introduction

Academics are expected and encouraged to cite all their influences, but they rarely do so (M. H. MacRoberts & MacRoberts, 1986, 2010). Sometimes one chooses to select the most relevant references to his/her work; sometimes the primordial reference for a phrase or idea is nowhere to be found. Hedging by surrounding an expression with quotes is one way of dealing with this issue. These quotes act as a signalling device for the reader: the quoted phrase was coined elsewhere and it would be misleading to think it is genuine here.

Eugene Garfield had long demonstrated his enduring interest for primordial concepts and historio-bibliography (Garfield, 1967) when he embarked on the quest of the primordial reference for the phrase ‘Publish or Perish’ . . .

G. Cabanac

University of Toulouse, Computer Science Department, IRIT UMR 5505 CNRS,
118 route de Narbonne, F-31062 Toulouse cedex 9, France

E-mail: guillaume.cabanac@univ-tlse3.fr

ORCID: [0000-0003-3060-6241](https://orcid.org/0000-0003-3060-6241)

Garfield's take on the primordial reference problem in 1996

An academic asked Garfield if he knew of the original reference for the familiar phrase 'Publish or Perish.' In a witty commentary published in *The Scientist*, Garfield (1996) recounts his attempts to chase this ubiquitous expression. He had recourse to various strategies involving a variety of information sources. He did a literature search, searched dictionaries (both in print and online, he notes), and several databases. A Stanford librarian even joined forces in vain to post his query on Internet newsgroups and electronic bulletin boards.

Crowdsourcing appeared to be more effective: word of mouth spread through Garfield's invisible college and a Yale librarian and editor of the *Oxford Dictionary of American Legal Quotations* dug up an occurrence of 'Publish or Perish' in a 1951 letter (Molarino, McLuhan, & Toye, 1987, p. 226). Later on, he found an earlier occurrence (Wilson, 1942, p. 197). Noting that Wilson was a sociologist, Garfield reached to Robert K. Merton, who had taught Wilson at Harvard University. Merton suggested that 'Publish or Perish' was an expression in the air in pre-war academe, but no earlier account of it was provided.

Contemporary authors of Garfield's chase endeavour (e.g., Tenopir, 1995, p. 575) were still quoting 'Publish or Perish' with no reference to its primordial source. Note that all of these authors (i.e., of the 1951 letter, 1942 book, and 1995 paper) cared to word the expression surrounded by quotes, as if to warn readers that they were not coining it and that they had failed to find the primordial reference for it.

In closing his commentary, Garfield (1996, p. 11) tossed a bottle into the ocean: "Perhaps one of *The Scientist's* readers can identify an earlier source for this common phrase." This is the challenge I have taken up two decades later using modern computing and networking capabilities.

Opportunities of text mining on scholarly materials

Nowadays various online services can help one track primordial references. For instance, Google Books digitised, OCRised, and tokenised the 8 million books fuelling the *Ngram Viewer* search interface. This amounts to 6% of all books ever published as of 2012 (Michel et al., 2011; Lin et al., 2012), that is a corpus including but not limited to scholarly books. Figure 1 suggests that the phrase 'Publish or Perish' appeared in print *in this corpus* around the late 1940s, which is in line with the findings reported in (Garfield, 1996). This expression gained popularity in the 1960s, peaked and plateaued in the 1970s, only to slowly decrease during the next decades.

Search engines focused on scholarly materials offer an other option to chase expressions appearing in the full-text of academic papers, whereas leading citation indexes are restricted to title, abstract, and keywords fields. Google Scholar, introduced a decade ago (Giles, 2005), is now an established and widely used academic search engine (Jacso, 2005; Harzing & Alakangas, 2016). Unfortunately, searching for 'Publish or Perish' and sorting results by date of publication retrieves papers added in the last year only, which has close to no use for primordial reference chasing.

The JSTOR digital library developed *Data for Research*¹ (Burns et al., 2009) which now indexes the full-text of 9 million papers dated between 1545 and 2014. The recent studies by McCain (2014, 2015) relied on this service to track down obliterations by incorporation (see OBI in Merton, 1988, p. 622), namely expressions such as "bounded reality" or "Nash

¹ http://dfr.jstor.org/?helpview=about_dfr

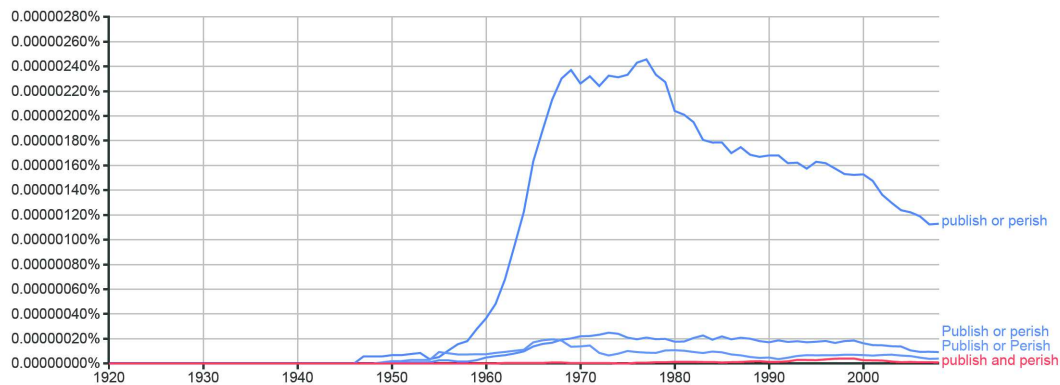


Fig. 1 Result of the *Ngram Viewer* for query [“publish or perish”, “publish and perish”] on the English corpus of Google Books comprising 8 million books (see <http://bit.ly/pop-ngram>).

equilibrium” appearing in the body of research papers with no (more) associated references in the bibliography.

Data for Research allowed me to further Garfield’s quest of the primordial reference for ‘Publish or Perish.’ The oldest occurrence *in this corpus* appears in an obituary by Bowman (1934, p. 180), who recounted that the late Professor Davis urged the Association of American Geographers to ‘Publish or Perish,’ back in a 1904 meeting. The longitudinal data shown in Fig. 2 correlate earlier observations made on mainstream books (Fig. 1) for the rise in popularity of this expression in the 1960s. Then, it seems that ‘Publish or Perish’ stuck in academics’ minds and texts. The sudden surge in popularity just before 2010 might reflect the release and massive uptake of the *Publish or Perish* software² (Harzing, 2010) that has been developed since 2006. In addition, the subject group breakdown suggests that the expression appears the most in articles classified in the Social Sciences.

Reflecting on Garfield’s recourse to crowdsourcing, I turned to Wikipedia and, to my utmost surprise, found there an even older reference to a 1927 journal article of sociology added³ on August 18, 2016 with a note about the source: ‘via Google Books’ hyperlinked to a 5-line snippet including the famous phrase.⁴ I requested a copy from a university library in Strasbourg and updated the incomplete Wikipedia reference, adding the page range for (Case, 1927–1928) that states on page 325:

... the system of promotion used in our universities amounts to the warning, “Publish or perish!” In the second place, publication in general ...

It is striking how this ninety-year-old critique sounds like many contemporary rants overheard in academe nowadays. *Plus ça change, plus c’est la même chose*. Plagiarising Garfield, I dare to hope that perhaps one the readers of *Scientometrics* can identify an earlier source for this common phrase.

Towards search methods for primordial reference chasing

These anecdotal accounts of Garfield’s chase for a primordial reference back in 1996, and my own two decades later, suggest a challenging research issue for the researchers in-

² <https://harzing.com/resources/publish-or-perish>

³ https://en.wikipedia.org/w/index.php?title=Publish_or_perish&diff=prev&oldid=735106196

⁴ <https://books.google.co.uk/books?id=qBAZAAAIAAJ&q=%22publish-or-perish%22>

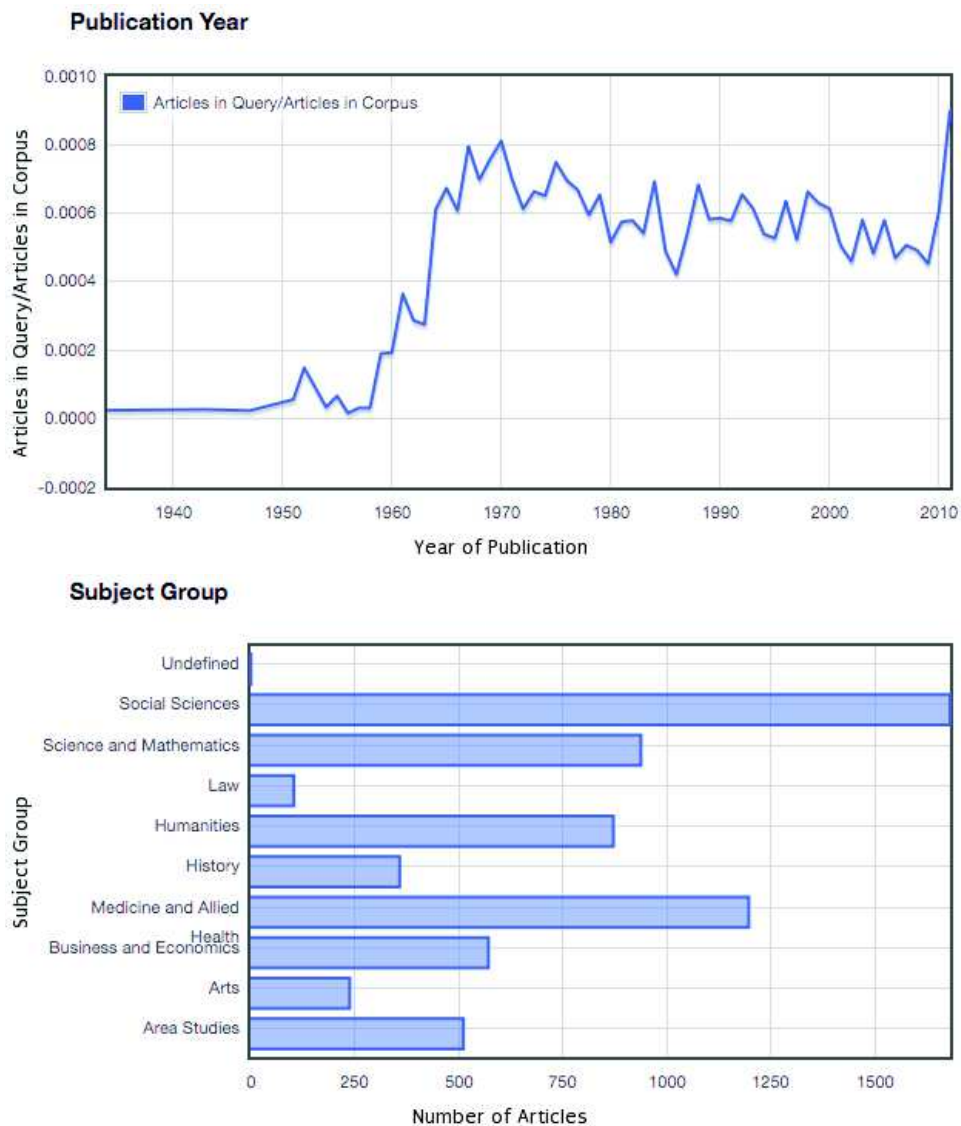


Fig. 2 Results of *Data for Research* for query ["publish or perish"] on the JSTOR DFR corpus of 9 million academic papers (see <http://bit.ly/pop-jstor>).

volved at the crossroads between bibliometrics and information retrieval. Tackling the ‘primordial reference chasing’ problem would contribute to tighten up the ties between these two communities of information scientists that have somewhat drifted apart from each other (see White & McCain, 1998, p. 345). Initially launched to foster re-connections, the *Bibliometric-enhanced Information Retrieval* workshop series seek to address such challenging issues requiring competence in both domains (Mayr, Frommholz, & Cabanac, 2017).

Garfield’s pioneering work on historio-bibliography (Garfield, 1967) led to the development of the *HistCite*TM software (Garfield, Pudovkin, & Istomin, 2002) allowing one to reconstitute the genealogy of papers chained by citations. With literature-based discovery as a framework (Bruza & Weeber, 2008), maybe time is ripe to rejuvenate and extend this legacy by moving from the citation level to the full-text level.

The following sections provide examples of variations in the primordial reference chasing problem that one might need to address.

Primordial reference for a phrase

Seeking the primordial reference for a phrase in a corpus of academic materials can be as simple as an exact string search. Take for instance the ‘invisible college’ (Price & Beaver, 1966) and the ‘Matthew Effect’ (Merton, 1968).⁵ Trouble arises when we are looking for a phrase used as a metaphor: it is ubiquitous but, in a certain context, it was used for the first time to describe a phenomenon or a concept. Let us consider here the use of ‘gatekeepers’ by Crane (1967) and the ‘Sleeping Beauties’ of van Raan (2004). Other phrases such as ‘salami slicing’ might even be more tricky to track in the literature — here the search engine should not return papers from the *Meat Science* journal if we are considering bibliometrics as the search context.

A challenging extension of this search task consists in retrieving the derivatives of a given phrase. Figure 1 shows the weak uptake of the ironic derivative ‘Publish *and* Perish’ whose eponymous primordial reference appeared in an insightful journal article (Hurt, 1961) according to JSTOR’s *Data for Research*. Another derivative recently coined by Ye and Bornmann (forthcoming) is ‘Smart Girls,’ in contrast to van Raan’s ‘Sleeping Beauties’ to qualify instant *versus* delayed recognition of research works.

Primordial reference for an eponym

Stigler (1980, p. 148) concisely summarised two empirical observations about eponyms:

“First, names are not given to scientific discoveries by historians of science or even by individual scientists, but by the community of practicing scientists (most of whom have no special historical expertise). Second, names are rarely given, and never generally accepted unless the namer (or acceptor of the name) is remote in time or place (or both) from the scientist being honored.”

Here a search task might be to: find the primordial reference coining a given eponym. For instance, who started calling a particular disease ‘Alzheimer’s,’ and in which paper? The interested reader is referred to (Cabanac, 2014, p. 1637) for a chase of ‘Hirsch’s index’ (Hirsch, 2005), an interesting case of an eponym coined and popularised without spatio-temporal distance. Derivative seeking is also an open problem: take the Northern and Eastern blotting techniques named after the eponym ‘Southern blotting’ in molecular biology (Thomas, 1992).

Stigler’s Law of Eponymy states that “No scientific discovery is named after its original discoverer” (Stigler, 1980, p. 147), but there is an earlier declaration of this principle in (Kennedy, 1972, p. 67). A true challenge here lies in the designing of a method to identify links between an established eponym and the same discovery made earlier in order to revise questionable or misleading attributions and inform the historians of science.

Primordial reference for a concept and re-discoveries

A third even more challenging search task consists in finding the references that could have influenced a given discovery or concept, provided that its author had known about a specific, earlier, reference. This relates to re-discoveries, like in the case of Edwards (2005) sending a

⁵ Disclaimer: the references provided here are not necessarily primordial references.

letter to the editor of *Nature* to stress that Sir Harold Jeffreys had told him, decades ago, that he conceived an indicator to measure his cycling prowess — this indicator being computed the same way as the *h*-index (Hirsch, 2005).

By text mining the digitised literature, an effective algorithm could perhaps trace the genesis of the Open Access movement (Tennant et al., 2016) back to Merton (1942, p. 122):

“The institutional conception of science as part of the public domain is linked with the imperative for communication of findings. Secrecy is the antithesis of this norm; full an open communication its enactment. The pressure for diffusion of results is reinforced by the institutional goal of ‘advancing the boundaries of knowledge’ and by the incentive of recognition which is, of course, contingent upon publication.”

In this passage about the norm of ‘Communism’ in the Ethos of Science, let us note the colloquialism ‘advancing the boundaries of knowledge’ and the quotes surrounding it, as if the author wished to refer to some source he couldn’t establish. . .

Concluding remarks

Garfield (1959, p. 461) coined the concept of ‘Unified Index to Science’ that he defined as “a single inter-disciplinary index to *all* documents, primarily periodical literature in *all* fields of science.” Today and according to recent estimates, this corpus would include the 2.5 million articles published each year in about 30.000 peer-reviewed journals (Ware & Mabe, 2015, p. 27) that are still active among the 300.000 journals ever recorded (Kaplan, Killough, & Thomas, 2012, p. 152). Conference proceedings, books, and grey literature also contribute to the mass of scholarly documents produced by the 7.8 million scientists worldwide (Soete, Schneegans, Eröcal, Angathevar, & Rasiah, 2015, p. 32).

Some online services seem to have been inspired by this vision of a ‘Unified Index to Science,’ providing search capabilities to scholars searching the literature and seeking primordial references like ‘Publish or Perish.’ Khabsa and Giles (2014, p. 3) estimated that Google Scholar was indexing the full-text of about 100 million scholarly documents. And yet, one may wonder how many primordial references are still not digitized, not indexed, and not yet accessible online. . .

References

- Bowman, I. (1934). William Morris Davis [Obituary]. *The Geographical Review*, 24(2), 177–181. JSTOR: [208785](#)
- Bruza, P. & Weeber, M. (2008). Literature-based discovery. (Vol. 15). Information Science and Knowledge Management. Berlin: Springer. doi:[10.1007/978-3-540-68690-3](#)
- Burns, J., Brenner, A., Kiser, K., Krot, M., Llewellyn, C., & Snyder, R. (2009). JSTOR – Data for Research. In M. Agosti, J. Borbinha, S. Kapidakis, C. Papatheodorou, & G. Tsakonias (Eds.), *ECDL’09: Proceedings of the 13th European Conference on Research and Advanced Technology for Digital Libraries* (Vol. 5714, pp. 416–419). LNCS. Springer. doi:[10.1007/978-3-642-04346-8_48](#)
- Cabanac, G. (2014). Extracting and quantifying eponyms in full-text articles. *Scientometrics*, 98(3), 1631–1645. doi:[10.1007/s11192-013-1091-8](#)
- Case, C. M. (1927–1928). Scholarship in sociology. *Sociology and Social Research*, 12, 323–340. Retrieved from <http://www.sudoc.fr/036493414>

- Crane, D. (1967). The gatekeepers of science: Some factors affecting the selection of articles for scientific journals. *The American Sociologist*, 2(4), 195–201. JSTOR: [27701277](#)
- Edwards, A. W. F. (2005). System to rank scientists was pedalled by Jeffreys [Correspondence]. *Nature*, 437(7061), 951. doi:[10.1038/437951e](#)
- Garfield, E. (1959). A Unified Index to Science. In *Proceedings of the International Conference on Scientific Information* (Vol. 1, pp. 461–474). Washington, DC: National Academy of Sciences — National Research Council.
- Garfield, E. (1967). Primordial concepts, citation indexing, and historio-bibliography. *The Journal of Library History*, 2(3), 235–249. JSTOR: [25540056](#)
- Garfield, E. (1996). What is the primordial reference for the phrase ‘Publish or Perish’? [Commentary]. *The Scientist*, 10(12), 11. Retrieved from <http://the-scientist.com/17052>
- Garfield, E., Pudovkin, A. I., & Istomin, V. S. (2002). Algorithmic citation-linked historiography-mapping the literature of science. *Proceedings of the American Society for Information Science and Technology*, 39(1), 14–24. doi:[10.1002/meet.1450390102](#)
- Giles, J. (2005). Science in the web age: Start your engines. *Nature*, 438(7068), 554–555. doi:[10.1038/438554a](#)
- Harzing, A.-W. (2010). *The Publish or Perish book: Your guide to effective and responsible citation analysis*. Melbourne, Australia: Tarma Software Research.
- Harzing, A.-W. & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804. doi:[10.1007/s11192-015-1798-9](#)
- Hirsch, J. E. (2005). An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572. doi:[10.1073/pnas.0507655102](#)
- Hurt, L. (1961). Publish and perish. *College English*, 23(1), 5–10. doi:[10.2307/373930](#)
- Jacso, P. (2005). As we my search – Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89(9), 1537–1547. JSTOR: [24110924](#)
- Kaplan, L., Killough, K., & Thomas, S. L. (2012). A 10 year collaboration—Still going strong, Ulrich’s and ISSN. *The Serials Librarian*, 62(1–4), 151–154. doi:[10.1080/0361526x.2012.652907](#)
- Kennedy, H. C. (1972). Who discovered Boyer’s Law? *The American Mathematical Monthly*, 79(1), 66–67. doi:[10.2307/2978134](#)
- Khabsa, M. & Giles, C. L. (2014). The number of scholarly documents on the public Web. *PLoS ONE*, 9(5), e93949. doi:[10.1371/journal.pone.0093949](#)
- Lin, Y., Michel, J.-B., Aiden, E. L., Orwant, J., Brockman, W., & Petrov, S. (2012). Syntactic annotations for the Google Books Ngram corpus. In *ACL’12: Proceedings of the 50th annual meeting of the Association for Computational Linguistics* (pp. 169–174). Stroudsburg, PA: Association for Computational Linguistics. Retrieved from <http://aclweb.org/anthology/P12-3029>
- MacRoberts, M. H. & MacRoberts, B. R. (1986). Quantitative measures of communication in science: A study of the formal level. *Social Studies of Science*, 16(1), 151–172. doi:[10.1177/030631286016001008](#)
- MacRoberts, M. H. & MacRoberts, B. R. (2010). Problems of citation analysis: A study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology*, 61(1), 1–12. doi:[10.1002/asi.21228](#)

- Mayr, P., Frommholz, I., & Cabanac, G. (2017). Report on the 5th International Workshop on Bibliometric-enhanced Information Retrieval (BIR 2017). *SIGIR Forum*, 51(1), 29–35. doi:10.1145/3130332.3130337
- McCain, K. W. (2014). Assessing obliteration by incorporation in a full-text database: JSTOR, Economics, and the concept of “bounded rationality”. *Scientometrics*, 101(2), 1445–1459. doi:10.1007/s11192-014-1237-3
- McCain, K. W. (2015). Mining full-text journal articles to assess obliteration by incorporation: Herbert A. Simon’s concepts of bounded rationality and satisficing in economics, management, and psychology. *Journal of the Association for Information Science and Technology*, 66(11), 2187–2201. doi:10.1002/asi.23335
- Merton, R. K. (1942). Science and technology in a democratic order. *Journal of Legal and Political Sociology*, 1(1), 115–126. doi:2027/mdp.39015008014428
- Merton, R. K. (1968). The Matthew Effect in Science: The reward and communication systems of science are considered. *Science*, 159(3810), 56–63. doi:10.1126/science.159.3810.56
- Merton, R. K. (1988). The Matthew Effect in Science, II: Cumulative advantage and the symbolism of intellectual property. *Isis*, 79(4), 606–623. JSTOR: 234750
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, ... Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 176–182. doi:10.1126/science.1199644
- Molarino, M., McLuhan, C., & Toye, W. (Eds.). (1987). *Letters of Marshall McLuhan*. Toronto: Oxford University Press.
- Price, D. J. d. S. & Beaver, D. d. (1966). Collaboration in an invisible college. *American Psychologist*, 21(11), 1011–1018. doi:10.1037/h0024051
- Soete, L., Schneegans, S., Eröcal, D., Angathevar, B., & Rasiah, R. (2015). A world in search of an effective growth strategy. In S. Schneegans (Ed.), *UNESCO Science Report: Towards 2030* (Chap. 1, pp. 20–55). UNESCO Reference Works. Paris. Retrieved from <http://unesdoc.unesco.org/images/0023/002354/235406e.pdf>
- Stigler, S. M. (1980). Stigler’s law of eponymy. In T. F. Gieryn (Ed.), *Transactions of the New York Academy of Sciences* (Vol. 39, 1, pp. 147–157). Robert K. Merton Festschrift Volume. doi:10.1111/j.2164-0947.1980.tb02775.x
- Tennant, J. P., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B., & Hartgerink, C. H. J. (2016). The academic, economic and societal impacts of Open Access: An evidence-based review [version 3; referees: 4 approved, 1 approved with reservations]. *F1000Research*, 5, 632. doi:10.12688/f1000research.8460.3
- Tenopir, C. (1995). Authors and readers: The keys to success or failure for electronic publishing. *Library Trends*, 43(4), 571–591.
- Thomas, K. S. (1992). The development of eponymy: A case study of the Southern blot. *Scientometrics*, 24(3), 405–417. doi:10.1007/bf02051038
- van Raan, A. F. J. (2004). Sleeping Beauties in science. *Scientometrics*, 59(3), 467–472. doi:10.1023/b:scie.0000018543.82441.f1
- Ware, M. & Mabe, M. (2015). *The STM Report: An overview of scientific and scholarly journal publishing*. The Hague: International Association of Scientific. Retrieved from http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf
- White, H. D. & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of Information Science, 1972–1995. *Journal of the American Society for Information Science*, 49(4), 327–355. doi:b57vc7
- Wilson, L. (1942). *The academic man: A study in the sociology of a profession*. New York, NY: Oxford University Press.

Ye, F. Y. & Bornmann, L. (forthcoming). “Smart Girls” versus “Sleeping Beauties” in the sciences: The identification of instant and delayed recognition by using the citation angle. *Journal of the Association for Information Science and Technology*. doi:[10.1002/asi.23846](https://doi.org/10.1002/asi.23846)