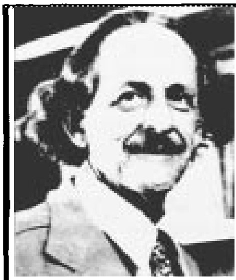
Essays of an Information Scientist: Creativity, Delayed Recognition, and other Essays, Vol:12, p.30, 1989 Current Contents, #.5, p.3-8, January 30, 1989



Current Comments®

EUGENE GARFIELD
INSTITUTE FOR SCIENTIFIC INFORMATION®
3501 MARKET ST., PHILADELPHIA, PA 19104

Citation Classics and Citation Behavior Revisited

Number 5 January 30, 1989

This essay discusses various aspects of citation behavior, as well as the process by which *Citation Classic®* papers are identified. In establishing citation thresholds to determine highly cited papers, it is necessary to control for numerous factors, including the dominance of large, high-impact journals. Often, our own analyses of research fronts and studies of most-cited authors will help us identify candidate papers. Also considered is the problem of influential and important work that, for a variety of reasons, may not be highly cited. We invite readers to inform us of any such uncited classics.

Richard C. Lewontin, Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts, has recently published one of the most interesting commentaries for "This Week's *Citation Classic®*" that we've had in a long time.[1] We've published over 3,000 such autobiographical reflections, representing the work of about 7,000 scientists.[2] It's a pleasant surprise when an author adopts an unusually informative approach to deal with this unique task.

It is not easy to compress one's thoughts into 500 words, as is required for a *Citation Classic* commentary. Many authors therefore provide little more than an extended technical abstract. The most successful commentaries take a more global view, no matter how specialized the basic subject matter may be.

Lewontin called me one day to ask if we could provide him with the citation frequencies of the paper he coauthored with Ken-ichi Kojima, Department of Genetics, North Carolina State College, Raleigh. Despite the information that we send to all *Citation Classic* invitees, he was perplexed by our selection procedure. Using the data I obtained for him from our files, he has provided new insight on citation behavior—a subject that has been a major preoccupation of mine for some 30 years.

I will not attempt to summarize here what Lewontin has already said splendidly. He provides an example of citation analysis that should be read by those who are perplexed by the sometimes inexplicable lack of direct impact of many important scientific discoveries.

Scientists often tell me that they have received inadequate credit or recognition for their work. Typically, this occurs when a well-known scientist "reviews" the work of the lesser-known author. At first, this makes the cited author happy, but elation diminishes when he or she notes that subsequent authors cite the review rather than the original "primordial" paper.

This type of citation behavior is often attributed to the "Matthew effect," the term coined by sociologist Robert K. Merton, Columbia University, New York, to describe the phenomenon whereby better-known scientists tend to get disproportionately great credit.[3] The degree varies from case to case but—as I will explain later—just a few dozen "missing" citations can affect the ranking of an important paper. Some worthy papers inevitably fail to be identified as *Citation Classics* simply because any system of selection involving an arbitrary cutoff must eliminate items just below the threshold. Such items, while failing to make an arbitrary cut, may nevertheless represent important and influential work. The problem is similar to that examined by sociologist Harriet Zuckerman, Columbia University, in *Scientific Elite*, her book on the Nobel Prize in America. Discussing scientists of Nobel class who have not actually won the prize, Zuckerman adopts the metaphor of the cohorts of 40 in the French Academy (an elite group that, like the ranks of Nobel Prize winners, happens to exclude many renowned and deserving scientists). Such Nobel-class candidates, as Zuckerman notes, may be said to occupy the "forty-first chair in science."[4]

Lewontin wants to know exactly how we choose *Citation Classics*. If my previous writings have not made this clear, then what follows may serve a useful, reiterative purpose.

Absolute Citation Frequency

We could have arbitrarily decided that any paper cited over a given number of times (400 citations is one possible threshold) would be designated as a *Citation Classic*. As Table 1 indicates, the number of papers that have achieved this absolute frequency is relatively small (that is, in relation to the estimated 30 million extant papers). In other contexts, of course, some 8,000 to 9,000 papers regarded as *Citation Classics* is hardly small.

If we made selections on the basis of absolute frequency, only articles published in a small core of high-impact journals would be selected. Some time ago, I called this "Garfield's law of concentration,"[5] which is an "extension" of Bradford's law of scattering.[6] A relatively small number of journals account for a large percentage of the papers published on any topic you can name—and for an even greater percentage of the high-impact papers. Consequently, regardless of field, a small number of journals would account for most *Citation Classics*, were they chosen only by absolute frequency.

Were we to use such absolute counts, we would repeatedly encounter papers from such superstar journals as the *Journal of the American Chemical Society*, the various editions of *Physical Review*, *Nature*, *Science*, the *New England Journal of Medicine*, *Lancet*, and so on. This results not only from the prestige of these journals, but also from their relatively large size. *Physical Review* is not the highest impact journal, but it publishes so many papers that it is bound to include superstar papers. Indeed, were we able to filter our data so that the top 1 percent of papers for each journal were not included in our calculations, we might find that the remaining 99 percent perform no better than papers from other journals with much lower impact.

In order to avoid complete domination by the high-impact journals, we decided many years ago that we would arbitrarily limit our initial choices to the top 10 to 20 papers for each journal. For smaller journals, we required that the citation threshold should not fall below 100. Keep in mind that we cover thousands of journals, many of which have never published a paper that has achieved that threshold. In the early years of *Citation Classics*, this approach worked well. But, later on, we found that life and science are not quite that simple.

Classic Papers, and Where to Find Them

As it is with journals, so it is with authors—a small percentage publish a large percentage of the papers in any field. When we chose the most-cited paper for a small journal, we often found that the same author had published one or more other papers on much the same topic—either in one of the larger journals or in one of the more prestigious review journals. We could not readily tell whether these papers covered the same ground unless the author or a colleague indicated the primordial paper.

It happens to be a fact of publishing life that new journals often define new specialties. That is why the most-cited papers for a specialty journal often emerge as classics. The editor of a new journal makes a valiant effort to find high-quality papers that may really determine the success or failure of the journal. Thus, we often find that the first volume or issue is apt to contain a *Citation Classic*. So the arbitrary selection procedure described earlier did not land us in trouble very often.

Table 1: Citation thresholds and cumulative citations at various thresholds for the 300,000 most-cited papers from the *SCI*®, 1955-1987.

Citation Threshold	Cumulative Citations	Cumulative Items
166,995	166,995	1
≥10,000	454,646	18
≥5,000	761,440	63
≥4,000	838,737	81
≥3,000	1,009,430	130
≥2,000	1,372,740	284
≥1,000	2,605,870	1,214
≥750	3,410,234	2,154
≥500	5,142,802	5,043
≥400	6,493,742	8,086
≥300	8,786,114	14,779
≥200	13,434,938	34,168
≥150	18,047,217	61,136
≥100	26,718,166	133,509
≥64	39,436,339	296,342

Alternate Sources of Candidates

As *Current Contents*® readers know, we have published numerous essays on the Nobel and other prizes, most-cited authors, and other related topics. Each of these editorial efforts invariably turns up candidate authors for *Citation Classics*. Sometimes our research-front analyses help us identify candidate papers for a specialized topic when they are core to a continuing series of annual research fronts. However, ISI® has not yet afforded me the luxury of creating a clustered research-front database covering 20 years or more of data.

Our clustering algorithms have been applied to *annual* files and to some limited three-year files in computer science, math, chemistry, and the earth sciences. Ideally, we need to create research-front cluster analysis for a 10-year period or longer. As with annual files, this analysis would identify several thousand research fronts. For each of these, there would be one or more core papers that would be the *Citation Classics* for that front. This method would turn up classic papers in smaller fields that are otherwise drowned out by the larger number of highly cited papers in large areas (clusters) represented by huge journals. Similarly, a research-front analysis of an earlier decade, such as that covered in the 1945-1954 *Science Citation Index*® (*SCI*®) cumulation[7] (now published in 10 volumes) would turn up many classics that are obscured by papers of more recent decades. Recently, we published the list of the 255 most-cited papers for that cumulation, but a much larger number would easily qualify as *Citation Classics*.[8]

Even if we do manage to cluster these 10-year cumulations, we do not expect to solicit or obtain commentaries for every one of these earlier papers. The analysis would be important to historians and those who would like to understand better the earlier specialty organization of science. Over 25 years ago, in a special study of the history of the genetic code (from Mendel to Nirenberg), my ISI colleagues Irv Sher, director, Development and Quality Control, and Richard J. Torpie, research associate, and I found that persons with excellent memory—Isaac Asimov, for one—can forget some key contributions to the cumulative process of science.[9]

Our procedure for selecting candidate *Citation Classics* is still heavily dependent upon our files sorted by journal. It gives us a special thrill to report that a particular paper is the "most-cited paper published by this journal." That description need not, of course, remain permanent, although I don't recall having to change it over the years. Maybe it's because old classics never seem to die (they just go on being cited regularly). Perhaps because of the Matthew effect, they don't even fade away.

The Vintage Factor

The data in Figure 1 show the number of papers from each year (1945 to 1987) that we have indexed in the *SCI*. Contrary to intuitive expectations, older classic papers do not have a citation edge on more recent ones. A paper published in 1945 would probably qualify as a *Citation Classic* if it had been cited 100 or more times. With each passing decade, however, that threshold has grown by about 100 citations. This, of course, corresponds to the overall growth of the literature. And, as I have said elsewhere,[10] even a superstar paper like Watson and Crick's on the double helix is seldom explicitly cited more than 1,000 times.

When there are twice as many papers published, the highly cited papers have that much more of a chance to be cited at a higher threshold. Theoretically, there is nothing that prevents a single paper or book from being cited by every new paper in a field. (For many years before *glasnost*, Marx, Lenin, and Stalin apparently achieved that kind of distinction in the USSR.) But in the world of real science, even the Lowry method (the most-cited paper ever[11]) "only" captures 1 percent of the citing papers. The growth of the literature has simply upped the ante for joining the top percentile.

How to Recognize a *Citation Classic*—Vintage or Recent

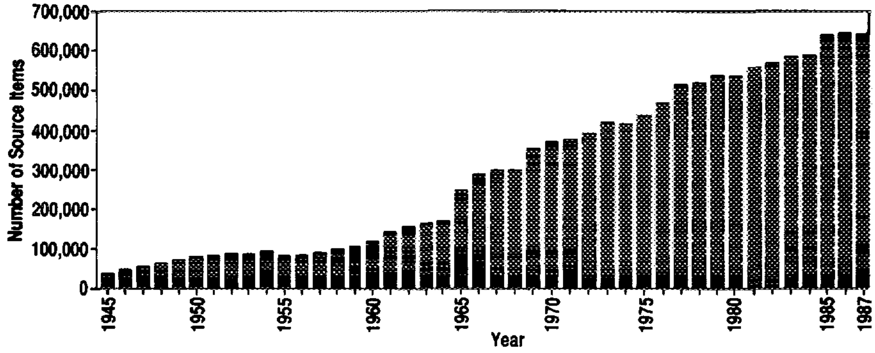
There are times when I find myself wondering about citation behavior in the past. For example, the chemist Sir Alexander Todd made many significant contributions to science, especially in the years prior to World War II.[12] His seminal papers have been cited above the average, but his work is evidently a prime case of "obliteration by incorporation," another Mertonian concept that describes how ideas, methods, or findings become so much a part of currently accepted knowledge that their original source is no longer explicitly acknowledged.[13] To understand the citation dynamics of Todd's work, you must go back to the earliest relevant literature to determine the citation impact of his work at that time.

That is why many of us are eagerly waiting to use the *SCI* for the prewar years. Given the full record of those years, one ought to be able to trace more precisely the impact of Todd's primordial work, year by year, until it was recognized with the Nobel Prize. It would be interesting to learn whether the award increased citations to his work. We could not find such an effect of other Nobel Prizes in more recent years.[14] In the past, science had usually progressed far beyond the frontiers recognized by the Nobel committees, even though the awards were not supposed to lag behind the research too much.

Many years ago, I published a paper suggesting that we might use a critical-path method for determining the lifetime impact of a paper or author.[15] I imagined the time when we would have access to a single integrated file containing data on most published papers, with each one being identified with the papers that cited it.

Starting with any single paper, we could trace through subsequent years to find the papers that had explicitly cited the primordial paper. However, using techniques such as bibliographic coupling as it now exists for

Figure 1: Year-by-year distribution of the number of source items indexed in the *SCI®* from 1945 to 1987.



the *SCI Compact Disc Edition*,[16] we could also trace the impacts of the citing works. In short, on the assumption that obliteration by incorporation is the usual phenomenon, we would thus eventually obtain a better measure of the impact of the original work.

Whether we would turn up many "classics" that somehow do not seem to make minimum citation thresholds seems to me to be an important issue. We know that a good many scientists of Nobel class have written papers that turn out to be *Citation Classics*.[17] But what about scientists of equal stature whose work is *not* highly cited? Are there many such authors? And if so, what circumstances or determinants account for such cases? Did citation ethics break down? Did the advances simply get incorporated by word of mouth or some other informal means? Although science is in general a cumulative process, the literature does not necessarily reflect the complexity of factors involved in the evolution of a field. Indeed, there may be significant gaps in the transmission of knowledge not easily detected by citation linkages. However, providing anything but anecdotal data is difficult.

The conclusion is simple—if you know of such cases, then please let us hear about the uncited classics that somehow did not receive their fair share of the currency that is the reward system of science. Stated another way, is it not a noble goal to identify those to whom we have not adequately paid our intellectual debts?[18]

I've been surprised and disappointed by the very small number of readers who write us to suggest candidates for *Citation Classics*. You would think that every scientist or scholar could name a favorite paper that he or she would personally acknowledge as a crucial contribution to a given field.

It may well be that the unheralded scholars, who have been overlooked by their peers, were indeed lacking in charisma or other factors that determine recognition by citation or by formal awards. But the extant literature is there for all posterity to examine. Barring an exercise in republishing every unrecognized earlier paper to ask whether it was "premature," we can only look to the personal and collective memory of living scientists to make up for flagrant omissions of major contributions.

While we are doting on some potentially overlooked colleagues, let me remind you that we have thousands of well-cited authors, past and present, who have not been recognized formally. In this sense, perhaps José Ortega y Gasset was right; it will never be possible to "elect" all deserving scholar-scientists to the status of Citation Laureate.[19] All such individuals should take comfort in knowing that their work has had some impact. In most cases, it will be difficult to measure precisely what that has been. But the *SCI* at least provides a point of departure for making an informed estimate.

* * * * *

My thanks to C.J. Fiscus, Christopher King, and Peter Pesavento for their help in the preparation of this essay.

REFERENCES

1. **Lewontin R C.** Excessive citations? A case from genetics. Citation Classic. Commentary on *Evolution* 14:458-72, 1960. *Current Contents* (3):16-7, 16 January 1989.
2. **Garfield E.** *Contemporary Classics in the Life Sciences*: an autobiographical feast. *Essays of an information scientist: ghostwriting and other essays.* Philadelphia: ISI Press, 1986. Vol. 8. p. 410-5.
3. **Merton R K.** The Matthew effect in science. *Science* 159:56-63, 1968. (Reprinted in: *The sociology of science.* Chicago, IL: University of Chicago Press, 1973. p. 439-59.)
4. **Zuckerman H.** *Scientific elite.* New York: Free Press, 1977. 335 p.
5. **Garfield E.** The mystery of the transposed journal lists—wherein Bradford's law of scattering is generalized according to Garfield's law of concentration. *Op. cit.*, 1977. Vol. 1. p. 222-3.
6. **Bradford S C.** Sources of information on specific subjects. *Engineering* 137:85-6, 1934.
7. **Garfield E.** The new 1945-1954 *SCI* cumulation provides unique access to the crucial postwar decade of scientific and technological achievement. *Current Contents* (27):3-10, 4 July 1988.
8. -------------, ed. 255 most-cited papers and books, *SCI* 1945-1954. *Science Citation Index ten year cumulation 1945-1954.* Philadelphia: Institute for Scientific Information, 1988. Vol. 8. p. 26-35.
9. **Garfield E, Sher I H & Torpie R J.** *The use of citation data in writing the history of science.* Philadelphia: Institute for Scientific Information, 1964. 86 p.
10. **Garfield E.** The articles most cited in the *SCI* from 1961 to 1982. 7. Another 100 *Citation Classics*: the Watson-Crick double helix has its turn. *Essays of an information scientist: ghostwriting and other essays.* Philadelphia: ISI Press, 1986. Vol. 8. p. 187-96.
11. **Lowry O H.** Citation Classic. Commentary on *J. Biol. Chem.* 193:265, 1951. (Barrett J T, ed.) *Contemporary classics in the life sciences. Vol. 2: the molecules of life.* Philadelphia: ISI Press, 1986. p. 87.
12. **Todd A.** *A time to remember: the autobiography of a chemist.* Cambridge, UK: Cambridge University Press, 1983. 257 p.
13. **Merton R K.** *Social theory and social structure.* New York: Free Press, 1968. p. 27-30; 35-8.
14. **Garfield E.** The 250 most-cited primary authors, 1961-1975. Part II. The correlation between citedness, Nobel Prizes, and academy memberships. *Op. cit.*, 1980. Vol. 3. p. 337-47.
15. -------------. Citation indexes in sociological and historical research. *Amer. Doc.* 14:289-91, 1963.
16. -------------. Announcing the *SCI Compact Disc Edition*: CD-ROM gigabyte storage technology, novel software, and bibliographic coupling make desktop research and discovery a reality. *Current Contents* (22):3-13, 30 May 1988.
17. -------------. Do Nobel Prize winners write Citation Classics? *Essays of an information scientist: towards scientography.* Philadelphia: ISI Press, 1988. Vol. 9. p. 182-7.
18. **Kochen M.** How well do we acknowledge our intellectual debts? *J. Doc.* 43:54-64, 1987.
19. **Ortega y Gasset J.** *The revolt of the masses.* New York: Norton, 1957. p. 110-1.