

Derek Price's Puzzles: Numerical Metaphors for the Operation of Science

Belver C. Griffith
Drexel University

Two of Derek Price's scientometric models are examined—one showing bibliographic references among papers in the self-contained specialty dealing with N-rays, the other showing exponential growth of the scientific literature. The author compares his own observations on the aging of the literature with those of Price, concluding that the rapid aging of scientific knowledge may be a relatively recent phenomenon. Also examined are the effects of the world wars on scientific growth.

The allocation of tasks among the speakers, now writers, gave me the opportunity to focus on Price's work on the measurement and modeling of science. Because so much of the present work in scientometrics and bibliometrics is directly based on his work, I decided not to review his writings or their impact, but instead to be highly selective, perhaps to the point of being eccentric.

My paper examines two of Price's most complex models, both appearing in the 1965 *Science* article, "Networks of Scientific Papers."¹ The choice of these models was made on the basis of several considerations. Each speaks to important features of science and imposes order on a substantial amount of data. Each exhibits a major talent of Price in displaying data graphically. The key feature is that each model displays the data so that visual mechanisms—human perception—do the bulk of data analysis without statistical procedure, or even much need for words. (It surprised me that the Yale economist Edward Tufte had to ferret out nineteenth-century French engineers and their graphs for a book on graphical display of data² when there was such a clever fellow as Price right in New Haven.) Finally, both of these models remain problematic even after nearly a quarter-century; for me, they still pose interesting questions and offer unexplored possibilities.

In *Science since Babylon*, Price has analyzed the history of N-rays and their discoverer Blondlot as a case study of the breakdown of normal mechanisms to avoid self-deception in science.³ That had apparently not satisfied his curiosity about this specialty. Figure 1 shows the pattern of references within the literature of this entirely self-contained specialty, N-rays. The containment was a result of the fact that the phenomenon was an artifact of poor laboratory controls and totally spurious. Accordingly, it began with the paper announc-

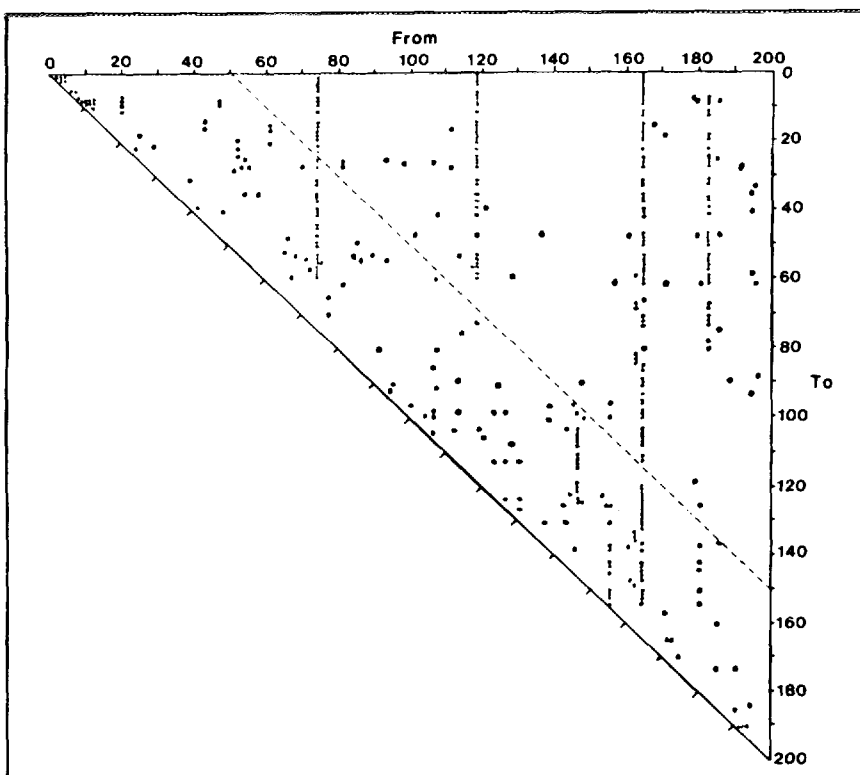
ing the "discovery" and ended with a letter to *Nature* explaining the errors in laboratory technique and a handful of papers that could not be retrieved from the journal publication process. To follow this plot, imagine the 200 papers in the specialty arranged across the top of the figure from the earliest one, at the left, to the last published paper, at the right, and similarly arranged down the right margin, early, top, to late, bottom. A dot represents a later paper, indicated on the numbered scale at the top, citing an earlier paper, indicated at the side. (It will help if you recall that a long vertical line of dots is a review paper.)

There is obviously a tendency for the dots to cluster at the bottom of the triangular plot. Thus most nonreview papers confine their citations to the fifty most recent papers, creating a band of dots that Price labeled as the "research front." Note that this representation of the concept of a research front is a relatively direct measure of scientists' behavior in basing current work on a layer of recent previous literature, going back only fifty papers earlier in time. Other writers have represented the "research front" as "cited document clusters" in document cocitation analysis. But representation of a cluster does not readily permit any analysis, except as to the popularity of the cluster, citations in current literature, or its coverage in the data base—and these are not of special relevance to the understanding of science.

Reviews occur periodically in the N-ray figure (vertical lines of dots) and clearly differ in "re-ency"—the coverage of the latest research findings—and in "comprehensiveness"—the coverage of the entire specialty. Price's graph thus displays measures of the importance of time both in the research process and in scholarly integration. To speculate, imagine a specialty with higher citation rates; Price's plot would develop the faint

AUTHOR'S NOTE: This article was originally presented at the memorial plenary for Derek J. de Solla Price at the annual meeting of the Society for Social Studies of Science, 19 November 1987, Worcester, Massachusetts.

Figure 1. Matrix showing the bibliographic references among 200 papers in the self-contained specialty dealing with N-rays. SOURCE: Redrawn from Price (1965); © 1965 by the AAAS; used by permission.



patterns of a glenurquhart plaid, the horizontal stripes marking high citedness of individual papers; the vertical, high scholarliness.

This one plot does all of the following: distinguishes between reviews and other papers, shows the periodicity with which reviews appear and compares them in recency and comprehensiveness, and defines a useful concept of a "research front" in that nonreview papers principally cite within a narrow band of recently published papers. Two such plots—of different specialties, of course—would permit us to compare them, over time, on all of these measures of process and structure. The problematic aspect of Price's model is thus its unique use, here, in just this one paper. The graph makes powerful distinctions, and it could be the natural complement to cocitation analysis, which has no inherent time dimension and loses continuity in the low overlaps of related clusters in successive yearly files.

Price developed the second model, Figure 2, when Eugene Garfield gave him access to data from the 1961 *Science Citation Index*[®]. It displays a variety of effects so clearly as to warrant inclu-

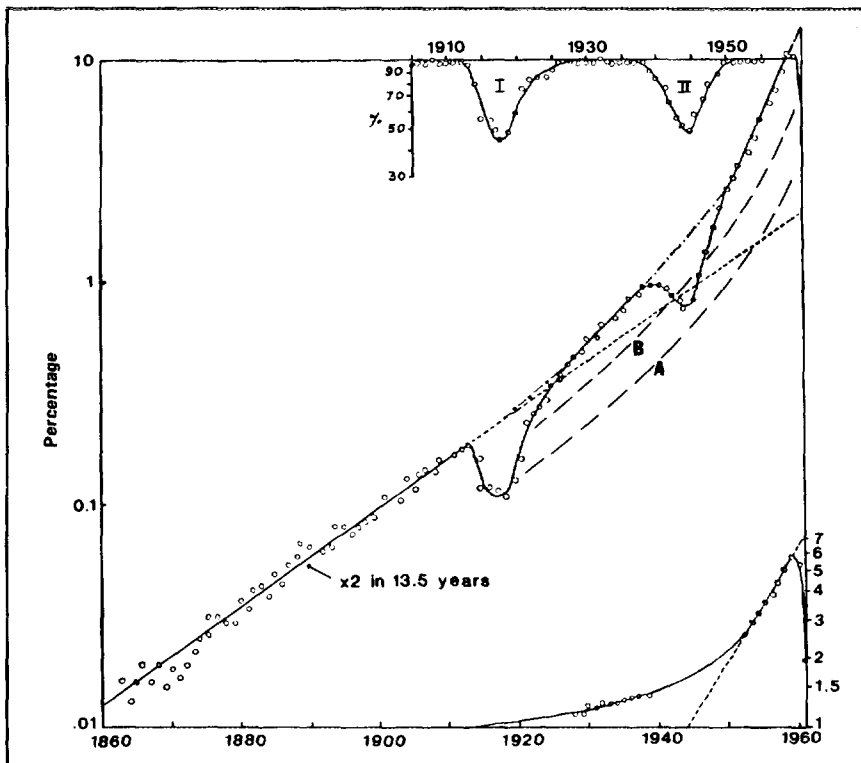
sion by Tufte.² The abscissa is the date in years; the ordinate is the percentage of all citations recorded in the 1961 *Science Citation Index* displayed on a logged interval scale (note marks at 10%, 1%, .1% and the size of intermediate intervals). Each data point is the percentage of citations given to documents with a specific annual date. This technique generates a plot on which exponential growth appears as a straight line.

This figure is the outcome of several effects, which are best explained separately. Data points on the far left indicate the percentage of citations received by older documents; citations increased regularly and exponentially as a function of annual date. Price speculated that this was a reflection of the number of items being published in those years and noted a doubling rate [13.5×2 , or about 5% growth per year] beside that portion of the curve. By Price's reasoning, there would be about twice as many citations to documents dated 1913 as to those dated 1900 because there were about twice as many documents published in 1913. On the basis of his and others' data on literature growth, he accepted this 5% annual

Figure 2. Percentages of citations in 1961 to documents published in the years 1862 through 1961.

NOTE: Base is all citations in the 1961 *Science Citation Index*®.

SOURCE: Redrawn and modified from Price (1965); ©1965 by the AAAS; used by permission.



growth rate as holding generally for the scientific literature and, consequently, for citations to that literature. He made that figure a baseline for viewing other effects in the data; it is shown in Figure 2 as a straight line made with small dashes.

Following the segment of the data that shows exponential growth, there are two major dips in data points and then a rise away from the literature growth baseline. Price displayed dips as deviations (from a best-fitting curve, judged by eye) at the top of the plot. In date, the dips mark the periods of World War I and World War II, starting about the time of the outbreak of major hostilities, rapidly falling to a low, and then rising to lie, again, precisely on the curve some years after the cessation of hostilities. There is marked similarity in timing and amount of greatest decrement in citations of materials published in World War I and World War II. Similarly, the timing and rate of recovery in citation counts are nearly the same for materials published after World War I and those published after World War II.

For the rise of data points far above the baseline, Price coined the term "immediacy." He attributed

this effect to the need of scientists for more recent material and their frequent citation of this material. Note, however, the duration of this departure from the growth baseline; it extends back for nearly fifty years.

To this point, I have followed the spirit of Price's presentation in the original article, adding emphasis to note the marked regularities in the data. Over the past few years, however, I have begun comparisons between his analyses and some of my own, the latter including published and unpublished material mostly on the obsolescence or aging of literature—the lessened citation and use of older literatures. There are some features of these data that appear obvious when they are pointed out but that took me some years (and, in one case, a new dean) to see. These build on one another in a different order from the discussion above, so let me start with "immediacy," first noting how I became concerned with the issues and acknowledging some important assistance.

In 1977, I recognized that the *Journal Citation Reports*®⁴ when combined with the Brookes methodology for studying obsolescence of

literature,⁵ would permit the analysis of changes in the use patterns, over time, in various journals and in the specialties they represent. This led to a journal article⁶ that, as one observer noted, proved definitively (except to several Englishmen) that the scientific literature becomes obsolete. In this study, we worked mostly with citations of documents from only the past decade; the literature in general showed a decrement in use of about 10% per year. An exponential model, with roughly that annual decrement, fit quite well. Some "hot" areas showed decrements up to 40% per year under certain circumstances, and the same journal could be a "newspaper" to its own authors, and quickly obsolete, and an "archive" to authors publishing in other journals, and durable in value. Eugene Garfield, Calvin Lee, and Henry Small helped complete that analysis by furnishing the same data as Figure 2, for items cited in 1974 and 1975. This provided an overall literature control for my data on individual journals. Later I received, with thanks, the data for items cited in 1986, as a control on the earlier data sets.

Recent examination of these long-term data sets shows some interesting features of "immediacy." The reader must recognize a counterintuitive aspect of the analyses, however. In these data, ever increasing use of more recent material is formally equivalent to higher rates of aging. More use of recent material implies, of course, less use of older material. "Immediacy," therefore, can also be described by a model of literature aging, that is, in terms of a decrement in use per year moving backward from the present.

The data show that "immediacy" spans about fifty years from the date of citation backward in time; and the elevation in use moves over time. Thus it starts in the teens (1915?) in Price's data; in the twenties (1927?) in the 1974-75 data; and in the thirties (1939?) in the 1986 data. (Of course, recent annual literatures receive about 350 times as many citations as the literature fifty years previous.) The exponential model for the decrements in the "immediacy" band in time works well but does not fit the very high citations of the most recent documents.

In broad overview, science as a whole looks to a very recent past of about 5 years for roughly half of all citations. (Recall "Price's Index," which he introduced in 1970.⁷) It treats the next 45 years—completing the 50-year "immediacy" period—all the same; citations drop 10% or so a year, with approximately half the decline due to growth. Citations to the literature of earlier years fall more slowly, use dropping 5% a year, with most but not all of the drop due to growth.

Citation rates seem to become unstable for material published before 1895.

Regarding the rate of 5% per year decrement for older literature, there seems to be about 4% growth and 1% aging. For example, documents from much earlier years are cited less in 1986 than in 1975 and less in 1975 than in 1974. For a future paper, I am generating a simple model more or less in line with the present argument. Separately, I repeated one of Price's own analyses and also obtained a similar, very small rate of change in the use of material of this age. In making this argument, however, I have a growing sense of taking unfair advantage: Price, for totally unfathomable reasons, felt strongly on this point that "old" literatures cease to age. He would certainly question my analyses with great fervor were he here.

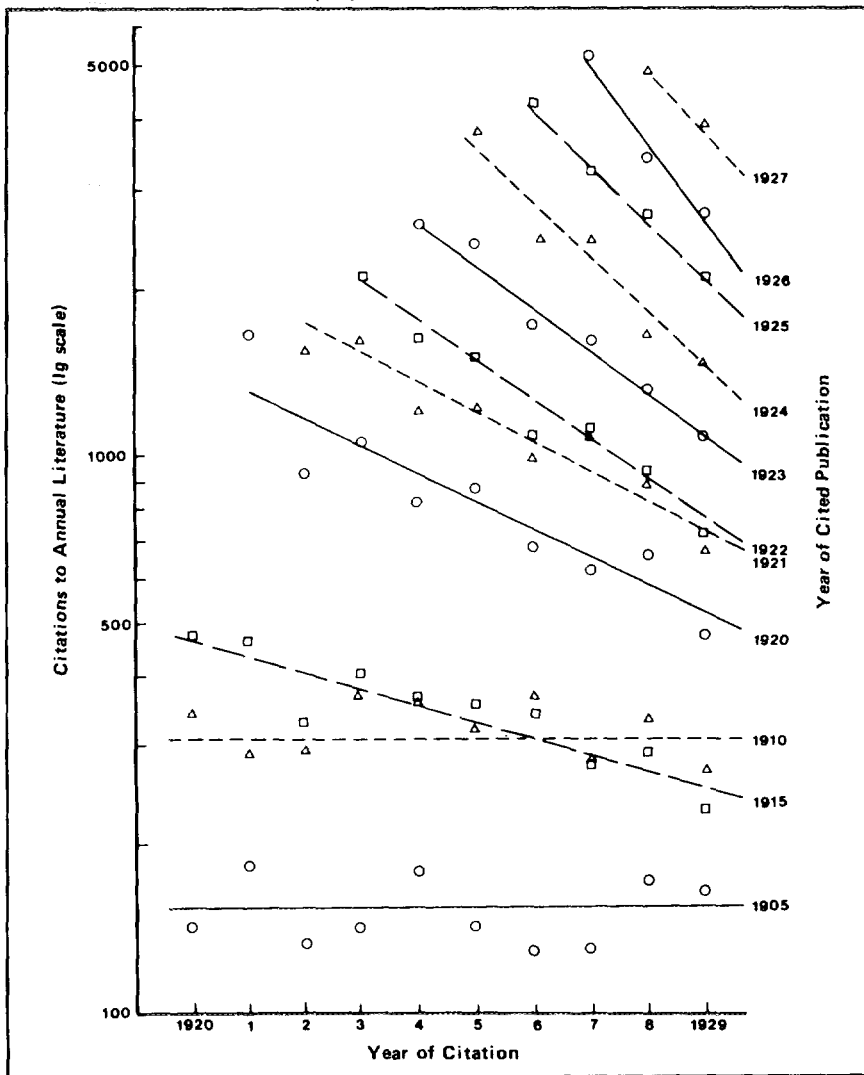
The most problematic aspect of "immediacy" is that it may have been invented by science in the 1920s. Figure 3 from Griffith and White,⁸ an analysis of historic data generated by Small,⁹ sparked this puzzling idea. The data are citations from 1920s physics journals to literature published before and during that period. The ordinate here is numbers of citations received, on a logged interval scale; the abscissa is the year of citation, 1920 through 1929; each data point is the number of citations given in a single year, 1920 through 1929 inclusive, to documents published in a specific year. Straight lines were fit by eye to the citations received by documents published in a single year; readers can judge those fits for themselves. Each line fanning upward in this plot represents a date of publication of the cited documents; the increasing slopes of the lines indicate ever greater rates of aging of the literature as its date of publication moves into and through the 1920s. Changes in the elevation of these lines, independent of slope, reflect popularity and growth in only the cited literature; the citing literature for each column of data points is the physics literature for a single year of the 1920s and is constant. Late in the decade 1920-29, physics appears a very "hot" field, indeed, showing rapid obsolescence of recent literature.

The problematic aspect of these data is that documents published in 1905 and 1910 did not age through the decade 1920-29 (bottom of Figure 3) even though, being at most 25 years old, they would have been well within the "immediacy" band of fifty years, observed for the literature of the 1960s and 1970s. In fact, "immediacy" in this data starts about 1915, just when it appears to start in Price's 1961 data. Thus it appears that this fifty-year elevation in the use of recent documents may have reached its full length only dur-

Figure 3. Number of citations given by the physics literature, 1920 through 1929, to documents published in years before and during the 1920s.

NOTE: The parameter of the lines is the years of publication. For simplicity, only every fifth year is shown for years prior to 1920. (The data points lie too closely together.)

SOURCE: Taken from Griffith and White (1982).



ing the last few decades, and that the rapid aging of scientific knowledge dates only from the 1920s.

A single, recent observation—the contribution of a new dean to my view of this model—caused me to give the dips associated with World War I and World War II new and closer inspection. After hearing me glibly state that the proportional decrement in citations was identical for material published in World War I and World War II, Dean

Lytle, totally new to these data, observed that this near identity between wars in terms of total effect seems impossible to anyone who had read history. How on earth could that happen in view of differences in times and circumstances?

A close examination of the effects of World War I and World War II at the top of Figure 2 will show them to be near perfect models of one another in decline, in greatest decrement, and in re-

covery. This remains puzzling, and speculation on this point becomes highly obtuse. Is science the price of admission to global conflict? It is indeed the case that most scientific nations join in and suffer the decline—and world science suffers the cumulative effect. Do global conflicts share a common overall dynamic? Must they, therefore, create the same patterns of decline and recovery? Even then, how could the overall effect in the two wars be exact models of one another, within a couple of percentage points?

The final and most problematic feature of Figure 2 is the recovery of citation rates to precisely those that would have been expected if no war had occurred. In this regard, it is easier to discuss World War I, in which the countries most affected—France, Great Britain, and Germany—were also the world's leading scientific countries. They lost a very significant portion of an entire generation of scientists; the ground war was mostly fought over highly developed portions of Europe, and there was a modest air war. Laboratories, libraries, and universities were damaged and destroyed.

Let us consider Figure 2 and why its data might be sensitive to events outside of science. *The data reflect the capacity of world science during a specific year to produce papers that are later cited.* War's diversion and destruction of personnel, resources, and facilities previously devoted to science would be expected to reduce that capacity, and did so, judging from the evident drop in citations to literatures published in 1915-19.

What, however, would we expect to occur afterward? Would it have been reasonable for recovery in capacity to have been from about the point of greatest decrement, as, for example, line A? That would, however, assume a total inability to take what science had lost and reallocate it. Such rigidity is unrealistic, and, therefore, line A is probably a poor model for a complex modern society.



Derek John de Solla Price

Perhaps one could imagine some degree of overcompensation, and capacity might rise rapidly because of immediate reassignments of people and goods and resume exponential growth from an intermediate point, like line B. Thus the capacity undergoes a loss, then partial (not full) recovery, before the long-term exponential process reasserts itself—seemingly a reasonable expectation. Instead, the system of science denied *all* tangible losses and returned quickly to the precise levels of capacity that would have existed without World Wars I and II. Somehow, there was no displacement at all in the growth of this capacity.

Although Derek Price never made clear all the sources of his profound private cynicism about the capability of funding to change science, I believe that he had also made this observation himself, although I cannot recall it in either conversation or print. What tangible gains can equal the tangible losses of war? It is still puzzling.

REFERENCES

1. Price D J D. Networks of scientific papers. *Science* 149:510-5. 1965. (Reprinted in: *Little science, big science...and beyond.* New York: Columbia University Press, 1986. p. 103-18.)
2. Tufte E. *The visual display of quantitative information.* Cheshire, CT: Graphics Press, 1983. 197 p.
3. Price D J D. *Science since Babylon.* New Haven, CT: Yale University Press, (1961) 1975. 149 p.
4. Garfield E, ed. *Science Citation Index: Journal Citation Reports.* Philadelphia: Institute for Scientific Information.
5. Brookes B C. Obsolescence of special library periodicals: sampling errors and utility contours. *J. Amer. Soc. Inform. Sci.* 21:320-9, 1970.
6. Griffith B C, Servi P, Anker A & Drott M C. The aging of scientific literature: a citation analysis. *J. Doc.* 35:179-96, 1979.
7. Price D J D. Citation measures of hard science, soft science, technology and nonscience. (Nelson C E & Pollack D K, eds.) *Communication among scientists and engineers.* Lexington, MA: Heath, 1970. p. 3-22. (Reprinted in: *Little science, big science...and beyond.* New York: Columbia University Press, 1986. p. 155-79.)
8. Griffith B C & White H D. Three approaches to citation data: author co-citation, document obsolescence and journal inter-citation. Presentation at the workshop, "Historical applications of citation data: *The Physics Citation Index, 1920-1929.*" 10 December 1982. Philadelphia.
9. Small H G. *The Physics Citation Index, 1920-1929.* Philadelphia: Institute for Scientific Information, 1981.