de Solla Price, Derek J.

Regular Patterns in the Organization of Science

Organon 2, 243-248

1965

Derek J. de Solla Price (United States)

REGULAR PATTERNS IN THE ORGANIZATION OF SCIENCE

Historians of science should be more sensitive than most to the way in which the advance of "hard" scientific understanding has removed many large topics from the realms of "soft" enquiry and speculation. The purpose of this paper is now to report, shorn of mathematical and statistical detail, the results of several investigations which attempt to understand in a scientific way the very behaviour of science itself. I believe that some of these results now appear so fundamental that they answer old questions and suggest new lines of enquiry, not only for historical explanation, but also for the special analyses of modern science that become necessary for planning the policies of governments.

EXPONENTIAL GROWTH AND SATURATION

It has been known for many decades, and indeed several times independently rediscovered as an empirical fact, that measures of the literature and the manpower of science grow exponentially so as to double their size in a characteristic interval of 10—15 years [1]. Though the doubling time varies a little from field to field and from country to country, and there are obvious anomalies attending the birth of a scientific field or country (and perhaps also their death), the behaviour in the large has been spectacularly regular on a world scale and for a time interval that extends from the time of Newton, when the device of the scientific paper was invented, almost to the present day, an in-

[1] For a bibliography of these see my *Science Since Babylon*. New Haven, Connecticut 1961, p. 101, footn. 3. See also: E. Wyndham Hulme, *Statistical Bibliography in Relation to the Growth of Modern Civilization: Two Lectures Delivered in the University of Cambridge in May, 1922*. London 1923; Fremont Rider, *The Scholar and the Future of the Research Library: A Problem and its Solution*. New York 1944.

terval of nearly 300 years corresponding to an increase of size of the order of a little less than 1,000,000.

Both the regularity which seems to transcend quite large social, political and economic changes over the wide range of application, and the magnitude of growth which enormously outpaces any human population explosion, argue for this being more than an empirical generalization. With suitable definitions we can now show that this is a consequence of the ability of old science to breed new. Though scientific papers vary greatly in their fertility, the distribution seems to have remained sensibly constant, so that the average paper produces one new citation per year, for every year after its publication. It takes an average of some 12 citations to make a new paper, hence the production of new papers is at a rate of *ca.* 8% *per annum*. In more detail, it appears that each paper, once published declines in its absolute utility at almost exactly the same rate as that at which the population of papers is growing. Thus its rate of citation will be half the initial one after an interval of 10 years, but by then there will be twice as many papers available to cite it, so the absolute number of citations stays almost constant [2].

The growth by a factor of a million since inception, brings science from being a rare phenomenon to one of the largest activities of the human race. The general explosion of science into our civilization has now reached the point where one is fast approaching saturation of literature, manpower, and money in the most-developed nations of the world. Whatever the nature of the ceilings involved, it would appear that in U.S., U.S.S.R., and in several countries of Europe the growth of manpower and literature in science has since about 1950 fallen progressively more and more short from the projection of the growth rate of the last three centuries. Though the growths are still very large, and the absolute size of science has become so huge as to be a matter of both pride and embarrassment to the nations concerned, it appears that some ceiling phenomena are being noticed, though at the onset of such conditions they were masked by the special circumstances of recovery from World War II.

The most interesting thing about saturated exponential growth is that it can readily be shown that if the process is reasonably normal then the transition time between free and saturated growth is of the order of three times the doubling period of free growth — independent of the size or nature of the ceiling. That is, we must expect that the

---

[2] I have analyzed the frequencies of citations in my paper *Statistical Studies of Networks of Scientific Papers* presented at the Symposium on Statistical Association Methods for Mechanized Documentation, National Bureau of Standards, Washington, March 17, 1964.

interval between the onset of deterioration *ca.* 1950, and virtual breakdown must be about 30—45 years. We are therefore at present one half or one third through the generation which separates the traditional three centuries of growth of science, from some new and yet unspecified way of life in which science is not free, at least in some countries, to continue its general pattern of behavior. Perhaps the big problem of this generation will be that of the "over-developed countries"[3].

It is, on the other hand, plain to see that one may expect the present growth of developing countries to permit science, on the world scale to continue its habitual doubling rate in spite of the stultification of the countries of greatest prowess. Clearly one must suppose that instead of the historical pattern of a shifting of scientific leadership from one great country to another we are now entering a period in which no great country holds anything like an absolute majority of the international capital of science.

NETWORK STRUCTURE OF RESEARCH FRONT SCIENCE

Examining in greater detail the way in which new scientific papers are built upon old ones, we find much that is of central importance to the historian and philosopher of science. Assuming that the references, included traditionally in the bibliography of any published paper, reflect (even if very weakly) the structural interconnections between papers, one may make use of the large corpus of computor handlings involved in the new and very powerful information technique of citation indexing. Statistics from such sources show clearly that two processes are involved. In the first the entire archive of accrued science gives rise to its quota of an average of one citation per paper per year, and this network is, in a sense, randomly distributed over the map of science, only rather rarely knitting together parts not previously well connected. In the second process, there is a strong preference for connections between new papers and those in the rather recent past; the half-life for this process is the order of two or three years.

Furthermore, this citing of literature still at the research front is far from random. Papers are joined together by multiple citation into clumps which exhibit strong internal interconnection within the clump, but much weaker connection from clump to clump. These clumps, which may be generated operationally in an information network, seem to correspond to the work of clusters of authors, about 100 authors (give or take a factor of two) forming a typical cluster which may often be identified with the Invisible College operating in the given field.

---

[3] Further analysis of saturated growth phenomena is given in my *Little Science, Big Science.* New York — London 1963.

The existence of this intimate interconnection at the research front is probably what constitutes the difference between the cumulation of science and the literature growth in non-scientific fields, and one may even arrange in an ordered spectrum the array of all literature-producing fields, set in order from science to non-science, by the proportion of research front structure in their networks of citation. This new operational definition seems rather more useful and provocative of fresh thought than the traditional use of such internationally misunderstood terms as *science*, *Wissenschaft*, and *nauka*. It is perhaps worth remarking in this respect that a great deal of the literature usually called "technological" seems to be of a non-scientific nature according to this definition. I have shown elsewhere that this phenomenon throws much light on the nature of technology, its relation to science, and its historiography [4].

INFANT MORTALITY AND UNEQUAL DISTRIBUTION OF SCIENCE

In order to generalize about empirical statistics by the use of theoretical concepts it is necessary first to examine the relations between scientific manpower and papers, and between papers and the journals in which they are published. In the course of this investigation it has now been found that several unexpected but simple regularities run through all data concerned with the distributions of such things as degrees of merit, usefulness, productivity and size.

In brief, the number of papers giving rise to n citations per year, the number of institutions producing n new doctorates a year, the number of authors publishing n papers per lifetime, and the number of journals containing n papers per year, all behave similarly. They follow the same type of distribution as that of millionaires and peasants in a highly capitalistic economy. A large share of wealth is in the hands of a very small number of extremely wealthy individuals, and a small residual share in the hands of the large number of minimal producers. Whether the exact form of the distribution is lognormal, exponential, a Zipf Law, or an inverse square has been a matter of much conjecture in each of the cases. What we now know is that all these laws are reasonable approximations to each other in the ranges studied, and that the gross phenomena are sensibly the same in all cases.

We know also that each of the distributions has been influenced by a far-reaching effect whose existence, though obvious once stated, has

---

[4] This is treated more fully in my paper *Is Technology Historically Independent of Science?* presented at the Symposium on the Historical Relations of Science and Technology, American Association for the Advancement of Science, Montreal, December, 1964, and published in "Technology and Culture", N. 3/1965.

not previously been recognized in cases of this sort. We refer to a process which may be the converse of the Matthew Effect[5]: "Unto him who hath hardly at all, is taken away completely". As instances of this, one notes that most journals that have published only one or two issues do not publish again, and that most authors who have published only one or two papers do not continue. Because of this law, science suffers from an enormous infant mortality. The growth rate of, say 7% *per annum* is due to a combination of a birth rate of perhaps 17% and a death rate of 10% each year. Over a short interval of time all journals publishing at all have only one issue, and all authors only one paper (to a first approximation), but over longer periods the great majority of journals and of authors come up again only rarely, and the bulk of the material published comes from a hard core which is a very profilic but small fraction of the total population. Typically about half of the published literature is produced by about the square root of the total number of journals or of authors, and for every ten per cent or so addition to this core one has to take in twice the number of journals or people[6].

Thus the world of scientific manpower and literature consists of a small hard core surrounded by an almost infinite population whose numbers increase exponentially as the magnitude and permanence of their contributions decrease at similar rate. Even if there is by no means a perfect correlation between the number of papers written by an author and the importance of his work, or between the significance of a journal and the amount it prints, any effect of this sort loads the scale still further. In short if there are for example 30,000 journals alive in the world, or 1,000,000 publishing scientists in a country then a selected list of about 175 journals or 1000 scientists would account for half of the bulk of the literature but probably for 70—80% of the important content.

A WEBER-FECHNER LAW OF SCIENTIFIC ACTIVITY?

In all the phenomena that have been analyzed so far, both in time series and in frequency distributions, it appears that the first and most

[5] Robert K. Merton, *The Matthew Effect and Visibility of Scientific Communication.* A paper presented at the Symposium on the Sociology and Ethics of Science, American Association for the Advancement of Science, Montreal, December, 1964.

[6] Material from unpublished but circulated papers: Donald de B. Beaver, *A Statistical Study of Scientific and Technical Journals*; John P. Britton, *The Productivity of Scientists: A Prelude to Manpower Studies.* Department of the History of Science and Medicine, Yale University, New Haven, Connecticut, November, 1964.

obvious approximations involve exponential functions in very simple ways, mostly in linear equations. Because of this we now conjecture that many of the measurable quantities used in our analyses in the form of simple head-counts of people, papers, and journals are rather analogous to the role of stimulus in the analysis of sensation in experimental psychology.

For example, if we are concerned with the productivity of authors, then it appears about the same step of "difficulty" for an author to take a transition from his first paper to his second, as from the second to the fourth, the 10th to the 20th, or the 100th to the 200th. Constant increase of difficulty corresponds to constant proportional increase in the number; alternatively it is easy to derive, the total difficulty is measured by the logarithm of the number of papers. Similar distributions yield the suppositions that the logarithms of the total number of authors or the total number of publishing journals give a measure of the stature or extent of the hard core literature in a field or in a country.

Finally, it can be seen that the normal exponential rate of growth corresponds simply to the linear growth of the logarithm of any measure of crude size. Thus, the obvious mathematization of the Matthew Law: "Unto him that hath is given" [7], yields an analogue of the Weber--Fechner Law which is at the foundation of experimental psychology, and perhaps from this flow a set of consequences and traditional controversies which may be just as fruitful for the mathematical analysis of science.

[7] Compare footnote 5.