The Proliferation of Scientific Literature: A Natural Process

J. M. Ziman

Summary. Primary scientific literature seems not to be growing at a greater rate than the scientific community it serves. The impression of excessive proliferation arises mainly from the differentiation of journals to accommodate rapid expansion in specialized fields of research. A large fraction of this literature is of marginal value, but should not be excluded from comprehensive archives for possible retrieval. For awareness of significant current developments, however, scientists depend on a small number of core journals whose quality is maintained by editorial selectivity and competition.

The impression that scientific literature proliferates unnecessarily, giving rise to severe problems for the individual scientist seeking information, is sufficiently widespread to deserve careful analysis.

Expansion and Specialization

The rate of growth of the literature of science is not known precisely; the rate of increase of serials seems to have been about 4 percent per annum through the late 1960's (*1*, p. 16). In recent years this seems to have slowed to a net figure (that is, one which allows for the discontinuation of many titles each year) of somewhere between 2 and 3 percent (*1*, p. 17). But a serial title is not a standardized unit, since the number of pages in an annual volume varies from journal to journal and may change appreciably from year to year. More detailed studies (*1*, p. 87) indicate that a collection of typical serials that grew in number from 1000 in 1960 to about 1700 in 1974 would have grown in shelf length from 67 to 124 meters in the same period. Thus we may surmise that scientific literature doubled in bulk in about 15 years during its major phase of expansion and that this expansion continued at a somewhat lower rate during the 1970's.

This increasing annual crop of scientific papers is produced and consumed by an increasing number of research scientists. The world figures on this total are almost impossible to assess with any accuracy, but they are consistent with the same maximum doubling time of about 15 years during the peak of scientific expansion, followed by a period of much slower growth in the past decade. Thus the growth of scientific literature has more or less paralleled that of the scientific community. Evidence for any significant change in the average number of pages per article is scanty, but it was estimated that the number of journal articles published per scientist or engineer in the United States fell by no more than 20 percent between 1960 and 1974 (*2*). This average is not a very precise statistical indicator of scientific productivity because of the very large variation from individual to individual, but it evidently has not changed drastically in recent years. Therefore, compared with earlier periods, there is no evidence of excessive proliferation in the amount of scientific literature produced by each scientist, or in the number of journals. It has been asserted that some individuals seek to publicize their work unduly by having it published in nearly the same form in several different journals. If this practice is common, it has not been proved so by quantitative studies.

Nevertheless, faced with an enormous, ever-growing body of literature in every discipline, the research scientist is bound to restrict his reading to an increasingly limited, specialized range of topics. Assuming that the amount that can be assimilated by an individual remains constant from year to year, it follows that this amount shrinks in inverse proportion to the whole body of literature. In other words, a scientist who could keep up with x categories of a particular field in 1960 might not manage to cover more than $x/2$ categories in 1975.

This argument assumes, however, that every branch of science is growing at the same rate. This is very far from the truth. Scientific fields are always changing in weight and significance. Old topics fade away, and new ones are born. The literature of an established field may take 50 years to double; a new field may expand almost explosively, doubling every 4 or 5 years for several decades (*3*). For a scientist in a slowly growing field, the pressure toward narrower specialization can be easily accommodated; in a fast-growing field with a literature growing by a factor of as much as 10 in 15 years, the effect of proliferation can be almost intolerable. In such a field the x categories in 1960 must now be shrunk to only $x/10$ categories in 1975, with grievous consequences in terms of fragmentation of knowledge. This has certainly been the experience of many senior scientists, giving serious cause for personal and collective concern.

Proliferation of the literature is not, however, necessarily a sign of ill health in science: it may be a natural consequence of scientific progress. This puts severe strains on the scientific information management system, which must be reasonably matched to the needs of the individual scientist who is both author and reader of scientific papers. One response of the system is further differentiation and subspecialization of journals. A journal, in order to accommodate a rapidly growing number of publishable papers within its subject range, may be divided into more specialized sections, which eventually speciate into effectively independent publications. But it is often difficult for this process to proceed at a rate sufficient to meet the demand for publication space, or for the scope of long-established journals to be enlarged to cover a gap that has developed between acknowledged disciplinary areas. This provides an opening for the funding of new journals—often by commercial publishers. Such journals (often very highly specialized or with novel interdisciplinary titles) may offer new outlets for numerous eager authors in the newly defined field of research.

Assuming that any one scientist can only buy and read a certain quantity of literature, it is obviously advantageous to have this literature packaged for separate purchase or perusal in appropriately specialized journals or journal sections. But this traditional process of evolution by speciation cannot meet all the needs of a subject whose literature is doubling every 5 years. It is not desirable for a sci-

J. M. Ziman is professor of physics and director of the H. H. Wills Physics Laboratory at the University of Bristol, Bristol, England BS8 1TL.

0036-8075/80/0425-0369$00.75/0

entist in such a field to be forced to narrow his range of interest so rapidly.

It is not possible, for example, to predict the divisions that will occur among the existing x categories. As experience with library classification schemes has shown, categories may be formed in several different ways. For example, scientists in an expanding field may tend to group themselves (i) according to the different classes of materials or organisms that they are investigating; (ii) according to different instrumental techniques, such as electron microscopy or nuclear magnetic resonance; (iii) according to different conceptual schemes (gauge theory, S-matrix theory, quantum chromodynamics); or (iv) in relation to the solution of particular practical problems such as the treatment of a disease.

But such groupings are somewhat tentative. It is only as a specialty matures into a stable discipline—that is, as it ceases to grow very rapidly—that the scheme by which it has been categorized can be trusted as a basis for a well-differentiated literature that is matched to the information needs of individuals. In the proliferation phase it is essential for the research worker to keep an eye on the whole range of tentative categories, seeking instrumental, theoretical, or technological information that is relevant to his own work. At the same time, he is concerned about the visibility of his work—perhaps this explains the allegedly growing tendency for a report to be published in several different journals, affectively addressed to quite different categories of readers.

Retrieval from the Archives

A major problem in dealing with the literature of an expanding field is information retrieval. Given the archival material in which the search is to be made, the task is reasonably well defined. Despite very considerable intellectual and practical efforts, the retrieval problem is far from solved, but new techniques to deal with it are being developed. We may say, indeed, that the growth of the scientific and technical literature in recent years has not been as rapid as the development of new and economical means for classifying and retrieving it in detail. Despite the apparently relentless growth in the bulk of scientific information over the past 10 or 20 years, the information contained by an archive today is probably more complete, better organized, and more easily accessible than formerly.

The research scientist making a search, however, is often distressed by the very large quantity of marginal literature that is stored in the archives and automatically retrieved for its apparent relevance (*4*). Such literature is characterized by the triviality of the results, the incompetence of the experimental technique, or the purely technological or commercial context in which it is reported.

It is often suggested that archives be drastically screened or pruned to eliminate storage of this sort of material. Unfortunately, such a process would be self-defeating if applied to the literature of science as a whole. It is precisely the hope of finding a previously unrecognized nugget in the dross that motivates the search. It is precisely because the value or relevance of such a scrap of information had not been realized that it was hidden away in an unlikely place.

Of course, it is imperative that every reputable scientific journal seek to maintain the highest standards of quality and relevance in the articles it publishes. It is also important that authors, editors, and referees not allow potentially valuable material to lie concealed in inappropriate publications. But what we call the scientific literature is open to wide interpretation. There is not—and should not be—any formal procedure by which the scientific community puts a seal of approval on a particular published work, as meeting specified standards of veracity and therefore being worthy of citation. It is simply a matter of practical judgment whether or not to store and catalog a great deal of other material in order to meet the diverse needs of authors and readers (*5*). This applies, for example, to many technical publications, whose primary purpose is to diffuse new technological developments but which may also carry significant original scientific information.

For many practical purposes a great deal of marginal literature could be excluded from archives without serious loss. The list of publications collected in a library or abstracted by a secondary service is always cut off arbitrarily at some level of quality or relevance. There is no great difficulty in raising this level to exclude material of doubtful value. For a local archive, such as a departmental library, this may be highly desirable for both economic and cognitive reasons.

But for a comprehensive archive, such as a national library or international abstract journal, this process cannot be carried very far without seriously compromising the value of the collection (*5*). The gain in average quality and relevance may soon be outweighed by the loss in comprehensiveness of coverage. A comprehensive archive in an active field of research may contain several thousand articles. Careful selection might well halve this number, but it would not be reduced to a level where formal retrieval techniques would not be necessary (*6*). A subject that has proliferated by a factor of 10 over a few years cannot be pruned by such a large factor without losing many potentially significant items of information.

Current Awareness

But the difficulty of dealing with the proliferating literature of an expanding field is not just the "needle in a haystack" problem of retrieving a few relevant items from a large mass of essentially irrelevant material. It is also the problem of maintaining current awareness over an expanding field. The individual scientist needs easy access to the obviously significant papers in a range of topics that exceeds his immediate research interest—to the various peripheral categories that might still be relevant to his specialty. Neither the speciation of journals nor formal archival searching provides the scientist with what he really wants: to see most of the very best articles over a fairly wide subject range as soon as they are published.

This is a severe demand, which cannot be met in full. It is not enough to exclude the weakest items in the marginal literature; it is necessary to select articles of high quality in manageable numbers. This can be achieved, in practice, by concentration on the core journals in each scientific field or subfield. It is now well established that articles are distributed very unevenly among the apparently relevant journals (*6*). A careful analysis of a large collection by quality, quantity, and relevance showed a minimal nucleus of no more than a half-dozen journals in which the most useful articles were concentrated (*7*). Regular perusal of a well-chosen set of such journals is probably the most effective way for a scientist to maintain a reasonable level of current awareness in a rapidly proliferating field of research.

It is true that use of different criteria for ranking journals in a particular field may result in somewhat different lists (*8*). In any case, such a small selection of core journals would be grossly inadequate as an "archive" for a major discipline, for which there would probably be several hundred relevant serials (*9*). But modern methods of selective retrieval and dissemination of information have not yet adequately replaced the traditional practice of "browsing" in the

quality journals for interesting new research.

Some of the labor of keeping up with a subject can be carried out by reading review articles and other secondary and tertiary material. This is also the function of conferences, summer schools, seminars, and other forms of unwritten communication. These means of information transfer should be continually improved and enlarged, but they are no substitute for the primary literature.

Quality Control and Competition

The wide range in the quality and relevance of scientific literature is not entirely dysfunctional. The literature can be sifted for items of current interest and significance by the conventional selection mechanisms of scientific information management systems. Good journals can thus be selected for regular perusal because they contain a high proportion of useful articles; more marginal material can be stored in comprehensive archives.

The status of a journal as a core source can be preserved only if it sets high standards of acceptability and maintains them by conscientious editing and refereeing. There is evidence that the majority of papers rejected by the core scientific journals are eventually published elsewhere (*10*). But this is no reason for relaxing the efforts of editorial boards and referees to achieve journals of the highest scientific quality. Authors choose journals by using the same criteria as their readers—quality and relevance. It is advantageous to an author to have his paper published in a recognized core journal because that is where it is likely to reach the most appropriate audience. The leading journals are as important to the progress of science as the leading scientists.

The literature of a field that is rapidly expanding in manpower and research facilities is bound to proliferate into a corresponding quantity of new journals and papers. It is very desirable, for both authors and readers, that these be of high quality. But this cannot be achieved by administrative decision. Core status has to be won in a long-drawn-out competition between various national and international journals and, even when no longer deserved, it decays slowly.

Scientific publishing is a risky venture. A new journal introduced to cover a newly differentiated specialty does not necessarily draw the best papers. It may prove a rapid success; or it may have to exist for a long time on articles that have been rejected by the established journals because they are of marginal quality in an overcrowded field. The viability and visibility of a new specialty or interdisciplinary grouping may seem doubtful for some years—even to those who are actively pursuing research within that category—so that many leading scientists may be unwilling to set the tone of a new journal by publishing in it as soon as it is founded.

On the other hand, learned societies, perhaps because of the subjects on which they are based, have sometimes been slow to expand and differentiate existing journals to provide outlets for new topics. The scientific community has often benefited from a commercial publisher who responded to this need by establishing a new journal, which in some cases became a core journal in the field. Even though not all such ventures are successful, the outcome does not seem to depend on whether the initiative came from the publisher or from the scientific side. Although the proliferation of poor journals is often blamed on irresponsible commercial publishers, there is no definite evidence to support this assertion.

Scientists who are involved in the publication of a new specialized journal must try to decide whether they are providing a medium for the communication of important new scientific advances or whether they are merely relieving the pressure on the core primary journals by offering more space in which relatively trivial papers can be harmlessly filed away. Given the apparent need for such outlets and the commercial facilities available, the marginal publication activity at the "free surface" of the scientific publishing world cannot be suppressed. But no scientist of any standing should take pride in having his name on a board of editorial advisers of such a publication, heedless of the quality of the work being published under that heading. Good research validates itself.

Summary

Many leading scientists feel that there has been a serious deterioration in the quality of the scientific literature in recent years. If this is true, is it to be attributed to the fragmentation of primary publications into many specialized journals of low quality? Is the presumed proliferation taking place only among the commercial journals, or is it also visible in the scientific society press? How might this trend affect the attitude of the societies, and their international unions, toward the commercial journals?

These questions were the seeds of this article. Unfortunately, none of them can yet be answered by reference to reliable evidence. It is extraordinarily difficult to confirm a subjective impression that there has been a significant decline in the quality of research publications over several decades. It is largely a matter of opinion whether any particular new journal was conceived as a result of publishing "push" or scientific "pull." The relationship between commercial publishing and the scientific community is so close and symbiotic that it is not easy to imagine a change of attitude that would be efficacious against this particular ill without also causing damage in other respects.

It may be wiser at this stage to recognize the deeper dimensions of the problem. The speciation of scientific journals is not a pathological sign; it is a natural accompaniment of the dynamic change and growth of scientific knowledge. Its ill effects can be mitigated by improved techniques and facilities for archival search and retrieval and by individual strategies of concentration on the core literature. Rapidly developing research areas not only pose the most problems for communication, they also provide the opportunity for new ways of solving these problems.

But short of a worldwide, learned-society monopoly over scientific publishing, with a stringent system of licensing papers for publication and citation, the phenomenon of proliferation cannot be eliminated by controlling the pressure to publish a paper nearer to its sources. Within the context of an open scientific literature, not constrained at the margins by expert peer review, there is no substitute for competitive standards of quality and adequate attention to the intellectual tasks of classification and indexing.

References and Notes

1. C. B. Wootton, "Trends in size, growth and cost of the literature since 1955," *BL R & D Report No. 5323HC* (1977).
2. D. W. King, *Statistical Indicators of Scientific and Technical Communication, 1960–1980* (King Research, Rockville, Md., 1976), p. 47.
3. H. W. Menard, *Science: Growth and Change* (Harvard Univ. Press, Cambridge, Mass., 1971), p. 52.
4. This problem is discussed from various points of view in K. S. Warren, Ed., *Coping with the Biomedical Literature Explosion: A Qualitative Approach* (Rockefeller Foundation, New York, 1978).
5. M. M. Cummings, in (*4*), p. 61.
6. K. S. Warren and W. Goffman, *Am. J. Med. Sci.* **26**, 267 (1972).
7. W. Goffman and K. S. Warren, *Nature (London)* **221**, 1205 (1969).
8. A. Singleton, *J. Doc.* **32**, 258 (1976).
9. J. Lloyd and P. Fenner, *Geotimes* **15** (No. 6), (1970).
10. A. S. Relman, in (*4*), p. 54.
11. This article was originally prepared as a report for the Policy Group on Scientific Information of the International Council of Scientific Unions and was approved for publication on behalf of that group. I am grateful to J. Lloyd, J. Meadows, and J. Martyn for their most helpful comments on an earlier draft.