Reading Behavior of Chemists*

DAVID E. GUSHEE
American Chemical Society, 1155 Sixteenth Street, N.W., Washington, D.C. 20036

Received July 18, 1968

Most discussions about the future design of chemical information systems emphasize the effects of computer speed on composition, storage, searching, and retrieval. Little attention has been paid to the behavior of the scientist (chemist) toward the learning process and the role of the literature in that learning process. The ACS has gained some insight into this behavior and thus urges caution before one leaps to the uncritical conclusion that the literature as now constituted is dead but has not yet fallen over. What seems more likely, when human behavior patterns are considered, is that the new and the old will co-exist side by side for several generations, and that the system's current emphasis on completeness will be replaced with selective compression.

It was here in Paris that the first recognized scientific journal, the *Journal des Savants*, appeared, in January 1665. This was followed two months later by the *Philosophical Transactions of the Royal Society* in London. All subsequent journals carrying reports of original research in the physical and social sciences continue in approximately the same format.

In the 303 years and 4 months of journal history, mankind has made almost no progress in effectiveness of communication of new scientific information. The scientists (and I include engineers and technologists in this term) still write their own reports of what they have done. In most reputable journals, these reports are evaluated by peers of the authors—that is, by other scientists who are (hopefully) as knowledgeable in the field of research as the author himself is. The peers seldom require competent linguistics in the author's writings. Even if they did, the editors—themselves scientists—would not be able to require the authors to meet significant linguistic standards.

When you speak to a scientist about the traditions of this approach to the publication of research, you hear that this approach is best. No one but the researcher, the scientists say, can present the research factually, clearly, and unambiguously. You hear about the objectivity of the mechanism and about the effective spreading of knowledge about progress with which a reader can trace the increase of understanding through time by using the reports of research as prepared by those who did the research.

There is a great deal of truth to this position. On the other hand, there are many things that are unsaid. First, it is quite clear that the literature is published as much for the authors as it is for the users of the information published. Authors prepare their own reports of their own research to establish priority, to make a name for themselves among their peers, to establish a position of prestige among their colleagues in other branches of scholarly work, and to leave a permanent tracing in history of the value of their existence.

The system over the three centuries of its organized existence has been policed by the integrity of the authors, the unwritten standards of the profession, the willingness of peers to examine objectively the authors' work prior to its publication (the referee system), and the requirement that a reader of the work be able to reproduce any experimental results reported.

But now, new variables in the scientific world cause one to wonder about the future viability of this traditional mechanism. In 1967, approximately 240,000 articles and patents were considered by Chemical Abstracts Service to deal with new knowledge in chemistry (Table I). At the current rate of increase of chemical knowledge, this number will double in about seven years, marking a catastrophic problem for any chemist or chemical engineer who wants to remain competitive in his awareness of the latest information that affects the success of his developmental endeavors (Figure 1).

The American Chemical Society, which produces *Chemical Abstracts*, also publishes in its primary journals about 4% of this new chemical knowledge. The Society spends close to 10 million dollars each year to publish this vast amount of information. As you can imagine,

Table 1. Abstracts Published in CA

Year	Papers	Patents	Total	Total To Date
1907–16	133,115	45,783	178,898	178,898
1917–26	162,145	47,741	209,886	388,784
1927–36	343,714	186,455	530,169	918,953
1937–46	355,784	142,693	498,477	1,417,430
1947–56	543,064	104,247	647,311	2,064,741
1957–66	1,291,600	262,811	1,554,411	3,619,152
1967	202,684	36,797	239,481	3,858,633

* Presented before the Société de Chimie Industrielle, 50th International Congress, Paris, France, May 30, 1968.

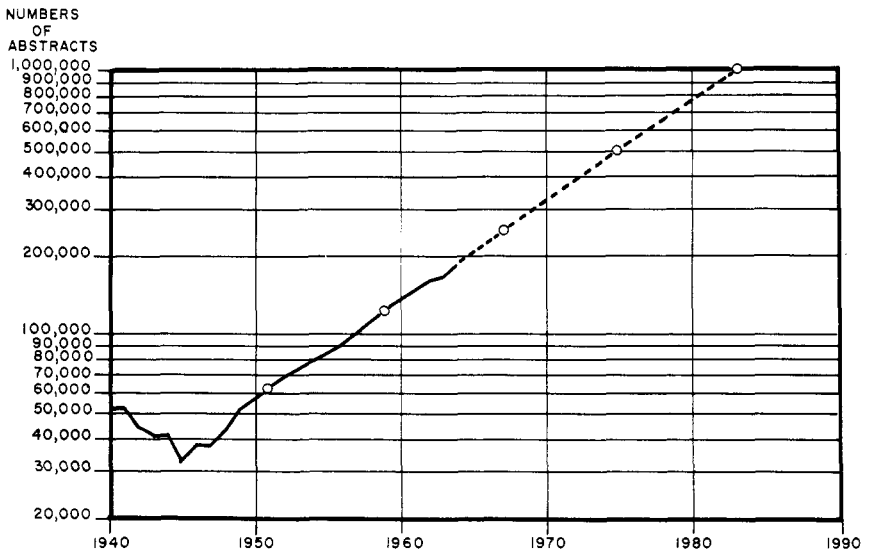

Figure 1. Growth of the chemical literature

a major problem is whether or not the consumers of this information are willing to pay us the 10 million dollars it costs us to produce it. This is merely another way of asking whether the three-century-old system still works.

Although the chemical profession is not accustomed to considering its fundamental literature in the same context as that of any marketing operation, it certainly has elements that are similar to those of producers of products to be sold. Up until recently, however, it has not been necessary for the caretakers of this operation to be very concerned, since the literature has in essence operated itself. With the vast and increasing amount of material to be published, the increasing costs of publishing, and the increasing demand on chemists for their time and money, this operation is no longer automatically self-regulating and self-sustaining. Rather, we find ourselves, even as publishers of only 4% of the world's chemical literature, in competition with other means of informing our audience and with the ultimate possibility that the audience would prefer not to be informed rather than to take the trouble to plow through the material which the profession as now constituted produces.

It is in this context that we ask ourselves the question: Is the primary literature of chemistry necessary? Surprisingly enough, there are no definitive studies that document the necessity of the primary literature being published in its present form. We all accept as an article of faith that the system as now constituted is the best system that could be developed. The scientific method, for example, is said to be based on this literature system, being defined as collecting information, forming a hypothesis, testing the hypothesis, and using the hypothesis in further observations or re-examination of other observations. The key point is that to collect the information in the first place, the literature must be consulted.

Another way of putting it is presented by science historian Charles Singer, University of London, who says: "The history of science shows that only those with a knowledge of how their predecessors have succeeded or failed have chosen wisely the phenomena to be observed and recorded."[2] Another observer-commentator on the scientific scene, Waldemar Kaempffert, then science editor of *The New York Times*, said, "Modern inventions as a whole are combinations of well-known elements that produce new results and serve new and useful purposes. Without a heritage, technology could not advance."[1]

Now you must admit that that is not a particularly profound documentation of the significance of the fundamental literature. When I accepted the invitation to speak to you today, I asked the *Encyclopaedia Britannica* to survey all the information in the sociological literature on the subject, "The importance of the chemical literature to chemical technology." It took them 13 weeks to make this literature search. What they returned was a great surprise to me, because it showed no references from the social sciences at all. The only documentation of the value or validity of the scientific literature has been made by scientists themselves. For the most part, therefore, we have no externally-supported proof that the literature as designed is the best way to collect, store, and transfer scientific observations from those who make them to those who use them. Further, of all the scientists who have examined their own literature, the basic references are primarily from the chemical world, although recently the physicists and psychologists have been doing some reasonable research into their own literature needs.

I do not want to dwell on the literature *per se*, but I think it is important to point out that the literature is driven by the needs of authors and not by readers, a point which I think has now been adequately demonstrated. A major problem is that the chemical profession itself has not consciously recognized this fact, so that attempts to add to the standards of the scientists some of the insights of information specialists, sociologists, and others usually meet with loud screams of anguish. Journal editors, recognizing that their success depends upon the good will of authors, are unable to police the literature by any standards other than those of those very same authors. The result is this explosive growth to which I referred in my opening comments. I think we all know intuitively that the idea content in the 240,000 abstracts of 1967 is not twice as great as the idea content in the

120,000 abstracts of 1960. Duplication, fragmentation, trivia, and just plain bad writing are the source of much of the growth of the literature in the past 30 years.

These are harsh words. However, communicating is an art separate and distinct from chemistry. Those attributes which make a good chemist do not necessarily make a good communicator. For this reason, the American Chemical Society has invested a good deal of thought, time, and effort in the past 15 years to systematize its understanding of the communication of ideas to chemists and chemical engineers. I doubt that we have found anything particularly new in terms of information science, linguistics, semantics, or anything of that sort. What we have found supports the general comment of other observers of mankind that chemists are people and respond like all people, with a few specific variations that are easy to identify.

Chemists are bound by tradition and, except for a few, resist changes in their practices. Further, they have a very unrealistic view of their own behavior, believing it is far more rational and objective than it really is. Chemists are not all equally motivated, for example, to search diligently for truth no matter how it is presented. They vary all the way from intensely oriented toward the literature to intensely oriented away from literature. They respond to journalistic stimuli in much the same manner as the public at large does. They reject in large measure dull writing, passive sentence construction, impersonal presentation of facts, in fact, most of the ideals of scientific and technical writing which we were taught to uphold and which we see reflected in our journals today.

We also corroborate the findings of sociologists and anthropologists that the rate at which a new idea penetrates any population varies with the effectiveness of dissemination of the idea. That is to say, within chemistry there are a few people who are always on the alert for a new fact or idea of significance. These few then form the centers from which the idea and information about it are disseminated to the next layer of chemists, and the next layer, and the next layer until eventually the new idea becomes a part of the standard knowledge of the standard chemist. The rate and extent of this process of dissemination of an idea vary with effectiveness of the communications devices used, despite the scientists' feeling that the idea will find its way into the minds of the population regardless of the techniques used, merely because it is a good idea.

Let me illustrate with a few examples. We have found in *Chemical and Engineering News*, for example, that the number of chemists who read a given issue can vary by as much as 25% from week to week. In general, although there is some random variation, the biggest reason for the differences in readership is the effectiveness of the design of the covers. A corollary of this is that an article in an issue with a "good" cover has 25% more chance of being read than an article in an issue with an ineffective cover design. The variables, incidently, are subject matter, composition, color, and relevance to the personal emotions and experiences of the individual chemists.

The cover is, however, by no means the only variable. We have found, for instance, that organic chemists read their literature in general with a keen eye for key words in titles, for specific authors of high regard, and for the significance to them of whatever structures or diagrams are available for scanning. We have yet to find a research article that no one has read. We also have yet to find a research article that everyone has read. And in general we find that somewhere around 10 to 15% of the subscribers of a journal read a significant part of any article, even in a journal with a hundred or more articles in it. For *Journal of Organic Chemistry*, this means that some 650–1000 subscribing chemists will have given some significant effort to any given article, plus some unknowable number of chemists who read but do not subscribe to the journal.

We have also found that when you ask a chemist how he reads a journal, he will frequently tell you that he goes to the table of contents, finds the articles he wishes to read, and then goes to those articles and reads them. There is no reason in principle to doubt this statement of behavior. However, when we survey chemists and ask them to tell us by looking at the table of contents which articles they have read, they are unable to do so. Therefore, we conclude that titles and authors are insufficient triggering mechanisms for recall. This may explain why *Chemical Titles*, although conceived on what appears to be sound principle, has never really become a widely used information retrieval tool.

We have also found that chemists tend to specialize their reading more than engineers or executives do. That is to say, chemists tend to pick out articles to read by looking at the table of contents and proceeding directly to the articles, whereas the engineers and executives tend to leaf through the magazines they read page by page. Since we already know that chemists, engineers, and scientists respond to the same artistic and journalistic stimuli that the lay public does, we now recognize that we are able to increase the readership of given articles by applying these journalistic variables to them. For example an item on a right hand page gets more notice than an item on a left hand page. A graph gets more notice than no graph, a table gets more notice than no table, although it is difficult to say whether a table gets more notice than a graph. A picture however gets more notice than either, and the bigger the picture and the farther to the right we place the picture, the more attention it will receive.

General titles get more attention than specific titles do, although if they become too general they again cause a loss of interest. There is clearly an art therefore in generalizing the content of an article in the title just the right amount, and not too much. We have further learned that when editors generalize too far, the readers soon learn to distrust the messages they are receiving in print, and the magazine soon begins to decline in credibility. Likewise, when the article titles are too specific, the readers miss a great deal of information that is relevant to them.

We have some information that implies that the longer an article is, the less it is read. This inference is confused by the presence among the data of many other variables. Nonetheless it does indicate that short incisive articles will be more effective in transferring information from author to reader than longer contributions will be. This

may explain why Communications, as currently present in *Journal of the American Chemical Society* and other publications, seem to satisfy readers as much as they satisfy the authors' need for establishment of priority. On the other hand there is evidence that if the contributions become too short, the reader cannot maintain an interest. This may explain why *Chemical Abstracts* section groupings have never become primary reading fare for any significant number of chemists.

Much more could be said. I have tried primarily to present the structure of the literature as we see it, some of the problems that threaten its destruction, and some of the information we have obtained from reader research and similar sources that indicate ways to overcome our problems and maintain a flow of printed information in the physical sciences. We know that chemists respond to stimuli in the same way that people in general do. We know that they respond to good writing, to significant general subjects, to compact writing, and to help from the author in understanding and interpreting the information being presented.

We know they have a range of willingness to read. We know they have a range of willingness to spend money to have information at their finger tips. We know that they place a higher value on some elements of information than they do on others. And we know that in the last analysis the learning process from the printed word must compete with the learning process by other means, such as by a telephone call to a friend or by the running of a new experiment. From these understandings and from others I have not mentioned, modifications in the nature of the printed literature of chemistry will occur. The pace of change is quickening here as in most other areas of modern life. Experimental journals, overthrow of traditional requirements or criteria, increases in costs and prices, and some entirely new variables will be added to the literature scene. The intriguing aspect of all this is that, for the first time in 300 years, the chemical world may be about to make a significant improvement in the effectiveness of its communications with itself.

But, on the basis of what we know about chemists' behavior, we can feel fairly certain that these improvements will come in a series of small steps over many years, not in a single massive redesign.

LITERATURE CITED

(1) Kaempffert, Waldemar, *Encyclopaedia Britannica* **12**, 543 (1963).
(2) Singer, Charles, *Encyclopaedia Britannica* **20**, 115 (1963).

Symposium on Redesign of the Technical Literature—Introductory Remarks*

CHARLES L. BERNIER†
The Squibb Institute for Medical Research, New Brunswick, N. J. 08903
Received October 4, 1968

Vannevar Bush described an automated desk library, the Memex, twenty-three years ago. Despite the progress of a quarter century and the very tangible results from investment of millions of dollars in information research, there are still too many words to be read and these are published in too many languages and too many journals. In the last twenty-five years, the technical literature has at least quadrupled, and a larger percentage of it is now in languages other than English. Increased specialization by the user to reduce the amount he must read loses vital information in multidisciplinary research—such as that on cancer or medical use of radioisotopes. Searching the literature for most of us is still uncertain, slow, and wasteful. Serendipity supplants complete searches. Results of search are often too many words to be read in the current forms of presentation. An increasing number of us are coming to believe that we cannot read all necessary technical literature as it is now presented, selected, and distributed. The verbal outputs of research and development as well as the secondary literatures, need re-examination. Would redesign of the technical literatures help to supply all of the right words in readable numbers? Could redesign of verbal or graphic outputs enable us to keep up and to use the record more conveniently? Our Symposium was designed to explore answers to these questions.

The significant contributions of this Symposium show that improvements in the technical literature can, should, and will be made. I am confident that another quarter of a century will not slip by without bringing many of the suggested improvements.

The primary literature seems to function as: 1. a random source of inspiration and serendipity; 2. an unorganized collection of data, largely unevaluated, and sometimes erroneous; and 3. a reward for authors. These three functions do not need replacement, but improvement; especially of those archaic formats that waste time of readers.

We need better communication ahead of the primary literature. While conferences are very important, immediate, written communication among those active in a subject field is also needed. Progress is often so rapid in active fields that published results are obsolete.

The critical review is usually much too slow and there are not enough—even of delayed reviews. We need immediate reviews, such as the microreviews of aphorisms that have been suggested. Microreviews, organized by subject and published even daily, should serve as experience-laden guides to the primary literature.

Repeated search for data and their redundant evaluation constitutes a serious waste of technical manpower—a waste that is exceeded, perhaps, only by the waste

* Presented before the Division of Chemical Literature, Symposium on Redesign of the Technical Literature, 156th Meeting, ACS, Atlantic City, N. J., September 1968.

Editor's note: Other papers from this symposium will be published in the February issue.

† Present address: State University of New York, Buffalo, N. Y. 14214.