STAT

MEMORANDUM FOR: [redacted]/HTO

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

"This material has application to FUTURE open source literature and future processing problems. Garfield is with ISI. I am currently monitoring a contract with ISI on open source processing. Lederberg, a Nobel Prize Winner, is currently a member of the DCI's STAP (Sci. & Tech. Adv. Panel)"

[redacted]

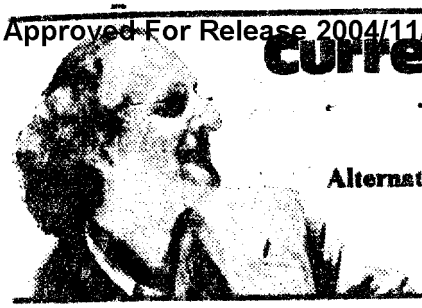
STAT

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

Date 1 Oct 79

FORM 5-75 101 USE PREVIOUS EDITIONS

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2



Current Comments

EUGENE GARFIELD
INSTITUTE FOR SCIENTIFIC INFORMATION®

Alternative Forms of Scientific Publishing: Keeping Up With The Evolving System of Scientific Communication

umber 36 September 3, 1979

Some time ago, I expressed some opinions on the future of the scientific ournal.[1] The primary point I made was hat the medium in which scientific arti- les appeared might change, but the contents would be essentially the same. Much has been said in recent years about the "paperless revolution."[2,3,4] But Joshua Lederberg, president of Rockefeller University, seems to have ynthesized it all in a paper which I ex- pect will have wide impact.[5] While communications and information scien- ists have grasped the technical signifi- ance of the electronic publishing revo- ution, Lederberg, as an accomplished ser, appreciates the impact a little nore than most of us. Hence the title, Digital Communications on the Con- fuct of Science: The New Literacy.'

Lederberg wrote his paper at the invi- ation of the IEEE for a special issue of heir *Proceedings* devoted to packet communications. Lederberg suggests hat electronic communications will not only speed up scientific information ex- change, but the new medium will also affect the quality of the messages con- veyed.

Lederberg's EUGRAM system in- volves a network of interconnected computers. The individual scientist pre- pares a scientific communication on a text-editing display terminal. Once the paper enters the system, it is immediate- ly retrievable by other scientists with terminals. Instant refereeing of papers is possible because you can send your EUGRAM to selected colleagues or ref- erees. This combines the features of an electronic mail system with Selective Dissemination of Information (SDI).

The system seems to resolve some problems associated with today's print- ed journal. These problems include the spiraling cost of printing and the ever- increasing number of scientific papers vying for limited space. Presumably, the electronic system will cost less than print journals, and we can assume that eventually more papers could be stored electronically than can now be pub- lished economically. Since all papers entered into the system would be re- trievable by any scientist, scientific pa- pers in the future might receive more in- terdisciplinary exposure than they do today.

It is tempting to contemplate in detail what Dr. Lederberg says. I suggest that the interested reader write for a reprint, especially since the paper was published in a journal that is not immediately ac- cessible to most of you.

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

The main point of my essay is to describe journals which typify the rapidly changing scene in the evolution of the journal.

At ISI® we are constantly evaluating journals. It is a strenuous but essential task. It is especially challenging because in spite of our comprehensive coverage, we must be selective. During the past six months alone, we've evaluated and reevaluated more than 1,000 journals. The result of all this work is only partially reflected in "journal coverage changes."

All this activity reflects one of the major quantitative consequences of big science. Science is big not only because there are large projects which produce papers with a dozen authors. Science is big because it is omnipresent, international, and still growing. It is inevitable that new journals proliferate like big science. But there is also a need for new kinds of journals. Some satisfy the special intellectual requirements of big science. Others reflect the quantitative needs.

The latter category of need has led to alternative publication media such as microforms or miniprint, which reduce the amount of print space used in publications. The former category includes those journals that present scientific knowledge in new editorial styles.

Since the electronic journal is still some years away, for the present, many journal publishers are turning to alternative media to cope with increasing costs. Publication in microform is one option pursued by journal publishers for some time now. The journals *Wildlife Disease* and *International Microform Journal of Legal Medicine* have published exclusively in microform since the mid-1960s.[6] Today, many print-journals also publish a microform edition for libraries with limited space. However, most of these are not intended to serve as an alternate means of original publication.

The obvious advantage of microform is simply that more papers can be published in less space and at less cost. However, microform journals have not gained wide acceptance within the scientific community. "The advantages [of microforms] all appear to be for libraries," says L.A. Page, treasurer of the Wildlife Disease Association and past editor of *Wildlife Disease*. In contrast, "most authors want to have their work in readily readable form."[7] You simply cannot browse through microform as you can with printed journals. Moreover, the reproduction quality of microform readers has not been very good, though it is improving.

It remains to be seen whether microfiche in particular will become a major medium of scientific publication before, during, or after the electronic revolution has arrived. One idea that seems to be catching on is the synoptic/microform journal. These journals only print synopses or summaries of scientific papers. The full papers are published simultaneously in a microfiche edition of the journal. This approach to journal publishing is similar to an idea suggested by Watson Davis as far back as 1933.[8] He proposed that synoptic journals be published by a central agency that would provide the full papers on demand—just as reprint requests are handled now.

A prime example of a synoptic journal which appears to be a successful experiment is the *Journal of Chemical Research*, which was started in 1977. Under the aggressive leadership of Dr. Helmut "Joe" Grunewald, this journal has been able to publish an average of 240 papers a year. The new journal caused some problems for us here at

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2
© 1979 by ISI®

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

udes only the key references cited. ıe complete bibliography is only avail-ıle in the microform or miniprint edi-ın. Eventually, ISI and The Chemical ciety, London, which publishes the urnal, worked out a solution. It is now ssible for us to process all cited refer-ces. Recently, the journals *Studia ophysica* and *Bulletin of the Geologi-l Society of America* switched to a nilar format. As in the case of the *urnal of Chemical Research*, the inted summary sections do not con-in all the references to the full articles. e are now working on arrangements ıich, with the cooperation of these urnals, will allow us to pick up these st references.

In a previous essay I discussed the use miniprint as a cost-cutting alternative r journal publishers.[9] The *Journal of emical Research* publishes a mini-ınt edition which it offers as an alter-tive to microfiche. However, not all ıblishers who have tried miniprint re satisfied with the results. In the rly 1970s, the American Institute of ronautics and Astronautics consid-ed publishing its journals in miniprint d offered sample miniprint articles to readers. The Institute abandoned the a, however, when a sizeable minority the readership responded negatively the miniprint samples.[10] Similarly, *Canadian Aeronautics and Space urnal* used to print synopses in normal pe size, but the full articles appeared miniprint at the back of the journal. ıis practice was discontinued about six ars ago.

The journals discussed up to this ınt all use alternative media to cope th the rising cost of publication. Some urnals, however, are experimenting th new ways of presenting scientific owledge. These experiments are at-tempts to cope with other problems in-herent in the journal system of scientific communication.

One type of scientific literature has been described by Senders as "fugitive" literature.[10] It includes papers of merit that are not suited for core journals because of their length or because they report so-called negative results. The American Psychological Association (APA) is coping with the growth of fugitive literature through its quarterly *Catalog of Selected Documents in Psychology*. The *Catalog* publishes abstracts of unpublished papers that would otherwise be lost to the scientific community. The APA offers reprints of the full articles to readers who request them. The cost of the reprint varies with each article.

Another experiment in scientific publication is the International Research Communications System (IRCS) established in 1973 by David F. Horrobin, University of Montreal; John Paul Eakins, of Imperial Chemical Industries; and Michael S. Buckingham, now managing director of IRCS. IRCS has reduced the lag time in communication to four weeks from the submission date of a manuscript until its publication, inclusive of refereeing by largely UK referees. Research findings are published as brief, 500-word notes.

Each article in the IRCS system appears in one or more of 32 print journals. Each journal or section covers a different medical specialty. IRCS also publishes three "key" journals, which present those articles from the entire compendium considered most important by the editors. All of the articles published in the 32 sections appear in a microfilm *IRCS Medical Science Library Compendium*.

One of the major problems in gaining acceptance with these new experiments

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

to publish their best works as brief communications or in microform. It becomes difficult for *Current Contents®* *(CC®)* to justify coverage until the quality of the material is on a par with the journals we now cover. This vicious circle may only be broken by testing out some of this material in *CC*.

The examples cited so far are journals which attempt to solve the space problem. If science is democratic, then everyone should at least have a chance to get in his or her 500 words! But this does not deal with the fundamental intellectual problem of science which frequently pushes in the opposite direction. There ought to be a way to foster more detailed discussion, especially when a problem is of vital interest to many people. A pioneering step in this direction was taken by the journal *Current Anthropology* in 1960. This journal fosters a system of open peer commentary. It anticipates in slower print form the "instant" peer interaction of Lederberg's EUGRAPHY.

The newest journal to use open peer commentary is *The Behavioral and Brain Sciences*, published by Cambridge University Press. This journal is the brain child of Stevan Harnad, formerly of the Rutgers Medical School, who is now the journal's full-time editor. Harnad consciously modeled *The Behavioral and Brain Sciences* after *Current Anthropology*.

Each issue presents about four "target" articles. Each article is accompanied by about 35 commentaries contributed not only by members of the immediate invisible college but also by outside peers. Authors who have written on subjects related to the target articles are prime sources of commentary. They are identified through the use of such current awareness tools as the *Science Citation Index® (SCI®)* and *Chemical Abstracts*.[11] Care is taken to insure that the commentaries represent a sampling of opinion from scientists throughout the world. Through open peer commentary, the knowledge imparted by the target article becomes more fully integrated into the entire field of the behavioral and brain sciences. This contrasts with the provincialism of specialized journals.

The commentaries or critiques presented in *The Behavioral and Brain Sciences* are not short quips—they average about 1,100 words. (Target articles average about 13,000 words.) Furthermore, the target author's response to the collective critique averages about 8,300 words!

Harnad's enthusiasm for *The Behavioral and Brain Sciences* is infectious. "Peer interaction is the real medium for the self-corrective aspect of science," he writes.[12] This belief is not unique, but the new editorial style makes it a reality. Indeed, most scientists are aware of the large number of errors even in the most prestigious journals.[5,13] The first words I stated in print about the *SCI* concerned the problem of corrections of all kinds which remained buried in the literature for lack of a means to bring them together with the original errors.[14]

In the case of *The Behavioral and Brain Sciences*, open peer commentary does not replace the traditional refereeing procedure. On the contrary, papers submitted to the journal are reviewed by about eight anonymous referees. Usually, three of the referees are experts in the field of the target article. The other five are experts in related

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

© 1979 by ISI®

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

fields who check manuscripts to verify interdisciplinary interest and intelligibility.

The most noticeable drawback of open peer commentary is the amount of time required for a submitted target article to be published. After the article is received, it must be sent to referees. Some rewriting is usually necessary after this has been done. Next, commentaries must be solicited and received. Then, the target author must write his response. Until now, this process has taken up to a year to complete. Harnad expects the time to be reduced to eight months once authors become more familiar with the journal's style.

ISI is now covering *The Behavioral and Brain Sciences*. The open peer commentary feature presented us with a real problem. How does one treat the commentaries? Are they part of the target article? If so, what does one do with the references found in the commentaries? Should we include the names of all 35 authors involved on the contents page? In the end, we decided to treat each commentary as an article in its own right.

All of the innovations described above are harbingers of significant changes on the horizon. As the price of paper and postage increases while the cost of computer storage goes down, we approach a critical point in the history of science. How many journals survive and for how long is an interesting problem for futurists to tackle. In the meantime we at ISI will continue to deal with each new innovation with an open mind. As you have seen, some innovations can disrupt our system temporarily, but we still encourage creative experimentation.

©1979 ISI

REFERENCES

1. **Garfield E.** Is there a future for the scientific journal? *Sci. Tech. News* 29:42-4, 1975.
2. **Cawkell A E.** The paperless revolution. Forces controlling the introduction of electronic information systems. (Part 1) *Wireless World* 84(1511):38-42, 1978.
3. ----------. The paperless revolution. Forces controlling the introduction of electronic information systems. (Part 2) *Wireless World* 84(1512):69-74, 1978.
4. **Lancaster F W.** *Toward paperless information systems.* New York: Academic Press, 1978. 192 p.
5. **Lederberg J.** Digital communications and the conduct of science: the new literacy. *Proc. IEEE* 66(11):1314-9, 1978.
6. **Herman C M & Davis D E.** Primary publication in microprint. *Bioscience* 14(4):27-30, 1964.
7. **Page L A.** Telephone communication. 25 July 1979.
8. **Davis W.** Project for scientific publication and bibliography. (Bernal J D.) *The social function of science*. Cambridge, MA: MIT Press, 1967. p 449-55.
9. **Garfield E.** Miniprint: is it a practical way to cut publishing costs? Or if you can read this, you can read miniprint! *Current Contents* (2):5-12, 9 January 1978.
10. **Senders J W, Anderson C M B & Hecht C D.** *Scientific publications systems: an analysis of past, present and future methods of scientific communication.* Toronto: University of Toronto, 1975. 175 p.
11. **Harnad S.** Telephone communication. 25 July 1979.
12. ----------. Creative disagreement. *Sciences* 19(7):18-20, 1979.
13. **Ingelfinger F J.** Peer review in biomedical publication. *Amer. J. Med.* 56:686-92, 1974.
14. **Garfield E.** Citation indexes for science. *Science* 122:108-11, 1955.

© 1979 by ISI ®
Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

Digital Communications and the Conduct of Science: The New Literacy

JOSHUA LEDERBERG, SENIOR MEMBER, IEEE

Invited Paper

Abstract—This essay is a personal perspective on the emergence of a new form of communication, optimistically called the 'EUGRAM'. This form is based on the convergence of economical digital communications with computer-aided facilities for file management, and protocols to facilitate the interconnection of users separated both in time and space. The EUGRAM is contrasted with the telephone, with the latter's demands on instant availability and the subjugation of the user to an almost uninterruptible stream of data. The EUGRAM is expected to increase the thoughtfulness of communication, the return of literacy in the efficient and precise use of language, and to enhance scientific discourse in many other ways.

INTRODUCTION

COMPUTER communication networks provide new tools and opportunities for the scientific community to share scarce computer-based resources. They permit a new form of informal communication between scientists and often provide motivation and reward for timely sharing of research results. In addition, computer-based support to large distributed segments of a scientific community is made possible via users and computers interconnected by computer controlled networks.

Today the most significant and useful form of computer communication is based on packet-switched technology which has been reduced to practice in daily support of some portions of the scientific community.

Two key elements of this technology base are:

1) computer-based user–user message capability, i.e., electronic mail plus the computer-management of text data.

2) sharing in the development, refinement and use of large, complex computer knowledge-based systems particular to a segment of science, which would not otherwise be widely available.

This essay is written from the perspective of an enthusiastic *user* of packet-switched communications. The system itself is here regarded as a black box that accomplishes efficient transfer of digitally encoded information in near-real time among terminals that interface both to human users and to computer-manageable files. The economical integration of user, file, processor, and distance-indifferent communication link is the novel capability of what I shall call a EUGRAM system. EUGRAPHY thus embraces not only electronic despatch of mail but also a panoply of computer-augmented text-handling tools and protocols. This account is informed by my experience over the last five years in the development of the SUMEX-AIM community for research in artificial intelligence related to biomedical science, which is more fully described in Appendix I. However, it will be primarily concerned with the expected impact of, and needs for, the elaboration of EUGRAPHY in the conduct of scientific research generally over the next 25 years.

Manuscript received January 9, 1978, revised May 15, 1978. The work on which the essay is based was supported in part by the Biotechnology Research Branch, Division of Research Resources, National Institute of Health, Grants RR-00785 and RR-612-06, and by the Defense Advanced Research Projects Agency, Heuristic Programming Project, directed by Prof. E. A. Feigenbaum.

The author was with Stanford University Medical School, Stanford, CA. He is now President of the Rockefeller University, New York, NY 10021.

A. Conduct of Science: Computers and Communications

The claim of science to universal validity is supportable only by virtue of a strenuous commitment to global communication. In the spatial domain, the canon of publication insists upon public awareness and criticism of avowedly new knowledge. This enforces the reliability of empirical reports and assembles them into common models of a real world. In the temporal domain, the archiving and retrieval of information sustains the discipline of novelty—assuring that we acknowledge, so as to be able to extend, the boundaries of 'human', i.e., universal knowledge.

The past twenty years have witnessed a growing self-consciousness about the structure of scientific activity, impelled in part by Malthusian concerns over the long term implications of a geometric increase at 0.25 dB/yr: a ten-fold expansion over the 40-year typical career of the scientist. Much more has been written than implemented about means of helping scientists keep up with the "information-explosion." One must acknowledge the utility of recent introductions of literature-searching and alerting services, many of which crucially depend on computer support and EUGRAM-like communications. On the other hand, it will probably be the cost-explosion of print media for scientific publications [1] that proves to be a more immediately compelling motive for fundamental reexamination of our methods of scientific documentation and communication. Designs for solving these problems reviewed long since [2]—must take into account that the media for communication also play a crucial role in quality control in science. The filtering procedures of the 'refereed journal' support the selection both of worthwhile reading, and of the workers whose established performance entitles them to the privileges of academic positions and social subsidy for their research.

Perhaps on account of these latter concerns, most of my colleagues in biomedical research would be loath to adopt many changes in the present system of print publication. In practice, frequent personal encounters [3] facilitated by

0018-9219/78/1100-1314$00.75 © 1978 IEEE

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

grant funding, jet aircraft, and invisible colleges [4], [5] seem to play an increasingly important role in the exchange of information within scientific specialities, but without any systematic inquiry as to the costs, efficiency, and equitability of these modes. Nevertheless, no piece of work, no claim to priority, is authentically recorded until it has appeared in public print in a respectable refereed journal. The long-distance telephone surely has its role also, but more for operational detail than serious intellectual discourse; and the use of the mails is as idiosyncratic as is the performance of the U.S. Postal System, except for ubiquitous broadcasting of xerocopied preprints of forthcoming publications.

In the face of this inertia, one should be skeptical about the marketability of new systems like EUGRAPHY, regardless of their technical virtues. Indeed, scientists may be the last to adopt them on a comprehensive scale, except for demonstrations that may arise from a) computer science, b) research mangement, c) military requirements, d) the ever graver collapse of conventional mails, and e) business applications like EFTS. With respect to c), we of course owe a great deal to the ARPANET as showing the way, and with the potential for a spillover into civil technology perhaps comparable only to jet engines. The sheer economy of EUGRAPHY, and the diffusion of microprocessors and displays into the laboratory and into everyday life, are bound to force an encounter with the challenges of new systems despite the traditional conservatism of the scientific establishment (with respect to its own way of doing business, and its attention to change outside the academic discipline [6]). Nevertheless, the histories of the medical and engineering sciences both show many instances where a reluctant marriage of *theoria* and *praxis* has engendered major enrichments of the basic sciences.

All the above notwithstanding, our own experience with EUGRAPHY at SUMEX-AIM has been extraordinarily good. Individual users, of course, rely upon it routinely for access to computer processing. More surprising was the utility of EUGRAPHY for research management, involving the exchange of texts even over relatively short distances—offices down the corridor or in nearby buildings. This phenomenon has provoked introspections about EUGRAPHY as a qualitatively different method of interpersonal communication from conversation, the telephone, the handwritten memo, the dictated letter, or the published report, and some speculations about the further evolution of EUGRAM's as part of scientific communication.

B. Comparing the EUGRAM with the Telephone [7], [8]

When telephone usage is limited to a few calls per day, and the connecting parties are reliably locatable, the telephone may indeed fulfill its image of instant, spontaneous communication. In current practice, beleaguered by time zone shifts, lunch hours, conferences, and competing calls, the reality of phone usage is exemplified by the employment of secretaries to make and receive the calls. The very instantaneity of the phone connection generates a queueing problem that defeats the basic motive. In due course, the two-way conversation may disappear, to be replaced by messages stored on tape recorders. The information density of speech may be viewed as very low, or very high, depending on how much of the burden is carried by the text, how much by inflection, phrasing, and other personal qualities. It may be only with respect to communications that have high affective content that audio channels can compete with digital channels, and to do this well may require better than the average channel quality than is now readily available between metropolitan centers. Even here, the enhancement of literary competence might go a long way to permit the EUGRAM to compete with the song. The EUGRAM, furthermore, has all the advantages of digital storage and accessibility to archiving, sorting and searching mechanisms that are far easier to implement, and require far less bandwidth than do voice messages. The EUGRAM itself can be composed quickly with a text-editor on the user display, where it is readily rehearsed, corrected and reedited before being transmitted. The same EUGRAM can be fanned out simultaneously to a large number of recipients, or it can be revised and perfected through several versions with similar broadcast, or with selective distribution. From the receiver's perspective, he has the advantage of a literate spatially oriented medium. In contrast to the time-fluent telephone, radio, or TV, he has the option of perusing his mail at his own pace, of interruption, backtracing, and cross-checking the text, even of marking it for reexamination and further rumination. He retains mastery of the use of his own time, and can coordinate attention to a coherently chosen set of tasks. He is liberated from the tyranny of synchronizing his own mental processes to those of the external actor. This freedom of course reduces the impact of that actor, just in proportion to the responsible autonomy it returns to the reader. In framing responses, entire messages or selected extracts together with added comments can be forwarded to others, or returned to the sender—lending focus to a 'discussion' and providing unambiguous texts for the development of a consensus. EUGRAM's can be filed and retrieved efficiently, or transcribed into hard copy as required. Text editors may be embellished with elaborate formatting aids, spelling correctors, even an on-line thesaurus to aid in composition. When quantitative calculations are in question, numbers can be mechanically copied directly from program outputs, avoiding pestiferous typographical errors. The same computer is likely to be the user's research tool and give access to shared data-bases: the EUGRAM's can then refer to common files by names that are themselves machinable. The user will also have access to other conveniences, such as desk-calculator-like programs for the checking of figures. He can even track the growing size of a EUGRAM-script (like this one) to be sure it fits into the assigned space. These word-processing capabilities can of course be consummated with hard copy sent through the mails, but with some additional effort, and the degradation of the machinability of the product at the other end. The paradoxes of instant telephony are most manifest when several parties are involved. In our experience, several weeks prior notice (or other rigid prearrangement) has been needed to schedule teleconferences if four or more people were required simultaneously. EUGRAM's to groups are sent in real time supported by conveniences like group labels. Stored in the receiver's file areas, EUGRAM's are exchanged among an active community like SUMEX-AIM within a few hours, often within minutes. Users also remain in ready communication with each other, via their respective EUGRAM files, even when either or both have travelled away from their customary homes. Lightweight portable terminals give any user full access to the system from any point which connects to the global telephone and other communications networks. Some facilities offer a fair amount of directory assistance, in locating and identifying the EUGRAM addresses of users; files may also be used to contain blocks of addresses that can be addressed by group names. At SUMEX-

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

AIM, publicly accessible bulletin boards are also available for broadcasting information or posting queries, without obtrusion, to a large audience. No doubt, 'junk mail' will become a problem in this medium, as it may in any other. However, the recipient has as powerful a technology for filtering unwanted messages as the broadcaster has for disseminating them. The struggle is more evenly matched, and there is then less economic incentive for abuses than applies, for example, to the distraction of one's attention by automated telephone sales technology.

Both for the management of the administrative affairs of the system, and for many of the research communications, EUGRAM's have become the preferred method of communication, provided they can be punctuated with occasional formal presentations, and more intimate encounters to help sustain the affiliations of the group. There is still plenty of personal style in the communications, and there is little problem evoking images of the warm bodies at the terminals. This intimacy can and should be supported by encouraging the occasional use of the EUGRAM system for arranging personal rendezvous. The trivial costs of such diversions are more than compensated by the enhanced efficiency of a worker who becomes adept at the use of EUGRAM's as if they were an extension of his own voice or handwriting.

C. *EUGRAM's and Complex Communications [9], [10], [11]*

One of the most controversial questions in social anthropology asks: "Is there a basic difference in modes of thought as between . . . "pre-scientific" and "science-oriented," "literate" and "nonliterate," . . ." societies [12]? The controversy is complicated by the empirical difficulties of measuring the cognitive styles of individuals independent of their social interactions and of the very media whose effects are in question. The evolutionist would have to interject that a certain neurological development was a precondition for literacy and presumably would have been subject to natural selection at least during the brief interval of human history since the invention of writing. Conversely, the oral tradition made its own demands on other centers in the brain. The only question is whether these cultural patterns have been sufficiently stable and durable to have had a significant effect on the differential evolution of the human brain in different cultures. Without going so far into the language/thought relationship, we can be categorical about the essentiality of writing for complex cognitive performances. The list—whether an inventory of baskets or grain, or a city telephone directory—is an externalization of cognitive activity that invites and sustains public use and scrutiny, and a form that has no effective analogy in the oral tradition. Indeed, it may have been the initial technological breakthrough in record-making preceding other forms of literature. A glance through the pages of this journal is evidence enough of the impossibility of assembling complex scientific arguments without the use of the written record. The manipulation of recorded symbols is a pale shadow of an internal cognitive imagination we hardly understand, but our most intricate intellectual exercises rely heavily on those external marks. In many instances, it might still be possible to read a journal article over the telephone and garner some degree of comprehension of the argument even without visible records: but consider how often we have to ask simple names to be spelled out and numbers repeated in phone discourse. Imagine then communicating a computer program of more than ten instructions over the telephone! Indeed, it is precisely for the sharing of such program source texts that EUGRAM's have been most manifestly indispensable for groups like the ARPANET and SUMEX-AIM communities. These program texts, which may reach hundreds of thousands of instructions, are among the most complex records of human logical effort—and more than any other production, the information is manifestly all in the text. However, they also typify the information content of other scientific efforts like mathematical proofs, structural analysis in chemistry, and other arguments. Some of these also resemble program sources in becoming almost impossible to criticize as written records alone, viz., without exercising them on the computer or in the laboratory. The recent demonstration of the four-color-map theorem comes to mind [13]. One of the facilities offered under SUMEX-AIM is the CONGEN system [14]. This is an aid to the organic chemist, offering him the computer generation of a hypothesis-tree of structures under given constraints. It can also be used as a verifier of claims of new structures, as a proof-checker. As an exercise in advanced organic chemistry, graduate students were assigned the verification of a set of structures recorded in the recent literature. Many of the proofs were found to be incomplete, usually for lack of tacit stipulations that were still plausible in the immediate context. We have no firm statistics, but perhaps one 'proof' in ten contained a substantive fallacy, unnoticed by the author and reviewers, that invited a critical reexamination of the conclusion. This suggests that organic chemical analysis has already become too complex for the existing media, that a significant part of the literature is shaky, and that computer-augmented proof-checking of complex structures should be part of the process of editorial review. The prevalence of statistical fallacies in the biomedical literature, often deeply rooted in careless experimental designs, has provoked much critical comment [15]–[18]. Certainly, it is responsible for a redoubled waste of resources, in the primary efforts, in faulty policy and practice, and in the further work needed for criticism and rectification.

Probably it is wrong to say that chemistry is so complex; to the contrary this finding is more likely a result of the simplicity and transparency of the logical argument in its proofs, which makes them more amenable to computer emulation. Outside of mathematics, very little scientific reasoning has been subjected to formal analysis and representation. EUGRAM publication now affords the opportunities and incentives to undertake more rigorous formulations both by providing more convenient media for depositing illegible proofs and offering access to symbol-manipulating machines to digest them. Increasingly, hardware engineers will find themselves companions to linguists and philosophers of science [19], [20]; they have long since shared profitable joint ventures with formal logicians.

Emergence of the New Literacy [8], [21]

The previous discussion declaims how the EUGRAM is a return to literacy, with some new forms and tools. The ease of its alteration saves some kinship of the EUGRAM to the oral tradition, with perhaps less social discipline but more effective tools to ensure the authenticity of the text. In fact, so much 'writing' is produced these days by dictation, with the most meagre and clumsy postediting, that these

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

tools may help bring the author closer to the well-tempered text he intends. Most tools are two-edged: the ease of inserting cliches and of conforming to system-defined formats may also hinder creativity. But this is like agonizing whether desk calculators will frustrate arithmetic skills. Some authors will balk at learning to type—even with all the facility of error correction afforded by every editor program. They can doubtless look forward within the decade to voice entry of rough texts that can speed up initial composition, and still leave scope for detailed editing. The author who does not interface directly with his own words with a text-display and editor is missing a powerful and precise organ of expression, which has no practical parallel in human communication today. Still, we can hardly surpass our inherent skills, though the wider availability of these compositional tools and challenges in education might help reverse the trend to illiteracy suggested by all recent statistics. Not every communication will or should be reduced to an unerasable EUGRAM. Lovers will not be deterred, even by the black box, no more than they are by the mails; but other intimate communications particularly some of the angrier ones—are better left to media where expletives can be deleted in hindsight. Even in scientific communication, there may be a place for a potential refuge: "I never said that?" in retrospection, namely to encourage some irresponsible imagination. This opportunity may be vitiated by the relentless accuracy of the EUGRAM, supported by new methods of encoding "signatures." Illegible handwritten scrawls will no longer offer a refuge of ambiguity. Nevertheless, while inscribed promises have more standing in court, voice-to-voice confrontation is less amenable to evasion at the moment: the journal editor will telephone a delinquent author when repeated pleas by EUGRAM have been ignored. Conversely, the poetic imagination may be less hindered in a literate medium than in immediate confrontation with other critical voices. Ambiguous phrases can be left in the record, when they would be challenged in the vocal stream. These very assertions are ones that might be difficult to articulate in a lecture: they reveal mostly how little we know of the uses of different media. Most of these remarks have concerned EUGRAM's between identified persons. The use of EUGRAM's for communication with archives opens up additional opportunities and foreseeable problems. In our experience at SUMEX-AIM, EUGRAPHY has been indispensable for the division of labor in drafting and criticizing complicated research proposals: 20 people may be closely involved in a product of 250 print pages. We have not secured a good system for tracking and interleaving successive versions, reducing a hairy tree of separate modifications to a coherent final form. Most nearly fatal is a cleanup reformatting that frustrates any simple line-by-line text comparison of deviant versions! Confusions of this kind in communal refinement of encyclopedic texts can perhaps be ameliorated with further software for documentation control. However, they reflect an underlying difference between EUGRAM's, manuscripts, and unit copying on the one hand, and letterpress on the other. Gutenberg's method lodges the major cost of publication in composing a definitive version of a master template. A side effect of the economic advantage is the focus on that version as a node of the intellectual commitment of the author, and of criticism by others. Communal revision over a EUGRAM network is likely to outpace the reaction time of individual critics: Scientist "A" will be entering his critique of Heisenstein's Field Theory version # 1764 when this has already been revised under the influence of "B" and superseded long since by version # 1769. The same fluidity of commitment may be self-aggravating if scientists are then unconstrained in what they enter into the archives, believing that their errors are erasable, and that they must compete for priority with less scrupulous colleagues. The blurring of nodes of publication will also greatly complicate the task of assigning due credit for intellectual innovation, although in principle there can be greater technological support (auxiliary files and the like) for documenting the participation of many minds. The advantage of this fluidity is, obviously, a possible mitigation of prejudice and rigidity of beliefs that may otherwise impede intellectual progress. The cost of nodal entry into letterpress also bolsters the gatekeeping role of editors and reviewers as trustees of the social investment entailed in that form of publication. This has already been eroded by the multiplication of commercial interests in scientific journals who receive large unacknowledged subsidy a) in the public funding of the underlying research, and b) in the asset of attention of the readership. Both of these have been exploited to the point that existing publication is fragmented to an untolerated degree: namely, in many fields scientists no longer accept the responsibility for awareness of every claim that has reached print, particularly if these have bypassed the recognized, peer-refereed organs of their discipline. Near-zero-cost entry into the archives of a EUGRAM system will aggravate that problem, but has the compensation of an easy technology for selective retrieval. The role of the trustees will be shifted from controlling what enters the archives to that of organized consultation about what is worth perusing. Controversial innovations may be more fairly evaluated if minority approval is enough to permit them to reach the visible record. The same technology can also be used to broaden the participatory base, and to reduce the grievous time lags and enhance the limited information flow that now characterizes peer review of research proposals used for the allocation of budgetary resources. The pros and cons of a wider base of 'voting' on one's colleagues' efforts can be roughly anticipated: in some sense more equitable distributions on the one hand; on the other, the factionalization of decision-making, political alliances, and the tyranny of the majority even in the most creative of individual activities. These dilemmas face us today; the new technologies will introduce a change of scale not of principle in the social monitoring of private thought. It is not just Big Brother we may need to fear, but the whole brood of our competing siblings. The enemy may also be within ourselves. Scientists generally are systematically socialized within the norms of their profession; nevertheless they must approach the raging floods of literature with some ambivalence [22]: there might be found the nuggets of insight that may help the investigator take a bold new step. There is also the fear of finding an anticipation that may destroy the novelty, and hence the entire utility, of months, years or decades of sweat and the pride of unique intellectual accomplishment. The designer of information systems can ill afford to overlook Mooers' Law that a "system will tend not to be used whenever it is more painful and troublesome for a customer to obtain information than for him not to have it" [23]. Some writers tend to be egregiously neglectful of citing the roots of their ideas, a self-serving amnesia that also obscures others' access to the overall picture. The neglect also impedes the efficient retrieval of connected

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

knowledge through devices like citation indexing. EUGRAM-based commentaries should facilitate the filling in of missing references by others, if the author has overlooked them, without making a major issue of the implied criticisms; and the anticipation of such corrections may deter the obliteration of the history of a subject. The cross-referencing and coding capabilities of bibliographic databases should also make it feasible for an author to exercise his historical responsibilities without excessively costly footnotes that may impede other uses of the entered material. In a similar vein, the systematic archiving of informal communications, including notes to oneself, surrounding the genesis of new ideas should facilitate the accurate reconstruction of the history of scientific discoveries—narratives that today are inevitably clouded with more restrospective myth than documentable substance. Altogether, we simply need to recognize that the new technology imposes fewer constraints *per se*, on the social structure of science, and that carefully designed new forms of social discipline will need to be established to meet the indicated functional needs.

The social innovations will doubtless evolve in response to microscopic pressures rather than as part of a system design, and their functionality will probably be tested on a time scale slower than continued technological inputs. Some of the needs and inventions can be foreseen; their main effect may be to facilitate another wave of illiteracy by the recruitment of still more elaborate devices for the human–bit interface. Reading and pecking are slow, and beneath the dignity of some professionals; voice response is even cheaper than the visual EUGRAM, and the technology for voice entry is on the way. Graphics already are an indispensable aid; there is no technological barrier to the integration of multimodal cable-TV (e.g., animated cartoons) with EUGRAPHY. Programming costs will return the initiative to the centralized broadcaster; hopefully, a few individuals will still insist on their own selection of intellectual fare and many will sustain bilateral conversation. The literate tradition can still be enhanced with improved designs of orthographic display, a wider menu of formats including color, perhaps even new alphabets and languages. Indeed, it is language itself that needs more constructive as well as descriptive investigation: our existing tongues have evolved in response to long outmoded technologies of communication, but we know too little of the underlying neurobiology to be confident how they might be improved. Such studies are also impelled by the prevalence of pathologies of language development that constitute a heavy burden on many children and their schools. A 26-character alphabet certainly bears no relationship to any system that would be systematically designed to enhance the speed and reliability of human communications [24], [25]. This discussion has intentionally focused on the difficulties and side effects that may attend the introduction of challenging new technologies of communication [8], [26]. Surely others will emerge as difficult to foresee as the impact of the internal combustion engine on the structure of cities. The problems should not obscure the constructive implications of steps towards the realization of an effective 'world brain,' which had already obsessed Leibniz, and which may be the defining attribute of technological culture: the efficient refinement and sharing of human knowledge [27]. We do well to question our moral capability of enjoying the fruits of such cooperation; but this is not to damn ourselves in advance, especially if we acknowledge that anticipating the human problems is a task of equal priority to engineering the hardware.

Appendix I

The SUMEX-AIM Facility

The Stanford University Medical Experimental (SUMEX) Computer was established in January 1974 with funding support from the Biotechnology Resources Branch, Division of Research Resources, of the National Institutes of Health. It constitutes the first national shared computing resource for medical research, exploiting current EUGRAM technology to serve a community of specialists doing research on the applications of artificial intelligence to medical (AIM) research. The building and sustenance of that community enjoys a priority equal to that of serving the computing needs of its members.

System Description

The computer facility consists of dual DEC Model KI-10 CPU's running under a locally developed dual processor TENEX operating system. It has 256K words (36-bit) of high-speed memory, 1.6M words of swapping storage, 70M words of disk storage, two 9-track 800 bits/in industry-compatible tape units, a dual DEC-tape unit, a line printer, a communications-network interfaces providing user terminal access. SUMEX-AIM may be accessed by local telephone lines, through the TYMNET and as a host over the ARPANET communications network. This set of network connections provides a richness of communication to the scientific community not otherwise possible.

Functions

Such a resource offers scientists both a significant economic advantage in sharing expensive instrumentation and a greater opportunity to share ideas about their research. This is especially timely in computer science, a field whose intellectual and technological complexity tends to nurture relatively isolated research groups. Each group may then tend to pursue its own line of investigation with limited convergence on working programs available from others. The complexity of these programs makes it difficult for one worker to understand and criticize the constructions of others, unless he has direct access to the running programs. In practice, substantial effort is needed to make programs written on one machine available on others, even if they are, in principle, written in compatible languages. In this respect, computer applications have demonstrated less mutual incremental progress from diverse sources than is typical of other sciences. The SUMEX-AIM project seeks to reduce these barriers to scientific cooperation in the field of artificial intelligence applied to health research.

Program (software) support will evolve from the basic system as dictated by the research goals and needs of the user. Initially, available programs include a variety of TENEX user, utility and text editor programs. The proliferation of INTERLISP as a language for AI-oriented programs was one of the principal reasons for selecting our system configuration. Other user languages include Snobol, Sail, Fortran-10, Bliss-10, Basic, Marco-10, Omnigraph, and Mlab.

Access to the system is divided between a group of Stanford University projects, led by Professor E. A. Feigenbaum, and autonomous nodes throughout the countfy. It is governed by an advisory committee representing those diverse interests as well as the funding agency.

Some Particular Facilities

Besides offering access to the running of programs, SUMEX-AIM maintains a number of devices to facilitate communica-

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2

tion among the users. Central to these are the mail-handling programs (SNDMSG, READMAIL, etc.) common to TENEX-ARPANET sites. In addition, the staff maintains a bulletin-board, in effect a common mailbox on which messages can be posted for public inspection and response, supported by various indexing and retrieval aids.

These devices have made possible a community effort in producing an encyclopedia of AI tools, the so-called AIHANDBOOK. This began as a production by graduate students at Stanford, but has enjoyed substantial input from other ARPANET collaborators, and is being readied for press publication.

CONGEN (CONstrained chemical structure GENerator) is an interactive proof checker that has been spun off from AI research on hypothesis formation in organic chemistry. Users provide a compositional formula, and the known constraints namely admissible *contra* forbidden substructures that may be present in the candidate molecules. The program generates all structures compatible with the given constraints, and presents them to the user for authentication. The usual result is an ambiguity, but with inspirations for additional constraints that can lead progressively to unique or nearly unique solutions. The procedures for assembling the candidates and efficiently applying constraints can be done with pencil and paper, but most chemists who have taken the time to learn the system find it a great advantage and can avoid numerous errors. Conversely, AI-research on CONGEN focuses on learning the human heuristics for converging quickly on the most plausible solutions, and in assembling the rules that will lead to performance that is both efficient and offers a congenial human interface. A large part of such efforts comprises aids for debugging, equally for program errors and human superstitions: most naive users do not carefully and fully exploit the constraints that are available to help in the solution of a problem.

ACKNOWLEDGMENT

Many more people have contributed to the ideas presented here by publication, by personal interaction, by telephone, and by EUGRAPHY—than I have been able to acknowledge in the available time and space.

REFERENCES

[1] F. F. Clasquin and J. B. Cohen, "Prices of physics and chemistry journals," *Science*, vol. 197, pp. 432-438, July 29, 1977.
[2] President's Science Advisory Committee, "Science, government and information," Washington: U.S.G.P.O., Jan. 10, 1963.
[3] E. C. Anderson, J. B. Arnold, C. Emiliani, W. H. Johnston, R. L. Miller, H. E. Suess, and V. L. Telegdi, "Back to the Homeric tradition." *American Scientist*, vol. 43, pp. 117-119, 1955.
[4] J. R. Cole and S. Cole, *Social Stratification in Science*. Chicago, IL: Univ. Chicago Press, 1973.
[5] D. Crane, *Invisible Colleges*. Chicago, IL: Univ. Chicago Press, 1972.
[6] W. van den Daele and P. Weingart, "Resistance and receptivity of science to external direction: the emergence of new disciplines under the impact of science policy," pp. 247-275 in G. Lemaine *et al.*, Eds., *Perspectives on the Emergence of Scientific Disciplines*. Chicago, IL: Aldine, 1976.
[7] I. de Sola Pool, Ed., *The Social Impact of the Telephone*. Cambridge, MA: MIT Press, 1977.
[8] I. de Sola Pool, F. W. Frey, W. Schramm, N. Maccoby, and E. B. Parker, Eds., *Handbook of Communication*. Chicago, IL: Rand-McNally, 1973.
[9] J. C. R. Licklider, *Libraries of the Future*. Cambridge, MA: M.I.T. Press, 1966.
[10] M. Turoff and S. R. Hiltz. "Meeting through your computer: Information exchange and engineering decisionmaking are made easy through computerassisted conferencing," *IEEE Spectrum*, pp. 58-64, May 1977.
[11] G. W. Arnold and S. H. Unger, "A structured data base computer conferencing system," in *National Computer Conference*, pp. 461-467, 1977.
[12] R. Horton and R. Finnegan, *Modes of Thought*. London: Faber and Faber, 1973.
[13] K. Appel and W. Haken, "The solution of the four-color-map problem," *Scientific American*, Oct. 1977.
[14] R. E. Carhart, S. M. Johnson, D. H. Smith, B. G. Buchanan, R. G. Dromey, and J. Lederberg, "Networking and a collaborative research community: A case study using the DENDRAL programs," pp. 192-217 in P. Lykos, Ed., *Computer Networking and Chemistry*. Amer. Chem. Soc. Symposium Series No. 19, 1975.
[15] N. D. W. Lionel and A. Herxheimer, "Assessing reports of therapeutic trials," *Brit. Med. J.*, no. 3, pp. 637-640, 1970.
[16] S. Schor and I. Karten, "Statistical evaluation of medical journal manuscripts," *J. Amer. Med. Assn.*, vol. 195, pp. 1123-1128, 1966.
[17] T. Dalenius, "Bibliography on nonsampling errors in surveys," *Int. Stat. Rev.*, vol. 45, pp. 71-89, 181-197, 303-317, 1977.
[18] I. J. Good, "Statistical fallacies," in *International Encyclopedia Social Sciences*. New York: Macmillan, 1968.
[19] C. W. Churchman and B. G. Buchanan, "On the design of inductive systems: Some philosophical problems," *British J. Philosophy of Science*, vol. 20, pp. 311-323, 1969.
[20] J. McCarthy, "Epistemological problems of artificial intelligence" in *Proc. 5th Int. Joint Conf. Artificial Intelligence*, MIT, Cambridge, MA, 1977.
[21] J. M. Nilles and F. R. Carlson, Jr., P. Gray, G. J. Hannenman, *The Telecommunications-Transportation Tradeoff: Options for Tomorrow*. New York: John Wiley & Sons, 1976.
[22] R. K. Merton, *Sociological Ambivalence*. New York: Free Press, 1976.
[23] C. N. Mooers, "Mooers' law, or why some retrieval systems are used and others are not," American Documentation, vol. 11., p. ii, July 1960.
[24] C. Cherry, On Human Communication. New York: Wiley, 1957.
[25] W. A. Rosenblith, Sensory Communication. Cambridge, MA: M.I.T. Press, 1961.
[26] G. Gerbner, L. P. Gross, W. H. Melody, Eds., *Communications Technology and Social Policy*. New York: John Wiley & Sons, 1973.
[27] M. Kochen, Ed., *Information for Action: From Knowledge to Wisdom*. New York: Academic Press, 1975.

Approved For Release 2004/11/04 : CIA-RDP90-00509R000100020003-2