electronic information handling

Edited by ALLEN KENT *and* ORRIN E. TAULBEE

Director, Knowledge Availability Systems Center
University of Pittsburgh

Manager, Information Sciences
Goodyear Aerospace Corporation

The Knowledge Availability Systems Series

electronic information handling

The **information explosion,** as the incredibly growing availability of data is termed, must not only be controlled but also needs to have its effects directed. Information handling by electronic means is the only feasible way to supply this direction, especially when the goal is to provide the means for making decisions.

To study the problems of information handling, authorities from education, industry and government were brought together at a national conference in the Fall of 1964. Jointly sponsored by the University of Pittsburgh, Western Michigan University, and the Goodyear Aerospace Corporation, the meeting dwelt on processing methodology in areas ranging from library science to military command and control.

The common thread binding these diverse interests is the support of decision making; the common concern is for the future. The forward-thinking analyses are thus presented in this volume under the headings:

- **analysis of the field**
- **end uses of information**
- **operational experience**
- **large-scale systems under development**
- **shortcomings of electronic systems**
- **planning**

Electronic Information Handling

Edited by ALLEN KENT
Director, Knowledge Availability Systems Center
University of Pittsburgh

and ORRIN E. TAULBEE
Manager, Information Sciences
Goodyear Aerospace Corporation

The Knowledge Availability Systems Series

SPARTAN BOOKS, *Inc.*
Washington, D. C.
MACMILLAN and COMPANY, LTD.
London

Library of Congress Catalog Card No. 65-17306

Sole Distributors in Great Britain, the British Commonwealth, and the Continent of Europe:

MACMILLAN and COMPANY, LTD.
St. Martin's Street
London, W.C. 2

Contents

III. END USES OF INFORMATION

IV. OPERATIONAL EXPERIENCES

Preface

A national conference on Electronic Information Handling was held on October 7–9, 1964, at the Webster Hall Hotel in Pittsburgh, Pennsylvania. Covering the rapidly burgeoning field of electronic information processing, the conference was cosponsored by the University of Pittsburgh, Goodyear Aerospace Corporation, and Western Michigan University.

In order to cover the spectrum of information handling problems, speakers were drawn from many fields of government, industry, and education. A correspondingly diverse audience of more than 400 persons, representing areas as varied as library science and command and control, were in attendance.

The papers presented, as reflected in the proceedings following, were organized into six sessions, on:

Analysis of the field
End uses of information
Operational experiences
Large-scale systems under development
Shortcomings of electronic information-handling systems
Planning for the future

The common thread running through the conference revolved about explorations of the field of information processing in support of decision-making requirements—decision making at various levels, in various environments, and for various purposes.

Acknowledgments

The assistance and cooperation of Western Michigan University, particularly Dr. George G. Mallinson, Dean, School of Graduate Studies, leading to the organization of the conference, is gratefully acknowledged.

Allen Kent
Orrin E. Taulbee

I. INTRODUCTION

1

Opening Remarks

THOMAS A. KNOWLES

President, Goodyear Aerospace Corporation

As an officer of the Goodyear Aerospace Corporation, I want to tell you how happy we are to join with you in this Conference, and to note the rather considerable *attendance and interest* which have been shown.

Perhaps it would be in order for me to explain why an industrial concern like ours is a party to an event cosponsored with two academic institutions, and how our particular company took the initiative, in this instance.

As you know, providing for our country's national defense and assisting it in providing health, welfare, and research support in areas of national interest involves a tremendous effort, a considerable portion of our national budget being allocated to these important projects.

With the need established, interest has been developed in a number of performing instrumentalities, some of them basically academic in nature, others in the nonprofit category, others in the form of specialty companies, and still others, like our own, as defense-oriented subsidiaries of large corporations working on the industrial scene.

While I cannot speak for all those organizations represented here that support research in such fields as defense and health, I know that they have undoubtedly developed a tremendous background of information-handling data, skills, personnel, and equipment either directly, or as by-products of other endeavors. In our own case, work on items like guided missiles, flight simulators, and space and warefare concepts has necessitated some knowledge of computers, memories, and other intelligence data-handling systems.

With a rather complex product line, our top management can hardly have a detailed familiarity with everything that is going on in all of these fields. Nevertheless, we do have the responsibility of endeavoring to steer the corporate course of action and to ration out our funds and facilities in accordance with some sort of a long-range forward plan, and to do this we talk frequently with those experts our company has recruited from the many technical disciplines, and from our many areas of effort.

In the harsh, competitive business environment in which we live, the various scientists and experts who come to us to ask for added personnel, funds, or facilities, must make a case for their programs in terms either of the national service we can render, or the volume of business which can be generated.

For a considerable period now the experts of our staff at Goodyear Aerospace have been alerting our management to the imminence of something which they refer to as an "information explosion" or "information revolution," and very frankly they have presented forecasts in the information-handling field which suggests that something tremendous and of significant national import is in the making.

And, while fascinating and intriguing prospects have been pointed out, some of us in management have found the problem so complex, the discipline so interrelated, the very techniques themselves in such an evolutionary form, that we have repeatedly pressed our people to bring more order and planning into the situation in order that we not make sporadic efforts in the field, growing like Topsy; but rather that there be some method and long-range continuity to our management approach and support.

The essence of what I have been able to gather from presentations thus far made to me is substantially this: the national importance of the subject hinges on the fact that in order to achieve our goals of social, scientific, and military progress, far better and more complete information is needed; and that the handling of such basic information is the common denominator of vital things like command and control, artificial intelligence, textual data processing, man-machine and automated library systems.

One also gathers the impression that we will need larger and more complete systems in the years ahead; new machine languages, and new hardware; and that any assault on the interrelated problems will require considerable more investigation of the theoretical and practical aspects, including the development of criteria for measuring comparative performance of systems.

Naturally, much remains to be done in educating ourselves and others about the needs and benefits of such systems; and it seemed to us that uniting the complementary capabilities of university and industrial organizations might stimulate rapid progress towards this end.

Since our people did not feel that substantial attention had already been given to the overall problem in any one place, it was our conclusion that it would be in both the national and our own interests if someone would gather together interrelated leaders in the various fields and disciplines, with a view to discussing just where we stand and just what should be done for our common benefit and advancement.

Because the mechanics of determining what things should be committed to memory or storage, how this should be done, and how fast they should be retrieved, could well be called out by specifications going beyond those applicable to the defense environment alone, it seemed to us that we should seek the broadest possible base for our discussion of what the field now has and what it should next provide.

In many ways such questions suggest the use of a broad and academic type of approach, for there is a responsibility to reach beyond and think in terms of more than any single classification of problems, or group of industries or services.

It was for this reason that we felt that we should endeavor to work with universities; and the selection of Pittsburgh and Western Michigan was prompted both by geographical proximity and by prior interest and leadership they had already exhibited in this important field.

So that is why Goodyear Aerospace elected to cosponsor this particular conference, and why we have joined with you in a sincere effort to inventory past accomplishments and to plan for the future. Doubting that our company interests and concerns are at all unique, I sense that all of us may have an opportunity to benefit.

2

Keynote Address

EDISON MONTGOMERY
Vice Chancellor—Planning
University of Pittsburgh

Until a week ago the Chancellor of the University of Pittsburgh, Dr. Edward H. Litchfield, was looking forward to talking to you at this time. Without warning, he received, through the Department of State, word that his Excellency Diosdado Macapagal, President of the Republic of the Philippines, had accepted a long-standing invitation to visit the University of Pittsburgh on October 7 and receive an honorary degree. The Chancellor was faced with the difficult choice of either not appearing before you this afternoon or precipitating a minor international incident. I am sure his choice to be host to President Macapagal is a fortunate one for United States foreign policy, although it will work a hardship on those of you who are in this audience this afternoon. With deep apologies, he has asked me to substitute for him and to give you the substance of the message he had prepared to open this conference.

Let me, therefore, join Mr. Knowles, President of Goodyear Aerospace Corporation, and Dr. Miller, President of Western Michigan University, who will be addressing you at tomorrow evening's banquet, in welcoming you to Pittsburgh and introducing this national conference on "Electronic Information Handling."

COVERAGE OF THE CONFERENCE

The topics to be covered during the conference are in the same area of interest that the University of Pittsburgh has assigned to a new part of the University, the Knowledge Availability Systems Center. This interest is not confined to a Center within the University. It has become a new university-wide philosophy.

Dr. Litchfield stated this philosophy in the Fall of 1962, and made it one of the major specific goals of the entire institution. He chose the term Knowledge Availability Systems to represent an activity far broader than "information retrieval," and to indicate concern with nothing less than the total problem of making knowledge available for desirable social purposes—currently and in the future.

Activities in this field had been pursued at the University of Pittsburgh before the establishment of a university-wide effort. Notable among these activities are:

1. The Health-Law Center, which has concerned itself with the storage on magnetic tapes of the statutes of many States, in order to accelerate their retrieval and thus facilitate legal research.
2. The Model Drug Prescription Project, in our School of Pharmacy, which has involved the electronic storage of drug prescription information for correlation with the side effects discerned by prescribing physicians.
3. The Crystallography Laboratory has been using computers to correlate data relating to crystal structures.

The Knowledge Availability Systems Center, established in September 1963 under the direction of Allen Kent, was charged with the responsibility of developing a program of research, operations, and teaching relating to the entire spectrum of information activities from the time information is generated until the time it is disseminated and put to use.

What has happened during the first year of activity?

1. A teaching program has been established which provides masters' and doctoral candidates with an opportunity to major in the emerging field of information sciences. Twenty-one credits are already offered in this program with about 250 students at the masters' level having taken, or now enrolled in the first course of the series. Three full-time candidates for the Ph.D. are already studying with the Center, representing, we are told, perhaps the total national crop of full-time students in this area.
2. In recognition of this strong start, the name of the Graduate Library School was changed on June 1, 1964, to the Graduate School of Library and Information Sciences to reflect our regard for the importance of this program.
3. The health sciences are represented in the new effort by the development of a Diseases Documentation Center, which will collect and interpret information, both published and clinical, relating to specific disease entities.
4. There has been substantial cooperative effort with Dr. Stafford C. Warren, Special Assistant to President Johnson, in drafting plans for a National Science Library System to cope with burgeoning periodical literature. This plan was presented publicly for the first time at a conference here at the University of Pittsburgh on the subject of Library Planning for Automation, held on June 2–3 of 1964.

5. A program for the spin-off of information developed through the national space program to industry in Pennsylvania and West Virginia is well under way. This operational KAS effort has been undertaken under contract with the National Aeronautics and Space Administration.
6. The Avco Corporation has made the University a gift of the Verac equipment. This hardware developed by Avco in collaboration with the Council on Library Resources permits the microreduction of records (at a reduction of 140 to 1) and their rapid retrieval.
7. We have received, on long-term loan, InSite equipment from the Beekley Corporation. This device permits ready searching of files using the peek-a-boo principle, but unlike other such systems, permits on-line printing of search results. One of the applications now being considered is that of class scheduling and registration.
8. The Photon, a computer controlled photocomposing system, has been acquired from the National Institutes of Health. The Computation and Data Processing Center has already, in its Project Upgrade, developed programs which involve automatic transfer of text from monotype and linotype paper tape to magnetic tape and which permit proofreading and editing of original manuscript composition through computing programming. With the aid of Photon, corrected manuscript may be set in a form ready for printing.
9. A detailed survey of the specialized information centers in this country has been completed in order to discern opportunities for developing a common, standard language that will permit interdisciplinary exploitation of the information stored.
10. The application of gaming theory to the investigation of relevance of IR systems is in progress. This program, supported by a generous grant from the National Institutes of Health, is looking into the use of a "heuristic information-retrieval game" to measure the behavior of users of IR systems in order to develop criteria for the system design.

I could mention many more things that have happened here, but suffice it to say now, that even in one year, starting with a new center, there are fifteen faculty and staff members now engaged in this program, involving the Graduate School of Library and Information Sciences, the School of Medicine, the School of Pharmacy, the Division of the Humanities, the School of Engineering, and the Division of the Natural Sciences.

Although we are gratified with the progress we have made in the field of the information sciences, there is a second group of reasons why we regard this conference as important.

COSPONSORSHIP OF THE CONFERENCE

You have noted that two major organizations have joined us in sponsoring and organizing this conference—the Goodyear Aerospace Corporation and the Western Michigan University—one a profit-oriented company, the second, another institution of higher learning. What circumstances have led to this rather unusual cosponsorship?

First, the profit-oriented company. One of the philosophies that Dr. Litchfield and his colleagues hold strongly is that a University must be a part of the community it serves. It must share in the responsibility for the economy of its region, as well as being responsible for intellectual activities. Developments within a university must be made available to the profit-oriented community that is our competitive society, but not just in a passive way—rather in a deliberate and planned program of transference of knowledge from the researcher to the industrialists.

Western Michigan University, of course, is also involved in higher education. It serves, however, a region in this country that is quite different from that of Pittsburgh. As an institution of higher learning in a more rural site and also reaching for a strong graduate program, it provides a field for experimentation in the information sciences happily complementary to that offered in Pittsburgh.

Cosponsorship of this conference represents a step toward initiating cooperative programs in this field among many similar institutions.

The technical and sociological problems to be worked out in this field are so extensive that no university can afford to be parochial in its efforts. It must seek relationships with other educational institutions as well as with industry.

And this leads into the third point I would like to make, as to why this conference is so very important.

WHY A CONFERENCE?

I suspect that many of you have read the recent article in *Science* entitled "Let's Run a Conference." This points out the popular trend toward running a conference when one has nothing better to do.

It is difficult for me to imagine anyone willingly or knowingly undertaking to punish oneself by holding a conference unless the reasons are clear and are pertinent.

Conferences are not a new business for universities.

The very nature of the educational process, which fosters research on an equal footing with teaching, has led to the elucidation and identification of new areas and fields, which later have become the entire subject matter of professional associations, which then take over the management of conferences on a regular basis.

But even then, as areas of investigation are pursued in the several specialties and subspecialties, each going its own way, it is often the university that discerns that the time has arrived to take stock, to review the several fields that are developing in parallel, to build bridges between these fields, and to redirect effort toward new goals.

It is those purposes that have stimulated us to arrange and to cosponsor this conference. The information sciences no longer concern only the traditional disciplines and professions. New fields of study have emerged with strange new names—information retrieval, artificial intelligence, bionics, mechanical translation, command and control.

We feel that the traditional and the novel must be related; gaps identified; and bridges built, so that research may go forward from a new platform of understanding.

The construction and reconstruction of such platforms are continuing tasks. Last week, work went forward on one in the library field, at the annual meeting of the Pennsylvania Library Association; earlier this week another platform was being built in the documentation field at the annual convention of the American Documentation Institute; and now another one is being constructed in Pittsburgh in the general field of "Electronic Information Handling."

Before I conclude, I should like to remind you of a paper published in 1955 by Dr. V. P. Cherenin in the Soviet Union. The paper was entitled "Certain Problems of Documentation and Mechanization of Information Search." Let me read several excerpts from a translation:

> ... The time is not far when a new revolution will occur in the storage and dissemination of data, similar to that which was produced by the invention of printing. It is difficult to guess how it will occur; nevertheless, by letting our imagination roam, it is possible to visualize the following information service of the future.
>
> ... All arriving and all existing data, after the necessary editorial processing and suitable exterior styling, are photographed at a considerably reduced scale on photographic film. Instead of large runs, only several copies of such microfilm are produced and are sent to one or several information centers. These centers transmit continuously over many waves all the data available in them at a tremendous sequence frequency of frames of microfilm, reaching, for example, a million per second. With such a transmission speed all data accumulated by humanity can be transmitted over many waves within a comparatively brief time interval—something like several minutes.
>
> ... Any frame of the microfilm can be received in any place on a special television screen equipped with a selecting device. All the instructions, classification schemes, table of contents of the microfilm with indication of the number of frames, and code designation required for the use of such a televisor are trans-

mitted at the start of the microfilm, therefore eliminating the need for using any kind of printed information.

...It is difficult to overestimate the flexibility and effectiveness of such an imaginary method of storing and disseminating data. Undoubtedly such a method or something analogous to it will turn out to be cheaper than the existing methods, when the volume of data will reach a definite limit. It goes without saying that, just as after the appearance of printing, the handwritten form of recording still remained in use, the appearance of a similar information service will still find a part of the data stored as before and disseminated in the form presently in existence. *Let us remark that, in spite of the fact that the information service of the future described above is quite fantastic, all the technical units required for its realization are in existence at the present time and being constantly improved.*

And now, ten years after this paper was published, we have not realized the objectives, even though, in the opinion of the experts, they are still valid.

The problems of information handling are becoming increasingly critical in more and more sectors of our society—in government, in industry, and yes, in the University.

Indeed, the need for rapid handling of information is so critical today that the University as the collector and imparter of knowledge is beginning to falter. This is a problem which must be solved, and solved rapidly.

3

What Do We Ask of Our Libraries?

JAMES W. MILLER
President, Western Michigan University

The distinguished conferees assembled here are certainly to be congratulated for the time, talent, and energies they are putting forth in these three days of meetings. It is most heartening to an academic administrator to know that this type of effort is being made to isolate the various facets and ramifications of the intellectual and technological problems involved in maximizing the efficiency and effectiveness of our libraries. As this audience knows, the simultaneous explosions of knowledge and population are plainly placing stress on the university community no less significant and no less intensive than the tensions being placed by these same phenomena on society as a whole. Nowhere on our campuses are we feeling more keenly the impact of an unprecedented explosion of recorded knowledge and the sheer impact of increased numbers of faculty and students than in our libraries.

As an administrator, I would hasten to add that in this period of stress there is too often a primacy given to the quantitative rather than to the qualitative aspects of our library problems. It is not, I believe, enough to think simply in terms of providing the same library services which we have offered in the past to the increased number of library users who are with us in the present. The user's time is a constant and so far as I know cannot be changed unless modern medicine is able to modify significantly our patterns of sleep and rest. Yet the sheer mass of printed material available to us is multiplying at an *exponential* rate. The user not only needs rapid access to vast accumulations of highly complex and diversified information, he needs real help to get quickly to material which has pertinence for his work. The user needs—yes, requires—considerably more help than our libraries are presently organized to give him in terms of discovering relatively quickly the relevance of specific pieces of library information and the pertinence of a particular piece of literature to other literature in the field of one's interest. It is pleasing to note that in this conference you are giving attention to *what* the librarian should be doing as well as attempting to become specific on *how* the librarian should do it.

The title of my address is meant to focus on the intellectual rather than on the technical aspects of the problems facing our libraries. Half facetiously and half seriously I might say that on the basis of the pattern of usage of some of our faculty and students we ask very little of our libraries. This is even true of personal library holdings which seem in some cases to have been acquired to impress visitors rather than to be read for comprehension and stimulation. The persons who gather a few or many books for appearance's sake remind me of Robert Burns' comments after he was permitted to browse in a Scottish lord's library only to find the pages of the books uncut. Burns wrote the following comment on the inside cover of a volume of Shakespeare's works:

> Through and through the inspired pages,
> Ye maggots make your windings,
> Oh, but respect his lordship's taste,
> And spare the golden bindings.

Recently an interior decorator, in what at first I could not believe was a serious recommendation, suggested that books on the shelves in my office be sorted so as to blend more aesthetically the colors of the bindings into the general color scheme of the office! Surprising as it may seem to you, this suggestion was serious and there and then I literally had to stop this person from physically demonstrating the point. Imagine being in the position of having to recall the color of the binding of a book that you might wish to examine or reexamine!

In general I think it fair to say that what we ask of our libraries is that they be organized, staffed and equipped to meet our needs. The question then is: What are our needs? Quite clearly our needs as individuals and our needs as institutions will vary. Neil Harlow in the September 1963 issue of *College and Research Libraries* delineates in a general way the levels of need for library services in academic institutions into three parts, namely: the levels of "college," "university," and "research." The libraries for the beginning student, which he calls the "college" level, would concentrate on general education involving introductory materials essential in the humanities, the social sciences and the sciences. At the "university" level, Mr. Harlow states the need in terms of the maturing scholar who should be provided with printed material emphasizing synthesis and the introduction to research. His third level, designated "research," is that library material which would be largely for the use of advanced graduate students, faculty members and the university's research staff. Whether you agree with this particular delineation of levels or not, the point is that thought has been, is being and needs continuously to be given to the question of what precisely are the needs that we are seeking to have our libraries serve. Without this type of examination it is fruitless and extravagant business to introduce expensive and complicated mechanized equipment into one's library. Many of us have complained about the buildings on our campus in terms of inadequacies and tend to blame the architect. In a majority of instances the fault is more likely to be with ourselves in that we have not developed clearly articulated programs. Having defaulted to the architect on the function of program, we blame him for what so clearly is our own inadequacy.

Ideally, in my opinion, we should ask of our libraries that their professional staff members be prepared and anxious to establish "intellectual camaraderie" with the faculty. Professional librarians can and should become fully involved in the education of students. With increasing enrollments and with greater emphasis and stress on independent study, librarians assume a significant and critical role in stimulating and assisting students in the use of library resources. As my colleague Dr. Russell Seibert, Vice President for Academic Affairs at Western Michigan University, stated in a recent article, "... every administrator should be permitted a few fond hopes. The fondest of those hopes is the dream of a library staffed with perfect librarians: librarians who love books and the contents between their covers; librarians burning with unsatisfied intellectual curiosity; librarians filled with the contagious enthusiasm for learning that will spark a student's interest without repelling him with too much bookish detail; librarians who are the soul of helpfulness, sensitive to the limits of, as well as the need for, assistance; librarians who are quiet-spoken and courteous, as respectful to those who are reading or studying as the mortician to the bereaved or the young mother of a sleeping child."[1] While the dreams of Dr. Seibert may never be fully realized, they are goals well worth striving to reach. No university can have a more valuable resource than technically competent librarians with broad cultural and intellectual interests dedicated and devoted to acquainting faculty and student with the resources of the university's library.

Again on an ideal basis we ask of our libraries that operations of its circulation of its own current holdings facilitate rapid search, location and acquisition of the material with which we need to work. We ask for adequate control of the books and periodicals on reserve. In fact, we ask ideally for a running inventory so managed that the frustrations and losses of time involved in finding finally that a book sought is in use, misshelved, being bound, lost, or not yet recorded would be reduced to minimal proportions. In a perfect organization I would suspect that there would be a sustained and systematic program of critical evaluation of the library's holdings in terms of what materials are either ready for disposal or retirement to some less costly storage area. Winnowing the rarely used and obsolete must be part and parcel of any system which seeks to be efficient, effective and economical. What we have been able to do in many areas in terms of records management I venture to say may have some general applicability for our libraries.

As our libraries grow and our student body increases we ask for a plan of new acquisitions designed to meet the unique needs of our clientele. We ask for rapid procurement, classification and cataloging along with bibliographies, indices, and reference services. Additionally we would ask for low cost and quick photocopying equipment. Ideally we would ask that a systematic screening be done to get into our hands pertinent data concerning the new acquisitions; this might include a table of contents, abstracts or other relevant information designed to offer helpful hints as to the contents of the new material. Duplicate copies of certain materials, microfilm equipment and adequate space and privacy in which to use the equipment are conveniences we would like to enjoy. Printed material which a particular library is unable to acquire for its own holdings should be accessible to the user by interlibrary loan, wirephoto, and possibly in the not-too-distant future, by electronic transmission.

Librarians of the character described in Dr. Seibert's remarks earlier in this statement, organization and procedures which are user-oriented, and a faculty prepared and willing to rationalize their relationships with the professional librarians and vice versa are, in my humble opinion, the basis upon which to build the library into the true "heart of the university."

On this last point we should, I feel, ask of our librarians and faculty that they meet on a regular basis—perhaps in faculty departmental meetings—to review current literature, discuss on-going and contemplated research on campus and consider ways and means jointly not only to promote the use of present services of the library and its study facilities but also to evaluate the effectiveness of present services and recommend new services to meet changing needs of both the faculty and student body.

In light of the growth in our libraries, the increasing amount of dissatisfaction being expressed by users, the enormity of the tasks faced by librarians to meet the twin cascades of an exponential rate of increase in printed materials, and a phenomenal increase of students and faculty, we must do as this conference is doing—namely, explore with vigor and enthusiasm every conceivable way in which our increasing and in many cases new needs can be served by our modern advances in technology. In any period calling for changes there are voices which will run the full gamut of the spectrum of thought in this area from the "Luddites" to the persons who see the millenium immediately within our grasp through the means of a fully automated library. Our solutions will likely be found somewhere between these extremes and possibly much closer to the fully automated extreme than with the "Luddite" group.

Libraries, it is clear, must be more than architectural structures filled with specific numbers of books, seeking ever to reach or overreach a specific quantitative figure of books per full-time-equated student. They should be fountains from which recorded knowledge can flow easily and quickly into the hands of our faculty and students and in a form economical for the user in terms not only of time but also of pertinence of each piece of literature for the purposes to which the student, scholar, and researcher wished to put the material. This is what the academic world asks of our libraries. Educators and librarians can be the planners. Electronics engineers must be active participants.

Some idea of what can be done is happening at Michigan's newest college, Grand Valley State, near Grand Rapids. For this institution Sol Cornberg has provided the latest in audio-visual equipment. The library includes 256 carrels, each outfitted with a microphone, two speakers, an eight-inch television picture tube, and a telephone dial. This plan makes available to the student any information stored in a "use attitude" or repository. Carrels could be placed anywhere, Mr. Cornberg points out, and need not be confined to the library.

Mobility of recorded knowledge is of particular importance as enrollment growth means physical facilities on the campus spread over larger and larger areas. The newer the residence halls on our campuses, the further they are from the library. By remote control, it should be possible to bring the information from the library to the student at his study area by means of wirephoto or closed circuit television. The latter might fit well into the student's learning habits. In most homes the youngster who used to curl up with a book has been replaced by one who stretches out on the floor in front of a television screen.

Electronics can do for education, learning, and research what it is doing for current events. It is possible for me to sit in my home and see—even as it happens—a gathering at Checkpoint Charlie in Berlin. I can watch—as it takes place—the Ecumenical Council in Rome. Recently I was able to see—as they contested—events in the 1964 Olympics at Tokyo.

Science, education and libraries can do the same thing for the printed word. It is in the realm of possibility that a student, professor, or researcher at Western Michigan University or at the University of Pittsburgh, or anywhere, could, through the magic of electronics, have access to needed material wherever it might be located. This science can do and it should be made possible at a feasible investment and cost of operation.

Knowing what we ask of our libraries, the attention of scientific minds can be directed to making such service a reality. With the assistance of competent staff people, library material can be classified, its relative pertinence to all other material noted and, in certain instances, recorded on tapes or disks in the interest of space saving. Means for making it available instantly by electronic control would be an integral part of any such system.

By no means does the use of scientific wonders suggest that our libraries become pushbutton operations. The type of librarian of whom Dr. Seibert dreamed would be of even greater importance. The human element would continue to be a prime consideration in developing, administering and servicing an outstanding library. Electronic assistance would allow time for in-depth performance of many library duties.

What we ask of our libraries will not happen tomorrow. We are looking ahead, but we must remember that the future is the present almost before we realize that the present is history. Man has ventured into outer space and is preparing for exploration of the moon. Rapid dissemination of the knowledge stored in our libraries is no less important, although not as spectacular. To science technology, the challenge is to help make our libraries current with this age of the atom and space travel so they can do what we ask of them before millions and millions of dollars are spent on new buildings which could become obsolete almost as they are opened.

REFERENCE

1. Russel H. Seibert, "Status and Responsibilities of Academic Librarians," *College and Research Libraries*, vol. XXII (July 1961), p. 255.

II. ANALYSIS OF THE FIELD

4

Forms of Input (Signals Through Nonnumeric Information)

ROBERT M. HAYES
University of California (Los Angeles)

INTRODUCTION

Traditionally, information systems have been characterized in terms of their dynamic properties, their internal decision processes, their information structure. Here, however, I am concerned with a somewhat different aspect—the form of the source of the basic data. We are all generally familiar with how diverse these sources can be—photographs, electroencephalographs, radar signals, audio and video recordings, telemetry, printed characters, punched media. My aim is to present these various sources within the framework of an integrated picture, based on two characteristic aspects of input—the one of *dimension* and the other of *formalization*.

The content of this talk can thus be summarized rather quickly: fundamentally, natural phenomena are multifaceted, both physically and intellectually. As a result, they are to some extent more complex than the processing equipment in an information system is capable of handling. To provide an acceptable input to the information system, some method must be used to reduce the natural complexity to the level of mechanical processes. We do this in a physical sense by reducing the dimension of the source; and we do it in an intellectual sense by increasing the degree of formalization in the source.

Before discussing these two aspects in detail, however, I ought also to comment concerning some other factors which, to a large extent, I am ignoring. Specifically, although the *physical* form of the input medium and the technology for recording on it are clearly most significant considerations in system design, they are not ones which really represent any intellectual problems. Thus, whether the input is from digital magnetic tape or punched cards may well determine how rapidly information can be processed or exactly what type of equipment will be used,[1] but it will not really affect what can be done with the information once it has been input, or what processing difficulties will be encountered in doing it.

Similarly, there are many technical problems related to the form of input which are involved in the actual handling of the information during the input process itself—problems in buffering, in code conversion (IBM twelve-bit code to internal six-bit code, for example), in format conversion (parallel to serial, for example), in timing and control.[2,3] Again, these are extremely significant in the actual design of the hardware system—and even, to an extent, of the programming[4]—but they also do not represent limitations on what can be done with the data once it has been input, or what processing difficulties will be encountered in doing it.

On the other hand, the two aspects I am concerned with today are fundamental in determining what can be done and how difficult it will be. Reduction in complexity is achieved by eliminating information content and by breaking up relationships implicit in the original data, which cannot be encompassed in the simplified data. The one prevents the information system from deriving results which depend upon the lost information; the other forces the information system to reconstruct the lost relationships.

CHARACTERIZATION OF INPUT BY DIMENSION

This aspect of the form of input views information in terms of its dimensions—of value and of space. For example, a photograph provides one or more dimensions of value (one dimension with a gray scale, several with a full color scale including hue, intensity, and brightness) as functions on a two-dimensional space; an audio recording provides a single dimension of value as a function on a one-dimensional space, etc.

A digital computer can handle only zero-dimensional data—sets of single numbers—and can therefore represent more dimensions only by the sequencing of those numbers. Present-day analogue computers are able to accept a single dimension of value—at least, on a single channel—on one dimension of space, by substitution of time for it. Recently, several "hybrid" machines have been developed which combine the continuous function processing of the analogue computer with the control and logical capabilities of the stored program digital computer.[5,6] This immensely extends digital computer capabilities, but still, more dimensions of space can be represented only by sequencing of the functions.

One can in principle visualize a type of processor capable of accepting information in two space—perhaps the photographic "dodger" is a primitive version of such a device.[7] But lacking such a capability, for the present multidimensional phenomena such as photograph must be processed by an input which provides some mechanism for reducing the dimensions to zero, or one. The process for doing so is conceptually clear: the data must be sampled at intervals in one dimension and scanned through the other dimensions. The result is a representation of a function on two dimensions, for example, by a sequence of functions on one dimension, where each function in the sequence represents a slice through the original function. By a succession of such a sampling and scanning—in each of the original dimensions—the data is ultimately reduced to simply a succession of numbers.

THE HARDWARE FOR SAMPLING AND INPUT

Obviously the simplest level of input, at least in the framework of our present discussion, is that which concerns the entry of discrete, essentially digitized data—alphabetic, numeric, binary. The variety of the corresponding input devices is almost too familiar,[8] but for the sake of completeness let me briefly review them: punched tape and corresponding tape punches and readers;[9,10] punched cards and corresponding card punches and readers;[11] digital magnetic storage, with a few types of recorders and many handlers and readers;[12,13] photographic binary recording and a few readers of it.[14–17] Summaries of the characteristics of most of the available commercial devices are listed in Tables 12, 13, and 14 of Becker and Hayes.[18]

Since these devices virtually all require manual entry at some point, much effort has gone into the development of mechanical devices to convert essentially digital information from nondigital form (such as printed images or pcm magnetic recording) into digital form.[19] But clearly, at this point, we are dealing with precisely the kind of multidimensional problem I have defined.

At the next level of complexity, the source is one-dimensional—in value, that is—and the input process requires conversion of analogue information into digital form. The variety of devices here, while perhaps not as familiar as the strictly digital equipment, is certainly not revolutionary.[20–25] The precise form from which any one of them takes is in large part a function of the nature of the source material—electronic "ramps," pulse counters, digitizing disks,[26] etc. In each case, the result can be considered as a "sampling" of the analogue signal at quantizing intervals. Traditionally, this has been viewed in terms of "round-off" error and its effects have at best been treated statistically.[27]

It is when we come to the next level of complexity, the continuous function of a single variable—usually time—that the applications become most interesting. In fact, virtually all of modern communication theory and control system theory is oriented toward this type of situation.[28–33] The equipment for sampling continuous signals is usually integrally associated with the digitizing equipment mentioned above.[34,35] However, in principle, one can visualize hybrid (analog-digital) computers which would function on samples from an original continuous signal source. For example, a computer memory of analogue form—supplementing the digital data and program memory—could store samples of varying size, which might later be further sampled and digitized under program control.[36]

The most general problem that seems within the present state of the art is that of handling images. For example, character reading equipment of the kind I have previously mentioned now exists, and several methods for analyzing the data resulting from them have been developed.[37,38] Probably the most significant applications at this level of complexity are just now beginning to appear.[39–55] The use of flying-spot scanners, previously applied to dodging and other methods of image enhancement, offers a powerful tool for digitizing images.[56]

The generalization of this concept of sampling to the case of three spatial dimensions is probably not a feasible concept as such. However, if we are content to accept some type of stereoscopic effect, there is existing electronic equipment which looks at two stereo photos with something like depth perception, follows terrain contour lines automatically, and traces out contour-line drawings.[57] The resulting electrical signals represent the images at cuts through the three-dimensional surface. Since the data about the terrain is in electronic form, as output from a cathode ray tube, it could be fed directly into a computer and used for terrain analysis without manual intervention.

In summary, the variety of input forms extends from simple keypunched data to digitized samples of analogue signals, to samples of continuous functions, to scanning of photographs and other images—and perhaps eventually to even more dimensions.

THE MATHEMATICS OF SAMPLING

Now there is nothing startling in this view of the forms of input. It is something which we all recognize intuitively and, in fact, have come almost to accept for granted. On the other hand, the consequences of this view are by no means obvious. In the case of digitization, these consequences would presumably be derived from an adequate theory of roundoff error. In the case of sampling of functions on one dimension, the development of a theory has had profound importance to information, communication, and control systems. The development of a comparable theory for image sampling will, I think, have similarly profound importance to our understanding of information processes. It therefore seems worthwhile to review the theory of the measurement of power spectra, particularly for the insight it may give to the problems which arise when we consider sampling of functions on more than one dimension.

This theory is based upon the concept that, while information may be conveyed by a particular signal (or function of time), this is solely because of the statistical properties characterizing it and the class of possible signals from which it comes. (Such an approach is, of course, consistent with the concepts of "communication theory," although it departs greatly from our intuitive concepts of information in its response-producing role.) The statistical properties we will review are not the only relevant ones, but they are usually the most useful ones. In particular, in almost every signal analysis problem, the autocovariance function, or its Fourier transform, the power spectrum, will be of prime importance.

Fundamentally, the power spectrum is based on the representation of the signal as a Fourier series; in this context it provides a picture of the relative contribution of each periodic component to the signal of interest (in fact, historically, power-spectrum analysis was called periodogram analysis).[58] From our standpoint, the significance of spectrum analysis lies in the insight it provides into the effects of sampling. Specifically, those effects are twofold: First, sampling limits the frequency which can be recovered to less than $\frac{1}{2}\Delta$, where Δ is the sampling interval.[59] And second, not only is it impossible to determine the contribution due to higher frequencies; in addition, the effects of these higher frequencies, through "aliasing" or "folding," alter the values of those frequencies which are within the limits. The significance of these effects has been well summarized by Blackman and Tukey:[60-61]

> We may logically and usefully separate the analysis of an equally spaced record into four stages—each stage characterized by a question:
>
> (a) Can the available data provide a meaningful estimated spectrum?
> (b) Can the desires for resolution and precision be harmonized with what the data can furnish?
> (c) What modifications of the data are desirable or required before routine processing?
> (d) How should modifications and routine processing be carried out?

The answer to the first question depends upon the spectrum of the source data; the response of the measuring (or sampling) instruments; the nature of the errors; and, as we have mentioned, the sampling interval. In particular, they will determine whether the effect of aliasing or of noise is so great as to make the data almost wholly useless.

The answer to the second question depends upon the resolution and accuracy desired, compared with the amount of data available and the number of separate pieces into which it falls. The answer to the third question depends upon the range of frequencies over which the spectrum is desired and estimates of the probable distribution of them, particularly with respect to the effects of folding. The answer to the fourth question involves the details of the technical processes of analyzing data of this kind and can be found in the Blackman and Tukey reference.[62]

It would be nice if the theory for sampling of functions on one dimension could be easily extended to two or more dimensions. For example, in traditional communication theory, the source is normally taken as a *sequence* of signals. This may be an appropriate view for an audio recording, for example, but not for a photograph.[63,64] To extend this traditional theory requires definition of basic functions comparable to the trigonometric, say, on two-dimensional regions, followed by the two-dimensional integral transforms comparable to the Fourier transform.[65] Unfortunately, two factors serve to complicate the situation: First, functions of the two variables are just inherently more complicated than functions of one variable, both as individual functions and more significantly as limits of sequences of functions.[66,67] And second, while the process of sampling a function on one dimension does not necessarily alter existing relations among values, the same process applied to a function on two dimensions *must* do so. The first factor can certainly be handled by appropriate extension of information theory and Fourier analysis to functions of several variables, but the second factor is fundamentally different.

In a very real sense, it is the second factor with which we will be concerned in discussing formalization, since it is formalization which provides the mechanism by which to define and easily to reconstruct relations existing in the original data. If we are to handle Gestalt with a digital computer, it must be through the formalization of the relationships implied by it.

THE FORMALIZATION OF INPUT

While sampling provides the method for reducing the dimensional complexity of natural phenomena, formalization is the method for reducing the intellectual complexity. I wish to propose a quantitative measure of the degree of formalization in a set of records. To do so, consider a record of N bits. The question asked is, How many different things can be represented by such a record? The answer, of course, is simply 2^N as a maximum. But now, suppose we format that record (structure it and formalize implicit relations). To be specific, we will divide it into f separate fields of n bits each. Can we now describe, and quantify, a measure of formalization? To answer this question, we still use as our criterion the number M

of different things which can be represented by such a record and then measure the degree of formalization by

$$C = \frac{\log M}{N}$$

For example, a fixed format allows the same n bit configuration to represent a different code when used in each of the f fields. Hence, $M_o = f \cdot 2^n$ and

$$C_o = \frac{\log f + n}{fn} = \frac{1}{f}\left(1 + \frac{\log f}{n}\right)$$

If we reorganize the record into one field of fg bits and f fields of $n - g$ bits, the first can allow the specification of the format, from a set of 2^{fg} formats, for the particular record; then within each format each $n - g$ bit configuration can represent a different code when used in each of the f remaining fields. Hence $M_2 = 2^{fg} \cdot f \cdot 2^{n-g}$ and

$$Cg = \frac{\log f + fg + (n - g)}{fn} = \frac{1}{f}\left(\frac{n - g}{n} + \frac{fg}{n} + \frac{\log f}{n}\right)$$

A different approach is to allow a set of role indicators, say 2^g; then the number of possible formats is again 2^{fg}. Each field will then have $n - g$ bits left for definition of a code within the format and within the role described by its role indicator. The total number of different codes is then $Mg = 2^{fg} \cdot f \cdot 2^{n-g}$ and

$$Cg = \frac{\log 2^{fg} \cdot f \cdot 2^{n-g}}{fn} = \frac{1}{f}\left(\frac{n - g}{n} + \frac{fg}{n} + \frac{\log f}{n}\right)$$

The effective power of either the format definition or the role indication approaches is therefore effectively the same. The difference in practice is solely one of processing convenience.

Normally, of course, we think of the number of formats 2^f or the number of classes of codes, 2^g, which the role indicators define, as relatively limited; but as g gets large and equals n, each configuration becomes a class unto itself. The result is the concept of "implicit" formats, where each n-bit configuration defines a table describing the formats in which it can occur, in terms of its occurrence in a given field. The actual format for a given record is then the logical intersection of the allowable formats for the configurations in each field. Then

$$Mn = 2^{fn} \cdot f$$

and

$$Cg = \frac{\log 2^{fn} \cdot f}{fn} = 1 + \frac{\log f}{fn}$$

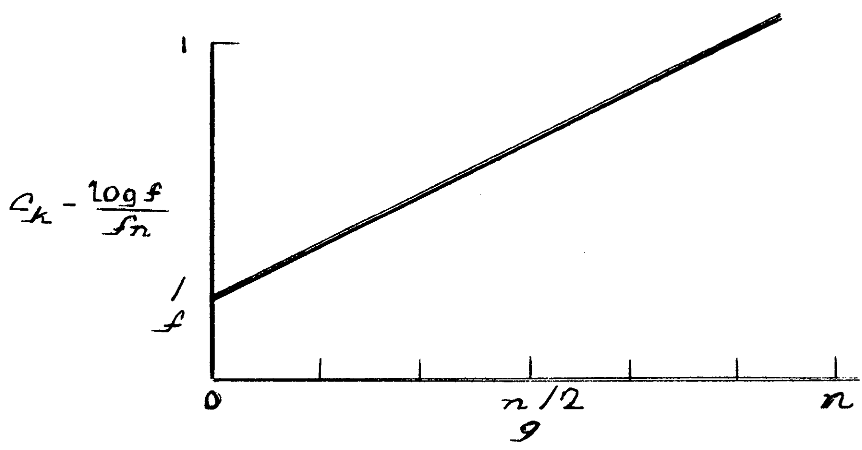
(Parenthetically, it might be asked how fn bits are able to allow definition of more than 2^{fn} different things. The point, of course, is that a record describes a relation among the f different fields, and although the number of relations among them cannot be more than 2^{fn}, the number of different codes being related certainly can be. Another parenthetical comment is that the number of codes in each case is a maximum. In practice, the actual number of codes will be very much less.)

The result is clear: given a record of fn bits, the degree of formalization of it is measured in terms of a single parameter, g—which can be interpreted either as defining the number of formats or the number of classes of terms—by the function

$$C_k = \frac{\log f \cdot 2^{fg} \cdot 2^{n-g}}{fn}$$

$$= \frac{g}{n} + \frac{n-g}{fn} + \left(\frac{\log f}{fn}\right)$$

Graphically:



From a practical standpoint, the significance of g is that it represents the number of different tables which must be stored and referenced in order to determine the meaning of codes within the record, and thus of the record as a whole.

Incidentally, this entire line of argument can be generalized, in very obvious ways, to include the effects of variable length fields and variable-length formats. On the other hand, it should be recognized that the programming problems in such generalized formats are enormously greater.

Now, turning to the relationship between input and format, it seems evident that complex phenomena occur at high levels in the spectrum of formalization which I have defined. A sentence, a photograph, a signal—each is at least at the level of an implicit format (in the sense I have defined it), depending heavily on context for both form and meaning. It therefore is difficult, if not impossible, for a computer to handle them without introduction of formalization, either through dictionaries of allowable forms or through external processing into a standard form.

To implement each of the stages in format formalization therefore requires the introduction of a dictionary—of the codes, of the formats, of the role indicators, of the terms themselves. In fact, the concept of format dictionaries may well be a fundamental one in the formalization not only of format but even of meaning. In particular, any format can lead to a nesting of formats—the terms appearing at the one level can imply formats which themselves consist of terms implying further formats, etc. Such a cascading of formats leads to further generalization of the format concept to even higher levels of complexity.

A final question should be discussed: How do we create a formalization? I think that the method of formatting provides one useful picture, but it's not the only one. Several approaches to different aspects have been proposed, each representing a variation of the mathematical concept of decomposition—or analysis into *fundamental*, critical components. For example, methods for file organization (classification) based on decomposition of the association matrix have been suggested.[68,69] At least one concept for decomposing item structure based on combinatorial assignment has been suggested.[70,71] The usual lattice model for vocabulary structure implies the possibility of lattice decomposition for creating a facet analysis.[72]

INTERNAL PROCESSES OF SYNTHESIS

Although internal processing as such falls outside the scope of this talk, there is such an intimate relationship between it and the basic input that I want to comment on that relationship. For example, data indentification is trivially simple if the input is well formalized (formatted), and can be extremely complex otherwise. File organization, similarly, is almost self-evident with formatted data and not at all evident with essentially free text. Therefore, the extent to which the input is formalized will directly affect the complexity of the internal processes. Now, this may be self-evident, but it is not at all evident how we choose the proper balance between formalization of input and complexity of internal processing. In the field of information retrieval, for example, investigation has tended to concentrate on either the highly formalized end of the spectrum—characterized by the several existing file management programs—or at the essentially implicit formats represented by language translation. Although much work has gone into definition of role indicators of various kinds, little has been done on the definition of flexible formats. I suggest that, because of the problem of balancing external formalization and internal complexity, serious consideration be given to the format approach.

With respect to the other factor in input—the dimensional one and the necessity for sampling—similar comments can be made. Much of the difficulties in character reading and pattern recognition are a direct result of sampling. It seems important therefore to develop an adequate theory for this area. One exists for signals, but for two-dimensional images it is a different matter. Again, the significance of the relationship between sampling and internal processing may be self-evident, but the mathematics of it—at least for images—is not at all self-evident.

SUMMARY

In summary, input, as I have considered it, is a process of transforming the physical and intellectual complexity of physical phenomena into simple forms suitable for processing by a computer. The methods for accomplishing the transformation are, respectively, sampling and formalization. Their characterization in mathematical terms is an essential first step to the understanding and solution of basic problems in the handling of information. My intent here today has been to describe these two aspects and indicate some directions in which the mathematical characterizations may develop. I wish particularly to emphasize the importance which image processing will play in the years to come and the value of formatting as a picture of formalization.

REFERENCES

1. Gibbons, James, "How Input/Output Devices Affect Data Processor Performance," *Control Eng.* (July 4, 1957), pp. 97–102.

2. Blumenthal, E., and F. Lopez, "Punched Card to Magnetic Tape Converter for UNIVAC," *Review of Input and Output Equipment Used in Computing Systems—Joint AIEE-IRE-ACM Computer Conference*, December 1952.
3. Leiner, Alan L., "Buffering Between Input-Output and the Computer," *Review of Input and Output Equipment Used in Computing Systems—Joint AIEE-IRE-ACM Computer Conference*, December 1952.
4. Frank, W. L., et al., "Programming On-Line Systems," *Datamation*, vol. 9, no. 5 (June 1963), pp. 28–32.
5. Lee, R. C., and F. B. Cox, "High-Speed Analog-Digical Computer for Simulation," *IRE Trans. PGEC* (June 1959).
6. *Proceedings of the Combined Analog-Digital Computer Systems Symposium*, Philadelphia, December 16–17, 1960.
7. Logetronic, for example, a type of electronic dodging equipment,
8. Aronson, Milton H., *Data Storage Handbook* (Instruments Publishing Company, Inc., 1962).
9. "All about Paper Tape," *Datamation*, vol. 5 (May-June and July-August 1959).
10. Perlman, Justin A., "Data Collection for Business Information Processing," *Datamation*, vol. 9, no. 2 (February 1963), pp. 54–58.
11. Gruenberger, Fred, *Computing Manual*, University of Wisconsin, 1953.
12. Kaufman, E. N., "Digital Signal Conversion," *Inst. and Cont. Sys.*, vol. 37, no. 2 (February 1964), pp. 117–119.
13. Wright, R. E., "How to Make Computer Compatible Tapes," *Control Eng.*, vol. 9, no. 5 (May 1962), pp. 127–129.
14. Recordak, *The Miracode System of Automated Information Retrieval*, Eastman-Kodak, 1964.
15. FMA, *User's Guide*, 1962.
16. Kuipers, J. W., et al., "A Minicard System for Documentary Information," *American Doc.*, vol. 8, no. 4 (October 1957), pp. 246–268.
17. Shaw, Ralph R., "The Rapid Selector," *J. Doc.*, vol. 5 (1949), pp. 164–171.
18. Becker, J., and R. M. Hayes, *Information Storage and Retrieval: Tools, Elements, Theories* (Wiley, 1963), pp. 308–316.
19. Fischer, George L., et al. (eds.), *Proceedings of the Symposium on Optical Character Recognition*, Washington, D.C., January 15–17, 1962 (Spartan, 1962).
20. Bower, C. G., "Survey of Analogue to Digital Converters," National Bureau of Standards Report #2755 (July 1953).
21. Burke, H. E., "Survey of Analogue-to-Digital Data Converters," *Review of Input and Output Equipment Used in Computing Systems—Joint AIEE-IRE-*
22. *ACM Computer Conference*, Dec. 10–12, 1952, pp. 98–105.
Fischer, P. P., "Analogue-to-Digital Converters," *Electro-Technician*, vol. 69, no. 3 (March 1962), pp. 165–168.
23. Fleischer, A. A., and E. Johnson, "Analogue-to-Digital Converter Capable of Nanosecond Resolution," *IEEE Trans. Nuclear Science*, NS-10 (1) (January 1963), pp. 31–35.
24. Klein, Martin J., "Analog-Digital Converters: An Evaluation," *Datamation*, vol. 4 (May-June 1958).

25. Suskind, Alfred (ed.), *Analog-Digital Conversion Techniques* (M.I.T. Press, 1957).
26. Husbey, H. D., and Granino A. Korn, *Computer Handbook* (McGraw-Hill, 1962), pp. 1829*ff.*
27. Beers, Yardley, *Introduction to the Theory of Error* (Addison-Wesley, 1957).
28. Barnes, John E., Jr., "Sampled Data Systems and Periodic Controllers," in *Handbook of Automation, Computation, and Control* (Wiley, 1958).
29. Blanger, C. G., "Sampled-Data Techniques in Complex Simulation," *Eastern Simulation Council*, Jan. 5, 1959.
30. Cherry, Colin, *On Human Communication* (Wiley, 1957), pp. 121*ff.*
31. Linvill, William K., and John M. Salzer, "Analysis of Control Systems Involving Digital Computers," *Proc. IRE*, vol. 41 (July 1954).
32. Ragazzini, J. F., and G. F. Franklin, *Sampled-Data Control Systems* (McGraw-Hill, 1958).
33. Salzer, John M., "Frequency Analysis of Digital Computers Operating in Real Time," *Proc. IRE*, vol. 42 (February 1954).
34. Hakimoglu, A., and R. D. Kulvin, "Sampling Ten Million Words per Second," *Electronics*, vol. 37, no. 6 (Feb. 7, 1964), pp. 52–54.
35. Hollitch, Robert S., and Albert K. Hawkes, *Automatic Data Reduction, A Catalogue of Devices Useful in Automatic Data Reduction*, WADC Technical Report #54-519, Part II, Armour Research Foundation (November 1954).
36. Walli, Charles R., "Quantizing and Sampling Errors in Hybrid Computation," *Proc. FJCC* (Oct. 27–29, 1964).
37. Greanias, Hoppel, Kloomok, and Osborne, "Design of Logic for Recognition of Printed Characters," *IBM Journal of Research and Development.*
38. Johnson, J. R., and N. Rochester, "A Simulated Machine for Recognizing Printed Numerical Characters by a Method of Lakes and Inlets," *IBM Journal of Research and Development*, Poughkeepsie, Nov. 30, 1953.
39. Clapp, L. C., "Optical Information Processing," *Int. Sci. and Tech.* (July 19, 1963), pp. 34–41.
40. David, E. E., Jr., and O. E. Selfridge, "Eyes and Ears for Computers," *Proc. IRE*, vol. 50, no. 5 (May 1962), pp. 1093–1101.
41. Davis, Malcolm R., and T. O. Ellis, "The RAND Tablet: A Man-Machine Communication Device," *Proc. FJCC* (Oct. 27–29, 1964).
42. Duncan, A. H., "Automatic Picture Digitizer," *Brit. Comm and Ele..*, vol. 9, no. 9 (September 1962), pp. 676–679.
43. Fulton, Roger L., "Visual Input to Computers," *Bus. Datamation*, vol. 9, no. 7 (August 1963), pp. 37–40.
44. Hargreaves, Barrett, et al., "Image Processing Hardware for a Man-Machine Graphical Communication System," *Proc. FJCC* (Oct. 27–29, 1964).
45. Holmes, W. S., and H. M. Maynard, "Input-Output Equipment for Research Applications," *Nat. Electr. Conf. Proc.*, vol. 18 (1962), paper 1539, pp. 509–517.
46. Jacks, Edwin L., "A Laboratory for the Study of Graphical Man-Machine Communication," *Proc. FJCC* (Oct. 27–29, 1964).

47. Julesz, Bela, "Binocular Depth Perception and Pattern Recognition," in Cherry, Colin (ed.), *Information Theory* (Butterworth, 1961).
48. Kirsch, R. A., et al., "Experiments in Processing Pictorial Information with a Digital Computer," *Proc. FJCC* (December 1957).
49. Krull, Fred N., and James E. Foote, "A Line-Scanning System Controlled from an On-Line Console," *Proc. FJCC* (Oct. 27–29, 1964).
50. Larsen, K., "Automatic Readings and Interpolation of Strip Charts and Film Records," *Proc. ISA*, Reprint 63 (March 1963).
51. Maffi, C., and E. Marchesini, "Semiautomatic Equipment for Statistical Analysis of Air Photo Linears," *Photogramm. Eng.*, vol. 30, no. 1 (January 1964), pp. 139–141.
52. MeFee, R. H., "Information Processing in Infrared Systems," *J. Soc. Photog. Inst. Engrs.*, vol. 2, no. 1 (Oct.-Nov. 1963), pp. 14–16.
53. Shaler, D., "Ultrahigh-Speed Microfilm Facsimile System," *IEEE Trans. Comm. Electr.*, vol. 82, no. 66 (May 1963), pp. 201–207.
54. Shonnard, J. R., "High-Speed Communication of Graphic Intelligence with Hard Copy Readout," *AIEE Trans. Comm. Elect.*, vol. 81, pt. 1 (61) (July 1962), pp. 176–178.
55. "Eye for Computers Is Quick as a Wink," *Business Week* (Sept. 19, 1964), p. 80.
56. Stein, Edward S., et al., *Factors Influencing the Design of Original-Document Scanners for Input to Computers,* National Bureau of Standards, August 19, 1964.
57. *Stereomat*, Benson-Lehner, Inc. (no longer manufactured, but some devices were delivered).
58. Carslaw, H. S., *Fourier Series and Integrals* (Dover, pp. 326*ff.*).
59. Whittaker, J. M., *Interpolatory Function Theory* (Cambridge, 1935).
60. Blackman, R. B., and John W. Tukey, *The Measurement of Power Spectra*, (Dover, 1958), pp. 50–52.
61. Tukey, John W., "Estimation of Power Spectra," in Rudolph E. Langer (ed.), *On Numerical Approximation* (Wisconsin, 1959).
62. Blackman, R. B., and John W. Tukey, *op. cit.*
63. Schade, Otto, "Evaluation of Photographic Image Quality and Resolving Power," *J. SMPTE*, vol. 73, no. 2 (February 1964), pp. 81–119.
64. Shaw, R., "The Application of Fourier Techniques and Information Theory to the Assessment of Photographic Image Quality," *Photo. Sci. Eng.*, vol. 7, no. 5 (Sept.-Oct. 1962), pp. 281*ff.*
65. Hu, M. K., "Visual Pattern Recognition by Moment Invariants," *IRE Trans. Inform. Theory*, IT-8, vol. 2 (February 1962), pp. 180–187; F. L. Alt, "Digital Pattern Recognition by Moments," *J. Assoc. Comp. Mach.*, vol. 9, no. 2 (April 1962), pp. 240–258; M. K. Hu, *Theory of Adaptive Mechanisms*, Syracuse University (December 1963), pp. 16–65.
66. Courant, R., and D. Hilkert, *Methods of Mathematical Physics*, vol. I and II (Interscience, 1962), esp. Chapter VI and Appendix to v. II.
67. Morrey, Charles B., Jr., *Multiple Integral Problems in the Calculus of Variations and Related Topics* (U.C. Press, 1943).

68. ADI-NBS, *Symposium on Statistical Association Techniques* (Washington, D.C., March 1964).
69. Borko, H., and M. D. Bernick, "Automatic Document Classification," *J. Assoc. Comp. Mach.*
70. O'Conner-Schultz, "Scan Column Index."
71. Prywes, *Multi-List Processing* (U. of Pennsylvania, 1963).
72. Hillman, Donald J., *Study of Theories and Models of Information Storage and Retrieval* (Lehigh, 1963).

5

Signals and Numerical Information—Interpretation and Analysis

A. J. PERLIS

Carnegie Institute of Technology

Inside a computer all information is numerical and this implies that its use and evaluation must be accomplished by numerical transactions within the machine. These transactions are generally organized and described by what we call programming languages.

It is a truism that a fool can ask more questions in an hour than all the wise men in the world can answer in a hundred years. The whole problem of information retrieval, I think, is related to that particular point. In this case, the wise men are the set of programming and formatting techniques that we are capable of bringing to bear, and the fools are the (so far, fortunately) largely mythical people who hope to sit at computers and ask any old question that comes along. Nevertheless, starts are being made in various places on various small problems to solve the problem of retrieval of information in those areas. Some of them are rather trivial, others more complex. All are partial and will undoubtedly remain so for the foreseeable future.

Mention was made of language translation by the previous speaker. The information that is in so far, from all fronts engaged in doing language translation on computers, is that effectively no progress has been made toward producing usable translations for technical people in various fields. The reason for this is, I think, summed up in a nutshell in the two words "context" and "semantics," and how they relate to one another. Semantics, or what meaning we give (either operationally or purely intellectually) to information received from some source has not yet been suitably formalized either in the field of logic or, unfortunately, in the field of computation, so that obtaining information as to the meaning of processing within a computer is a very difficult proposition, and at the moment we can say that very few positive results have been obtained outside of a few restricted areas. These restricted areas are those where the classification of information has been going through a sifting process for centuries, and I refer to certain restricted parts of mathematics. In these restricted parts of mathematics where the information transformations are arithmetic in nature, an increasingly useful amount of information is obtained and it is the success in this area that has led people to predict the ultimate success in other areas which at the moment share nothing in common with the first area, except possibly that both are the products of human minds.

What I am about to say contentwise has to do with what experience and success we have had in processing information in these rather restricted areas. The basic problem is that too many of the approaches to computers are what we might call problem-oriented. Now, what I mean by problem-oriented is that someone or some group assumes that a problem can be described in a certain way, such as "A library can be operated on a computer—it's all bits, has fast input, has a higher-speed printer than any I've ever seen before. We can get photographic output, and in a few years, I'm told, we'll get photographic input."

We have a problem—how do we store a library in a computer? Stated in this way, such problems always can obtain partial solutions, which ultimately fall far short of the dreams of the proposers, but are at the extreme limit of the abilities of the people who actually achieve them. Some people look at tasks not as problems but as procedures, and all the success we have had in computation to date has been because certain specific areas have been attacked from the standpoint of obtaining procedures. Indeed, all computation is based on procedures. It is only when we are able specifically to describe procedures that we get any mileage at all out of computers.

How do we describe procedures? The first place to start is with the concept of data representation or format, in the words of the previous speaker. This, I think, is the key to all successful use of a computer in any problem, be it information retrieval, Monte Carlo work, or simulation of traffic systems. The basic key is data representation.

What is the data that we choose to use in a computer? How big is it? What is its precision? What do we wish to do with it?

In the outside world we have one picture which is a very heavily contexted dependent picture of information and hence data representation which is constantly organized and parsed, if you will, by our mind. The first stage in the use of the computer is in effect to deduce the appropriate, approximate information representation that is going to take place inside the computer. Now immediately we can eliminate this problem by stating that all information inside the computer is a string of zeroes and ones.

But it is precisely because we do not have to say that, and can still have the computer process at a rapid rate, that we are able to make progress in the use of computers in information problems. For example, real numbers are abstractions in the outside world and approximations in the inside world, but in certain problems they are the natural data to be used in describing the data—the natural format to be used for describing data. For example, scientific computations are of this form. Those of you who have had sufficient experience in computing will recall the history of computation where we started out with numbers that were merely integers; then we had an internal set of transformations programmed, if you will, which allowed us to represent approximate real numbers by pairs of integers, and thereafter deduced a set of operations on these pairs that were natural for the operations on reals. All of the internal transformations were then procedurized once and for all.

Those who came along later and worked on alphanumeric information were aware of the fact that these too were represented as strings of digits which, however, could be procedurized as soon as we knew what the operations were that we wished to perform upon them. Lately, we find that computers are being considered—one computer has even been constructed—whose basic data representations are what we call list-structures. The class of problems for which we need these structures, as the natural internal data, is a more complex, and indeed, a newer class of problem than those for which real numbers were sufficient.

Other forms of data representations will be found in time. It may indeed turn out that the basic importance of a computer in the intellectual life of mankind is through the fact that it places a problem before the mathematicians of our society to develop a whole host of new arithmetics—arithmetics which allow us to manipulate in the same way that the piano postulates provide us with the basis for manipulating an arithmetic or, if you will, ordinary integer arithmetic, that will enable us to manipulate in the same natural easy way trees of information and list-structures of information which at the present time are handled by means of nonformalized procedures. The real intelligent use of computers in information retrieval and other problems will await the solution of at least this problem.

Now for each data structure that we happen to deduce as appropriate for our problem there is a natural set of operations which seem to occur and the understanding that one has of a particular problem is, in a large part, determined by how totally he is able to define the set of natural operations. In arithmetic we all know what they are. When we get over to more complex structures like matrices or lists, we find that other operations have to be added to the compendium in order for us to say in a precise and in a brief way the basic computations that we wish to have performed.

It is the job of the programmer at the present time to find the set of internal procedures which will carry out the transformations from one representation, the natural one that we as users would like to possess, and the unnatural one, the one that the computer actually does possess. All information processing inside computers—or almost all—is concerned with these procedure transformations.

Now once we have decided on a data structure and the basic operations for a problem, the next step, if the problem is large enough (and one which, so to speak, assures continuing support) is the definition—at least, this is the usual chain of events which takes place—of a language. I think this is the second act of intellectual import which has occurred in the past several years with respect to computers. It is the fact that we now design or invent languages almost on a moment's notice. Language, instead of being something which is studed *au naturel*, is now designed to fit a specific purpose in a computer and there is no limit to the number of such languages that can be designed for specific purposes.

Why design a language? Well, that's a good question. People who have already designed one always ask it of those who are about to start. The reason, of course, is to cut down the amount of explicit relationships we must explain to a computer if the number of such explicit relationships is large, either because it is large in a single problem or because the number of users who have to so express themselves is large. As soon as that situation occurs, along comes the need for the design of a language, just to increase the flow of communication between man and machine. These languages all follow much the same sort of path. They proceed from internal representation of desired data, to applications of the appropriate operations, to the creation of sentences, and from sentences, the creation of programs, the specification of the sequencing rules by which these programs are to be executed, the specification of a library by which these programs are to be accumulated and indexed and accessed, and the imbedding of all of this in a kind of operating system on a machine. If one looks at a large part of the intellectual effort now going on in the United States in the so-called programming area, one finds that it is involved in one or more of these areas and not much else.

What does it mean now in these terms to recognize information inside a computer? If we wanted to be very blunt, we are able to recognize information only when we can make a selection of one of two programs to be executed as a result of this recognition. Thus recognition is a selection process of one of two programs. This isn't very helpful because all complex problems are ultimately broken down this way. No complex problem would probably ever be programmed if it had to depend on such a recognition definition.

Let's consider one specific problem. There has been proposed from time to time the development of so-called information-retrieval systems. An information-retrieval system depends on, it seems to me, several things. One is a corpus. This corpus is the set of facts and relationships which is stored in the computer. Second, there is a set of allowable queries. Third, there is the processing of these queries to produce the desired information. This is really what we mean by an information-retrieval system. If we knew exactly what we wanted we would merely ask for it by its place in a directory. It is because we do not know exactly what we want that we cannot ask that.

What can we ask for? Here we come across a very critical problem—the problem of education. It is quite possible to take a mass of information and deduce from this information a set of allowable relationships and from this a syntax and a semantics of a language, in whose terms you can ask questions of this corpus and no other. Very few people have attempted to do this thus far. There have been a few first steps, but it seems to me that this is really what is required to solve the information-retrieval problem. Given the corpus, the number of relationships is ever increasing. Given the input language, it starts out simple and gets more complex, as we learn more and more about our abilities to parse these grammars. And finally, on the education side, it becomes more and more essential that we ourselves learn this language independent of English or whatever other language we use in order that we can make use of this mechanism.

Thus, when we talk about information-retrieval systems, we can break it down, I think, into several disjoint parts and several problems. First, there is a problem of accessing this corpus of information. It is clear that we do not wish to access it in most computers by direct table look-up. Somehow or other we have to derive a code for information from which we can deduce approximately where to find what we are looking for. This code will inevitably not be a constant code. It will inevitably lead to redundancies. That is, there will be several pieces of information which fall roughly in the same ballpark. It is inevitably a case that we will miss some information. No code will be perfect if it's going to be interesting. Having devised this code, there is then the problem of transforming questions, appropriately written, into sequences of codes and sequences of procedures which pick out, in some sense, the best candidate or candidates from this corpus. This means to me as a programmer that the information-retrieval problem can probably, at least in certain worthwhile instances, only be solved by both passing the corpus through a prescanner, human, and by teaching people to ask questions in a fixed and generally context-free way.

All of the experience with language translation to date has shown that we get very little information out of language translations, precisely because the computer and the processing techniques we have are context-bound and the languages on which we seem to operate are not. Several experiments in language translation and in information retrieval have given us a glimmer of hope that partial solutions can be obtained, but these solutions are going to depend in large part on the development of explicit languages in which terms we ask questions of this corpus.

I am not particularly interested myself personally in the information-retrieval problem. It is one of these big messy problems that's going to take a long time to solve, and there are lots of smaller, nicer, nonmessy problems which are more easy to solve. And it is of course the case that this field, like all others, dare not wait for formalization to commence. It should also not expect to get good solutions for some time to come. It should certainly not expect to find a solution in hardware. What recognition you are going to be able to buy and what processing you are going to be able to buy, you are going to be able to buy through programming, and not through hardware. Toward that end I would like to recommend that all of you who are in the information-retrieval field become very familiar with the subject of mechanical languages and become very familiar with the subject of statistics. The mechanical languages will teach you how to format queries and programs. The statistics will teach you how to organize a corpus.

Finally, with respect to the one big information-retrieval problem—language translation—one of the things that seems to come up as soon as you dig a little deeply into this problem is the fact that there isn't one. We find that as usual we have overemphasized an urgency which only existed by virtue of extrapolations. There does not seem to be any great shortage of human bodies to translate information these days. There does not seem to be any urgent need for machine translations on a production basis. I leave with you a question which I cannot answer, though I have a sneaking suspicion what the answer is. Is there an urgent need for total mechanical systems at this time for doing information retrieval, or is it possible that the most rational information-retrieval systems we can create at this time are complexes of man and machine in which the man part is by far the biggest and most important? There is no prior reason why all problems, merely because they are large and because they involve millions of bits of information, have to go on computers. They seem to start that way but gradually sanity and the size of the tasks cause them to be replaced.

There is a basic law with respect to computing which states that anything you want to pay enough for can be done on one or more machines, currently on order.

6

Mechanical Resolution of Linguistic Problems

A. D. Booth

Dean of Engineering, University of Saskatchewan and Professor at Large, Autonetics, Western Reserve University

The optimum information retrieval system is one which I should like to call a symbiosis of man and machine. Men do some things very well that machines do very badly. One should not use machines for such purposes. So, if you expect a champion for the machine, you won't find him here. I ought to say that in the University of Saskatchewan and occasionally in the University of London I lecture on the use of computing machines on numerical analysis. I always preface my remarks by the statement that: "Machines are the last refuge of the inept," which ought to put them into perspective.

On the other hand, having bowed to Dr. Perlis on that subject, I should dispute him when he says that no progress has been made in machine translation. This, as a matter of fact, is quite untrue. Depending on the level at which you want to consider the translations, some progress has been made. There are quite decent programs for translating English into Russian. I suspect there are some programs in the United States for scientific translation of Russian into English, and there are certainly some programs, because I was concerned with part of the writing of them myself, for the translation of French into English. These work and, if you wanted to look at the output of a machine doing this sort of work, it would be rather doubtful whether you could distinguish the output from that produced by a human being. However, I suspect that Dr. Perlis' remarks were in the nature of being provocative and not supposed to be a statement of fact.

By way of an introductory remark, I want to tell a story. It has been remarked of academics that they are good for two hours of speechifying, although somebody else remarked in the same context, "That's what *they* think." I'll try not to take two hours, but anyway, let me tell you a little story. A few years ago I was invited to read a paper at a conference that was held in a place called Alpbach in the Tyrol. This conference had some highbrow title like "Language, the World, and its Philosophy." I looked at this with horror, but it provided me with a means of getting a free holiday to a rather nice place. I said I'd go. When I got there I was completely horrified. There was a collection of very long-haired professors, obviously of enormous erudition and of a mental caliber I couldn't compete with, and I was set down to open the proceedings. Of course I noticed this beforehand and had come prepared with a text constructed by one of our computers on the subject: "Cybernetics and the World." I had programmed the computer to do the sort of thing that Shannon did originally: produce a text by taking a word from random from some page in a book on the subject of cybernetics, then finding some other page on which the same word occurred and taking the next word on that page, then going to some other page selected at random, and so on. This way I constructed twelve minutes of fairly plausible text. At the meeting, I noticed the simultaneous translators making a fine go of this and they were nodding and the audience was sitting in the front row looking intelligent and saying "Mmm, mmm, very profound." At the end of this performance, I took the parliamentary utterances of various Ministers in the British Parliament for successive days of one week and took the second sentence of each pronouncement, irrespective of the Minister. And I finished up with this. It read very well and was a really high-powered speech. Then I turned to the president of this meeting and I said, "I am sure, sir, that you will appreciate the profundity of those remarks." I am afraid that this was a bit unfair because he turned to me, and in a very audible voice said, "Yes, that was a very fine account of the subject." At this point, of course, I did the sort of thing that all comics do—I turned to the audience and said, "Well, gentlemen, you will be interested to know that there was no meaning whatever in that twelve minutes of discourse." The front row of the audience rose and left like a black cloud; the remainder of the audience were rather young people, and when we came to get our groupings of young men for the classes which we were giving later on, I am delighted to say I got about 95 percent of them. The graybeards, I'm afraid, didn't get to first base.

Well, now to come to something more serious. I think I have entertained you for five minutes; let me now deal with the subject of mechanized linguistics.

I'm going to try to give you a view of the structure of this operation because there are some important things in it, whether Dr. Perlis' remarks have much justice or not. There are some important things we can do; there are some important ideas in this field, and it's worth describing them. You'll see that at many points I make contact with some of the remarks of Dr. Perlis on things like structure. First of all, a remark about the machines themselves. I am not one of those people who believe in building gadgets. You may almost paraphrase Wittgenstein and say that whatever can be done can be programmed on a computer. Therefore, you shouldn't build a special machine. You ought to be quite sure what you want to do before you build a machine. The structure of computing machines as they exist at the moment really divides itself into two depending largely on the type of storage involved. This is rather important because whatever the future of computers is going to be, and this isn't by any means certain in some of our minds, present computers are, in a sense, unfortunate because many computers have adequate amounts of storage to contemplate attacking problems of language, but have this storage arranged in what I might call a hierarchial structure. The computers have a very small amount of very high-speed store, a rather larger amount of medium-storage sometimes, and quite often a great deal of very low-speed storage. On the other hand, there are the ultraexpensive computers, which have all of their storage on immediate access media. Now the way that you think of language in connection with a computing machine depends very largely on the structure of the machine with which you are concerned.

Actually, right at the very beginning of processing any data, whether linguistic or otherwise, derived from a list, involves deciding whether the statistics of the data are of paramount importance or whether the importance is secondary. Let me quote an example that makes this point. If you have a machine which is operating such a simple thing as a dictionary or look-up procedure there are many ways of using this, from the very simplest (which Dr. Perlis mentioned) in which you address the item of information by the code word of the unknown word, if you like. If you want to look up *et* in the dictionary, you find the code number of *et* (e.g. $e \equiv 05$, $t \equiv 20$, so that $et \equiv 0520$) and in the storage position having that code number, you find the translation *and* or whatever the equivalent is in the language you are concerned with translating it into. This type of storage is completely unworkable for very good reasons concerned with the structure of language. For example, if you take words of less than or equal to ten letters in English, it turns out the number of possible words is slightly over 10^{14}. The number of actual words in English is about 10^6. To those of you who are not clued up on these big numbers, this means that if you wrote down these words in a list, on average there would be about 10^8 blank spaces between each entry in your list of words. It would not be a good idea to have a store unit in this sort of way. This is an elementary example.

Consider next the dichotomy of storage in present machines, the fact that you can have hierarchial storage or immediate-access storage. For hierarchial stores, it turns out that probably the best way of proceeding is to consider the statistics of your word list and then to sort the input text into some order before presenting it to the computing machine. On the other hand, with random-access storage the best argument suggests that you needn't concern yourself with these statistics, you just go straight to the list and, if you have an appropriate look-up procedure, whether this is by a method which involves a treelike structure, of the sort you heard about a moment ago, or whether it involves a simple partitioning of the list doesn't matter too much. Both of these methods are workable and reasonably efficient. But you do have to know quite a bit about the machine you are going to have available in the future before you start committing yourself to large amounts of work in this particular field. This is, if you like, a preliminary word of warning.

While having said this about language statistics, or data statistics, what sort of pieces of information do you want? There exists one very general law that applies to language particularly (it was discovered, in fact, originally as applying to language) but also applies to almost any list of information one can write down in some structurable order. This law is known as Zipf's law. I don't know why it's called Zipf's law because, although Zipf ennunciated it in the 1930s and made a great stir, it was first enunciated by a Frenchman called Estoup about 1919. This Estoup law states that for ordinary language, and for a lot of other things as well (numbers of entries in telephone directories under each name, for example), if you arrange your list of entries in terms of their rank—that is, the most frequent entry has rank 1, the second most frequent, rank 2, and so on—and if for each entry in this list you put down the frequency of occurrence of this word, then rank times frequency is constant. It's a very important law for look-up procedure analysis, and for mathematicians, too. Because whatever one may think to the contrary, mathematicians have not been completely oblivious to the need of considering the effects of structure on function. One of the situations you can analyze is this. If you want to operate a dictionary, would it be a good thing to plan the dictionary so that the most frequent word in the language is the first entry in the dictionary, the next most frequent word the second entry, and so on? The problem is then to determine, for this ordering, whether or not looking up words in a frequency-ordered dictionary is better than looking in a dictionary in monotonic increasing order of word magnitude expressed as a code number. It turns out that the answer is that this dictionary is unworkable; that the normal dictionary is better used with binary partitioning. However, one of the things mathematicians got interested in was wondering if there were any laws of occurrence of data for which frequency-structured dictionaries would be better than any other variety. It turns out rather interestingly that if the Zipf-Estoup law wasn't (rank $\times$ frequency $=$ const.) but instead (rank^n $\times$ frequency $=$ const.), $n > 2$, then it is more efficient to use a frequency-order list than it is an ordinary dictionary. This is one of the sorts of information that any respectable person working in the field of language data processing ought to consider for himself before he starts. It's certainly no good going blindly to a computer, mechanizing some wonderful idea derived from hot air, and then wondering why your system is inefficient. You should investigate these efficiencies before you start. This is the basis of the remark I made earlier that the numerical calculation on computing machines is the last refuge of the inept. You can do quite a lot without using a machine, some of it purely mathematical.

We have thus decided that we must consider our computing machine and the lists to be used. That leads to discussion of what I might call the mechanics of linguistic statistics. You notice that the title of my talk (which incidentally I more or less approved because I would have hated to have been put down as talking about machine translation, in which I frankly don't believe) is "Mechanical Resolution of Linguistic Problems." It starts with the mechanical resolution of problems of linguistic statistics. Here again one begins with the problem of how to get the data into the machine. As far as I can see from the program, you're going to hear a number of ways in which data can be presented to the machine. The classical way is to present it on punched cards, and the classical way may be the best, but I doubt it. In the first place, a decent punched-card producing machine with a typewriter input costs a great deal of money, so generally you have to rent it. So for this reason, although the punched card is not a bad way of putting machines in, it certainly isn't a very economical way.

The second direct form of input is by a punched paper tape. This is very attractive because modern electric typewriters can produce tape as a by-product, so that the typist does your letters and at the same time produces a machinable record on punched paper tape. Tape is also very important in that many books are produced by the monotype process, and the monotype rolls used in producing books, can in principle, at least, be read into a computing machine.

Incidentally, on the subject of tape and cards, I might remark that of tape doesn't involve great redundancy because you don't leave a large space between words. You put a space symbol and go on to the next word. On the punched card, you have the difficulty of deciding in advance the format of the information you are putting on, and this quite often leads to the undesirable situation in which you plan for words with a certain maximum length, although many words do not have the maximum length at all. In English they have average lengths of five letters, and you are quite likely to waste quite a bit of the surface of the card. (This doesn't bother the punched-card manufacturer!)

The two other forms of input which have merit are the direct character reader and the spoken word. Many workers, including the Russians, regard character readers as very important, and certainly they are for any language which does not use a Roman script. The Russians are working on Chinese characters. So far I haven't heard the results of this work, but in 1960 they had a prototype reader.

Finally—and this sounds something like a physics text—the spoken word is a quite good method of input to computers. You have all seen things like *Shoe Box* into which you can speak the digits 0 through 9 and out of which you can obtain a suitable digital input for a computing machine. Actually, spoken-word input is probably not too useful for normal data processing but is quite likely to be useful for cataloging and stockkeeping, operations of all sorts in the areas where one does keep stock, and this goes from libraries to stocks of shoes in a shoe factory.

So much for the basic mechanisms. Now for two of the tools of mechanical language data processing. Many people say, "Let's sit down with a classical conventional dictionary and a classical conventional grammar, start from scratch, and see if we can work out a program to do a machine translation." My own concept is that the method to be adopted should be quite different. Machines are useful, whatever one may say to the contrary, in symbiosis with men; and an ideal symbiosis of machine and a man is in producing the basic material on language for use, if you like, in making a dictionary or making a grammar. Our own machine translation work has been based from the beginning on the notion that we use the machine to help us get the data which we want. Specifically, I view machine translation as a highly structured operation. The structure is twofold—the structure of the words themselves and the structure of the grammar. Machine translation works in a hierarchial process, starting with a list of words represented, from the point of view of analysis, not by a conventional dictionary starting with the word "a" in English and ending with "zymurgy," but rather by a dictionary starting with the most frequent word and then the next most frequent word and so on. If you are working out the program for a machine, it's a good idea if the first time you demonstrate the machine it doesn't fall down on the simplest sentence merely because somebody started with an obscure portion of a complicated dictionary of a technological subject. You first must produce a frequency or ordered list of words. Of course, this has been done by people like Dewey, but it pays to do it again when dealing with scientific material, and you do it on the machine. Having produced a structured list of words we then get to work putting in the relevant data about these words using a human operator and starting with the most frequent word. You then know that at any stage you are likely to deal with quite a large amount of the material in the text. The same thing goes for the grammar.

I can tell you a story here. Years ago when we were beginning to translate French into English, I went to the Professor of French at our College in the University of London and asked him what was the most frequent difference between word order in French and English. First he disclaimed any knowledge of this; then he came up with something obscure, which I have never been able to find in any French text, and which I suspect was something deriving from his speciality, Medieval French. We did eventually get the answer to this one—the most frequent ordering difference between English and French is, in fact, the inversion of the order of nouns, adjectives, and adverbs, and the next most frequent is pronoun-verb structures. We derived these pieces of information by analyzing sentences on a computing machine, using a combination of the linguist and the computer to produce this statistical data. Thus our program started off from zero on the assumption that we could do word-for-word translation (which of course we can't) and then worked its way up through an increasing list of complications—for example, the noun-adjective-adverb situation, the pronoun situation—eventually ending up in what we call MT6, which was quite a potent program. In Saskatchewan at the present time, we are applying just these principles to the analysis of the combination of English-French. English is most interesting in a number of respects, chiefly because it is the most ungrammatical language in the world, which makes it rather attractive.

I think I've talked long enough, but let's say a word or so about machine translation. We've heard something about its limitations. What sort of things can machine translation do? At various levels, I would maintain—other people's opinions notwithstanding—that machine translation can be useful. For example, if you merely translate the scientific nouns and verbs in a text, with no attempt whatever to do anything about their relation to one another, the result is very useful indeed to a human scientist. Perhaps some of you don't believe this but the fact is that many scientists who do not have access to a translating machine—I suppose this means, at present, all scientists because there are no translating machines doing this sort of work—and who are not skilled linguists start off merely by looking at the text to find what they conceive to be technical words and then looking these up in a dictionary. Quite often they go no further than this and say "Well, obviously this paper is of no interest." At this level, even word-for-word translation, with no particular assistance with the grammar, is useful. A machine *can* do it; it does at least save the scientists from looking up words in a dictionary. Of course one can go considerably further than this. If you are prepared to specify your field of interest and your language, it doesn't take too long (using the machine-man combination for the rules and the word lists) to produce a rough machine translation. There are a lot of lacunae in this. The disadvantage of statistical ordering is that the machine does not deal with all of the words. It makes no claim to do this. It will deal with the hundred or thousand or ten-thousand most frequent words, but when the word is not one of the most frequent, the machine first makes a check that something isn't merely wrong with the works (which all good machines ought to be programmed to do), then says "Well, this word is not in my list of words," and outputs it in original form with a note to some human being to look it up in the dictionary or to consult a colleague. Machines are useful at this level.

I can't help remarking as a little *jeux d'esprit* that one of the amusing things that people sometimes talk about is to do literary equality translation on machines. There are some bogus characters who say that we can do literary translation on machines, and while this is completely false in the general sense we *can* do something—and this something is quite amusing for a reason which I'll try to explain. Supposing that we want to translate Shakespeare into Goethe. We first make a list of all of the sentences that Shakespeare ever wrote, which is a fairly trivial operation, machinewise. Next we get a human being to go through this list, just as in making a telegraphic abstract or any other indexing operation, putting alongside each sentence certain category numbers which indicate the area of human endeavor into which the sentence falls—for example, boy meets girl, or boy loves girl, boy falls in love with girl, girl jilts boy, and so on. Having done this, we do exactly the same thing for Goethe, and now have two lists of sentences, each of which has associated with it some category numbers which effectively tell what the sentence is about. When we present our Shakespearian corpus to the machine, it looks up the Shakespeare sentence in the list, finds the category number, and goes to the list of Goetherian utterances. It will probably find several Goethe utterances of the same sort so it flips a coin, or, machinewise, consults a table of random numbers, picks out the equivalent of Goethe, and says "This is what Goethe says about the situation Shakespeare has described." When finished, we have exactly what Goethe said about the Shakespearean situation.

We've actually tried this on a small scale and there's one most interesting consequence. In using machine analysis of word statistics and structural occurrences, we can usually detect whether or not an author has been faked. For example, we've recently done some work on the authenticity of certain Johnsonian fragments from newspapers, in which word statistics show quite clearly whether or not he was the author of a particular fragment. When we do this particular analysis on a text constructed in the manner just described—that is, taking the actual utterances of A about the situations described by B, the interesting thing is that the word statistics in the utterances of A are now correct for A. You can no longer do literary detection on it. This is rather fascinating because it does give a means of rewriting a few sonnets of Shakespeare or a few new Shakespeare plays and getting away with it. The literary detectives won't be able to operate, because the words are what Shakespeare (or Goethe) actually wrote.

That is just an aside but it is one of the things which a study of the structure and statistics of language makes possible. It is in a real sense machine translation because we are creating an artifact. We can go even further and make the selections from Goethe rhyme in the proper way; the possibilities are endless.

Finally, I thought I ought to say something about recent work, such as that done by Bar Hillel and Chomsky, the two oracles of Israel. Bar Hillel has been described by various people as the leader of the destructive school of machine translation. He wants to knock you on the head. He goes around producing counterexamples as to why one cannot do machine translation. Quite frankly (being brought up to regard any problem as a source of irritation until I have solved it) I go around saying just how you can solve Bar Hillel's paradoxical problems, most of which are quite trivial. However, he has done some good work. One good thing he did was to upset the Wittgenstein hypothesis. I mentioned a paraphrase of this earlier; what Wittgenstein actually said was interesting, particularly for anyone concerned with information science. He said, "What can be said, can be said simply." Oh, that this were written on the hearts of authors!

Bar Hillel, being the devil's advocate, examined this hypothesis in the context of a rather restricted grammar and showed that the hypothesis was wrong. In fact there exist utterances of infinite complexity in any language in this artificial language group—and by extension, in all natural languages. These sentences are not reducible to any simpler form. Later, Shamir and Bar Hillel advanced the interesting hypothesis that there exists a reduction alogrithm that can be applied to sentences in a certain restricted class of grammars in which there is hope that some subset of natural language may fall. Bar Hillel and Shamir showed that there exists an algorithm for the reduction of sentences to sentences of canonical form or of minimum complexity. A sentence may indeed be of infinite complexity, but, in this event, we can show that it can be reduced no further. If a sentence is just badly put together, the algorithm gives a formal means of reducing it to a sentence of minimum complexity. The importance here is that, by taking a number of documents, we can in principle reduce the contents to minimum complexity and form the union of this information for all documents. The effect is to produce an output which contains all of the original material in the original documents but none of the redundant material.

I can't help concluding with a piece of statistical information derived from a survey I make of the computer engineering literature for 1960. I was doing this as a survey article for a British journal and the interesting result was that, in 1960, there were very approximately 10,000 pages of periodical publication in the field of computer hardware. The original material in this 10,000 pages could be described adequately in 40 pages. A plausible inference is that the exponential growth, or information explosion, is a figment of the imagination. The growth is much more nearly linear. The moral of this should, I think, be left to university presidents to unravel!

7

Pattern Recognition*

LEONARD UHR
Mental Health Research Institute
The University of Michigan

The perceptual mechanisms of living organisms have developed around wavebands of energy that are commonly emitted by objects in our physical world: the eye around vibrations of subatomic particles, the ear around vibrations of molecules. The purpose of perception is to reduce the signals that the mechanism senses—that is, this energy as it would affect a typical physical object like a photographic plate or a sounding board, and to judge whether it belongs to any of a class of signs that are of interest to the organism because they suggest acts that it should take. The judgment that some part of the flow of experience belongs to such a class is the act of "pattern recognition." Thus pattern recognition is the decision-making process that assigns to some experiences (carved by this very decision out of the total flow of experience) some internal meanings. (For the moment, by "meaning" I simply mean some set of connections.)

A bit more formally, pattern recognition is a many-one mapping from a very large set of arrays to a relatively much smaller set of names. The word "mapping" should be taken in an intuitive rather than a mathematical sense, for it simply indicates that some set of transformations has to be made to get from the raw input data in the array to the choice of a name. If we *had* nice mappings, there would be no pattern-recognition problems. Since we are talking about inputs from a physical world, we are always talking about arrays that contain discrete sets of data. This is so because the primitive quanta in the physical world are discrete and because any sensing mechanism has a maximal resolving power (uncertainty at the level of physics, where we resolve objects with objects of their own size; the "jnd"—just noticeable difference—at the level of psychology).

The need for and value of pattern recognition comes about *only* when some economy is effected by the recognition process. Such economizing does in fact take place in most real situations, where the objects, whose

*This paper is an edited portion of a manuscript in progress, tentatively titled "Computers and Discovery." It was prepared for presentation at the 6th Annual IBM Conference on Bio-medical Electronics in Poughkeepsie, Oct. 6, 1964, and at the Conference on Information Handling in Pittsburgh, Oct. 7, 1964. Preparation of this paper, and the author's personal research discussed therein, was partially supported under NIH grant M5254 and ARPA contract SD 266.

energy emission must be recognized, themselves are affected by coherent forces that bend, stretch or otherwise deform them. And, of course, the position of the object vis-à-vis the observer leads to the whole set of linear transformations (as they change their positions in three-dimensional space). Since the organism's problem is to continue to recognize an object as itself, even though it has undergone some linear transformation or some acceptable deformation, the organism will sense many arrays to which it will assign the same meaning.

If we thought of pattern recognition as an abstract problem, we would have to allow for situations in which this reduction from many input arrays to relatively few names could not be made. For example, each input array might simply mean something different, as indeed it does in nonredundant codes of the sort frequently seen in man-made information processors. The operation and address codes in computers are good examples. A worse example is the arbitrary, random assignment of each possible array to a name. Because there is no simplifying set of transformations that will turn one member of the set of arrays with a particular name into the other members, each array would have to be identified completely as itself, in effect named as itself, and then a table would have to be used to assign the class name.

The word "perception" would seem to be somewhat broader than the word "pattern recognition," since the former refers to the entire process of transforming the raw data of the stimulus into the recognition, the attribution of a class name. But there is certainly great overlap between these two words as they occur in common usage. Perception tends to emphasize the earlier transformations that regularize the input data, making the different examples of the same pattern in some sense more similar to one another. Pattern recognition tends to emphasize the final step when the instance is given a name.

I will use the words "input," "sensed data," "instance," and "array" more or less synonymously for the material presented to a pattern recognizer; "measurement," "characterizer," "transformation," "operator," and "mapping" for the steps that the pattern recognizer takes; and "pattern," "name," "class," and "output" for the result. At times, distinctions between these near-synonyms will be noted, but their similarity would seem to be the salient feature about them.

The large body of pattern-recognition research that has arisen in the past ten years in the interdisciplinary area between psychology, psychiatry, mathematics, engineering, and physiology that is variously called "cybernetics," "artificial intelligence," "systems sciences," "communication sciences," and "information-processing sciences," among other names, has been largely concerned with a particular simplified version of the general problem of perception and pattern recognition. This has been the problem of the assigning of the appropriate class name to an isolated two-dimensional array of discrete symbols. The bulk of the research has been on recognition of letters of the alphabet and, occasionally, other visual patterns. Most of the rest of the work has been on the recognition of spoken words or phonemes. A scattering of work has investigated recognition of other arrays, such as Morse code and diagnostic symptoms. A good bit of research that has gone on under other names, such as "concept formation," "language processing," "learning," "memorizing," "prediction," and "decision-making" is closely related and, in fact, often investigates the same problems.

Virtually all of this research handles the problem of naming a static, isolated matrix whose primitive symbols are discrete and clearly discriminable. The primitive set of symbols usually contains only the two values *black* and *white* (or *0* and *1*) in the case of visual patterns, or a small range of intensities (typically from 0 through 7) in the case of auditory patterns. A primitive symbol will, then, reflect the amount of light at a given spot in a two-dimensional picture, or the amount of sound energy of a given frequency at a given time. When I say that each primitive symbol can be perfectly resolved, I am talking about things that are often very tiny, of the magnitude of the individual spots on a TV raster or the amount of light that subtends a single cone in the retina. Thus there might well be thousands of such spots in the single pattern to be recognized, even if this pattern were merely a simple curve.

Psychology has amassed a great number of particular facts as to the interactions of the many factors involved in even the simplest perceptual acts. But it has not developed anything in the way of a coherent theory of how the crucial recognition toward which the entire perceptual process leads actually takes place. We are variously told that the brain compares its ideas with the incoming percepts, that the percept calls forth the memory trace, that the brain recreates the pattern until there is no more mismatch, and that this process is the idea, and so on. But what do words like "compare," "idea," "trace," or "recreate" signify?

But we now have a large number of computer programs and analog computers (and remember, these are equivalent, simply being alternate methods of representation) that do in fact recognize patterns. For want of anything that we could seriously call scientific theory, that was more than suggestive verbiage, these programs must be taken seriously as the first attempts toward developing a good theory. For they are, in fact, theoretical models of the traditional sort. They may well be bad models, in that they are inelegant, without great power, or (but this is the case surprisingly infrequently) contraverted by the empirical data. But bad theories, with their power to make things clear and lead to their own downfall, are far better than no theories at all.

THE STRUCTURE OF PROGRAMS FOR PATTERN RECOGNITION

The problem of pattern recognition seems to fall into several relatively clear-cut processes. The pattern must be characterized; each characterization must be assessed for its implications; the set of implications from a set of characterizations must be combined into a single decision.

The characterization stage can probably be subdivided, although the distinction is not altogether clear, between the set of transformations that *preprocess*, by regularizing the pattern, and the set of transformations that are more like measurements or characterizations. Thus, roughly, normalizing for size, sharpening of edges, and filling in of irregularities are part of the preprocessing phase; identification of angles or loops, comparisons between different parts, and identification of *significant* strokes are part of the measurement phase.

Each of these stages has two aspects: the mechanism that performs, and the genesis of this mechanism. Programs that have built-in mechanisms may well be pertinent to perception in the developed organism; programs in which the mechanism develops over time and experience may also be pertinent to learning, maturation, and evolution.

The mature, performing pattern-recognition program operates as follows: First, it performs a set of measurements on the array of symbols that is the pattern. A measurement consists of a set of specifications as to where to go, in terms either of the matrix as frame of reference or relative to other symbols in the array, to find a subset of symbols, and how to evaluate this subset. Most measurements actually used map the presence or absence of a match of the characteristic being searched for in the matrix onto the values 1 and 0 (present or absent). Thus, the process of performing a single measurement or characterization is indeed a mapping. The output of this set of measurements is a new array of symbols (the *names* of the characteristics that were *found*) that may or may not be connected one to another in a matrix (such as the input matrix) with each symbol connected to its physical neighbors, or in some other graph. Now the program may or may not perform a set of measurements (either the same set or a new set) on *this* array, producing yet another array. This process may continue for a set number of steps, or until no more characteristics are found, that is, until an entire set of measurements gives outputs of 0.

Within this general framework the sequence of measurements can vary widely. Some programs make only one set of measurements; some make two or three, or some small fixed number; some continue to measure until no more transformations are effected. The sets of measurements also vary, from the set that contains a single measurement the output of which directs the choice of the next measurement; to sets that contain many measurements, some unique to this set and some held in common with others; to the iterated use of the entire set. Finally, the final use, the final significance, of the presence or absence of a particular characterization also varies. A characterization is, in effect, a summarizing statement as to the presence or absence of a set of symbols in a certain configuration in the array. Thus a characterization summarizes information got from previous characterizations, since this set of symbols is simply the output from previous characterizations. (The original matrix is really the output from the—for the computer—trivial characterizing step of assessing the presence or absence of each of the first-step characterizers—the primitive symbols that are built-in known symbols so far as the computer is concerned—and listing the name of the characterizer that was found. Remember that these programs typically do with 100 percent accuracy their fine discrimination of just noticeably different objects.)

When only one measurement is made at each point, and the choice of measurement is contingent upon the outcome of the previous measurement, we have a simple sequential tree. To the extent that many measurements are made simultaneously (in the sense that no decisions intervene), our program has a parallel structure. But note that in a certain sense this is simply a technical matter of scheduling.

In general, a sequential tree of measurements is more efficient, since it makes only those measurements that are indicated. It is faster because it makes fewer measurements, but slower because it must continually decide what measurements to make. Thus, optimum overall processing time will depend upon the speed of measuring vs. the speed of deciding. In the strictly sequential tree without any redundant branches, a single mistake will ensure that the program is wrong. That is, such a tree is as strong as its weakest link. But redundancy can easily be introduced into such a tree, either by having many paths to the same final decision or by having the decision made at each node lead to more than one node—in effect, by making the tree more parallel.

This whole structuring of the sequence of measurements seems very attractive in terms of our feelings about the human brain. We have suggestive anatomical and physiological evidence that there are parallel structures in the brain (e.g., the cones, lateral, geniculate, cortical projection areas, and, indeed the entire visual system), and there are sequential structures in the brain (e.g., the several layers in the cortex and the sequence of structures in the visual system just described). Compelling logical arguments for a parallel-sequential system can also be made: (1) sometimes time is important, sometimes space; (2) parallel inputs will speed up processing, since they handle simultaneously what must otherwise be done sequentially, and therefore to the extent that there is space (in this case, body area), they should be used; but sequential operations will also increase economy for exactly the same reason that the binary search of 20 questions increases economy, and therefore to the extent that there is distance—from the surface of the body to its center—they should be used.

Introspectively, and this is also loosely supported by some evidence (that I find inconclusive and hard to interpret) from the psychological literature about people's abilities to process things sequentially and in parallel, we have a rather strong feeling that complex decisions are made in several stages. Certain facts or things noticed lead to a vague, usually unverbalizable, feeling as to what might be there and at least to some extent direct the search to find further evidence that will further confirm, or deny, this hypothesis. It is here that we use vague words like "expectation," "set," "tendency," and "hypothesis" for a process that apparently goes on in the brain when it perceives, remembers, forms concepts, and even problem-solves, a process that really is strikingly similar to the standard process of scientific experimentation and induction. Thus our introspective and intuitive gut feelings are that the brain works in a parallel-sequential fashion, and logic, good design principles, and experimental evidence all tend to confirm this feeling. What we would really like a program to do, then, is to make a few *first glance* measurements of a sensed pattern, decide on the basis of these measurements what to look for next and where, and continue this process until it is sufficiently certain to make a decision. There should also be general expectations, from long-term and immediate past experience, as to what general type of pattern to expect, and there should be flexible costs and values attached (again depending upon past experience which has shown what each pattern implies) that will affect how careful the program is in choosing to decide upon less than certain evidence. Most of these requirements can only be met by a program that learns from past experience, is engaged in a several-step dialog of action and reaction with its environment, and has a sufficiently rich need-value system. So it is unrealistic to require them of the simple pattern-recognition programs being described at this point. But the sequence of tentative decisions and directed further looks is quite within such programs' abilities. Few programs in fact follow such a sequence, I strongly suspect because of considerations of economy given the peculiarities of the techniques of programming. This is unfortunate, since, ideally, the program should be a function only of the computer being created, rather than of the particular general-purpose computer, or the programming language, being used. Unfortunately, pattern-recognition programs with any great power are still such difficult programs to write, and come so close to taxing the powers of present-day computers, that such compromises are usually made.

But we should keep clearly in mind that our reasons for wanting a parallel-sequential pattern recognition model are not really firmly grounded, that this is still very much a matter for conjecture. So rather than insist that programmed models be parallel-sequential, we should ask of these models that they exhibit to us the strengths and weaknesses of each type of model. After all, if the brain is in fact of a certain sort, it has become that way for some very good evolutionary reason.

The ordering of a sequence of measurements is quite a subtle thing, about which we know little. It is equivalent to the breaking up of a single, very complicated function that maps from one set to another in one step (the completely parallel tree) to a sequence of simpler functions that effects this mapping in several steps, just as in the 20-questions game, there are certain questions that are good to ask early on, and certain that become important, or even meaningful, only much later in the game. What we think of as "preprocessing"—the processes that tend to regularize the different instances of a pattern class, that tend to make of the pattern class a more compact set in the space of measurements that will then be applied—contains the measurements that should be made first.

The actual choice of the set of measurements to be made seems to me the crucial problem of pattern recognition, and, indeed, of many aspects of intelligence. Loosely, it is the uncovering of those things that are important. Each measurement is in a very real sense an hypothesis that the output is correlated with the desired decision. The choice of characterizing measurements for pattern recognition is thus very similar to the choice of hypotheses to be tested in that series of experiments known as *science*. That is, at this very early point in a rather mundane and simple process, we run head-on into the problem of hypothesis-formation, or *discovery*. How do we get a good set of measurements?

TYPES OF CHARACTERIZERS

The problem of finding a good set of operations with which to characterize an input instance of a pattern is, then, equivalent to the problem of finding a good set of variables with which to characterize some empirical domain. Nor is it apparent which is the more difficult problem. In both cases, and indeed in all interesting cases, the number of *possible* variables is overwhelmingly large, far too large for any exhaustive examination of all of them ever to take place. This is so whether the set of possible variables is finite or infinite. For example, in the simple pattern recognition problem there is only a finite number of possible measurements. Since the input array is finite, a finite enumeration of all possible configurations of symbols in the array will include all possible characterizers. That is, there are v^{n^2} possible characterizers in an array with n cells, each ranging through *v* values. For example, a 20×20 0-1 matrix will have 2^{400} possible characterizers; a 100 by 100 $2^{10,000}$, a 20×20 matrix whose values range between 0 and 7 will have $2^{3^{400}}$.

All these numbers are far beyond the bounds of computability by enumeration, and, more important, they completely violate the fundamental reason for the existence of a pattern recognizer—the need for quick response in order to satisfy a need system. So the exhaustive algorithm is worthless. Here is a good instance of the mathematician's criteria of solvability being quite useless to the scientist. We are very simply thrown back on intuition. As Peirce has pointed out, it is a very strange and beautiful thing that intuition has worked so often, both in the intuition of evolution that has developed living systems, and the intuition of the scientist who has, time after time, hit upon the right hypothesis. This, Peirce suggests, is the deep meaning of the scientific faith that nature is simple, and, therefore, the simple hypothesis is preferable. If nature were not simple, we could never come to understand it. Put another way, we have come, and we will come, to understand those things about nature that are sufficiently simple. But *simple* does not have any absolute meaning; rather, it always refers to an understander, a system that, having much the same structure, finds some other system like it, hence simple.

We are here in a marvelously circular situation, one which, if it could be understood, might well hold the key to many of our problems. Animals, and above all, man, have evolved as a function of nature. We have evolved precisely because to a certain extent we could understand nature and, through this understanding, gain what we needed. This means further that nature was *understandable*, and understandable not merely by some superintellect with great powers of reasoning, but by an evolutionary process that could move only in remarkably small steps relative to the apparently large increases that were effected. Thus the very grain of our beings is adapted to nature, *understands* it in profound ways that are far beyond our conscious intellect. Our mind, then, when it considers something to be *simple* to be, intuitively, right, may well be talking from this substrate, and it certainly behooves us to listen carefully.

A second great source of inspiration in our search for good hypotheses, in the crucial discovery phase of our enterprise that we are now discussing, is our introspections. Let us avoid arguing about the worth or respectability or even reality of introspective evidence as scientific data. All of us *do* make statements, "I am tired," or "I see a fuzzy edge." Many of us, if we are asked to discuss how we see objects or decide among horses, will give answers that may include such statements as, "objects have edges; angles and loops are important; the interrelations between lengths and slopes are important." We might also say that these are the "meaningful" characteristics.

Now it probably would be hopeless to start running introspective experiments to find general structures of perception across people; and it would probably be equally unilluminating to objectify the hunches got from such introspection into behavioral experiments. The former were of course done ad nauseam until the behaviorists blew them apart and started doing the latter. But now we have an entirely new approach. We can objectify the hunch, not by clever experiments whose operationalization, the cleverer it is the more suspect it must be, because there are always too many factors potentially operative; rather, we can put the hunch into the form of a computer program, and simply see how well it works—that is, how well it predicts all the perceptual phenomena for which it is appropriate.

I have developed this argument for intuition and introspection at some length for several reasons. First, these are in fact the tools that good scientists have always used, and they have been used in the most successful and interesting pattern-recognition programs. We talk about the mathematical-deductive aspects of science at great lengths, and teach experimental methodology ad nauseam. We pay lip-service to the need to develop hypotheses in the first place, but then mutter quietly that, because this is so mysterious, we will be silent. But this has actually had the effect of making many people forget that the first steps are the crucial ones; and, worse, it has made many people antagonistic toward hypotheses that cannot be justified on methodological or mathematical grounds. But these are merely the trappings that come afterwards to clean up discoveries. Second, an extension of this argument, a number of mathematically oriented people who have worked on or examined problems of intelligence and dynamic model-building have deplored the lack of firm theoretical (in the mathematical sense) foundations. But no empirical science has developed in such a way; and the science of brains, which has the most complex of all problem domains, one that has been completely intractable to mathematics, is probably the least likely one to do so. Third, at least until recently the common cant among psychologists and other soft scientists has been that introspection is a useless tool, and that it may not even exist, and that intuition is such a vague concept as to be completely meaningless. From these condemnations, fallacious as they are, comes an even more unfortunate next step—to refuse to use anything that comes from such quarters, that does not come from "reputable" sources, namely deduction and (interobserver) objective experimentation. But the canons for the acceptability of evidence and ideas are perfectly simple. We should accept what *works*—what is valid. We ask for circumstantial evidence, such as the graduate degrees attained by the observer, his sanity and respectability, his skills in technique, his biography and previous successes, his method of collecting his observations, the number and type of people who believe them, the status of his theory, and so on, only because these things tend to be correlated with the goodness of his observations and generalizations. But when the fruits come, it is only these that we must examine, as objectively and dispassionately as possible. Now at the present there is a strong (and in many cases misguided) prejudice against evidence, hunches, or whatever gained by introspection. This is justified for those types of evidence that can not be so gained, and this might well be a very large part. But whatever is valid must be admitted.

PLAUSIBLE PROPERTIES

Introspection is only one source for the characterizers that have been used in pattern-recognition programs, and those people who have used it have not always been aware, or willing to say, that they have. But my impression is that those programs that have been motivated by an introspective examination of how pattern recognition goes on display a surprising similarity. They are the programs that have been named "characteristic features" programs. Typically, such a program looks at a handful (from 5 to 60) of characteristics that its programmer felt were meaningful in that they convey useful information about the pattern to him. Note that in most cases there has not been any objective assessment of this; it is not known until the program is written that in fact they do. Such programs measure characteristics like straight lines and curves in certain positions of the matrix and in certain relations one to another, loops, angles, and joins.

Some programs that have been lumped into this group make preponderant use of characteristics of the same type of complexity, but ones that have been chosen more because of the ease of programming them (for example, the number of line segments in a column or a row of the matrix) or their mathematical respectability (for example the moments).

Good examples are such programs as those of Grimsdale, Sumner, Tunis, and Kilburn (1959), which decomposes patterns into their *meaningful strokes*, and then compares the graph of strokes so formed with graphs already stored in memory, Bomba (1959), which looks for similar strokes, but without so much care in assessing their interconnections, Unger (1958), which uses a greater miscellany of the sort described above, and Doyle (1960), which tries to use the best characterizers from previous models. (Note that a "characterizer" is much the same as a "heuristic" in game-playing programs. For example, the work of Grimsdale's group chooses a natural set of heuristics in terms of what the human appears to do in much the same way as the work of Newell, Shaw, and Simon (1960) chooses a natural set of heuristics, from observing humans and asking them to introspect, for game playing and theorem-proving. Unger's choice, which is not limited to the intuitively plausible, might be likened to Gelernter's (1960) criteria for his geometry theorem-prover's heuristics. Unger's and Doyle's choice of, essentially, a good bag of tricks, is quite similar to Samuel's (1959) choice of as powerful as possible a set of heuristics for a checker player.)

TEMPLATES

Another type of program, one that is often embodied in an analog machine, is the template matcher. Typically, a photographic plate with a stencil of the pattern (usually a typed letter) is matched with the reflected light from a pattern to be recognized. A disk that contains all the letters of the alphabet may rotate very quickly, with a photocell behind the target that integrates the light from the pattern that passes through each letter, essentially giving a correlation between pattern and letter. Then the machine will choose the name of the template that passed the highest percentage of light. The equivalent of this simple and cheap gadget in a program for a digital computer is a very time-consuming and cumbersome process of matching the individual cells in the input matrix with the individual cells of a stored template—a good example of an awkward, inappropriate and misleading, yet possible, simulation of the appropriate analog.

Often such a machine will have an optical system that sharpens or fuzzes the image of the pattern in such a way as to normalize or jiggle or perform some other appropriate transformation, one that could be described only by an extremely sophisticated mathematical equation or digital program.

Logically, the trouble with such a machine is that it will not handle slight variations from the template, except to the extent that the optical gadgetry gives sophisticated transformations. This method has been investigated chiefly in the context of building practical commercial machines for such applications as check and record reading; and the criterion that is typically set the designer is 99.4+ percent accuracy. This is often achieved with sufficiently carefully printed letters in a sufficiently standardized type font. But superficially one reacts by saying, "Ah, but they have made their problem sufficiently easy by controlling their patterns." However, it is not at all certain that this is, indeed, the case. Since results are so close to 100 percent accuracy, a more powerful and "sophisticated" program cannot clearly do better no matter how perfect its results; and, unfortunately, the developers of machines have rarely if ever been willing to conduct or publish tests with messier patterns that would show their machines to be less than almost perfect. But whereas I was one of the people who assumed for a long time that these template machines were obviously doing well because they were tackling a much simpler problem, I would not be at all surprised to learn that they gave comparable results in comparable tests.

Often when the philosopher or psychologist talks of an "image" or "idea" that is stored in memory and "recalled" by the incoming stimulus, beneath the verbiage a template is all he seems to mean. When we see a template plain, in a clearcut description or, worse, a photographic contraption, we tend to recoil. But we want more objective tests than emotional reactions, and the truth may very well be plain, if not homely. It is fairly obvious that what I will name the "silly template" will not do. The silly template is the template that matches only when all its little quirks and irregularities must also match. In the computer program it is easy to have a silly template, since a standard matrix intersection program can ask "Are these two matrices *identical?*" with great ease, but finds it extremely difficult to understand, much less ask, "Are these two matrices *similar?*" But the photographic plate, being analog, has its saving 5 percent inaccuracy built into its very grain; and all that one must do to improve this inaccuracy even further is to hire sufficiently unskilled craftsmen.

The *template* program becomes much more interesting with the simple extension of making templates that are not the patterns—for example the letters, themselves—but, rather, are the *strokes* that compose these patterns, as done in a machine developed by Rabinow (1957). In fact, this can even give a saving in the number of templates needed, when there are fewer strokes than letters. Now such a machine needs a little bit of explicit logic, for it must decide upon a letter because of the appropriate combination of strokes. For example, the graph of the program developed by Grimsdale's group would be entirely appropriate. In fact, this stroke template machine is, after all, almost identical to the Grimsdale program, which is generally accepted as being one of the very most powerful and intuitively and psychologically satisfying of pattern recognition programs.

1-TUPLES

Another type of program looks at the individual spots or cells in the input array, and asks, for each cell, which patterns the symbol it contains implies. Put another way, the individual cells are the characterizers of the pattern. For convenience, I will call this the "1-tuple method." Typically, for each pattern to be stored in memory, a probability contour map is developed by a program that looks at a sample of the pattern. The size of the sample is often determined by traditional statistical considerations, so that a sample sufficiently reliable to serve some specified purpose is used. In fact, this is a method that lends itself well to straightforward statistical analysis, and often the program then continues to perform a factor analysis and develop an optimal, or sufficiently good, discriminant function for the prediction of the different patterns.

An extension of the method should now be obvious. Rather than examine every cell in the matrix, redundant cells can be eliminated, so that only the small subset that gives good discrimination need be looked at. Both the exhaustive and the nonredundant methods have been used extensively.

Again, this would seem to be an overly simple scheme. But the published results (admittedly unsatisfactory) of comparisons suggest that no clear-cut superiority of more powerful methods has been demonstrated as yet. It seems more than reasonable to expect that this method will eventually be shown to be limited and weak. First, it is easy to construct patterns on which it will fail—in general, those patterns in which *interactions* between the spots are important. And one would be tempted to say that the very word "pattern" entails the requirement that several things are in a relation, are interacting. So one could almost argue that, when the 1-tuple method succeeds, we have merely demonstrated that we shouldn't have honored our problem set with the name "pattern."

On the other hand, it might very well turn out that many, or even most, of the stimuli that we commonly *do* call patterns *can* be handled by such a method. Certainly they *cannot* if individual instances of patterns vary widely, but one could use a set of *preprocessing* characterizers to regularize these instances and make them appropriate for a second-stage 1-tuple recognizer.

The 1-tuple method is probably close to what associationist philosophers and psychologists had in mind. Stochastic learning theory, when it talks about real world problems that must be sampled, often seems to be talking about such a model. The one program that has been written, by Marzocco (1961), to embody a stochastic learning model makes this assumption explicit.

GESTALT CHARACTERIZERS

Several models have been developed that purport to examine the "Gestalt" characteristics of the "whole" pattern. Some, such as Uhr's (1959), find characteristics of the sort used by the Grimsdale group, and then examine them in relation one to another. For example, the relative positions, sizes, curvature, and so on are computed. Note, however, that such a pattern-determined rather than matrix frame-determined scale, if this is what we mean by the vague word "Gestalt," is also approached by such simple normalization procedures as drawing a minimal rectangle around the pattern and expanding it until the rectangle fills the matrix. We simply do not know enough about the Gestalt, and, of course, this term may very well refer to several different phenomena. The author of at least one 1-tuple program model refers to it as a Gestalt-sensitive model because it looks at and summates over the entire pattern. Another simple model, developed by Nieder (1960), sums the distance between all pairs of points on the contour of a pattern. This is a type of 2-tuple model, one that does take pairwise interactions into account. But again, it seems preferable to reserve the idea of the Gestalt, at least in its most powerful use, for a model that looks at very high-level interactions. Nieder's model is very similar to the "Gestalt" models previously developed by Rashevsky (1948).

N-TUPLES

A generalization of the 1-tuple method is to use as characterizers *n*-tuples, where *n* is sufficiently greater than 1 to be sensitive to whatever interactions actually do occur in the patterns being recognized. Bledsoe and Browning (1959) investigated programs that used randomly generated 1-,2-,3-,5-,9-, and 11-tuples as their characterizers. Such a model stores the correlation of *every* configuration on the *n*-tuple with each pattern. It thus multiplies exponentially in its memory storage requirements as *n* increases: a 1-tuple needs 2, a 2-tuple 4, a 3-tuple 8 stores. Bledsoe and Browning found that performance improved until the size of *n* was around 5 or 7, and then tended to decrease. This particular result may be specific to their specific model, with its total set of characteristics. But it is interesting to speculate whether their results suggest the degree of interaction and complexity to be found in patterns, or at least in relatively simple patterns like the letters of the alphabet. After all, there is no reason to think that the level of interaction is so high that everything really affects everything else, and hence the pattern. Put another way, the "Gestalt" may well be a mixture of several Gestalts. We know in fact that there is a good deal of redundancy in patterns in the real world, and we have a good understanding why this is helpful and even necessary (for example, to give error-correcting codes that combat noise). And we know that the brain cannot handle very high levels of interaction. So the Gestalt is certainly something less than an interaction of *all* of the parts. Now the question might be posed, how many more than seven parts are ever involved?

Note that the probability contour method, when a choice of only subsets of the cells is made, is in ways equivalent to the *n*-tuple method. A subset is an *n*-tuple, but it stores information about only one, or at most a few, of the possible configurations. Similarly, a template is actually a very large *n*-tuple, but, again, it stores information about only the template configuration, in which the cells within the pattern are filled. When the template allows for a partial or loose match, it becomes a very complex set of smaller *n*-tuples, that is, all the possible combinations of filled and empty cells that would lead to the choice of this template.

AN ALPHABET AND LANGUAGE FOR DISCOVERY

Indeed, *any* characterizing measurement could be described as a set of n-tuples. This becomes obvious when we look at the problem from a slightly different point of view. Whatever our characterizer might be, it will partition the set of all possible configurations of the cells in the array into those that it accepts, assigning "1" as their value, and those that it rejects. But each of the configurations that it accepts is simply an n-tuple, an array of 0's and 1's. So the total set is simply an "or'd" collection of n-tuples. A good characterizer is a very simple description for a large collection of good, equivalent, n-tuples.

Now remember that the number of configurations of the matrix is an astronomically large figure. It would serve no practical purpose to describe all of our characterizers in this standard way, or to ask a program to make an exhaustive search through all such configurations. So we are back to the same problem, one of getting a sufficiently simple and sufficiently powerful, hopefully a near-optimally simple, set of characterizers. But the n-tuple formulation gives us a relatively convenient space within which to ask a program to help us in this search.

The space of possible characterizers is overwhelmingly large. If we use a characterizer such as "Is there a concavity?" we know of no way of representing this in terms of other, simpler, characterizers. What we would like is a formulation such that the space of all possible characterizers can be composed from some relatively simple set of *primitive* characterizers by using some simple and well-defined set of composition rules. The 1-tuple is just such a primitive, and combination of n-tuples into larger n-tuples just such a composition procedure. That is, we can set up a space for searching for good characterizers by using the space of all n-tuples. Or, better, we could ask the program to compose the members of this set from a simpler already-formed set, starting with the 1-tuples.

Put another way, our problem is to find a convenient and efficient set of descriptions of the patterns we want the program to recognize. Then the program need simply see whether each description is valid for each pattern. "Description" is simply another name for "characterizer." Now we need a convenient language within which to write such descriptions. The language must be rich enough so that all necessary descriptions are *writable*. But it must also have some elegance—we do not want to write each description as a separate and unique entity. We want a language with a relatively simple set of primitive symbols—its *letters* and combining rules—that will allow it to develop the necessary set of words and sentences—the characterizing descriptions.

For pattern recognition, the values 0 and 1 would seem to provide us with a good set of letters. Our primitive 1-tuple specifies the position of a 0, or a 1, in the matrix (or, alternately, with respect to some other position—either some fixed point on the matrix or the position relative to some *other* n-tuple). The combining rule might simply be, T_i plus T_j gives T_{n+1}, when previously the program contained T_n tuples. Such a rule will allow any pair of 1-tuples to be combined into a 2-tuple, and, generalizing, any n-tuple to be formed by successive application of the rule on the appropriate sequence of pairs of n-tuples.

Such a procedure gives both a method for examining the space of possible characterizers, and also an overall heuristic guideline for the direction that this search will take. In general, the search is from tuples where n is small to larger tuples as needed, starting with the 1-tuple. It is not at all obvious that such a procedure will work, and there are no arguments that *compel* us to choose it. It seems reasonable, however, on several grounds.

First, we are, remember, in the standard dilemma of science and induction: the empirical domain that we would like to organize is overwhelmingly too large for exhaustive methods. At best, we can only try and hope; we will never have guarantees. Second, the use of the smallest possible n-tuple seems in harmony with science's guiding principle, simplicity. Third, and this is probably one of the factors that underlies simplicity, economy also dictates a small n-tuple. Fourth, this seems to be close to nature's method of evolution.

So let us consider a program that tries to find a minimal, near-optimal set of characterizers, without having any characterizers programmed in. The model now is a model for generating and testing new characterizers. The model-builder is now looking at problems of discovery and induction. It is not at all obvious whether any search in such a large space will work. Even with the sort of description of the space outlined above, and the overall method, typical in induction, that orders the possible descriptions according to some criterion of simplicity, and examines the simplest descriptions first, the potential space still seems overwhelming. But this is identical to every real-life and scientific situation: the space of potential correct inferences is always overwhelmingly large. One can only try, and take the consequences.

A program written by Uhr and Vossler (1963) that was written in the spirit of this argument, although it differs a great deal in the details that were added in order to give more direction and power to the search, turned out to do surprisingly well, despite the fact that it started without characterizers, but only with the ability to generate and then test characterizers as needed. Essentially, this program assumed that the search was by a nerve net of the sort that we see in the eye, and certain evidence from the behavior, physiology and anatomy of the eye was used to specify some of the program's detail. For example, the search space was cut down to those n-tuples that would be plausible for the type of nerve net that was posited.

It is difficult to say with any degree of certainty how well this program's performance compares with that of other programs. Indeed, there is virtually no comparison evidence for any pairs of programs, an unfortunate circumstance that to a great extent results from unimportant differences in the format of input data—for example, the size of the matrix or the exact columns on the cards in which the matrix must be punched. But there is every reason to think that after only three to ten learning trials it performs at at least as high a level as most other programs. This despite the fact that it must develop its set of characterizers as a function of its experiences with a few instances of the pattern set. This is extremely encouraging, since it suggests that a space that on the surface appears to be overwhelmingly large can be searched successfully in a reasonable length of time, when only a few weak heuristic assumptions are made.

Several other programs that attempt to discover a good set of characterizers have been programmed, by Roberts (1960), Kamentsky and Liu (1963) (who, essentially, choose a best set from a larger prechosen set), and Prather and Uhr (1964).

A discovery program is an especially promising program, because its essence is that it handles problem domains that have *not* been preanalyzed by the programmer. That is, the programmer has not intuited or in some other way developed a set of characterizers that he knows, or thinks, will work. He has not developed an adequate theory of the empirical domain to be analyzed, thus leaving to the program the relatively mundane task of applying this theory. Rather, the program, in a very real sense, is beginning to help in the development of the theory. The programmer's task now becomes one of giving the program rich possibilities for good languages, tests, and methods for building such theories.

We would expect such a program to be more general in its abilities. Since it is not developed with a specific pattern set in mind, and since it purports to be able to discover appropriate characterizers for no matter what arrays, so long as they are characterizable (for example, the colors, *red*, *green*, and *yellow*, could not be characterized by a two-valued device that responded only to intensity of light), it seems only reasonable to expect, and even to demand, evidence of such generality. In fact the program of Uhr and Vossler has been tested for its ability to learn to recognize a variety of different patterns wider than used for any other program known to the author. These patterns included handprinted and handwritten letters, handwritten Arabic letters, hand-drawn pictures of simple objects like cars and trees, pictures of simple objects like shoes copied from a mail-order catalog, cartoon faces, photographed faces, a variety of randomly generated meaningless patterns, and *spoken* speech. The program achieved 100 percent or almost 100 percent, success on the known examples of all these patterns, and 50 to 100 percent success on the unknown examples. In a number of comparison experiments, the program did as well as, or substantially better than, human subjects. The attempt was made to train the human subjects under as favorable conditions as possible. It is impossible to equate such performance between the computer model and the human being, because there are still so many aspects of the situation with the human that have not been modeled satisfactorily. But it seems interesting to note that the computer model does so well, on this relatively difficult task, one that has, until recently, seemed too complex to be modeled.

Remember that *pattern recognition* is, basically, the application of some reasonably small set of measurements, from some overwhelmingly large set of possible measurements, to examples of patterns that have been inscribed in arrays. Each specific model is a particular choice of some subset of measurements that is suspected to be adequate. The problem is much too large to be solved analytically, or, even though finite, by exhaustive enumeration. Nor can the set of patterns be explicitly described for any interesting domain. Therefore it is not possible to choose an optimum set of measurements, or even to know how far from optimum any particular set of measurements may be. The best that we can do is to compare sets one with another.

Because the space of possible measurements is so large, the thought of a search through the space of possible measurements has, at first blush, seemed ridiculous. Evidence from certain types of search, as best exemplified by the "perceptrons" that have been studied by Rosenblatt (1958) and others, in which excessively long training sequences result in relatively weak performance, has tended to confirm this feeling. But the perceptrons that are mathematically analyzable, hence studied, are not capable of attaining the rich variety of structures that one would expect to be powerful. If indeed a wide variety of different preprogrammed models of pattern recognition do quite well, attaining much the same level of performance, it seems reasonable to posit that the set of possible measurements, although horrendously large, contains a sufficiently large subset of information-bearing measurements so that a sufficiently powerful subset of measurements can be drawn from it without too much analysis, whenever the designer of the model uses a reasonable amount of thought and care. Put another way, the space of measurements, although too large for exhaustive search, is sufficiently rich in good measurements that can be found in likely places. *Intuitive* concepts, such as those we hold about meaningful characteristics, an alphabet of strokes, and so on, are themselves the fruits of natural, only partially conscious experiments that each of us, and the evolutionary process, has made on the environment. The information gleaned from these experiments is sufficient for our purposes, for it does, in fact, give sufficiently powerful subsets.

Now it is not so surprising that a model that attempts to discover and generate its set of measurements will succeed. For one of the prime requirements of an evolutionary development of pattern recognizers is that the measurement to be found be sufficiently simple, with respect to the mechanism that is searching for it, to be *find*able. And, indeed, we find in several models good indications that even discovery programs can attain the fairly good level of success that is typical of pattern-recognition programs.

We would expect of discovery programs a greater generality of abilities over *different* pattern sets, since these programs have not been designed specifically to handle particular problems. And, once again, we find this to be the case, so that, in at least one instance, the same program successfully learns to recognize either visual or auditory patterns.

DIRECTIONS FOR FUTURE RESEARCH

In many ways the simplified pattern-recognition problem is, indeed, simple; but in other ways it has been greatly complicated by the things that it leaves out. If patterns were in a more natural context of other patterns, the very difficult new problems of delineating and isolating the particular problems, of segmentation and figure-ground, would confront the model builder. But a great deal of additional contextual information would also be at his service, once he was capable of handling the situation; for patterns would no longer have to be recognized entirely from themselves. Rather, there would be much additional information in their contexts. If patterns existed over time, so that they changed, moved, and, in general, were transformed into themselves by whatever natural forces controlled their universe, there would, once again, be an enormous amount of additional information available to the model.

For example, in the recognition of continuous handwriting, the problem of identifying the individual letters, when they are now interconnected, is still beyond the power of present-day programmed models. One program (Uhr and Vossler, 1963) that attempts to turn such a continuous pattern into a set of isolated patterns gave evidence of being able to perform with much greater than chance but far less than perfect success (around 50 to 60 percent success). One would hope for much better results, and, in fact, most people feel that pattern recognition programs are doing a totally unsatisfactory job when they give such a performance. But it is not certain that they are. First, we should ask how well human beings do with such materials. We know that people make many mistakes in recognizing individual letters from cursive script when they are taken out of their context. In fact, we could even argue that the whole purpose of the handwriting process is to speed up communication, at the expense of cutting down redundancy, until the minimum-effort, but still-readable, product has been achieved. This is why many of us are the only people in the world who can read our own writing. And, since there is a good deal of information in the context, simply in the contingent probabilities of parts of sentences, and letter *n*-grams, recognizability of the individual letter can be sacrificed.

If patterns existed over time, the model would be able to make use of (or even to learn) the concept of identity over changes, and very quickly build up a coherent picture of the ways in which the specific instances of a pattern class are related. This suggests something about the type of measurement that should be used, since it would be well for the set of measurements to be similarly ordered. Such a procedure, in which patterns grow larger and smaller, move laterally or rotate in the third dimension, would make it quite easy and relatively straightforward, for the model to develop measurements that reduced different instances to an invariant with respect to the linear transformations. Nonlinear transformations, and smoothings with respect to noise, could similarly be learned. This, then, would incidentally be a situation in which the "preprocessing" measurements were relatively naturally built up first, and hence segregated from the identification measurements.

Thus the attempt to enlarge pattern-recognition programs to handle more aspects of the total perceptual problem will, in addition to complicating the problem, make available to the model a good bit of additional information, and to at least some extent make the problem easier to solve.

It is a shame that so much work has gone into the simple pattern-recognition problem and virtually none into its extensions. To some extent this can be explained by the fact that each extension probably increases the size of the program, and, possibly, therefore, the effort needed to develop the program, by a factor of at least two or three—when these are already complex programs that often push the limits of the ability of existing computers. But at least a beginning could, and should, be made. Probably a more important reason is the fact that most pattern-recognition research, especially as performed with adequate programming help and computer facilities, has been, essentially, an applied effort to develop specific gadgets to handle specific problems under specific limitations of time, space, and money.

In general, what would be needed in the way of a more complete pattern-recognition program might be as follows. Rather than accept as inputs only isolated 0-1 matrices, the model should accept a continuing stream of n-valued inputs, where n equals at least 8. This stream should extend indefinitely in two, or even three, dimensions. It should not merely be presented to the model; rather, the model should be able to direct its glance at new parts of the stream, much as an animal can move his head, or even his entire body, to take a look at something that might be important. Now this already is well beyond the capacity of present computers, with their limited memories and virtual inability to act upon anything other than an environment that they have simulated internally. More reasonably, we might ask that the program accept an n by "potentially infinite" matrix, continually sliding into the program's "gaze," which is itself an $n \times m$ matrix. The model could then be given some ability to shift its gaze, so that the particular part of the input matrix it is looking at at any moment is a function of the decisions that it has made on the basis of information it has gained and knowledge it has stored, as well as being a function of how fast the simulated universe unrolls itself. Such a situation could then be interpreted in the following different ways. First, the incoming experience might be a continuous array with two spatial dimensions, such as a complex aerial photograph or continuous handwriting. This would allow the program to take advantage of contextual information. Or, second, the experience might be a one-dimensional string that continues over a second *time* dimension. This would allow the model to take advantage of the redundancy of a pattern as it endures and changes into other forms of itself. But if we asked the program to handle anything very interesting in the way of patterns when time is introduced as a third dimension, we would again be posing a problem that is probably too large for existing computers to handle at all satisfactorily. (This is not to say that such problems should not be posed; on the contrary, they seem to me among the most interesting and hopeful for current investigation.

REFERENCES

1. Bledsoe, W. W., and I. Browning, "Pattern Recognition and Reading by Machine," *Proceedings of the Eastern Joint Computer Conference* (1959), pp. 225–232.
2. Bomba, J. S., "Alpha-numeric Character Recognition Using Local Operations," *Proceedings of the Eastern Joint Computer Conference* (1959).
3. Doyle, W., "Recognition of Sloppy, Handprinted Characters," *Proceedings of the Western Joint Computer Conference* (1960), *17*, pp. 133–142.

4. Gelernter, H., "Realization of a Geometry Theorem-proving Machine," *Information Processing*, Paris: UNESCO, (1960), pp. 273–282.
5. Grimsdale, R. L., F. H. Sumner, C. J. Tunis, and T. Kilburn, "A System for the Automatic Recognition of Patterns," *Proc. IEE*, Part B, (1959), *106*, pp. 210–221.
6. Kamentsky, L. A., and C. N. Liu, "A Theoretical and Experimental Study of a Model for Pattern Recognition," IBM Research Paper RC-933, (May 10, 1963).
7. Marzocco, F. N., and P. R. Bartram, "Statistical Learning Models for Behavior of an Artificial Organism," Second Annual Bionics Symposium, Ithaca, New York, (Aug. 30, 1961), Report SP-464, System Development Corporation.
8. Newell, A., J. C. Shaw, and H. A. Simon, "Report on a General Problem-solving Program," *Information Processing*, Paris: UNESCO, (1960), pp. 256–264.
9. Nieder, P., "Statistical Codes for Geometrical Figures," *Science*, (1960), *131*, pp. 934–935.
10. Prather, R. and L. Uhr, "Discovery and Learning Techniques for Pattern Recognition," *19th Annual Meeting of the ACM,* Philadelphia, Pa., (1964).
11. Rabinow, J., "Optical Coincidence Devices," U.S. Patent No. 2,795,705, (June 11, 1957).
12. Rashevsky, N., *Mathematical Biophysics*, (Chicago: Univ. of Chicago Press, 1948).
13. Roberts, L. G. "Pattern Recognition with Adaptive Network, *IRE Conv. Rec.*, (1960), *8* (Part 2), pp. 66–70.
14. Rosenblatt, F. "A Probabilistic Model for Information Storage and Organization in the Brain," *Psychol. Rev.*, (1958), *65*, 386–407.
15. Samuel, A. L., "Some Studies in Machine Learning, Using the Game of Checkers," *IBM J. Res. Devel.*, (1959), *3*, pp. 210–229.
16. Unger, S. H., "A Computer Oriented Toward Spatial Problems," *Proc. IRE*, (1958), *46*, 1744–1750.
17. Uhr, L., "Machine Perception of Printed and Handwritten Forms by Means of Procedures for Assessing and Recognizing Gestalts," in *Preprints of 14th Assoc. for Computing Machinery Meeting*, Boston (September 1959).
18. Uhr, L., and C. Vossler, "A Pattern Recognition Program That Generates, Evaluates and Adjusts Its Own Operators," in E. Feigenbaum and J. Feldman (eds.), *Computers and Thought* (McGraw-Hill, 1963), pp. 251–268.

III. END USES OF INFORMATION

8

Expressed and Unexpressed Needs

HENRY W. BROSIN, M.D.
Director, Western Psychiatric Institute and Clinic
Department of Psychiatry
University of Pittsburgh

THE PLACE FOR HUMAN COMPONENTS IN AN INFORMATION-RETRIEVAL SYSTEM

It is a privilege to have the opportunity to meet with you today in order to stress again some of the crucial needs for the human components which must be taken into consideration in designing information-retrieval systems (IR). The issues are familiar enough through repetition in numerous journals in and out of the communication field, as well as in textbooks, conferences, symposia, the reports of the so-called Crawford (1962), Terry (1962), Weinberg (1963), and Visscher (1963) Committees and of the American Psychological Association.[1,5,17,18,19] There is no doubt that if the price is right, the mechanized components of the IR systems can do anything now conceived as desirable at the automated level. The revolution in computer technology is in being and we can look forward to many startling developments within the next two to four decades. However, will the expert manpower required for appropriate processing of input, storage and retrieval be equal to the challenge? Even if the technological victories are as impressive as predicted or hoped for by 1999, will there be information specialists who can help the user expand his concepts and broaden and deepen his search?

SOME CURRENT DEFICIENCIES

My background and interests are those of the biomedical and the behavioral sciences communities. Since I have no formal experiments or surveys to report, I will report discussions and experiences utilizing the clinical case (anecdotal reports) or natural history methods.

There is no need to review here the still unmatched potentials of the human organism, with its 10^{10} elements in the central nervous system, both physiologic and psychologic as a receiver, storer and coder-decoder of information, for this has been done by Quastler (1955), Broadbent

(1958), and Miller (1963), among others.[3,9,12] Not only is it the largest system known to us, but it is the most flexible, and the utilization of these properties is the central theme of my essay. By sound planning, in accordance with evolutionary and educational concepts, we must develop experts who will help the scientific community gradually learn how to react more intimately with the various machines which are becoming available. But even after we have consoles in our home studies or *talking typewriters* which learn to help us correct our errors, we will probably have need for human intermediaries at some stages of the process. Since many of these consoles will probably not be *generally* available for the behavioral sciences for several decades, many of us would hope that, in addition to and not in place of, there will be careful research planning about improving IR systems with resources now becoming available. We can do better work in spite of the inevitable cultural lag, and the fact that current guesses are not as convincing as well-controlled experimental studies or high-level logical studies of abstract systems. As you know, the problems of information retrieval in the behavioral sciences are in some ways more acute than in the physical. Rapidly developing subject matter, new subject areas, and new interdisciplinary needs make identification of author and information processor responsibilities much more difficult, for the needs of users become more complex and elusive, while the material becomes more widespread over several disciplines and more difficult to identify. Changing concepts and nomenclature increase the complexities.

This is not to minimize the numerous technical problems in designing appropriate hardware, nor the logical problems involved in making the IR systems maximally effective. Most investigators and scholars, both as producers and users, would welcome the new massive instrumentation. However, there are many who have grave doubts about the assumption that the most internal problems of an information center or library would be satisfactorily solved if modern computer techniques *now available* would be installed. Good service may not be as easily purchased as computers. Quantity is not enough, and effective methods of selection must be found if instruments are to be useful. Capacity, speed, and other technical problems are not the limiting problems for handling the 10^{14} characters calculated to be the total sum we need to automate. It is another task to provide essential information as needed in appropriate forms to investigators, scholars, teachers and students, each with his own needs, both formal and idiosyncratic. If time permitted it would be worthwhile to delineate the differences in need of each of these categories at different stages of the career of the person and also of his project. I would support the claim of the active investigator for the highest priorities in designing IR systems. However, it is the experienced investigator who has already developed a reasonably good system for his own purposes, often depending upon the informal channels for current information, as shown by the APA studies (1963), who has the least need for a new system.[1] The inexperienced investigator, and the interdisciplinary scholar, teacher, and student have much more complex needs which are not easily met. It is from these areas that approaches to "unexpressed needs" become obvious from autobiographic accounts.

One citation from the field of anthropology, which is representative of much that we need to correct in all of the behavioral sciences, including psychology and psychiatry, will tell the story.

> Each subject has its own peculiar library problems, and anthropology has some especially serious ones. In the first place, the systems of organization used in most general libraries in the United States make it exceptionally difficult for anthropologists to find the literature of their field.... [These systems] were devised and put into practice many years ago when anthropology was generally visualized as a very small subject, and its point of view was familiar to few readers. The result is that traditionally and in current practice books which are written from the comparative point of view are catalogued and shelved with books which are not, because of some similarity in subject matter discussed. In most general libraries the literature of anthropology is scattered from religion and philosophy to warfare and marine transportation. This situation may have the advantage of calling the attention of an occasional reader from another field to anthropological contributions related to his interest, but it creates undeniable difficulties for anthropology students. . . . Most libraries use subject headings of the Library of Congress, because these headings are printed on Library of Congress catalog cards and are also available in a bulky manual. Unfortunately, Library of Congress subject headings are designed to help the "general reader" who knows no anthropology, and the categories which are familiar to students are either not represented at all or appear under unfamiliar names.[14]

It is undoubtedly redundant before this highly informed group to spend more time on the current inadequacies of indexing, classifying, abstracting and cataloging which are only too well known to all. However, I will mention a few which can furnish us with lessons for the future, the most pressing of which seems to me to be the need for subject specialists as catalogers. The Library of Congress places a book by a psychiatrist, Paul Federn, entitled *Ego Psychology and the Psychoses*, under the subject heading "egoism," and the subject cataloging is accepted by university library catalogs across the United States. For most practical purposes, it is lost to professional workers under this heading. I am happy to say that most psychological writings are better indexed via *Psychological Abstracts* than those in other behavioral sciences, but it will take much workmanlike skill and many years to correct the current situation. I find amusing the summary by Hans Peter Luhn, a leader in the field of information retrieval by computers, who "remarked, on looking over rough versions of the figure (of public dissemination), that the contemporary information retrieval approach was like sending stale bread to China via air express."[1] This is said about the field of psychology where the public dissemination is probably superior in the sense that a larger portion reaches more interested persons, and faster, than in other behavioral sciences where a three-year lag is all too common.

Many of you may have been led to believe that the MEDLARS System utilizing the National Library of Medicine's new *Medical Subject Headings* (3d ed., January 1964) would solve many of the old problems. It has distinct advantages for the older, conventional medical areas, but is a great disappointment to those in the biomedical community who require information in the other behavioral sciences. The problem is a complex one to which there are no easy answers. It may be argued that there should be a separate *"Index Psychologicus":* There is good evidence currently that under the leadership of Dr. Martin Cummings of the National Library of Medicine, much hard work is now being done to improve the retrieval potential of existing systems for the behavioral sciences.

PROBLEMS IN INDEXING, CLASSIFICATION AND CROSS-REFERENCING

Some critics claim that basically it is not feasible for any system of subject headings to be really satisfactory and propose such new techniques as the KWIC (Keyword in Context) Index recently used by the National Conference on Social Welfare.[10] This is an automatic coding device or "title permutation indexing" which is a combination of word and machine indexing. The total operation is performed automatically and the title and related bibliographic data have been key-punched for use as input by the computer. Titles are amplified by editorial insertion of keywords which help identify the content of the document.[10] This thesis that no system of subject headings can ultimately be satisfactory is supported by the failure of *Index Medicus* to mention more than a few score key psychiatric concepts (with conspicuous omission of those associated with psychoanalysis) or to provide coverage of administrative and forensic psychiatry. It fails to coordinate older subject headings, such as "mania," with proper cross-references. Furthermore, there is persistent confusion of terms from psychosomatic medicine with those of conversion hysteria, and similar misunderstanding of the new use of old words, or attempts to fit new technical terms under old headings, such as placing "narcissism" under "egocentricism." There is a failure to link related topics in psychiatry, or to link areas such as psychosomatic medicine with appropriate headings in the autonomic nervous system. "Psychoanalytic interpretation" is a heading used to cover a wide variety of subjects from history, literature, biography, to clinical work and dream interpretation. The failure to make appropriate linkages prevents the highly desirable dissemination of significant and relevant experiments from neuroanatomy, biochemistry, neurophysiology, clinical neurology and allied disciplines to psychiatrists, psychologists, and other behavioral scientists and vice versa. It also delays transmission of vital findings from the basic scientists to practicing clinicians and vice versa, where the analogy between basic scientists to engineers may occasionally be useful. Here is another approach to the problem of exploring for and identifying "unexpressed needs."

Another example of the failure of current IR systems at a higher level of abstraction may be seen in drug evaluation. Only gradually are workers in the field of evaluation of drugs with human subjects becoming aware of the manifold difficulties in establishing genuinely useful "control" series, even though the placebo phenomenon has been known since Hippocrates and the bibliographic coverage is somewhat better than any field in psychiatry.[11] The tragedy of thalidomide is a good example of the cost of delayed transmission. Here is a good example of an unexpressed need due to traditional thinking and attitudes but many similar examples could be found in the well known diseases.

PROGRESS IN STUDYING THE NEEDS OF SCIENTISTS

There has been considerable growth in sophistication in the IR research community recently regarding new studies which attempt to establish some solid facts about how scientists really seek, find, and utilize information at various stages of their careers and during various phases of the development of their projects. The early assumption that the primary mission would be accomplished if the predominantly relevant *published* articles pertaining to an investigator's immediate project were made easily available has been altered considerably. Due to the considerable delay in reporting and dissemination in *appropriate* journals, the lack of proper addressing through inadequate indexing and classification, the inadequate comprehensive coverage in serial abstracts and reviews, it is apparent that other methods must be found to help the investigator in his primary task. It is a sad commentary that chance may play a large role in an important article becoming publicly visible. Much could and should be said about the central importance of critical reviews written by the best people available, as found in Germany and the U.S.S.R., but even here the bias of the reviewer may play a vital role and steps must be taken to build in devices to protect against the loss of significant contributions.[1]

A most significant advance came about when IR investigators became aware that beyond the technological and logical problems of IR proper was the problem of the type of questions which were being proposed to the IR system. Kent, Swets, Swanson and Clapp have each examined the techniques which partially solve the problem of obtaining data from a record in answer to a particular request.[4,6a,15,16] Kent is presenting at this conference his proposal called "The Information Retrieval Game," which should arouse considerable interest.[6c] Clapp proposes "Associative Chaining as an Information Retrieval Technique," which also has merit in regard to the vexing question of what a request for information may really mean in depth. He finds that in some situations, "the answer to a query is not a single item but a collection of items organized on the basis of the original question." While not proposed as the ultimate IR method, he suggests that it is a useful "step in another direction, which will bring us closer to our ultimate goal—the design and construction of wide class useful information retrieval systems."[4] While the report I have does not present the terminal results, the success to the date of publication (November 1963) convince him that the chaining concept—"that answers to a query must be constructed from several items so as to span the question —will eventually be incorporated into the next generation retrieval systems."[4]

Kent's examinations of the basic assumptions go much deeper into the individual ways of perceiving nature, or into the paradigms which each of us has as "fundamental hypotheses or models in respect to which thinking occurs. As in all perception, a shift from one hypothesis to another may occur at any moment, and unpredictably.[6c] The provisions for handling surprise, novelty, and even the "irrational" as an anticipated part of the work to be done by the system, is in itself an innovation. This examination and statement of the nature of the requestor's hypothesis is more in line with biological models and is deserving of serious attention. I do not know without direct experimental experience whether the "game-theory" technique will prove useful in long term exploration, but believe that trials in appropriate areas of IR activities will be worthwhile because there may be relatively delimited sequences which can be studied with considerable benefit. The weakness of most game-theory models, as you know, is that new postulates or rules of the game must be written to provide for new contingencies, and some operations become too complex for such analysis.

The views cited above are consonant with the position taken by Kessler and his colleagues at the Lincoln Laboratory,

> that the evaluation of new ideas and components must be made in a system environment and not in terms of parameters unique to each component. For this reason it is important to develop a measure or estimate of "system goodness" or figure of merit. . . . A distinction is made between scientific message units and their mode of propagation. The message units (scientific talks and papers) are considered adequate for their functions, but they are encountering increasing losses and delays in propagation. . . . Valid directional indexing should be sought in the operational history of the author and the intended reader. . . . A scientific paper is a reflection of the operational history prior to publication. We now extend this concept and say that a scientist's information needs are also determined by his operational background.[7]

He suggests deriving an index of a consumer's information needs from extensive examination of the scientist's work habits, publications and his own statements concerning these components.

I regret that my limited acquaintance with the field as well as limited time prevent me from citing other relevant authors on this theme. Although a significant number of my references are only one year old, and most of them have been published within 3–4 years, I suspect that I am not quite up-to-date in this wonderful field with its unusual acceleration.

I believe in the exploratory value of the natural history method and the clinical case method which have served us so well in the pioneering stages of several disciplines, and would therefore suggest that much more use be made of the autobiographical methods to determine the working habits of scientists. Very few men can write well about themselves, and certainly not in depth. Perhaps St. Augustine and Pascal deserve special accolades because almost alone they came close to revealing clearly some of their motivation, whereas even such a braggart as Benvenuto Cellini missed genuine insights. However, if responsible scientists worked systematically collecting freely written autobiographies focussing on attitudes and work patterns as well as developing questionnaires and other measures, there would become available rich source material and insights for designing new experiments. As a psychoanalyst, I must add that much about any man cannot be written, e.g., Freud's own "interpretations" of his own dreams.

THE CHALLENGE TO THE INFORMATION SPECIALIST

There is no substitute for the exercise of intelligence in controlled experimentation or research scholarship. Computers and IR systems can only do what they are programmed to do and are no substitute at this time for personal mastery of scientific material or creativity. However, IR systems conceivably can be designed and implemented for a more intimate interaction between living men who are biological organisms and the computers and systems in such ways that the ends and not the means will be paramount.[13] Instead of merely purveying facts accurately, quickly and at low cost per bit, information specialists should take their built-in, intrinsic, proper place in the scientific and total academic community so that they may participate in every phase of the scientific or humanistic process from its early beginnings to accomplishment. As a biologist and former engineer who is interested in thinking about thinking, it seems inevitable that information scientists should be able to help create a worldwide intellectual and social climate through active participation and leadership in the scientific and other academic communities not only through research, but by being educators who influence profoundly those around them in all departments of the University and the community at large, including industry, government, and the world of affairs.

It seems entirely feasible to a biologist who subscribes to a belief in cultural evolution, that information specialists should be leaders in the effort to enhance man's intellectual powers through the use of prostheses or tools which are extensions of himself. Few can doubt that we have gained considerably in our ability to abstract much better and manipulate propositions more quickly in the approximately one million years of our existence as *Homo sapiens*, through the development of language—i.e., communication systems. It is a legitimate expectation that in an improved intellectual climate, with better mastery of our material, the talented men of the future will be able to achieve somewhat higher orders of abstraction in a framework of improved logics in many fields. We have now a remarkable example in physics, and we can hope that a similar epoch will emerge in the behavioral sciences.

CONCLUSIONS

1. Information retrieval is an intellectual and not merely a mechanical operation. Its ultimate goal is to help to provide that creative leisure for talented men which will be of greatest benefit to the total community.
2. Research in IR methods which take into account psychological and sociocultural factors of experimentalists, authors, processors of information and users at all stages of their careers and of their projects, is urgently needed.
3. The members of individual disciplines must take a much greater interest in helping the IR experts design systems. Scientists cannot expect good results based upon abstract designs with little or no research on user needs. Multiple research centers for IR systems

with private and local, as well as Federal grants, are desirable to provide the diversity needed.

4. Manpower rather than technology will probably be the limiting factor in designing and maintaining genuinely useful IR systems, even in 1999. Furthermore, we urgently need more IR specialists now, who have a reasonable mastery of a particular field, to do research, design and help operate new indexing systems, promote better abstracting which may prove to be the second biggest need after good indexing, and assist all publishing channels to do a better job using the newer concepts expressed at this conference. It would seem reasonable that all large professional organizations organized around disciplines with professional journals, should attract IR specialists with the equivalent of a doctoral training in the discipline to help the editors and the membership to take advantage of the newer IR concepts and technology. Since both the material and the IR methods will alter significantly in the next few decades this should be an ongoing process.
5. Information specialists should take their proper place in the academic community as investigators, scholars and educators in the teaching-learning process, and establish balanced programs in which technology and the human components each have their appropriate functions.

REFERENCES

1. American Psychological Association, Reports of, Project on Scientific Information Exchange in Psychology, Vol. 1, Washington, D. C. (1963).
2. Bolt, Beranek, and Newman, Inc., "Research on Concepts and Problems of Libraries of the Future," Final Report to The Council on Library Resources, No. 1101, Cambridge, Mass. (November 1963).
3. Broadbent, D. E., *Perception and Communication* (Pergamon, 1958).
4. Clapp, L. C., "Associative Chaining as an Information Retrieval Technique," Report No. 1079, to The Council on Library Resources, Bolt, Beranek, and Newman, Inc., Cambridge, Mass. (1963).
5. Crawford, J. H., "Scientific and Technological Communication in Government," Report on Scientific and Technological (STINFO) Activities to Dr. J. B. Wiesner, Special Assistant to the President, U.S. Department of Commerce, Office of Technical Services, AD299545, Washington, D. C. (No date given but became available in 1962.)

5a. *International Conference on Scientific Information*, 2 vols., National Academy of Science—National Research Council, Washington, D. C. (1959).

6. (a) Kent, A., *Information Retrieval and Machine Translation* (Interscience, 1960); (b) Kent, A., *Textbook on Mechanized Information Retrieval* (Wiley,

1962); (c) Kent, A., "The Information Retrieval Game," chap. 25, this volume.

7. Kessler, M. M., "An Experimental Communication Center for Scientific and Technical Information," MIT, Lincoln Laboratory, No. 4 G-0002, Lexington, Mass. (March 31, 1960).
8. Law, A. G., and Richman, A., "Processing Psychiatric Research Data," *Data Processing for Science and Engineering* (Jan.–Feb. 1964).
9. Miller, J. G., "The Individual as an Information Processing System," in W. S. Fields and W. Abbott (eds.), *Information Storage and Neural Control* (Thomas, 1963).
10. National Conference on Social Welfare, "KWIC Index to NCSW Publications," 1924–1962, Ford Associates, Columbus, Ohio (1964).
11. *Psychopharmacology Abstracts*, vol. 1, 1961, Philadelphia, prepared by Medical Literature, Inc. for Psychopharmacology Service Center, U.S. N.I.M.H., Bethesda. (Last issue published, vol. 2, no. 12, December 1962, which appeared in mid-1964; index for 1962 not published; no 1963 issues have appeared.)
12. Quastler, H., in *Human Performance in Information Transmission,* U. of Illinois Report No. R-62, 1955.
13. Richmond, P.A., "What Are We Looking For?" *Science*, vol. 139, no. 3556 (1963), pp. 737–739.
14. Rowe, J. H., "Library Problems in the Teaching of Anthropology," in David G. Mandelbaum et al. (eds.), *Resources for the Teaching of Anthropology* (U. of California Press, 1963), pp. 69–70.
15. Swanson, D. R., "Searching Natural Language Text by Computer," *Science*, vol. 132, no. 3434 (1960), pp. 1099–1104.
16. Swets, J. A., "Information Retrieval Systems," *Science*, vol. 141, no. 3577 (1963), pp. 245–250.
17. Terry, L. L., "Surgeon General's Conference on Health Communications," Nov. 5–8, 1962 (Washington, D. C., U.S. Dept. of Health, Education, and Welfare, Public Health Service (February 1963).
18. Visscher, M. B., "Communications Problems in Biomedical Research," NAS-NRC, Washington, D. C. (Oct. 31, 1963). Contains an excellent 285-item bibliography.
19. Weinberg, A. M., "Science, Government and Information," a report of the President's Science Advisory Committee, The White House, Jan. 10, 1963 (Washington, D. C., U.S. Government Printing Office, 1963).

9

Scientists' Requirements

WALTER M. CARLSON
Director of Technical Information
Department of Defense

Within the broad framework and sweeping scope of this conference, it is especially pleasant to be discussing the reason why we are here—to serve the user. It is axiomatic that no product is sold successfully nor any service used economically in the long term without customer satisfaction. This axiom seems to have been widely ignored in this country's work on electronic information handling, and these remarks will therefore be devoted to an examination of information systems from the customer's point of view.

As a point of departure, there are a few of my own basic considerations and definitions that need to be stated in order to minimize confusion.

First, it is my conviction that there will be no electronic information-handling systems for use by this nation's scientists for a matter of decades. We now have electronic means for communication of messages and data. We now have electronic means for processing of data into new formats by predefined procedures. We now have electronic means for processing of data about documents or the textual content of documents. We do not have, nor are they in sight, any electronic means for extracting meaning from data, signals, or text. Since the concept of information specifically refers to the extraction of meaning from data, signals, or symbols, I shall try to use the word "information" only in its proper context.

Second, the information requirements of scientists will be the only subjects discussed in this paper. Interesting but completely different problem areas involving an engineer's requirements or involving document handling, data processing, or library automation will be included only to the extent needed to provide a comprehensive picture of the central theme—the needs of scientists for information.

Third, the use of information by scientists is a richly discussed and a largely unexplored topic. Despite the many volumes on the subject and despite the energy devoted to defining, designing, and installing systems for "better information handling," no one has come forward with an authoritative statement of the basic mechanisms involved. Without a clear idea of the "why," there can be no rational selection of the "what," and there can be no practical description of the "how."

Fourth, the role of the scientist is to produce new knowledge, which is, in itself, information for use by others. This definition of a scientist and the implied relationship to the information he uses and generates does not set aside the tasks he performs in doing calculations, in designing and running experiments, in making calibrations, in supervising technicians, in attending administrative meetings, or in selling his projects to management. It simply means that these functions are part of the complex picture of the scientist that also includes the functions in which we are chiefly interested today: talking or writing to a colleague to find out "what's new," reading the literature, thinking about current problems, writing or delivering reports on work progress, and seeking the technical advice of supervisors or co-workers. I am under the impression that these latter functions, which can be broadly classed as scientific communication, occupy more than one-third of a scientist's time devoted to technical matters at work or away from work, and that they are as important to him as anything else he does.

Fifth, last, and perhaps most important, there are no measuring tools available for telling us either that present systems are inadequate or that any proposed improvement will change our scientific productivity. It seems inconsistent and misleading to discuss "information sciences" in the same context as the uses of information. How can we be scientific about the handling of information as long as we lack the means to spell out in quantitative terms the most elementary aspects of human conversion of data, signals, or symbols into information? To me, a requirement is an identifiable need or prerequisite derived from a knowledge of current conditions or from an estimate of future conditions. Without the ability to quantify our knowledge or estimates, previous statements about requirements have assumed the status of sheer speculation. Charles Bourne addresses this critical issue in more detail in his paper. But there is little doubt that this lack of certainty over hard, measurable facts on user needs has resulted in the well-known attitude of individual scientists and the scientific community toward new approaches to data processing and document handling—they don't want them.

Accordingly, all of the major conclusions of this paper must be evaluated as one person's set of speculations. As indicated earlier, however, these speculations are as closely related as possible to the viewpoint of the scientist-user.

The process we are dealing with is traditionally considered to occur in four stages in the human mind: (1) observation (or acquisition); (2) gestation (or mulling-over); (3) correlation (or synthesis); and (4) confirmation (or making-sure). In an operational sense, there can be little doubt that every worthwhile idea and every worthwhile use of data or documents involves these four stages in the mental process of a scientist.

In the dimensions of space and time there seem to be no limitations. Each of the four stages can occur virtually anywhere; in fact, the third stage, often dubbed the "Aha" point, is often suspected of occurring at the shaving mirror more often than anywhere else. Furthermore, there are recorded instances of two or more decades elapsing between observation and correlation, and there is nothing to prevent all four stages occurring within a matter of milliseconds or to prevent portions of the first two stages overlapping each other in time.

For today's purposes, the information requirements of scientists are tied to the observation and confirmation stages. As an engineer trained in physical processes, I shall avoid trying to make comments about the intricate processes by which the human mind is able to order and reorder complex signals to extract meaning and to synthesize entirely new and previously unrecorded information. Nor does it seem necessary to review here what is known about the brain's capacity for storing apparently unused signals for periods approximating a lifetime.

Certainly, for present purposes, the organization of data, signals, or symbols to serve a scientist in his acquisition of existing knowledge is a sufficient challenge to keep us all busy for a long time.

It seems generally agreed that a human being goes about the complicated business of acquiring data, signals, symbols, and their documentary forms with two definitely different but interrelated purposes. The first purpose is general and the less tangible. He notices things, he reads documents, and he talks to other people to satisfy his never-ending curiosity about the world in which he lives. The observations he makes may or may not bear any known correlation to anything he has stored in his mind. But the important point is that he does observe, he does attach meaning, and he does store enormous quantities of new material. And we all have had the occasion to notice that our most creative scientists have had an outstanding capability to observe and store isolated data that have no bearing whatsoever on current interests or past associations. In the documentation field, this first purpose is usually embraced by the term "current awareness."

The second purpose is specific and more readily studied. A human being has a problem to be solved or a task to perform. If he is unable to reach some desired objective with the information resources stored in his head, he has to search for data, signals, symbols, or their documentation that can be converted into the additional information he needs. He looks in his files, he talks to his colleagues, he looks up anyone he believes to be specializing in the subjects involved, he sends for any reports he has heard about, and in a remarkably small fraction of searches he asks a librarian or information specialist to help him. He performs this search to expand his capability for solving the problem or performing the task. He usually received much of the material he needs in a reasonably ordered form when he undertook the job, but he wants more—how much more he does not know at the observation stage. In the confirmation stage, however, his search is highly specific. He has drawn conclusions, formed ideas, gained insights, or postulated hypotheses. He now wants to see what he can find in the work of others to check himself. He wants to see if he can follow and extend the reasoning of others on the basis of his own correlations. It is here that he has his greatest desire for fast, accurate, and comprehensive retrieval of recorded information.

One obvious speculation that develops from the logic of the "model" just formulated is that scientists make literature searches chiefly in the last stages of their work on a particular problem. It is only reasonable to expect that people tend to search when they know what whey are looking for.

In any event, it is quite important to the electronic implementation of scientific literature searching that the accuracy of this speculation be tested. This means that we must have much more data than we now have on how scientists *actually* acquire information to replace the comfortable —and apparently wrong—traditions that serve as the justification for so many of our procedures now.

This lack of data about what is really happening now has been a matter of priority attention within the Department of Defense for the past 18 months. Some of our experiences are germane to this discussion of the end uses of information.

Based upon the extensive experience of earlier studies of how information is used, the DOD study has started with at least two basic tenets:

1. If we intend to find out what technical people actually do in acquiring information, we should be careful to assume nothing about their habits at the outset.
2. No data-gathering procedure short of personal interviews of a fairly large sample is likely to produce data of the type and quality needed for answering the question of how people acquire and use information.

While it will be some time before the results of the DOD survey are available, some valuable lessons have been learned. For example, the early designs of the interview assumed that we could find out about both the general and the specific purposes of acquiring information. We learned swiftly through pilot tests of the interview that the general, or "current awareness," mode is beyond the reach of practically realizable interview procedures for large samples. Thus the survey is restricted to the specific, or "task-oriented search," mode, which seems to be quite manageable with a semistructured interview that relies heavily upon the respondent's being able to provide key data about items of information he acquired to do a job.

Another lesson learned is that technically trained people welcome, almost aggressively, the opportunity to discuss their information-gathering habits, and it is vitally important that the interviewer have enough technical training to be able to follow the answers, capitalize on unexpected leads, and draw reasonable conclusions of his own about the truth of the statements being made by the respondent. Very sharp differences arose in the quality of the data obtained during the pilot testing, and these were traced directly to the ability of the interviewers to understand what they heard.

A third lesson learned was a gratifying confirmation of the first of our basic tenets. It turns out that technically trained people have almost no formal instruction in the use of the available information resources, and their ingenuity in developing ways to find what they believe they need results in a great variety, great effort, and a general sense of satisfaction with the way things are now. It is becoming quite clear that any assumptions that might have affected the interview procedures could have caused serious problems in finding out what people *actually* do when they sense a need for more scientific information to do their job.

It is our intention to publish the results of our survey as soon as possible after it is completed.

The real key, the real determining factor in the long run, however, is how the scientist-user will accept newly developed data and document systems created "for his benefit." If the customer does not buy the new systems in the sense of making good use of them, our country could waste huge sums of money. These new approaches, especially those involving electronics, are expensive. On the other hand, if the customer is happy with the newly offered services and uses them to advantage, we may trigger an era of scientific development that will transcend anything we have seen to date.

Thus the challenge to all who would invent, design, install, or operate electronic information-handling systems for the benefit of our nation's scientists is to motivate the scientists to use the new systems. It might pay to look at the present state of this motivation, despite the lack of good measuring tools. Furthermore, let us examine the motivation as individual incentives controlled by a system of rewards and penalties.

As a summary statement, it appears that the penalties accruing to an individual who seeks information are more persuasive than the rewards. The situation might be typified by a brief examination of two circumstances in which rewards dominate and three in which penalties appear to control the individual's behavior.

Recognition for making a unique addition to human knowledge is a reward avidly sought after by most scientists. This recognition is accorded in the form of patents or in some form of publication, such as journal articles. The issuance of a patent carries a warranty that the work embodied in the invention has not been published previously by someone else. In a less restrictive sense, the acceptance of a refereed article by a reputable scientific journal carries a similar warranty. The incentive to reap the reward of recognition is strong. The scientist and his legal or editorial associates place a heavy demand on the available literature search resources. The requirements are high specificity and complete coverage, but there is usually plenty of time (months or years) to conduct the search.

Satisfaction with one's own performance is a reward that motivates a large segment of the scientific population. The competition from his peers provides one of the strongest incentives for excellence to which a scientist responds. Accordingly, he spends a fairly large fraction of his time using all available modes of technical communication to maintain an active, and highly personal, intelligence network in the field of his specialty. The scientist wants to avoid repeating work completed by others, but he also wants to know enough details about the successes and failures of the others so that he can build upon them with his own knowledge and competence. The requirements are for high specificity, for a very short time between a technical event and the circulation of data about it, and for two-way oral communication whenever possible.

While these are strong incentives, they lack one of the better known ingredients—money. When we look at the situations where money enters, the incentives are less favorable to extensive use of the available information.

Project cost controls are a good example of the situations in which penalties operate to inhibit the acquisition of information. Since information is not recognized specifically as a resource, neither its acquisition nor its dissemination appears as a project cost item. Under these circumstances, the scientist who invokes the use of new or specialized literature services finds that he is under pressure to conserve project funds by cutting down on expenses not covered by the project estimate. When he has the choice of reducing his own man-hours on the project to pay for the otherwise unbudgeted literature services, the scientist reacts to such a penalty in his own self-interest, and he tends to conserve project dollars to cover his own salary.

Research-program goals are sometimes applied in a manner which penalizes efforts to acquire complete data on a subject. Undue emphasis on commitment of budgeted funds can and sometimes does result in authorization decisions that are based upon financial rather than technical considerations. The scientist who insists upon a complete review of the literature on the subject before a new project is initiated finds himself unpopular with the manager whose goal is to get funds committed and who places this goal above technical verification of need for the project. The penalties that can operate in situations of this sort are quite persuasive. Before concluding that this is a rare situation, you may wish to reflect upon how many research managements now insist that a complete literature review is a prerequisite to authorizing new work.

The trend toward viewing research as an institution carries serious implications for scientists who are "information-minded." The increasing national investment of money and vital manpower in scientific research places heavy pressure on the administrators of research organizations to maintain those organizations. One of the consequences of this pressure is to conduct research for the sake of conducting research. While this practice has not become widespread, it serves as a major deterrent to technical communication on two counts. First, the people conducting research for its own sake do not wish to be told that the problems they are working on have been solved by someone else. Second, these same people are reluctant to see their own results circulated and applied because someone might get the idea that the problems assigned to them had been solved. In either event, the overtone of job security carries a penalty for effective information transfer, and this potentiality should be given very careful study in evaluating a scientist's requirements for information.

Perhaps it would be more accurate to relate the scientist's requirements and his administrative environment as components of a single entity. Certainly the individual scientist working at his own pace on tasks of his own choosing is becoming rare in our technical economy, and data or document systems designed to serve a vanishing breed are not likely to find a very large market.

A concluding summary statement of the foregoing comments would be a strong plea for quantitative, detailed study of the scientist's use of data, signals, or symbols and their documentary forms before an investment is made in electronic systems to serve him. While it is not yet clear how to evaluate the degree of inefficiency introduced by existing procedures, it seems quite clear that the evaluation must be made in terms of the rewards and penalties that accrue to the individual scientist. After all, he is the customer to be served, and he will "buy" only those new approaches that will help him and not hurt him in his current environment. And, as we have seen, it will be profitable to recognize the complexities of that environment in a highly pragmatic manner. Information is approaching the status of a commodity, and commodities are tested in the market place, not in theoretical discussions.

10

Some User Requirements Stated Quantitatively in Terms of the 90 Percent Library

CHARLES P. BOURNE

Stanford Research Institute
Menlo Park, California

INTRODUCTION

Librarians, publishers, and information system engineers have very little verified information and few guidelines to describe the user's specific requirements for information. Such information is needed to properly design or evaluate the information systems. To date, most of the statements of requirements have been rather subjective, and often reflect opinion rather than actual fact. Relatively little objective data have been obtained. This is probably due in large part to the fact that there are extremely difficult methodological problems in trying to determine and state user requirements in a meaningful manner. This paper suggests an approach or point of view that might help this situation by providing a method of phrasing the statements of user requirements in a more convenient and meaningful manner. This paper also furnishes several examples of such statements, and discusses the techniques and data that support these statements.

In this paper, attention is initially focused on the information requirements of workers in the field of science and technology, with no serious attempt made to include workers in other fields. However, it seems quite likely that the approach, and perhaps even the stated principles, could be extended and generalized to cover other fields of knowledge.

THE BASIC APPROACH: THE 90 PERCENT LIBRARY

The basic approach or point of view suggested here is first to envisage the library users as a composite or aggregate collection of people with a great variety of interests, approaches, needs, habits, and idiosyncracies, and then to ask the basic question, "What does the library have to do to satisfy 90 percent of this population's needs?" That is, what periodicals should be acquired so that 90 percent of the periodicals they use and make reference to are available? What literature searching speeds shall be provided in order to meet the response times required for 90 percent of the requests? By taking this point of view, our attention is focused on the actions or services necessary to satisfy a specified fraction of the user population. In this way, no attempt is made by the designer or operator to satisfy every possible request or need that might occur. Both the system designer and operator thus openly acknowledge that, in some instances, some users' needs will not be fully met. However, this approach keeps the library from being overdesigned or from going to extreme efforts in an attempt to make it all things to all people. Past experience by many types of organizations (e.g., transportation industry, retail sales) indicate that a disproportionate effort is usually required to raise the system performance from a capacity to satisfy some high fraction (e.g., 90 percent), to satisfying 100 percent of the user requirements. The libraries are no exception to this rule. The point of diminishing returns is such that it is probably more effective to run an information service at something less than a capability for 100 percent satisfaction of the users' requirements. The figure, 90 percent, is used in this paper as an example. Any other figure could of course be used, established by the people responsible for the design, operation, and support of the library. It would seem that many libraries in fact already subscribe to this principle even though it may not be stated so explicitly. For example, few, if any, local libraries try to duplicate the holdings of our national libraries in order to immediately fulfill any local request, but instead assume that they can satisfy "some reasonable fraction" of their requests from the local collection and handle the remainder in some other way.

This approach of stating requirements of performance measures in some numeric terms has certainly been used before in many types of applications. It may even be practiced to some extent in some libraries. However, it is mentioned and reemphasized here because it forms the basis for the descriptions of requirements to follow.

The question of whether the library should be designed to serve a large fraction (say 90 percent) of the general user group, rather than the remainder that provides the exceptional requirements is another and separate topic, not to be included in this discussion.

IS IT MEANINGFUL TO STATE USER REQUIREMENTS IN SUCH TERMS?

The answer to this question is "yes" for some requirements, but certainly not for all of them. Consider the following statements as examples of requirements that could be stated in these terms.

Ninety percent of the information needs of a given user population are satisfied by:

1. Books that are less than ____ years old.
2. Periodicals that are less than ____ years old.
3. Retrospective search speeds of less than ____ days.
4. Document delivery speeds of less than ____ days.
5. A collection of less than ____ chosen journals and less than ____ chosen books.
6. A current-awareness service that periodically furnishes information at intervals of not more than ____ days.
7. A reference retrieval service that provides not more than ____ percent irrelevant material with the search results.

Such statements might be posed as general principles, or, more precisely, as hypotheses to be tested, and with the specific missing numbers determined empirically for separately defined user populations. There are indications (discussed later in this paper) that the specific numbers might not differ greatly between different user populations. Thus it might be possible to use a formulated set of requirement statements and the accompanying empirical data (expressed as a single number or range of numbers) as standards for the design and evaluation of information systems and services. The specific numbers could be continually modified as time goes on (similar to the development and maintenance of critical tables) to reflect the acquisition and analysis of more empirical evidence and changing user needs. This approach is sensitive, of course, to the argument that the empirical data may reflect current use patterns (habits) rather than actual need, but this may still provide better statements of goals or requirements than are currently available. It should also be noted that the exact figure stated for the specific requirements will be tempered by practicability. The stated "needs" will change as technology makes improvements possible.

WHAT SPECIFIC REQUIREMENTS CAN BE STATED IN THIS WAY AT THIS TIME?

As mentioned earlier, many of the published statements regarding user requirements are really statements of opinion, or hypotheses, and are not statements that have been backed up by reasonable amounts of supporting evidence. It would be extremely helpful if data could be collected, organized, critically reviewed, and presented in a way that supports statements of user requirements. The general statements below, and their supporting evidence, are presented as a start toward this objective and as an example of the suggested approach.

General Statement No. 1 (use of materials of various ages): "For the majority of users in most fields of research, a specified fraction of their needs for literature can be fulfilled by literature that is younger than some given age."

First Specific Example of General Statement No. 1 (age of journal material used in science and technology): "For the majority of users in most fields of science and technology, 90 percent of the needs for journal articles can be fulfilled by journals that are less than 30 to 50 years old. The exact number depends on the subject field."

After the general statement has been made, other more specific statements can be made for various special cases, such as different subject fields and user populations. There may of course be arguments regarding the methods used to obtain the data, and disagreement about the value or validity of the actual numbers used in the specific statements such as the one above. The numbers could be modified when more evidence is collected and critically analyzed.

An example of the data that could be used to support the first general statement and its first specific example appears below. They were assembled from the reported results of 50 different studies concerned with the use of literature as a function of its age. These data are plotted as cumulative distributions in Fig. 1. Some of the studies were based on actual use records of libraries (i.e., circulation records), but most of them were based on the ages of articles that were cited as references in the articles of leading technical journals. The data have some measurement error due to many factors, but can serve as a reasonable approximation.

The supporting data came from studies of a wide variety of subject fields, including:

Botany [1*]
Ceramics [2]
Chemistry and Chemical Engineering [1,3–9]
Electrical Engineering [10–12]
Entomology [1]
Geology [1,13]
Mathematics [1,14]
Mechanical Engineering [15]
Medicine [16–21]
Metallurgical Engineering [22]
Petroleum Research and Technology [23,24]
Physics [1,3,12,25]
Physiology [1]

*References are listed at the end of the paper.

Zoology [1]
Other general technical fields [26–32]

The collected data also covered a wide span of dates. That is, some studies reflect the use patterns of 1962, whereas some studies reflect the use patterns of 1899.

Second Specific Example of General Statement No. 1 (age of book material used in science and technology): "For the majority of users in the medical field, 90 percent of the needs for books can be fulfilled by books that are less than 20 years old."

Few data were collected[21,33] to support this second specific example of the first statement. The data that were analyzed are plotted as cumulative distributions in Fig. 2.

General Statement No. 2 (number of sources of materials): "For the majority of users in most fields of research, a specified fraction of their total needs for literature can be fulfilled by literature from a specified number of sources."

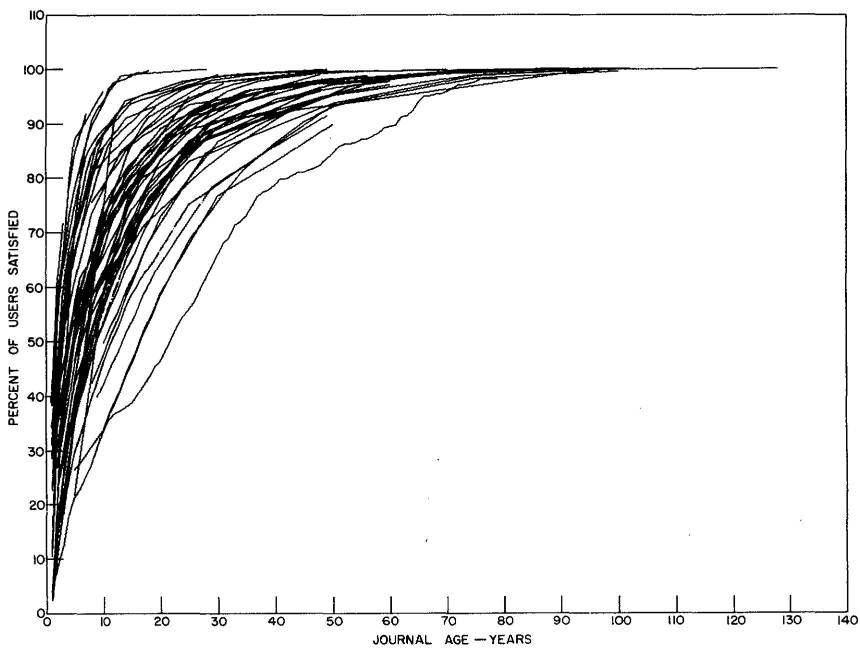


Figure 1. Distribution of journal use by age—science and technology in general.

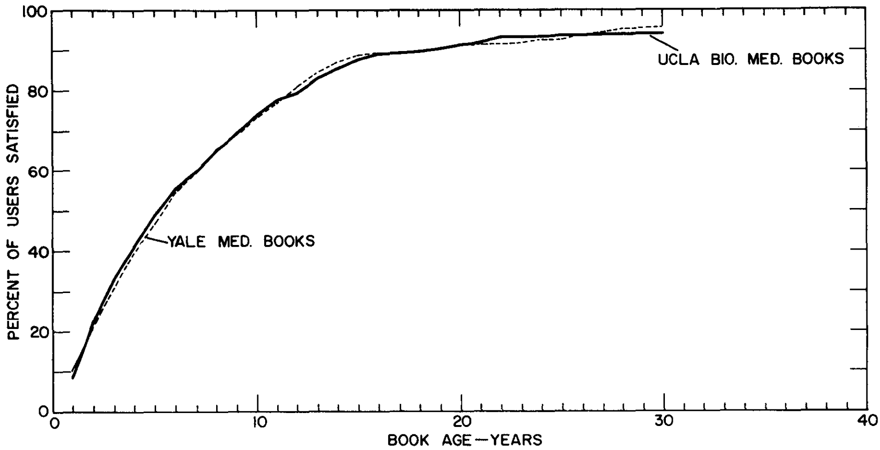


Figure 2. Distribution of book use by age—medical field.

First Specific Example of General Statement No. 2 (number of journals required in science and technology): "For the majority of users in most fields of science and technology, 90 percent of needs for journal articles can be fulfilled by 100 to 1,000 chosen journals. The exact number depends upon the nature and scope of the subject field."

The data to support the above general statement and its first specific examples were assembled from the results of 27 different studies that were concerned with the number of journals required to satisfy particular user populations (both authors and library patrons). The data are plotted as cumulative distributions in Fig. 3, and represent the following subject fields:

Biochemistry [34]
Chemistry and Chemical Engineering [4,5,6,35]
Dentistry [36]
Electrical Engineering [11,37,38]
Geology [13]
Mathematics [14]
Mechanical Engineering [15]
Medicine [16–20, 38–40]
Metallurgical Engineering [23,41]
Petroleum Technology [23]
Physics [35,42,43]
Physiology [44,45]
Other general technical fields [46–48]

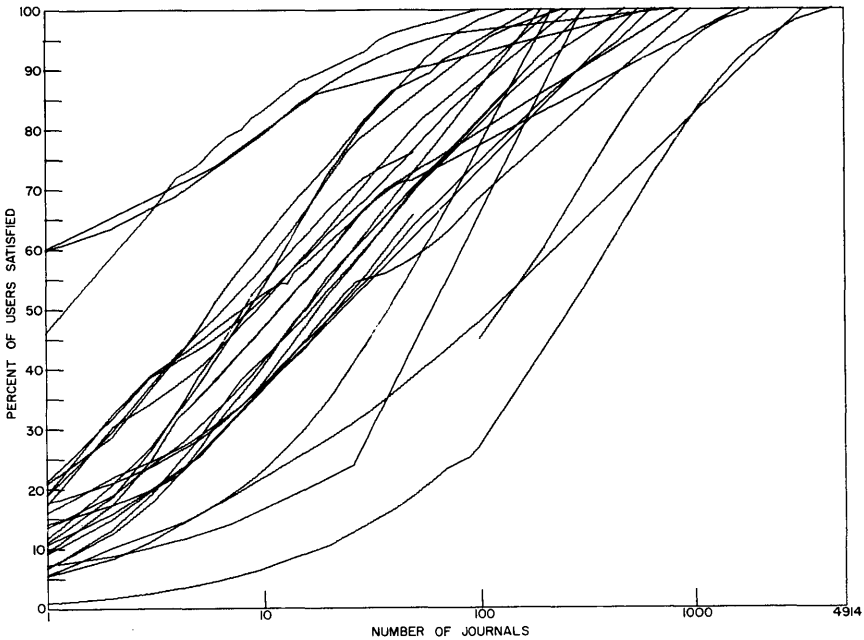


Figure 3. Distribution of number of journals required—science and technology in general.

General Statement No. 3 (speed of retrospective searches): "For the majority of users in most fields of research, a specified fraction of their total needs for extensive retrospective searches can be satisfied by a system that provides the search results not later than some specified time interval after the request was made."

First Specific Example of General Statement No. 3 (search response time for electronics research engineers): "For the majority of engineers doing electronics research work, 90 percent of their needs for extensive retrospective searches can be satisfied by a system that provides a list of relevant references from 2 to 15 days after the request was made."

The supporting data for General Statement No. 3 and its specific example are shown in Fig. 4.[49]

SOME ADDITIONAL COMMENTS ON THE MEASURED DATA

A close look at some of the data that were used to construct Fig. 1 (use of journal literature of various ages) disclosed patterns that seem to contradict some of the earlier studies on this subject. The contradictions center on two main points, and are discussed in more detail below. This is

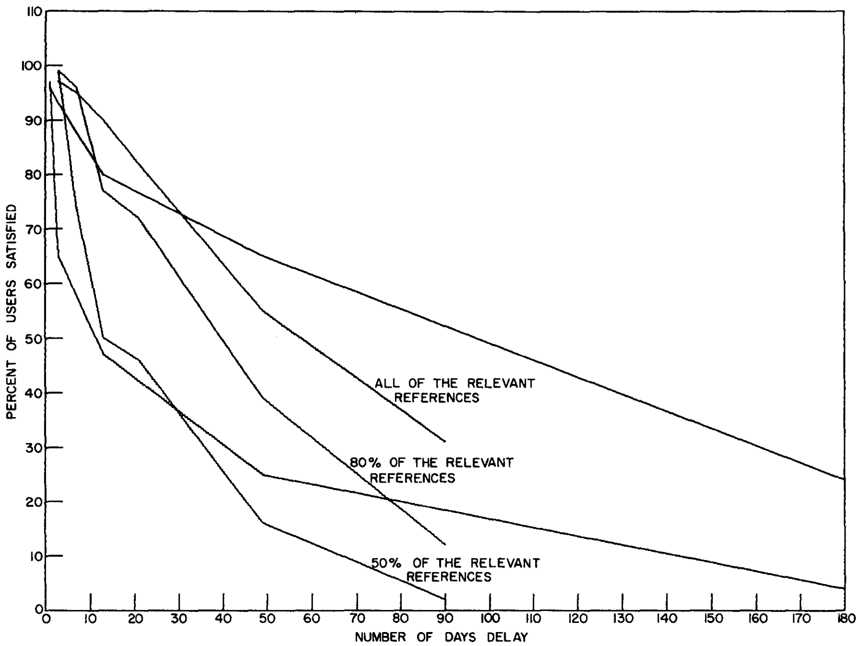


Figure 4. Required reference retrieval speeds—electrical engineering field.

admittedly a digression from the main theme of this paper, but it is related to the methodology for determining the quantitative statements, and is included here for completeness.

CITATION COUNTING VERSUS TRAFFIC COUNTING

Several authors (including most of those who have performed citation counts themselves) have suggested that as a method, citation counting was less accurate than measuring the recorded usage or circulation patterns. The inaccuracy has been attributed to many things, such as the difference between time lags that occur between publication and citation and time lags that occur between publication and library circulation. For example, one seldom finds citation counts that include references that are one month old, whereas one often finds circulation records that include one-month-old items. Some systematic error is also due to the rounding off of date of publication and citation, using figures for the years but not for the months. Additional error is due to using the nominal date of publication, rather than the date that the author wrote the manuscript and used the references. Also, there is some error because citation counts are influenced by the fact that there were fewer articles published in earlier years. It is also argued that the user population represented by the citation count method (i.e., the authors in the source journals) are different from the users represented by the library traffic or circulation count. All these points suggest that we might expect some systematic difference or bias between the results of the two approaches. However, the data collected here seem to support the view that there is no obvious difference in the results obtained by the two techniques. The curves that represent the traffic study approach are rather uniformly distributed throughout the entire range of curves shown in Fig. 1. Figures 5 and 6 show data for request patterns and citation counts, respectively, for a mixture of subject fields, and represent specific subsets of data taken from Fig. 1.

THE FUZZY HALF-LIFE

Several authors have suggested that perhaps there is something that might be called a "half-life" constant for technical literature, and that such a constant can be determined and shown to exist as a descriptive measure of a particular subject field (e.g., "...chemical literature has a half-life of 7.2 years"). The half-life is often interpreted as the time during which one-half of the currently active literature was published.[50] How-

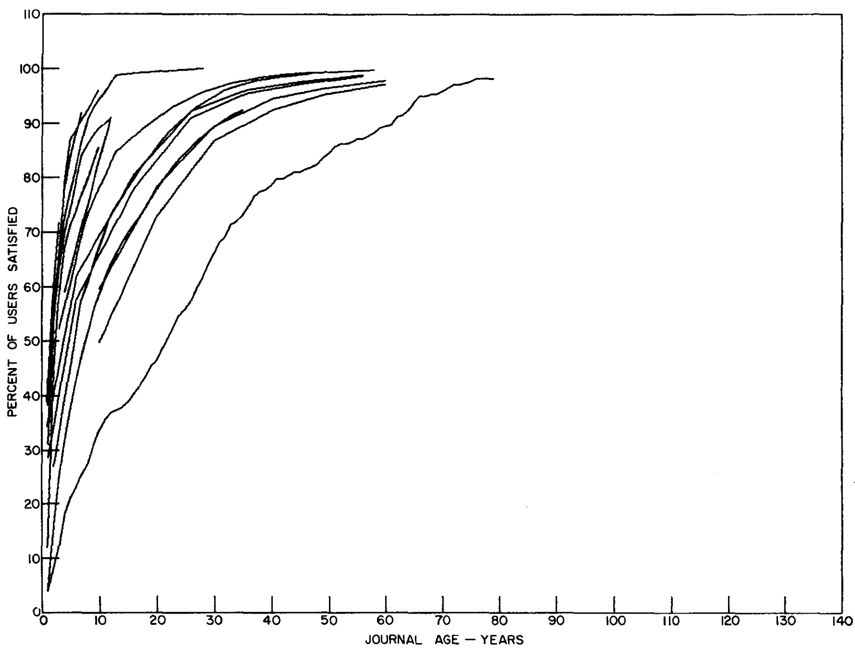


Figure 5. Distribution of journal use by age—as measured by actual library requests.

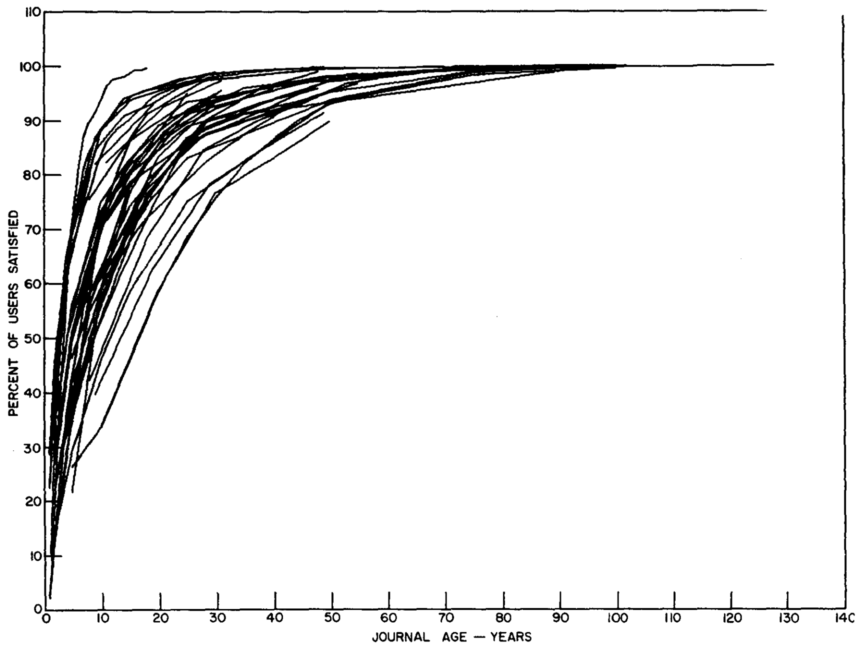


Figure 6. Distribution of journal use by age—as measured by citation counts.

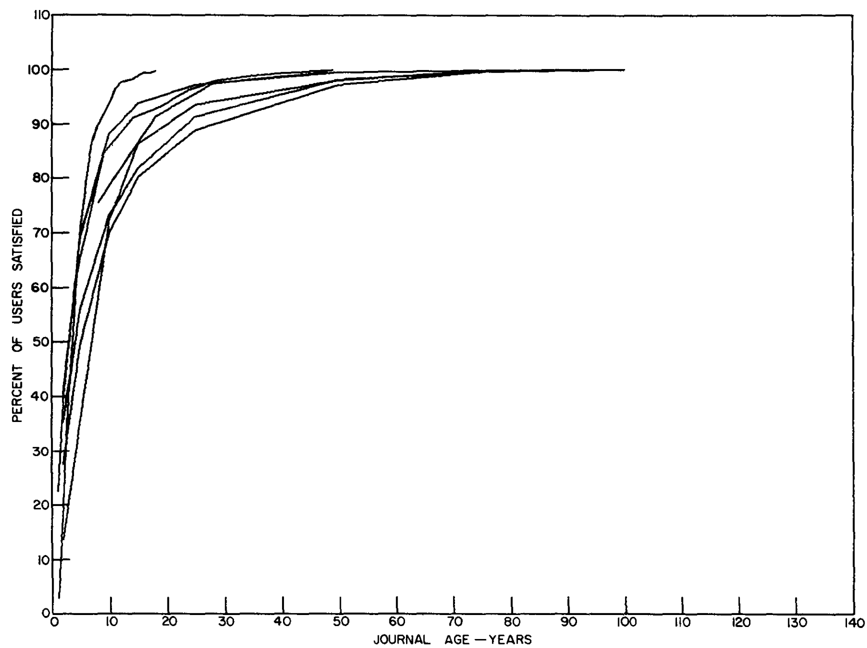


Figure 7. Distribution of journal use by age—physics field.

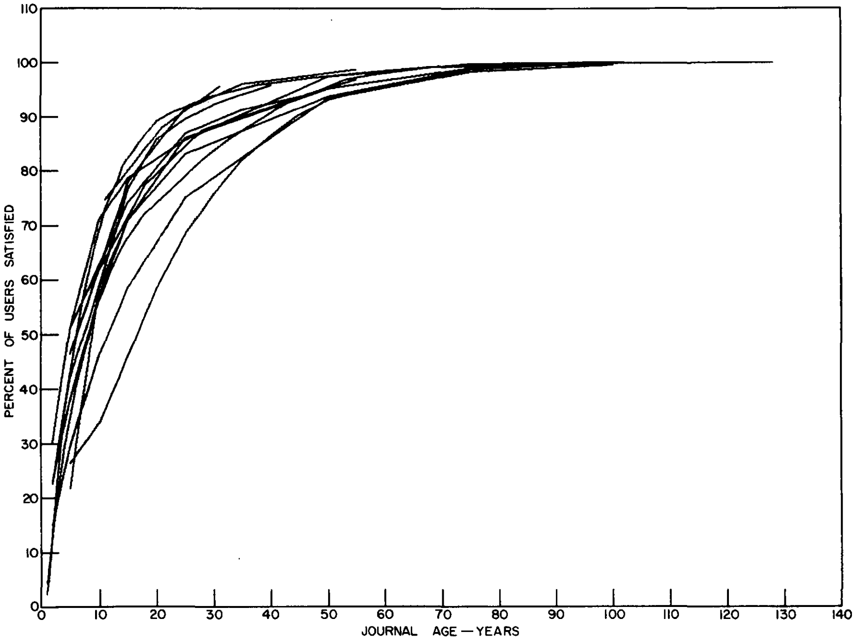


Figure 8. Distribution of journal use by age—chemistry field.

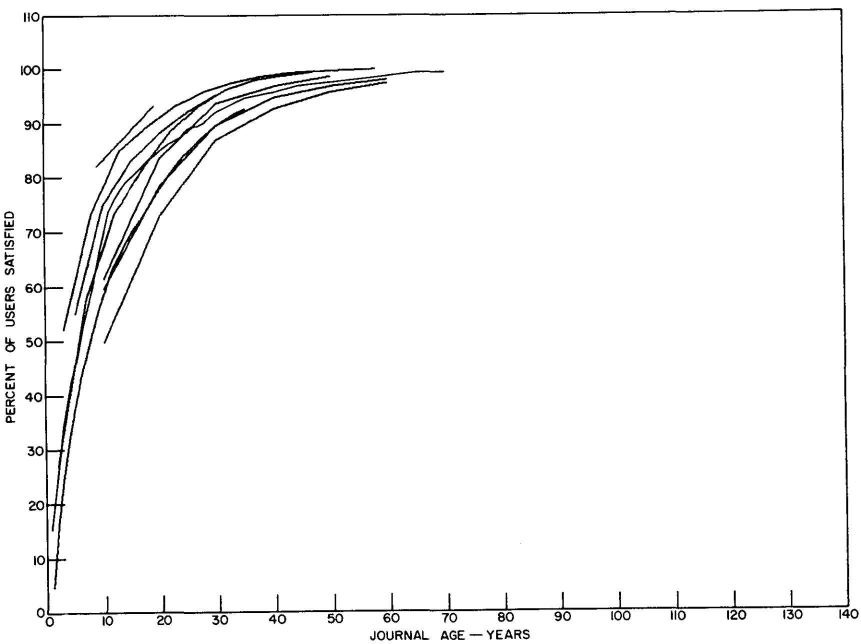


Figure 9. Distribution of journal use by age—medical field.

ever, most of the reported half-life studies were apparently made with only one sample or one specific user population, so that there was no indication of the great variance that might be possible with different samples or different test conditions, or different interpretations of the scope of the subject field.

Figure 7 shows what might be considered seven different half-life studies made in the field of physics.[1,3,12,26] Figure 8 shows twelve different half-life studies for the field of chemistry.[1,3-9] Figure 9 shows nine different half-life studies for the field of medicine.[16-21] The striking thing about all of these illustrations is the great variance possible in the value that could be quoted as the "half-life" constant for that field. The curves represent a smear of possible values for a specific field, so that the half-life figures now

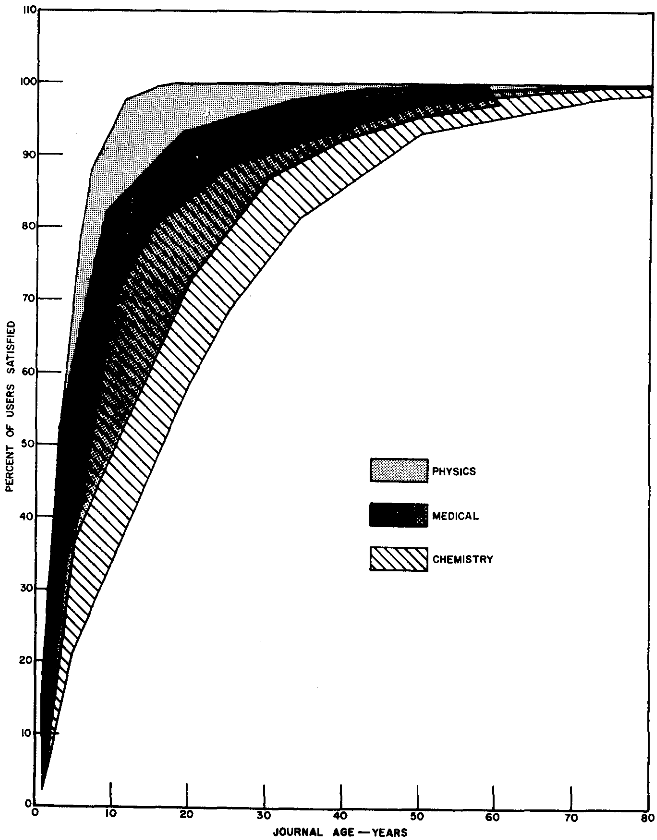


Figure 10. Distribution of journal use by age—composite patterns for physics, chemistry, and medicine.

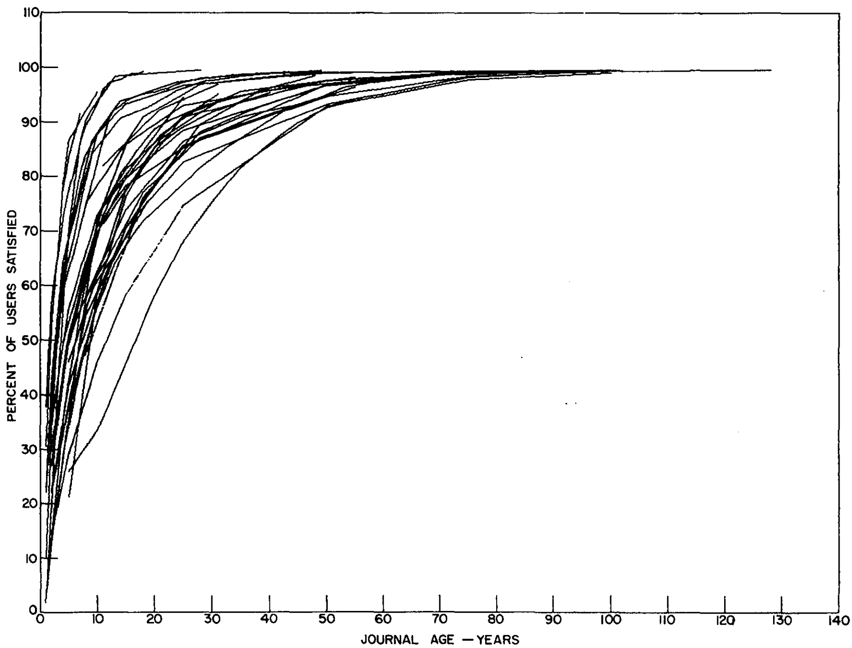


Figure 11. Distribution of journal use by age—physical sciences and mathematics.

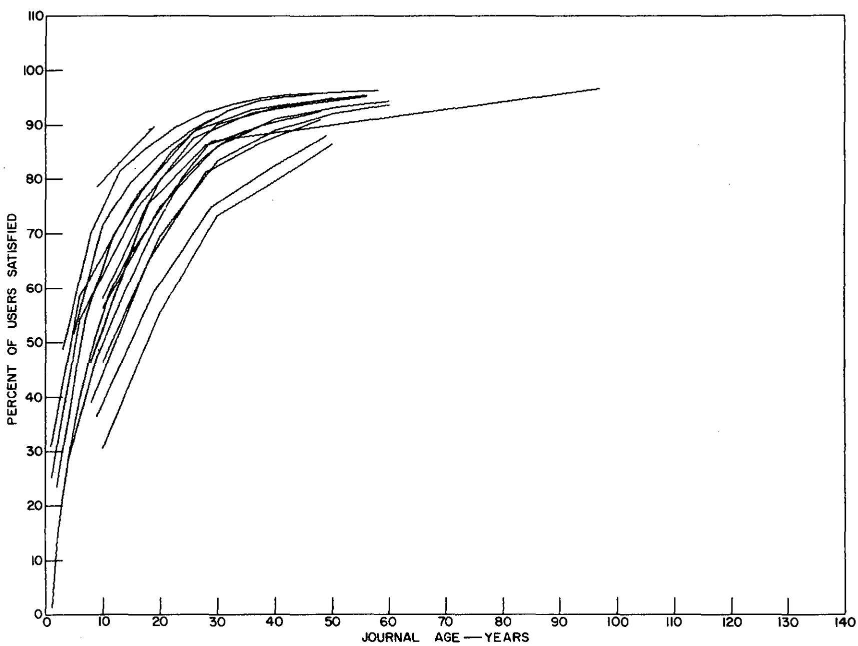


Figure 12. Distribution of journal use by age—natural sciences.

take on a probabilistic rather than a deterministic manner, and we now talk of half-lives in terms of "variance" and "best estimates" and "confidence figures." Variance in these examples did not seem to be related to the size of the sample or the particular year that was studied.

The smears for the subject fields (see Fig. 10 for the superimposed curves for chemistry, physics, and medicine) are so great that they almost completely overlap each other when superimposed on the same curve. Because of this, it is difficult to think in terms of readily identifiable differences in half-lives for various subject fields. There certainly are differences, but they are not dramatic differences. Even the contrast suggested by some people between the half-lives of literature in the physical sciences (Fig. 11) and those of literature in the natural sciences (Fig. 12) loses its

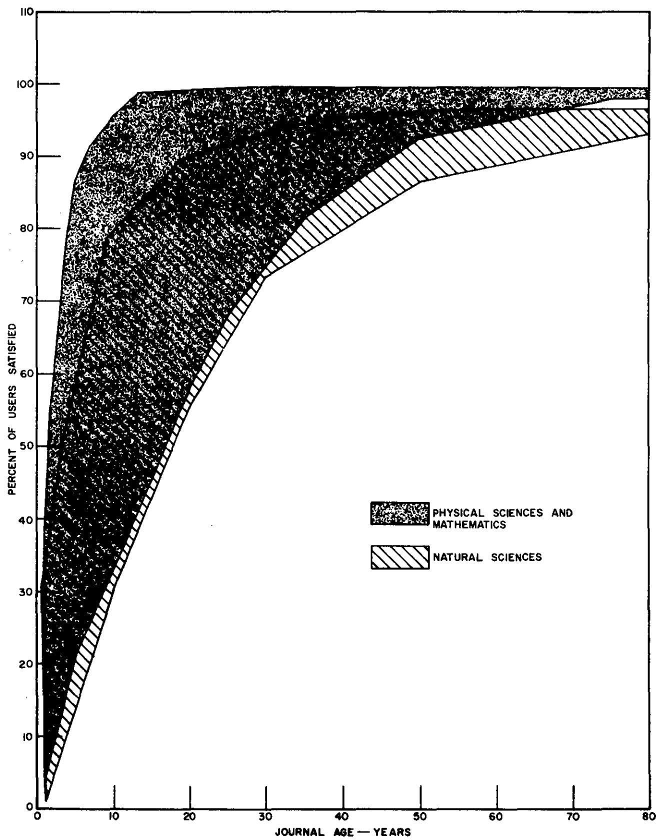


Figure 13. Distribution of journal use by age—composite patterns for physical and natural sciences.

impact when viewed in terms of their variance or smear (see Fig. 13). The net result of these observations seems to be that we have what might be considered very "fuzzy" half-lives, rather than easily discriminated constants.

SUMMARY

It appears to be both possible and reasonable to make some statements of user requirements in terms of what is required to satisfy a specified portion of the user population. Several general and specific examples were given to support this stand and others could easily be suggested. There is the possibility that requirements, when stated in this manner, might not be significantly different among different user populations except for the specific numerical value associated with them for each user population. This relatively simple mechanism for stating requirements provides a useful tool for the system designer and the evaluator of library systems and service.

REFERENCES

1. Brown, Charles Harvey, Scientific Serials: *Characteristics and Lists of Most Cited Publications in Mathematics, Physics, Chemistry, Geology, Physiology, Botany, Zoology, and Entomology*, ACRL Monograph No. 16 (Assoc. of College and Reference Libraries, Chicago, Ill., 1956).
2. Westbrook, J. H., "Identifying Significant Research," *Science*, vol. 132, no. 3435 (Oct. 28, 1963), pp. 1229–1234.
3. Fussler, Herman H., "Characteristics of the Research Literature Used by Chemists and Physicists in the United States," *Library Quarterly*, vol. 19 (1949), pp. 19–35.
4. Gross, P. L. K., and E. M. Gross, "College Libraries and Chemical Education," *Science*, vol. 66, no. 1713 (Oct. 28, 1927), pp. 385–389.
5. Burton, Robert E., "Citations in American Engineering Journals, Part I: Chemical Engineering," *Amer. Doc.*, vol. 10, no. 1 (January 1959), pp. 70–73.
6. Smith, Maurice H., "The Selection of Chemical Engineering Periodicals in College Libraries," *College & Research Libraries*, vol. 5 (June 1944), pp. 217–227.
7. Barrett, Richard L., and Mildred A. Barrett, "Journals Most Cited by Chemists and Chemical Engineers," *J. Chem. Educ.*, vol. 34, no. 1 (January 1957), pp. 35–38.
8. Patterson, Austin M., "Literature References in *Industrial & Engineering Chemistry* for 1939," *J. Chem. Educ.*, vol. 22 (October 1945), pp. 514–515.
9. Patterson, Austin M., "Journal Citations in the 'Recueil,' 1937–1939," *Recueil des travaux chimiques des Pays-Bas et de la Belgique*, vol. 59 (1940), pp. 538–544.

10. Coile, R. C., "Periodical Literature for Electrical Engineers," *J. Doc.*, vol. 8, no. 4 (December 1952), pp. 209–226.
11. Roa, Gundu, "Scatter and Seepage of Documents on Radio Engineering," in *Documentation Periodicals: Coverage, Arrangement, Scatter, Seepage, Compilation*, S. R. Ranganathan and A. Neelameghan (eds.), pp. 167–180, Report, Documentation Research & Training Centre, 112 Cross Road 11, Malleswaram, Bangalore, India.
12. Hooker, Ruth H., "A Study of Scientific Periodicals," *Review of Scientific Instruments*, vol. 6 (November 1935), pp. 333–338.
13. Gross, P. L. K., and A. O. Woodford, "Serial Literature Used by American Geologists," *Science*, vol. 73 (June 19, 1931), pp. 660–664.
14. Allen, Edward S., "Periodicals for Mathematicians," *Science*, vol. 70, no. 1825 (Dec. 20, 1929), pp. 592–594.
15. Burton, Robert E., "Citations in American Engineering Journals, Part II: Mechanical Engineering," *Amer. Doc.*, vol. 10, no. 2 (April 1959), pp. 135–137.
16. Kurth, W. H., "Survey of the Interlibrary Loan Operation of the National Library of Medicine," Report, U.S. Dept. of Health, Education, and Welfare, Public Health Service (April 1962).
17. Jenkins, R. L., "Periodicals for Medical Libraries," *J. Amer. Med. Assoc.*, vol. 97 (Aug. 29, 1931), pp. 608–610.
18. Hunt, Judith Wallen, "Periodicals for the Small Bio-Medical and Clinical Library," *Library Quarterly*, vol. 7 (1937), pp. 121–140.
19. Sherwood, K. K., "Relative Value of Medical Magazines," *Northwest Medicine*, vol. 31, no. 6 (June 1932), pp. 273–276.
20. Barnard, Cyril C., "The Selection of Periodicals for Medical and Scientific Libraries," *The Library Association Record*, vol. 40 (November 1938), pp. 549–557.
21. "Activity Statistics for a Large Bio-Medical Library," Part II of the "Final Report on the Organization of Large Files," Advanced Information Systems Div., Hughes Dynamics, Inc., Los Angeles, Calif. (April 30, 1964).
22. Burton, Robert E., "Citations in American Engineering Journals, Part III: Metallurgical Engineering," *Amer. Doc.*, vol. 10, no. 3 (July 1959), pp. 209–213.
23. Mote, L. J. B., and N. L. Angel, "Survey of Technical Inquiry Records at Thornton Research Center, 'Shell' Research Limited," *J. Doc.*, vol. 18, no. 1 (March 1962), pp. 6–19.
24. Cole, P. F., "The Analysis of Reference Question Records as a Guide to the Information Requirements of Scientists," *J. Doc.*, vol. 14, no. 4 (December 1958), pp. 197–207.
25. Burton, R. E., and B. A. Green, Jr., "Technical Reports in Physics Literature," *Physics Today*, vol. 14, no. 10 (October 1961), pp. 35–37. See also a letter to the editor about this report by A. O. Cezairliyan and P. E. Liley in *Physics Today*, vol. 15, no. 4 (April 1962), p. 58.
26. Urquhart, D. J., and R. M. Bunn, "A National Loan Policy for Scientific Serials," *J. Doc.*, vol. 15, no. 1 (March 1959), pp. 21–37.

27. Urquhart, D. J., "Use of Scientific Periodicals," *Proc. International Conf. on Scientific Information*, vol. 1 (National Academy of Sciences, Washington, D.C., 1959), pp. 287–300.
28. Cole, P. F., "Journal Usage Versus Age of Journal," *J. Doc.*, vol. 19, no. 1 (March 1963), pp. 1–11.
29. Randall, Gordon E., "Literature Obsolescence at a British and an American Aeronautical Library," *Special Libraries*, vol. 50, no. 9 (November 1959), pp. 447–450.
30. Graziano, Eugene E., "Interlibrary Loan Analysis: Diagnostic for Scientific Serials Backfile Acquisitions," *Special Libraries*, vol. 53, no. 5 (May-June 1962), pp. 251–257.
31. *Science Citation Index*, pp. IX, Institute for Scientific Information, Philadelphia, Pa. (1961).
32. Davis, Earl H., "Use of Periodicals at Long Beach Public Library," *Wilson Library Bull.*, vol. 11 (February 1937), pp. 397–398.
33. Kilgour, F. G., "Recorded Use of Books in the Yale Medical Library," *Amer. Doc.*, vol. 12, no. 4 (October 1961), pp. 266–269.
34. Henkle, Herman H., "The Periodical Literature of Biochemistry," *Med. Library Assoc. Bull.*, vol. 27 (1938), pp. 139–147.
35. Martin, M. W., Jr., "The Use of Random Alarm Devices in Studying Scientists' Reading Behavior," *IRE Trans. on Engineering Management*, vol. EM-9, no. 2 (June 1962), pp. 66–71.
36. Hackh, Ingo, "The Periodicals Useful in the Dental Library," *Med. Lib. Assn. Bull.*, vol. 25 (1936), pp. 109–112.
37. McNeel, J. K., and C. D. Crosno, "Periodicals for Electrical Engineers," *Science*, vol. 72, no. 1856 (July 25, 1930), pp. 81–84.
38. Neelameghan, A., and M. V. Ranga Rau, "Seepage of Documents in Medical Electronics," in *Documentation Periodicals: Coverage, Arrangement, Scatter, Seepage, Compilation,* S. R. Ranganathan and A. Neelameghan (eds.), Report, Documentation Research & Training Centre, 112 Cross Road 11, Malleswaram, Bangalore, India (1963).
39. Morse, E. H., "Supply and Demand in Medical Literature," *Albert Einstein Medical Center J.*, vol. 8 (October 1960), pp. 284–287.
40. Kilgour, Frederick G., "Use of Medical and Biological Journals in the Yale Medical Library," *Med. Lib. Assn. Bull.*, vol. 50, no. 3 (July 1962), pp. 429–449.
41. Bloomfield, Masse, "A Survey of the Information Habits of Atomic Energy Material Scientists," *Sci-Tech News*, vol. 16, no. 4 (Winter 1962), pp. 150–151.
42. Kessler, M. M., "Technical Information Flow Patterns," *Proc. 1961 Western Joint Computer Conference*, vol. 19 (Institute of Electrical & Electronic Engineers, 1961), pp. 247–257.
43. Kessler, M. M., and F. E. Heart, "Analysis of Bibliographic Sources in the *Physical Review* (vol. 77, 1950 to vol. 12, 1958)," Report R-3, Massachusetts Institute of Technology, Cambridge, Mass. (July 13, 1962), AD-282 697.
44. Mengert, William F., "Periodicals on Endocrinology of Sex," *Endocrinology*, vol. 18 (1934), pp. 421–422.

45. Morgan, Melvin B., "Characteristics of the Periodical Literature of Physiology Used in the United States and Canada," *American J. of Physiology*, vol. 191 (1957), pp. 416–421.
46. Scott, C., "The Use of Technical Literature by Industrial Technologists," *IRE Trans. on Engineering Management*, vol. EM-9, no. 2 (June 1962), pp. 76–86. This paper was previously published in the *Proc. Int'l Conf. on Scientific Information*, pp. 235–246 (National Acad. of Sciences, National Research Council, Washington, D.C., 1959).
47. Hoyt, J. W., "Periodical Readership of Scientists and Engineers in Research and Development Laboratories," *IRE Trans. on Engineering Management*, vol. EM-9, no. 2 (June 1962), pp. 71–75.
48. Bonn, George S., "Science-Technology Periodicals," *Library Journal*, vol. 88, no. 5 (Mar. 1, 1963), pp. 954–958.
49. Bourne, C. P., et al., "Requirements, Criteria, and Measures of Performance of Information Storage and Retrieval Systems," Stanford Research Institute, Menlo Park, Calif. (December 1961), AD-270 942.
50. Burton, R. E., and R. W. Kebler, "The 'Half-life' of Some Scientific and Technical Literatures," *Amer. Doc.*, vol. 11, no. 1 (January 1960), pp. 18–22.

11

Health Sciences (MEDLARS)

MARTIN M. CUMMINGS, M.D.
Director, National Library of Medicine
Bethesda, Maryland

In my presentation today I intend to offer a critique of our operating MEDLARS system, sharing with you my view of this unique reference retrieval system.

I'm sure most of you know that MEDLARS is an acronym for Medical Literature Analysis and Retrieval System. The system has been in operation since January of this year although the input to the system began a year earlier. I would like to review briefly the history of MEDLARS' development and recall the long range objectives of this program which were directed toward improvement of the management of the biomedical literature.

The immediate objectives of MEDLARS are: first, the rapid dissemination of lists of current publications in the medical field, including the monthly publication, *Index Medicus*, and other regular recurring bibliographies in more specialized areas such as cancer and heart disease. Second, the bibliographic control of the medical periodical literature available for rapid retrieval in response to subject-oriented queries of our computer files. We call such searches demand-bibliographies. Third, the wide availability of the MEDLARS data base to other libraries and research institutions which may duplicate the retrieval capacity of this system and make more specialized use of the contents of the file within their own research programs.

MEDLARS was developed under contract with the General Electric Company's Information Systems Operations in three phases. Phase 1, the preliminary study and design, lasted from July 1961 to January 1962. This phase included development of a basic set of specifications for equipment, programs, and personnel required to implement MEDLARS. Phase 2, a detail design, began in January 1962 and included equipment procurement, computer programming, and detailed procedure development. Phase 3, systems testing and implementation, overlapped Phase 2 and included equipment installation, file conversion, detailed testing of the data-processing portions of the system and a period of preliminary operation. Phase 3 ended in August of this year

The following items of automatic data-processing equipment are now operating in the MEDLARS system. Thirteen paper-tape typewriters, Friden Flexowriters, for preparation of the computer input; a Honeywell-800 computer for editing, sorting, compressing, merging, storing, and formatting data for subsequent printing; and a special computer-activated optical printer called "GRACE," which is an acronym for Graphic Arts Composing Equipment, used to convert the computer output into high-quality photocopy for publication purposes.

I want to point out that Mr. Montgomery yesterday referred to the acquisition of a Photon printer by the University of Pittsburgh [Chap. 2]. This is not the same equipment that I am referring to. I had an opportunity to speak to him today. The Photon equipment at the University of Pittsburgh is a punched paper-tape-driven instrument with a speed of eight characters per second, whereas the instrument (GRACE) which I refer to is a computer-driven phototypesetter with a speed of 300 characters per second. I thought you might be interested in seeing the layout of the MEDLARS hardware. Figure 1 shows a portion of the computer facility. Figure 2 shows how the new GRACE equipment is linked to the Honeywell computer. An operator stands at the GRACE console. The component at the left contains the photocomposing flash tube matrix.

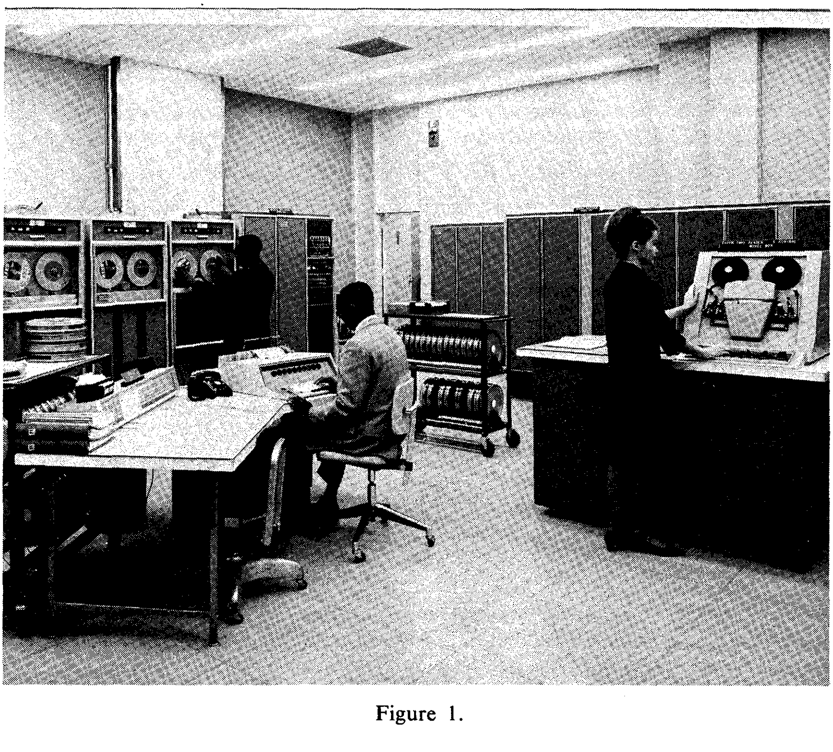


Figure 1.

Figure 2.

The MEDLARS system has been logically subdivided into three component parts: an input subsystem, a retrieval subsystem, and a publication subsystem. The input subsystem joins the scientific and linguistic talent of 20 trained literature analysts to the tremendous processing capabilities of the computer. Medical periodicals and journals, after check-in of the serial record, are forwarded to the index unit where the professional indexers classify the subject content of each article in the journals by assigning subject headings from the Library's controlled Medical Subject Headings List of 6,400 terms called "MeSH." Each article is printed under an average of three subject headings in the monthly *Index Medicus.* Additional headings (up to 32) may be assigned for storage on magnetic tape for use in the retrieval subsystem. The indexers also translate titles of foreign literature papers and transliterate those in non-Latin alphabets. Journals with indexer data sheets are next processed by the Flexowriter operators who prepare a paper-tape record for computer input. This basic unit record includes the article's title, author names, journal reference, and the subject headings assigned by the indexer. After verification of the Flexowriter hard copy, corrected tapes are batched and spliced for entry into the computer. The computer input programs are run once a day. At the present time, more than 700 articles per day are entered into the system. These programs edit the input extensively, reject improperly prepared unit records, and build the two major data files: the compressed citation file, which is used in the retrieval subsystem, and the processed citation file used in the publication subsystem.

Currently 150,000 articles from 2,400 medical journal titles are processed annually and added to the computer file. This input is expected to grow to 250,000 articles from 3,000 serial journals by 1969. More than half of the articles appear in foreign journals, requiring a massive translation effort.

The retrieval subsystem is initiated when a medical researcher, teacher or practitioner requests a demand bibliography. Such requests are forwarded to a staff of search specialists who have had extensive training both in indexing and the logic of machine retrieval. The searcher formulates the request in a logical statement, intelligible to the computer system. The search parameters include the subject heading, journal titles, specific languages, author names, year of publication, and computer entry date. Formulated search requests are punched into paper tape, proofread, and batched for computer processing. This system has the capability of performing 90 to 100 demand searches per day. The demand search computer programs have been designed to match a group of search questions against every record in the compressed citation file. The demand bibliographies which result from this search are printed in any one of a variety of output formats by means of report generator programs. Demand bibliographies are normally printed on the computer's high speed printer.

I would like to show you several examples of the type of computer printout prepared in response to a demand search inquiry. One format which we use is shown in Fig. 3. It is a 3 × 5 card which gives the author, the citation, and indicates that the article appeared in Japanese medical literature. On the right you can see it also acknowledges the fact that it has been translated from Japanese. Listed below are the major descriptors which should portray the content or the concepts contained within the particle article.

Here is another format (Fig. 4) where the printout is arranged in a slightly different referenced order with the journal, volume, page, and year appearing before the author and title. Again, within the parameters of the search request, there appear the major subject concepts contained in the article. This search was in response to a request from the Food and Drug Administration, asking for a certain type of drug toxicity.

I should tell you that I was extremely careful in selecting these examples since the variability and depth of indexing may range from a minimum

16092

TAKAHASHI K

(STUDIES ON OPEN HEART SURGERY. I.
CLINICAL AND EXPERIMENTAL STUDIES OF RIGHT
CARDIAC BYPASS)

(JAP)

SAPPORO MED J 23;217-37, MAR 63

BLOOD PRESSURE DETERMINATION; DOGS;
EXPERIMENTAL LAB STUDY (A); *HEART,
MECHANICAL; *HEART SEPTAL DEFECTS,
VENTRICULAR; *HEART SURGERY; *TETRALOGY
OF FALLOT; THORACIC RADIOGRAPHY

Figure 3.

of three to a maximum of 32 subject headings. I think these are quite fair and representative examples.

Each working day, punched cards are entered into the computer, telling which recurring bibliographies or which citations for *Index Medicus* are to be compiled on that particular day. The computer selects the appropriate

CYCLOPHOSPHAMIDE TOXICOLOGY.
CITATIONS ARE GROUPED AS FOLLOWS
CLINICAL STUDIES FOLLOWED BY EXPERIMENTAL STUDIES.

MER J SURG 106;777-82, NOV 63	FEIND CR, HERTER F, MARKOWITZ A IMPROVEMENTS IN ISOLATION HEAD PERFUSION. CARCINOMA, EPIDERMOID; CHILD; CHORDOMA; *CYCLOPHOSPHAMIDE; *HYPOTHERMIA, INDUCED; *ISOLATION PERFUSION; *JUGULAR VEIN; MELANOMA; NCI (2); NEOPLASM THERAPY; PROGNOSIS; RETINOBLASTOMA; RHABDOMYOSARCOMA; SARCOMA, OSTEOGENIC; TOXICOLOGIC REPORT (4); *TRIETHYLENE MELAMINE
NN OTOLARYNG (PARIS) 80:5-12, JAN-FEB 63	LEROUX-ROBERT J, FORETTE B (ACTION OF CYCLOPHOSPHAMIDE IN THE TREATMENT OF O.R.L. AND CERVICO-MAXILLO-FACIAL EPITHELIOMAS)(FR) *CARCINOMA, EPIDERMOID; *CYCLOPHOSPHAMIDE; *EAR; *FACIAL NEOPLASMS; *LARYNGEAL NEOPLASMS; *MAXILLARY NEOPLASMS; *NECK NEOPLASMS; NEOPLASMS; *NOSE NEOPLASMS; TOXICOLOGIC REPORT (4)

Figure 4.

citations from the processed citation file, performs a complicated task of page composition, and prepares a magnetic tape file of print records for the phototypesetter, GRACE. Four issues of *Index Medicus* have been produced by the GRACE printer, and the revised medical subject heading list will also be produced by the GRACE printout. A little later I will give an example of the quality of a GRACE printout.

GRACE is a revolutionary computer-driven typesetter printing from a font of 226 characters, upper and lower case, onto positive photographic film or paper, and operating at a speed of approximately 300 characters per second. It represents the only system currently capable of delivering high-quality typography directly from a computer at computer speeds. GRACE converts digital information from magnetic tape to characters on photographic film. The exposed film is developed by an automatic film processor, inspected, cut into page-sized sheets, and packaged for delivery to a printer. The resulting film masters are used directly for platemaking, printing, and binding of the final publication.

The output printing load is expected to increase from 290 million characters this year, to 590 million in 1969. The use of GRACE in the Library has reduced our composing time from 25 days to 16 hours for each issue of *Index Medicus*. Its photocopying power has been estimated by the Government Printing Office to be equivalent to that of 55 linotype operators. Figure 5 shows a sample of a page of *Index Medicus* which reveals the improved image quality and readability of the text compared to the ordinary monocharacter of a regular computer printout. It also shows how a page of *Index Medicus* is organized.

Since MEDLARS has been in operation for only eight months, it is impossible at this time to narrate a full history of the operational experience. However, some comments can be made on the basis of results to date. The basic data-processing system design appears to be adequate to accomplish the original MEDLARS objectives. All of the bibliographic publications have been tested and are now in production. The demand-search capability is now being thoroughly evaluated, particularly through consumer evaluation of our products. We await, as I am sure many others do, the development of more precise measurements of recall and relevance for evaluation of our system. We are using, internally, a modification of the Cleverdon technique of measuring recall and relevance and we are pleased with the results to date.

Several problems have been encountered during this first year of operation. They relate mainly to preparation of input data. First, the recruiting and training of scientific indexers is a recurring problem. A professional indexer must have an extensive background of knowledge in the life sciences and, in most cases, must also have an excellent foreign language

INDEX MEDICUS

Acute cardiac emergencies. Borland DM.
Hahnemannian 99:8-10, Mar 64
[Oxalosis. Clinical picture, morphological findings, pathogenetic problems] Gasser G, et al.
Deutsch Arch Klin Med 209:257-76, 9 Jan 64 (93 ref.) (Ger)

OXAZINES (D2)

Antiinflammatory potency of 2-(beta-chloroethyl)-2,3-dihydro-4-oxo-6-amino-(benz-1,3-oxazine) HCl(A350), phenylbutazone and acetylsalicylic acid in carrageen - induced edema. Arrigoni-Martelli E, et al.
Med Exp (Basel) 10:164-8, 1964

OXEDRINE (D4)

Leakage of transmitters in salivary glands. Assarson N, et al. **Brit J Pharmacol 22:119-25, Feb 64**

OXIDATION-REDUCTION (H)

Enzymic oxidation of ethanolamine by beef serum. Hayashi M, et al.
Chem Pharm Bull (Tokyo) 12:223-7, Feb 64
A cytochemical localization of reductive sites in a gram-negative bacterium. Tellurite reduction in Proteus vulgaris. Iterson W van, et al.
J Cell Biol 20:377-87, Mar 64
A cytochemical localization of reductive sites in a gram-positive bacterium. Tellurite reduction in

Figure 5.

capability, since 75 percent of the articles indexed for *Index Medicus* come from journals written in any of 30 or more foreign languages. Success in search and retrieval is directly proportional to adequacy and consistency in indexing. Although no complete test of the system's retrieval capability has yet been made, as I indicated earlier, we are highly encouraged by the results of measurements of relevance and recall.

I would agree with Dr. Brosin that medical subject headings constitute the major problems of any system such as ours. Glossaries and thesauri cannot be static if they are to reflect the advances of science. Our system, however, is designed to accept new terms, when they appear in the literature, as provisional subject headings. Often, we have as many as 2,000 provisional subject headings entered into the computer tapes over and above those which appear in the printed medical subject headings list. I am not in agreement with Dr. Brosin's critique of the relationship of software to hardware, with specific reference to the field of behavioral sciences. I submit that it is extraordinarily difficult for psychiatrists to communicate with computers when psychiatrists have difficulty communicating with psychiatrists. Quite earnestly, I view the major deficiencies in the medical subject headings list of the Library to fall in three areas: first, in the field of dentistry; second, in the field of behavioral sciences, as pointed out by Dr. Brosin; and third, in the field of drugs and chemicals. These deficiencies were recognized early and appeals were made to the professional societies representing these disciplines to assist the Library in updating the descriptors within these areas. We have had a vigorous response from the dental profession through the American Dental Association. They have provided two experts in the field who have been working with us. As a result of this effort, more than 200 new specific dental terms will be introduced into our Medical Subject Headings List.

In the field of drugs and chemicals, we have had a very warm response from the Food and Drug Administration and there have been discussions with *Chemical Abstracts* to attempt to introduce more specific, more comprehensive terms in this important area. However, so far, we have had no response from the National Institute of Mental Health, which was requested to provide advice and assistance in this area. We plan to seek assistance from the American Psychiatric Association.

Librarians alone cannot develop authoritative medical subject headings lists. This is a task to be shared with the biomedical community. For this reason, I have come to the point of view that either the World Health Organization, or the Medical Division of the National Research Council of the National Academy of Sciences should undertake to standardize medical nomenclature and classification, not only for the National Library of Medicine, but on behalf of all groups concerned with the management of biomedical literature.

Another weak point in the MEDLARS input subsystem has been the utilization of punched paper tape. Correction procedures using the paper tape are very cumbersome, and it has been difficult to keep the registration of the tape within the extremely small tolerance allowed by the paper-tape reader of the computer. Difficulty has also been encountered in recruiting and holding Flexowriter operators, who must type complex medical terminology on special equipment and yet are still classified as clerk-typists according to Civil Service standards. However, the Library is convinced that paper-tape is superior to punched-card processing for the MEDLARS program, and we look to remote control console direct entry or optical scanning as a better long-range solution to the problem of input.

Another serious problem connected with MEDLARS has been the shortage of trained search specialists. This has necessarily limited the number of searches which can be formulated. Hence, full machine capability has not yet been approached. In fact, we reached only about 25 percent of the machine's operating capability due to the limited size of our search staff. It is hoped that this problem can, in part, be alleviated through the decentralization of MEDLARS. A contract has been negotiated with UCLA for the reprogramming and reconversion of Honeywell tapes for use on IBM 7090 and related equipment. We plan to establish six or eight university-based regional MEDLARS centers so that the means of access to, and retrieval of, the literature will be shared freely and extensively with the entire biomedical community.

Despite the problems mentioned above, we believe MEDLARS is unique in several respects. First, it is the only system of this type operating in a research library in the medical field. It is also the only large-scale reference retrieval project based on a research library, thus providing both bibliographic control and access to the documents themselves. The problems of system engineering have been adequately solved, proving an operational reality, with an average of 700 new documents being processed and put into the files each day. The total store of articles indexed is now 240,000. I think you would agree that the other unique feature of MEDLARS is its revolutionary printing capacity. We consider MEDLARS as only a first step. It will be constantly studied and revised to keep pace with new technical developments.

The National Library is now actively involved in research and development directed toward the use of data-processing equipment for other library procedures such as acquisitions and cataloging. We hope to be perceptive, if not sensitive, to the consumer requirements. In this context, we have developed program plans to support specialized information centers through MEDLARS services. The use of the system for support of medical education, continuing education, and the practice of medicine awaits exploitation.

IV. OPERATIONAL EXPERIENCES

12

Conjectures on Information Handling in Large-Scale Systems

GEORGE W. N. SCHMIDT
North American Air Defense Command

Conjecture implies formation of an opinion or judgment upon insufficient evidence. After twelve years of experience in the military application of computer systems in the areas of command and control, simulation and intelligence, the best I can do is conjecture. Certain specific information handling problems have been solved. Others await solution and will require either or both software and hardware techniques development.

Basically those information handling problems which are considered to have reached a reasonably successful level of solution are exemplified by:

1. The financial problem, represented by the payroll processing by the various military finance offices. The data is well defined. Individual's serial number, grade, length of service, marital status, dependents, etc. The only field that can cause a storage or retrieval problem is the individual's name, as it is alphabetic and variable in length.
2. The personnel problem which is now partially automated at the records center in which the service records of Air Force personnel are now maintained, and personnel assignments processed.
3. The supply functions which are being mechanized at base level in order to speed up the resupply and inventory control functions.
4. The aircraft control and warning function as exemplified by SAGE (Semi-Automatic Ground Environment). This system processes the returns from surveillance radars to arrive at a position of the aircraft by latitude, longitude, altitude and time. This data is correlated by the computer program with the flight plan as filed with the FAA. The data which correlates with the FAA flight plan is reported as known friendly that which does not correlate is declared either hostile or unknown and identification procedures are initiated.
5. The Ballistic Missile Early Warning System in which radar returns are processed by both wired and stored program logic. The wired logic establishes the validity of the returning signals as coming from a real object in space and also converts the return to azimuth, elevation, range and range rate data form. The stored program logic is used to generate azimuth rate and elevation rate data and to perform

> the discrimination tests which eliminate nonthreatening objects from the reporting system. The data relating to those objects which are classified as threatening objects is formatted into 63 bit messages by the program and passed over communications to the Display Information Processor at Colorado Springs. The Display Information Processor program decodes the message and computes the alarm levels, time to go to soonest impact, and the parameters to be passed to the ICONORAMA equipment to drive the display of impact and launch ellipses.

Those information handling problems awaiting solution are those which require the processing of narrative text, photographic indexing and interpretation.

The problems which have yielded to solution are those that have a common characteristic, well-defined organization and structure that can be readily formatted. Those problems which are presenting the most difficulty also have a common characteristic, a complex organization and structure which is permeated with exceptions and is not amenable to formatting.

I feel there are these two basic classes of data available for exploitation, formatted and unformatted. Examples of the formatted data are BMEWS data which because of its origin, radar data, can be formatted at the source. It is no problem to handle the more than 6.3 million messages a year and present the data to the user in summary displays. Other sensors can collect data and furnish it in formatted form for processing. Several of these record their data in a typical magnetic tape format, i.e., 556 bits per inch density, 112.5 inches per second speed with a 10-second record length. Using 100 word per minute teletype lines to transfer this data, if error-free communications were possible, would require only 17 hours, 37 minutes, 30 seconds per record. More of this later.

Examples of unformatted data to be processed are incident reports, i.e., descriptive narratives of objects seen or nonstandard activities; scientific treatises; proceedings of symposia and other technical meetings; other information of this kind and photographs which must be indexed for retrieval and also interpreted.

In both the formatted and unformatted classes there appear to be two categories of information-processing requirements. One could be called "real-time," the other "deferred." To permit intelligent argument, in the Greek sense of argument, I should define my terms. "Real-time" information handling requires update of the data base, response to queries, and summarization of the data so that the user may react to the changing conditions and affect the environment from which the data is collected—i.e., the data is being processed concurrent with the operation. "Deferred" information handling requires update of the data base, response to queries, and summarization of the data ex post facto so that the user may perform detailed analytical studies to establish criterion measures, patterns and new techniques.

Capability to do "real-time" processing implies that there is available a history of data in depth relating to the problem. Based upon this file of data, the necessary criteria and patterns for quick-look analysis can be established and narrative statements relating to the "real-time" problem can be retrieved. This leads to the problem of the structure of the file.

Several techniques have been used experimentally. In almost all the approach has been to establish a dictionary of terms, their synonyms and some code to represent them. Documents are scanned by people who select the meaningful words and encode these words for inclusion in some formatted field, record, or file so that a search can be made of the formatted portion which will then constitute the retrieval control.

Because word-by-word encoding has proved to be not entirely satisfactory, this technique has been expanded to include phrases or as sometimes stated, "keywords in context." Again the process is one of human interpretation of what is significant in the document. As encoders change and as individuals' moods change, the index capability changes introducing inconsistencies which will degrade the retrieval capability.

The English language being what it is, things such as prefixes, suffixes, tenses, etc., present the indexer and the file definer problems of the type related to unformatted data. With the field length varying from one letter to more than 25 letters and irregular verbs requiring cross-referencing to their root, a voluminous dictionary of terms would be required.

Perhaps another approach to the problem could be investigated. Eliminate the human cataloguer or indexer from the system. Rather than look for the significant words or phrases, establish a machine search technique which would identify the "nonsignificant" words, i.e., the, and, but, that, etc. There are probably fewer of these in the English language than the other type of words; and, therefore, a much more limited dictionary could be used for an initial screening of a document to form the basis of both indexing, storage and retrieval. "Nonsignificant" words appear to constitute approximately 50 and up to 65 percent of most documents. The remaining words could then be catalogued by their location within the document and some formatted file of these words be generated as the retrieval control.

Any index of this type information will be large. One of the applications with which I am working will require the capacity to store between 200 and 300 narratives a day with an historical depth of not less than one year and preferably two years to improve both "deferred" and "real-time" analytical capability of our analysts. The indexing problem is tremendous and the structure of this index in order to permit ready access to the desired data without serial search of the entire file to locate the data is desired. Tape files with chronological addition of the data to the file generates a tremendous amount of tape spinning with the associated inefficient use of the central processor.

This has led to the consideration of disk files, tape files, and bulk core memory. During the investigation there has been much emotion and little fact upon which to base our decision. We have sifted through much of the emotion and as much fact as we could find. Our "guestimates," conjectures, if you please, indicate that there are some areas of data retrieval where tape will outperform disk for the retrieval of information for processing purposes. The controlling factors seem to be the record length and its relation to the track length for recording on the disk. Our initial feeling with the announcement of large-volume disks was one of elation. We now have tempered that elation and realize we need more data relative to the payoff crossover point definition between disk and tape. One of the applications in which we see the greatest payoff for disks is that of sorting formatted data for purging, merging and updating of the file.

The announcement of large-size core memories—in excess of 200,000 words—by several manufacturers is interesting and many applications in information handling can be seen. Large speedups are possible because bigger batches of data can be processed without repeated input-output interrupts. Large core memories should allow larger, more sophisticated, greater depth of cross-referencing in the index for retrieval.

In the application in which I am most interested, several individuals are required to have access to the data base. Under the standard techniques of executive and monitor control the first one in with the highest priority would be the first one to have his job processed, with the resultant queuing problem.

The area in which preliminary investigation shows the greatest payoff for large-scale information handling systems will accrue is multiprocessing capability both in hardware and software because several analysts may then concurrently be serviced. Several organizations are now operating such systems either experimentally or in a limited operational situation.

Some sort of hybrid configuration of the computer with multiprocessor capability and an associative memory device appears to be desirable—the associative memory to be the index or library catalogue which would be computer generated by a technique similar to that previously discussed. The request for data would be processed by the associative memory device which would furnish to the central processor the acquisition control data whereby the data could be extracted or the desired documents retrieved. The associative memory device would be a job set-up preprocessor and, effectively, a peripheral unit.

Earlier a data-collection system was mentioned which required a large amount of time for data transmission. Before any large information-handling system can be automated to the degree required to handle the "real-time" and "deferred" requirements, some way must be found to summarize the data at the collection point. One technique is to place a data processor at the collection source. This was done at BMEWS. Secondly, some form of error detection or correction system must be designed into the communications system and terminals. Until this is done, human intervention between the collection source and the input to the data file will be required with the resultant slowing of the system response time in satisfying the "real-time" requirement.

Most systems today require pro forma sheets from which the keypunch operator punches cards which in turn are verified on another keypunch.

We are looking toward elimination of the card punch requirement by substituting a keyboard with a monitor readout so that the catalogue keypunch operator can correct as he punches and get the data more directly to magnetic tape for insertion in the data base. Eventually, as programming techniques are developed, the cataloguing can be automated to a large extent. These same type consoles will be available to our analysts for the insertion of their queries.

The organization with which I work is out at the far end of the line—that is, we use the techniques and hardware you people design in an operational environment. We are not aware of all the techniques under study and do not always know where to go to get the information. Perhaps some organization such as the Knowledge Availability Systems Center might act as the central facility for information relative to information-handling techniques. This, in itself, would present an interesting information-handling problem in the area of unformatted data handling.

In this rambling presentation, however, are the basic elements upon which I framed the conjectures which follow:

1. Except for the volume of data involved, formatted files constitute no serious problem to any programming group.
2. Insufficient specific problems related to the handling of unformatted data—i.e., narrative text—have been solved in detail to permit the techniques to be expanded to the general case.
3. Where multiple sensors feed a central file, some summarizing or screening technique at the collection site is required to reduce the communications requirements and prevent cluttering of the central file.

4. Error-detection and correction codes in communications systems will be an absolute necessity before any automated indexing and file generation system will work.
5. Some system for the interchange of information on the status of techniques and hardware development in the information handling is required.

13

Large Systems

FRANK L. HASSLER
Defense Communications Agency

INTRODUCTION

As technology has provided ever more capable electronic computers, communication methods, and sensing elements, system designers have been working to implement information systems on a scale commensurate with the tools.

The purpose of this paper is to examine in general the experience obtained with large systems. In this examination, the word "system" means the composite of sensing elements, communications, and automatic data-processing (ADP) equipment, personnel, and procedures used to accomplish the broad functional mission of the complex. All the system examples used will contain all of these components, but emphasis will be placed upon the ADP aspects of the system.

In a discussion of experience with large scale systems, a distinction will be made between systems with known, repetitive functions, sensor based systems, and command systems. Each type is characterized by different degrees of complexity, cost, uncertainty, etc., and the differences create marked variations in performance.

SYSTEMS OF KNOWN REPETITIVE FUNCTIONS (CLASS I)

Systems with known repetitive functions are exemplified by library systems, inventory control and accounting systems, or systems performing scientific computation. The ADP support tends toward scheduled run, batch processing complexes.

System costs may range from one to one hundred million dollars and will in most cases represent a saving over costs for a completely manual system to perform the same function. For example: A complex of small computers on a regional basis to handle central accounting for a firm with up to 10^7 transactions per month might cost more than $50 million.

The startup time for systems in this class may range from one to two years. This is based upon the assumption that the functions are well known, and that programming time and hardware implementation times are about equal. Finally, it is assumed that some means of data inputting is already existent in a form that requires little modification.

The degree of automation is usually high for such systems, at least in terms of data organization, computation, and formatting of outputs in useful form. Sophistication of data inputting is also possible but not widely used at present.

The utilization of the ADP support to the system is high in the sense that it is easy to tailor it to the expected loads and it is relatively easy to add new capacity when required. As a result, high design efficiencies are possible.

The performance of a Class I system is good to excellent in the sense that the information processing is precise and rapid. As a result, some applications can be undertaken that are not feasible with manual methods.

When comparing Class I systems as defined here with other types it must be remembered that these systems are the least complex. Functionally, the logical operations performed usually require one to three men in a manual system. While the system may handle many problems, the problems generally are not interrelated and data correlation is low. Technically, the system complexity depends upon the load and degree of automation of the data-input subsystem.

SENSOR BASED SYSTEMS (CLASS II)

The majority of sensor based systems serve military applications. Examples are: BMEWS (Ballistic Missile Early Warning System), the SAGE Air Defense System and NUDETS (Nuclear Detonation Detection System). Missile range instrumentation provides a nonmilitary example. These systems have many highly sophisticated electronic sensing elements, elaborate data communication subsystems and large, rapid computers.

Costs for sensor based systems are very high. BMEWS probably cost about $1.0 billion. SAGE costs are more than twice as great. In comparing costs with other classes it should be remembered that the quoted costs are total system costs, the bulk of which are for sensors and communications.

Startup times are long. BMEWS, begun in the fall of 1957, took more than three years to become fully operational. SAGE required four to five years. For NUDETS, three years was required to implement a prototype installation.

In sensor based systems the degree of automation is very high. In most instances automation is essential if the system functions are to be performed within a meaningful span of time.

The utilization of the system is high to perform the function for which it was designed. However, in military applications, the functions of operational importance often change markedly. Modifications of design functions or provision of added capacity for sensor based systems are performed only with the greatest of difficulty. This is even more pronounced for the ADP aspect of the system.

The performance of the systems is generally good from the point of view of technology. That is, the systems do perform their designed functions rapidly and accurately in a real-time mode that would be impossible with manual methods. Performance is generally more questionable from an operational point of view because of the tendency of the systems to become obsolete in a rapidly changing world. For military applications in particular, not only do the operational functions change but also threat changes have had dramatic effect upon the vulnerability of the system, and hence upon its usefulness.

The complexity of sensor based systems is significantly higher than in the case of Class I systems. Functionally the complexity would require the equivalent of ten to twenty people in a manual system [e.g., two radar operators, two communications officers, a track analyst, a weapons specialist, a weather officer, etc.].

The technical complexity is far greater than in the previous case. The data rates are more rapid, the processing timing requirements far more stringent, the logical complexity far greater, etc.

Given the complexity of sensor based systems, cost cannot be viewed as a negative aspect of experience. Complex technical performance is costly. It is probable that design efficiency or clever use of technology would have only second order effect on cost.

Similarly, within reasonable limits of available technology, startup times are governed by the lead times in equipment design and acquisition. For example: in BMEWS, communication construction times were generally the pacing items, not radar development.

Given the complexity of sensor based systems, performance, particularly for nonmilitary applications, can't accurately be counted as negative. The cost of obsolescence is the price of progress. In hardware, general-purpose design has long been used to combat change in functional requirements. In computer programming, general purpose data handling procedures are somewhat newer and are being used to lengthen the period of useful operation.

The crucial point constantly under debate today between system critics and defenders is "whether or not we must have complex sensor based systems to begin with?" The critics insist that in view of the cost and time taken for what is provided, some theoretically less capable approach might have provided as much performance with much shorter time delay and for far less cost. It would appear in some cases that the critics are winning the argument, for the automated approach is being augmented with or abandoned in favor of methods employing decentralized, less automated information handling.

COMMAND SYSTEMS (CLASS III)

Command systems are exemplified by military staff organizations that support a commander in the performance of a command mission. In the case of NORAD (North American Air Defense) the mission is primarily air defense of the North American Continent, with SAC (Strategic Air Command) the mission is strategic bombing; with the National Military Command System (NMCS) the mission is strategic direction of the U.S. Armed Forces. Command systems contain elements of sensor systems, force reporting and management systems, and staff information processing and presentation systems. The ADP support in command systems can be of two types. In the first type, ADP is used in Class I applications by various staff elements. The size, cost, and complexity of the ADP support depends upon the number of applications developed. In general, the many separate ADP applications are integrated by the staff, not by the ADP support. Thus, for the first type of support the discussion of Class I systems holds for the ADP aspects of the system.

In the second type an attempt is made to significantly automate many of the system functions. Thus, in addition to numerous Class I applications, much of the resulting output is processed, integrated, evaluated against criteria provided by the staff, and displayed in summary form by machine. The second type of ADP support can presumably be arrived at in two ways, either late in the life of a Class I type of ADP-supported system, or by intentional design at the outset. To date only a few attempts have been made to implement command systems with ADP applications of the second type. The remainder of the discussion relates primarily to these attempts.

Total costs for command systems vary widely between systems, ranging from a few million to several hundred million dollars. For one small computer installed in existing space used as a data-storage and retrieval system supporting the staff, the cost is towards the low end of the scale. A system with a large command post, special protective construction, and extensive communications, may cost in excess of $100 million. A system with several alternate sites, internetted with communications and operational procedures, may easily cost several hundred million dollars. In all cases, the costs of the ADP complexes need not greatly exceed those of Class I systems.

In comparing the total costs of command systems with costs of other types of systems, caution should be exercised because significant cost elements are not normally counted. For example, the costs do not usually include associated sensor systems or the cost to the subordinate commands of acquiring the data required by the command system.

Startup times for command systems are long. An example of a system employing the first type of ADP support is that of USSTRICOM. At USSTRICOM, a computer was installed within a year, but two years were required to provide a data-retrieval capability to support a predominately manual staff operation. Today, computations are being programmed to relieve the staff of the more routine processing loads, and procurement is being initiated on the remaining elements of the system. The NMCS has followed a pattern of development similar to that of USSTRICOM. For systems employing ADP of the second type, four to five years are required (as far as we know).

The degree of automation in command systems ranges from moderate to low. In systems where the mission has existed for some time and is subject to a certain measure of mathematical definition, both data-storage and retrieval functions and data-processing functions are performed. In the case of newer commands or systems with large uncertainties (particularly those at higher echelons), data-storage and retrieval functions are automated first, and only at some later time (perhaps) are the processing functions done by machine.

To date, reliance upon the ADP support to the command systems has usually been only moderate. For recent systems, the ADP support is actually under development while installed in the user facility. To date in these cases development has not progressed to the point where ADP utilization records can be compared with those of other systems.

Performance from the standpoint of operational employment is acceptable for system applications with minimum functional uncertainty. When the functions are vaguely defined or where they vary, experience has been poor. From the standpoint of technology, application greatly lags the development of tools.

The complexity of the command system is as great or greater than that of the sensor-based system. Functionally the ADP complex usually supports directly an operation center staff numbering twenty to thirty. Indirectly the ADP complex often supports a much larger staff with more widely varying functions. On the other hand, the system usually does not receive data at the frequency of a sensor-based system.

The various negative factors of experience with command systems make this the least attractive type of system to automate from a cost/effectiveness view. The primary reasons for the negative results appear to be the uncertainty inherent in command environments, and the lack of ability for automated systems to quickly adapt to changing functions.

It would be ideal if system lead times could be made dependent upon equipment-acquisition schedules. To approach such a goal, system designers have recently been preoccupied with the problem of generalizing computer programming. Then instead of a system-design process forced to follow classical methods (Fig. 1), the "new design" method (Fig. 2)

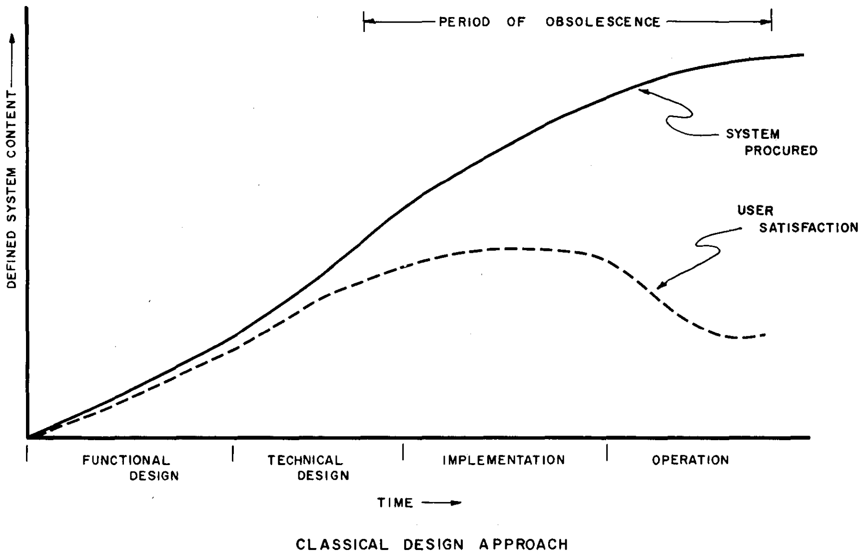
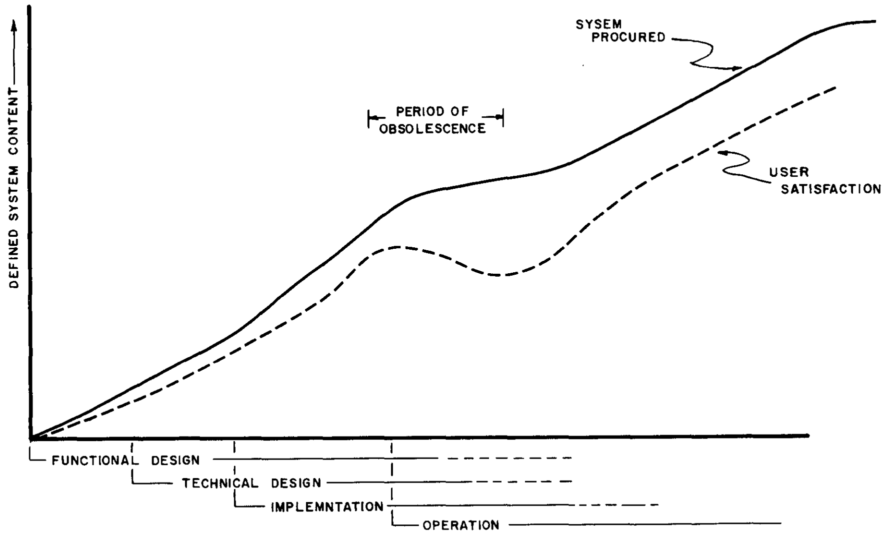


Figure 1.

would permit more nearly parallel development of hardware and programming subsystems.

With the classical approach, a period of intense analysis was begun to define in ever-increasing detail the functional content of the system [functional design]. After the jobs were defined, sized, and analyzed for interrelations, the technical design was begun leading to equipment and program specification followed by periods of implementation and operation. During this sequence, the user, heavily involved at first in job definition, becomes increasingly discontent. As time passes, more and more design compromises are built into the system, and in addition, his appreciation of his mission begins to deviate from his early projections. As a result, by the time his system is operational, he can ill afford the additional loss of projected capability that occurs when trying to make a paper



THE "NEW DESIGN" APPROACH

Figure 2.

specification function in the real world. Furthermore, to change his system, he must go back to the point in time early in the design cycle, change some of the early "frozen-in" decisions, and work the process through again. The result is a long period of obsolescence, and reliance upon a manual system.

With the "new design" approach, much of process of functional and technical design can be overlapped. In some respects this is just a tacit admission of what the technician did all along. More importantly, design and implementation can be overlapped. With a knowledge of generalized programming techniques, important factors bearing upon equipment selection can be tackled early and equipment acquisition initiated. Because the key to generalization is to construct the basic data-processing functions independent of the specification of operational function, program development can begin earlier, borrow more from other systems, and readily accommodate variations in operational function to be performed. As a result, user discontent is less pronounced. He may still have to suffer some loss of desired capability when faced by some hard technological facts. However, he is not additionally constrained by a need to seek premature definition of his functions, and he can reserve the right to change his mind within reason. He still faces disillusionment when he compares the product with the specification, but not to the same degree, and he can implement corrective changes in a much more reasonable time frame.

The recognition of the need for a new design approach began several years ago, and much progress has been made in this direction. While no one current operational system fully qualifies as an example, several have one or more important elements required for general purpose design.

The key issue in system design, however, is not tool design but application. Hand-in-hand with the recognized need to adopt a new design approach for tools there is a need to address another major problem area where inadequate attention is usually paid, the area of data definition. Before a system has operational value one must have tools to manipulate data, and data with sufficient information content. It is this last area that is most often neglected in command system development today. The neglect stems from two primary causes. First, the uncertainties inherent in system functions makes this area a most difficult one in which to work, often requiring tedious and costly analysis, definition, experimentation, modification and not infrequently a good deal of political negotiation before satisfactory solutions are hammered out. The second cause stems from the growing reliance upon the new design approach. Since the technician can make the hardware and program development increasingly independent of functional detail, he has begun to withdraw from this area. He exerts less pressure upon the user to develop it, claiming rightfully that the area is the responsibility of the user, and he no longer employs a large amount of technical resource in the area.

To adequately plan for large systems it is necessary to understand the magnitude the problem data definition represents. It is not a major problem for a base commander to keep track of the status of his aircraft by type. However, if status must include data of significance to logistic support planners, and data to support force allocation planning, etc., the data records begin to get cumbersome. In the NMCS it is not uncommon for a file record to contain four or five subsets of data to support different functional aspects of file usage where each subset contains ten or more data fields.

To generate such a file from the beginning is an exceedingly time-consuming task. It may take three months or more of initial operations analsis to determine areas requiring support. Having defined the general purpose and content of a file, three or four months of detailed analysis are required to establish the file format, a dictionary of terms, and to establish a suitable file vocabulary. General coordination with all concerned parties of draft file specifications can consume one or two additional months. Generation of the file at the data sources can require another two to three months—followed by a period of data consolidation, file generation, and analysis of what went wrong, lasting perhaps another two months. Subsequent modification of reporting procedures and a second generation phase to get a usable file brings the total time for file generation to between 14 and 17 months. The effort involved can run in excess of six man-years per file. Certain economies can be practiced by formatting data in machinable form from readily available manual files at the expense of additional resources required to generate the data. Added economy can be had by borrowing data already put in machine form somewhere else.

As a result of the difficulty encountered in constructing useful data files, it may not be surprising that systems like USSTRICOM or NMCS have had equipment complexes and programming routines long before there was data of major operational significance in the system. Nor is it surprising that in the early phases of system operation where data development has only begun that the capability provided by the system can be easily matched by efficient manual methods.

At this point in time it would seem that there is no effective solution to the problem of data definition that does not require a sizable investment of time and resources in operations analysis.

The term "evolutionary design" has become the vogue recently, at least in the Washington area, to describe an orderly design progress that advocates a learn-as-you-go policy in easy steps. Such a policy could be implemented by combining technical design activities employing the "new design" method with a substantial program of data definition.

Unfortunately, in some recent system developments, undue emphasis has been placed upon the uncertainty in command environments, and the tendency has been to use uncertainty as a rationale to defer planning for the systematic introduction of new capability. The result has been uncontrolled system growth generally at a rate less than could be reasonably obtained.

Assuming that a more positive approach is adopted and applied, particularly to command-system development, the major obstacles of uncertain environment and a resistance of the ADP support to rapid change in function can be substantially reduced. Even so, systems would continue to be expensive and would continue to require long times to implement—not, however, out of proportion to the complexity of the functions they would be designed to perform.

Since the pressures for central management that motivate command-system development appear to be relatively unchanging, the only other apparent alternative to large-scale investment in complex systems for command lies in redefining some of the philosophy of centralized management with the goal of reducing the complexity of system functions.

One example of a possible change in philosophy might be embodied in a system that keeps status on what subordinate element has what responsibility and what supporting system capability to carry it out. Such a file, if it reflected current status and contained adequate directories, could greatly ease the problem of executive problem definition and delegation of authority to execute assigned responsibility. It would imply that the tools to provide operational solutions to problems should be placed in the hands of subordinates close enough to the problem to work on it effectively. Such a system would probably also require a major advance of management science to insure that the risk in operating in such a decentralized mode was reduced to an acceptable minimum.

SUMMARY

In general, particularly for systems with military applications, costs are high, startup times are long, and functional performance often leaves something to be desired. However, the degree to which this is true varies markedly with the type of system under consideration.

Because of the characteristics exhibited by large military systems, their development has increasingly come under the scrutiny of high-level groups in government. These groups usually reflect user desires for high performance, short startup times, and lower costs. That these groups are not highly pleased with the development of large systems is apparent judging from the reductions in support of some of the programs, and the fact that most large-scale systems with major ADP support were initiated prior to 1960.

The apparent conclusion to be drawn is that large-scale systems that rely heavily on ADP support are bad. However, costs are not disproportionate to the complexity of the functions desired, and startup times are not excessive when compared to similar times for completely manual systems of similar complexity and scope.

Furthermore, performance is very different for different classes of system. Some of it has been very good. In those cases where performance is poor, much can be done to improve the situation. To insure ADP support responsive to uncertain and changing environments it is necessary that ADP programs be generalized as much as possible. Much technological effort is currently being expended in this area.

Of far greater impact in the successful design of ADP support, the problem of data definition and acquisition must be approached as the highest priority item and successfully solved. It is this problem that lies at the core of system application. Recent actions by the Department of Defense have directed the user to take a greater role in the development of his system. To this proper enhancement of the user role, the technical implementer must join a major portion of his resources in a direct attack on the problem through analysis and experimentation. It is possible that these steps may have to be coupled with fundamental changes in concepts, particularly in command applications, before long-range difficulties can be resolved.

In the current situation, problem definition in terms of the data to be used by the system, will be the barrier to increasing use of automation in large systems. It is likely that the near future will see the initiation of few if any truly large-scale command systems employing a high degree of ADP support. Instead, efforts will be focused on the search for simpler, less complex, faster to implement but possibly less adequate methods for solving system problems. Automated support, particularly in command systems, will be largely confined to Class I applications.

Mr. L. D. Earnest of the MITRE Corporation suggests that ADP may develop along the lines of a public utility. This would seem reasonable for systems of the Class I type. Large-system experience supports this view. People with definable jobs and data sources use the ADP service provided. Operators of the ADP facility provide for system growth on the basis of extrapolation of usage records. For applications where ADP is premature the user would like to wait until adequate data definition is accomplished. With ADP utilities he *could* wait, secure in the knowledge that the ADP support would be available when required.

14

Command and Control

JIRI NEHNEVAJSA
Professor of Sociology
University of Pittsburgh

I will not attempt to analyze concrete operational experiences in the area of command and control systems. Such an evaluation calls for data on distributions of performances relative to system performance criteria. Even if available, these data would not be altogether appropriate for a presentation at a general conference of this type.

As a consequence, this paper will be limited to the consideration of certain problems which I consider particularly salient in terms of all, or at least many, command and control systems. These are problems which have significant bearing upon the behavior of the operational system but are not, at the same time, identical with what might be viewed as specific operational experiences.

Furthermore, I propose merely to highlight some of these problems without subjecting them to the detailed analytical scrutiny which each singly may well deserve.

The concept of control implies a capability to monitor an on-going situation and to compare its properties with the characteristics of some corresponding intended state of affairs. This involves, of necessity, some effort at predicting the probable course of events over an appropriate time horizon.

The notion of command, in turn, implies that information on the relation between actual and intended situations and processes permits an evaluation which leads to the determination of appropriate courses of action. It also means a capacity to communicate decisions to those who are expected to execute them as well as to those whose own actions will be affected by the decision in any significant manner. The command concept also entails the idea that the execution of a decision, as well as its effects, come to be monitored, and the nature of the feedback leads to reenforcing the initial choice or toward a reassessment and a new decision.

A few thoughts now about issues associated with the control functions of the systems.

The intended situation is generally some plan. On one end of the spectrum, this may be a war plan providing for patterns of force deployment under varieties of likely circumstances and for usually several alternative objectives. On the other end of the spectrum, this may be a plan implicit in any specific decision in that its objective, too, is to produce some desired state of affairs or to prevent some unwanted system state from occurring.

The difference is one of levels of complexity. But it is far more than that at the same time. One kind of plan refers to an environment which as yet does not exist. Another one is responsive to the here-and-now in a more direct manner. In military systems, of course, the interaction of these issues is quite direct and quite crucial. At any given time there exists some range of intended or desirable situations which *ought* to prevail right now to make for optimal transition to the nonexisting war environment if it ought to become realized in the next moment. Thus, one set of situational control functions is instrumental to major future objectives.

Now data pertaining to the characteristics of a given intended state of affairs may be provided in varying levels of detail. Generally, the greater the level of detail and specificity in the definition of the situation that ought to prevail, the greater the likelihood that in *some manner* the actual situation will deviate from the model. If plans are provided only in generalized form, the greater the likelihood that potentially serious discrepancies between plan and reality will go undetected with severely degrading effects upon the system as a whole. How to strike a balance remains unsolved unless one is willing to accept diffuse user satisfaction or dissatisfaction as the main criterion.

Similarly, it is not altogether obvious whether plans as profiles of intended situations and processes are preferably generated within a given command and control system or whether they are better viewed as an input into the system which could come from any appropriate source as a fait accompli. The former approach taxes the system heavily in that it must also involve complete planning capabilities. The latter approach alters the fabric of authority, at least at the highest levels of the organizational hierarchy, in that certain accustomed discretionary powers simply disappear.

In any event, no plans can genuinely provide for all contingencies, so that situational and on-the-spot replanning must be almost assumed as the rule rather than an exception. Replanning and planning, of course, are the same processes but viewed from a different point of departure. The problem of off-line and on-line activities (and *their* interaction) becomes quite fascinating.

An actual situation keeps changing. Furthermore, the variables which are used to describe the on-going situation change at different rates and with dissimilar predictability. There is some time delay, no matter how apparently trivial, between acquisition of data by sensors and its generation in the form of a usable output. The profile of an actual situation at any given time has, therefore, two important and limiting characteristics: for one, it refers to some past situation in any case and not to the situation of the moment. Secondly, the individual descriptors of this actual situation are of varying obsolescence because of their different rates of change, different modes of acquisition and processing. The implications of this problem have really not been studied, and my suggesting it here as a serious problem does not prejudge the alternative outcome of appropriate studies. But time-tagging of information items has not been attempted on the whole in any systematic manner, nor do we know how this relates to the confidence which a decision-maker has in the information at his disposal.

A discrepancy between the actual and intended state of affairs signifies some system problem. One issue along these lines has to do with the relative magnitude of deviation between intended and actual values which can be detected due to the system modes of data acquisition, and the magnitude which can be processed as a function of equipment capabilities. This is largely a technical problem.

The second issue has to do with some threshold magnitude of discrepancy which establishes a boundary between tolerable and no-longer-tolerable departures of the actual from the desired state of affairs. This, in turn, is chiefly a policy problem.

The third issue has to do with the possibility—or better yet, the fact—that cumulative effects of otherwise tolerable discrepancies may not be tolerable. The criteria for making such choices seem lacking at the moment.

The last issue along these lines has to do with the possibility that joint effects of otherwise singly tolerable discrepancies may not be tolerable. The criteria both for design and operations choices are largely lacking at the moment.

Before I mention some of the overall system problems, a few remarks more specific to the command function seem appropriate.

A discrepancy which constitutes a system problem can be resolved either by altering the nature of the actual situation or by modifying the specifications of the intended state of affairs or by both to some extent. The main issue has to do with the determination of the conditions under which it is necessary or preferable to seek to alter the actual state of affairs and bring it into harmony with the intended state, and those circumstances under which it becomes necessary or preferable to adapt the characteristics of the intended to the actual situation.

Generally, command and control systems lack the capability to provide data on projections of the most probable consequences of a given decision before it is firmed up, communicated, and its execution begun. Some such testing can be accomplished in simulated environments, but it raises the most serious methodological questions as to sampling of decisions, circumstances, and decision-makers to yield some confidence in the generalizability of the results to actual operating environments.

Indeed, it would seem at least theoretically possible to develop system capabilities to identify decision options appropriate for a given situation, to identify the probable immediate consequences of each alternative choice, and to identify the probable longer run consequences of each choice. But this raises the most serious question as to whether there would be anything left for the human decision-maker to decide.

I am not prepared to argue altogether that this may be undesirable under all circumstances. Yet even this is a more complicated problem than one concerning the role of men in the total process, or one that simply concerns the efficiency of allocating various functions to machines and others to men. The point I am willing to make, however, is somewhat as follows: even if feasible, computerized decision-making per se is not really quite computerized. What happens is simply a drastic redefinition as to who makes the decisions, and thus a revolutionary modification in existing patterns of authority. In effect, a data-processing specialist or a programmer will make a set of permanent decisions in the place of a decision-maker normally expected to make them.

This may be an improvement or not. But in any event, the importance of this shift cannot be overemphasized, and its implications certainly must not be overlooked. This is underscored by the tentative observation that much less attention is paid to the training of programmers in anything but programming than the corresponding attention which goes into processes whereby our society elevates certain men into significant decision-making roles. And I will be the last one to underestimate the centrality of the decisions which are made quite routinely by programmers of even very low professional calibre.

To argue that the decision-maker can control what is being done on his behalf seems to me somewhat unrealistic. For one, there are individual styles of decision-making and these are not as readily transferable from person to person as are occupancies of various positions and roles in our social system. Secondly, we know very well that decision-makers may be unable to verbalize, or verbalize in a manner directly understandable to the data-processing specialist, the criteria which actually guide them in using information and in reaching conclusions on the basis of it. Thirdly, in complex systems we are speaking of hundreds of thousands of programming instructions generated in segments and subsegments by whole teams of data-processing specialists. It does not seem possible to comprehend all this very adequately any more than it seems likely that given decision-makers could effectively channel the development of these enormous information-handling systems.

Command and control systems are complex. They are also significantly real-time systems. They are expensive to design, install, maintain, and operate. They are expensive to modify, and despite the fetish made of flexibility, often too rigid to permit even small fixes without major effort.

Some consequences flow from these simple observations. First, the complexity tends to be so staggering that the system user must continue relying on the system designer throughout the life-cycle of the system except for routine utilization. This is not implied as a critique. Rather, I am suggesting that this signifies the arrival of new partnerships, and the necessity for these partnerships might as well be recognized at the outset. There is, I firmly believe, no such thing as the system user taking over a complex command and control system as a terminal package. The marriage of system user and system designer continues and this might as well become an aspect of system planning.

Nor is it quite feasible for the system user to be his own designer. In theory this sounds perhaps plausible. In reality, some system is in existence which the user is quite busy employing on an on-going basis right now. He cannot suspend his operational responsibilities of today while developing a system for tomorrow. And I daresay that he cannot do both.

The cost associated with command and control systems is still another matter. It amounts to commitment. This tends to mean that once a development program is initiated, there are sufficient emotional, political, and other reasons to see it through even if alternative systems or alternative configurations became available. This holds above all in the area of equipment procurement, and the problem is accentuated by the fact that far too often equipment is acquired long before the realistic stage of system development would warrant it. Many systems are designed around hardware, and this normally means some off-the-shelf hardware or some modified equipment already fully available.

I should add that many research laboratories, too, are designed around hardware with similar consequences. In both instances, instead of identifying the problem and the resulting equipment requirements, the problem and all other requirements are constrained by the hardware which, after all, must justify its cost.

This issue is, indeed, coupled with off-the-shelf thinking. Truly, an on-going battle rages between those who prefer to approach problems by blue-skying and those who prefer improvements of an existing situation. Clearly, this is not an either-or problem, for if it were it might have already been resolved. It is obviously safer to avoid radical departures from current thought. It is therefore both safer and easier to simply superimpose modern equipment upon previously manual functions without significantly altering these functions, or even questioning their viability. The probability of success is greater, but the consequences of succeeding somewhat less than spectacular.

In the area of man-machine interactions, perhaps the major problem revolves around the determination of the type, amount, and timing of information which the human decision-maker is to receive, and at the same time, the determination of the information which he may have access to, even though it need not be presented to him under most circumstances.

Men are on the receiving end of an enormous quantity of information already. in fact, too much of it, as it is. There does not seem to be much point in automating and speeding up this flow, and thus even increasing the effective amount per unit time. Selectivity rather than all-purposiveness would seem more appropriate both in terms of access to data and of its actual presentation to decision-makers. It is consequently of great importance to identify the information which particular decision-makers ought *not* to receive.

Information which people say they want is often not the same as information they want. The information they want is generally quite in excess of information they need. At a given level of the decision-making hierarchy, an effort to provide detailed data on all aspects of the system and its operations would tend to lead to centralization of decision functions. At least, it would degrade the use of imagination which goes with autonomy and fairly clear responsibility at more subordinate levels within the organization. No systematic data presently exist on relations between system outputs, the actual decisions in operational contexts, and the actual consequences of such decisions. The problems of determining these information needs therefore remain quite serious.

The notion of real-time monitoring implies a system capability to be operative around the clock. This requirement seems to be always present, and it is the more critical the more the command and control domain of responsibility has to do with rapidly changing events rather than relatively slower ones. Indeed, some fallback provisions are an important ingredient of command and control systems. These may be provisions to return to some version of pre-electronic data-handling modes. Or else, multiplexing of the core equipment and the appropriate communications linkages may be used as an alternative.

Relatively little systematic thought has been actually given to multiplexing of equipment between and among various systems rather than hardware duplication or multiplication within each system. Although this alternative may seem quite appealing, its consequences are not altogether clear. It may, for instance, involve using the same kind of equipment across a variety of systems and this has something of the effect of monopolization in the hardware production and distribution field.

The same kind of an issue holds regarding intersystem compatibility of equipment, program languages, and resulting procedures. Yet, some degree of compatibility is of great relevance because of the interfaces which invariably exist among several command and control systems, if not all of them.

This is further complicated by the fact that various systems are, at any particular point in time, in different stages of development, or else in different stages of their life cycle. In the rapidly changing field of data handling, these time differences in and of themselves make adequate compatibility of past with present, and present with future, systems quite difficult.

The sociological and social psychological components of systems and their operations are also rather central in the eventual capacity of the systems to act on their objectives. Existing organizational forms significantly constrain the range of choices which are open in system design and utilization. Major departures from prevailing cultural patterns within an organization, such as the military establishment, may be so threatening as to make even good solutions less than acceptable. The problems associated with phasing people out of one type of working environment and an accustomed set of behaviors into another environment are ample and they are rarely in the direction of upgrading, rather than down-grading, system performance.

I would now like to bring my discussion to a close on a somewhat different theme. I have singled out a number of problems associated with development and utilization of command and control systems. This has led me to the exclusion of the tremendous progress which I believe has been made in the course of the past two decades or so in the conceptual, methodological, and hardware aspects of these systems. Nor must we be oblivious of the fact that starting from scratch, numbers of people from various disciplines have developed a truly impressive know-how such that it at least provides assurance that past errors are unlikely to be repeated. These individuals are heavily concentrated in relatively few organizations, but they are here and they were not here only some ten to twenty years ago.

Enough progress has been made to justify thinking about the expansion of command and control concepts to areas in which such notions have not generally been employed. To mention but three important areas:

for one, there exists potential use of command and control thinking in conjunction with the conduct of the nation's foreign policy. Secondly, and in a somewhat similar vein, command and control concepts would seem to be suited rather well to the generation of global foreign aid planning, execution, and progress monitoring.

Outside of government, the third major area has to do with large-scale industry. The steel industry is probably an excellent example in that the timing, quantity, and especially quality of produce must be not only closely planned but also closely monitored, and the effects of severe discrepancies reverberate through the nation's economy as a whole. Other areas could be similarly discussed with potentially interesting implications.

In some sense, the military command and control systems serve as a central prototype for certain forms of information-handling problems now and in the future. These are systems involving large quantities of data, and major requirements on speedy access to prestored information. At the same time, they entail the need for real-time monitoring and real-time testing of actual against planned-for situations. If all the problems can be adequately solved in conjunction with command and control systems, I submit to you that problems in the development and utilization of other information-handling systems with fewer taxing time requirements appear much more amenable to successful resolution.

Today's thinking, finally, ought to be oriented to the mid 1970s, and today's implementation to the early 1970s. In the development of hardware, this has become a fairly customary orientation. But I am convinced that we must extend it to all aspects of this newest and most fascinating area of knowledge availability systems.

V. LARGE-SCALE SYSTEMS UNDER DEVELOPMENT

15

New Mathematics for a New Problem

ORRIN E. TAULBEE
Manager, Information Sciences
Goodyear Aerospace Corporation

INTRODUCTION

Perhaps you have wondered what is the new problem with which we shall be concerned in these pages, and secondly, after the problem is expressed, what new mathematics has been developed that is applicable to the problem. Let me say at the outset that the principal concern of my discussion is with classification. What is new about this problem? It has been around since the dawn of civilization in one context or another. My primary reason for referring to it as new is that there is new emphasis on this problem as our information-handling systems increase in complexity. Throughout our discussion we shall explore some of the ramifications of this significant unsolved problem but we will demonstrate certain results take a positive step toward finding satisfactory classification schemes. schemes.

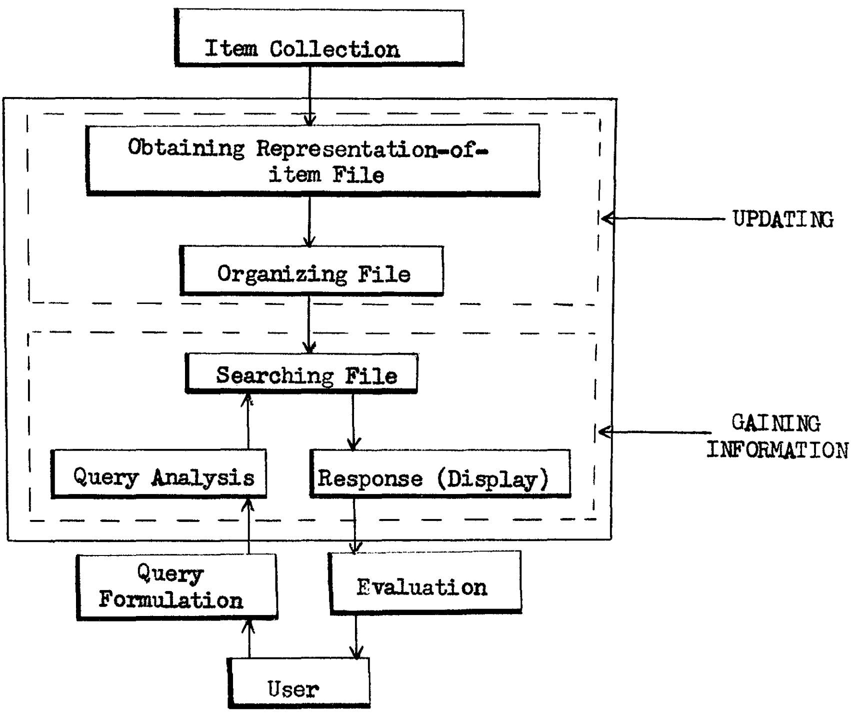
CLASSES AND CLASSIFICATION

Before beginning a discussion of classification, one must concern oneself at least to some extent with the notion of classes. It is not our purpose here to delve into the philosophical considerations of what classes are, but in case one is interested he should consult Ref. 7. Nor is it an easy question to decide generally what the concept of a class should be and in particular what a class should be in the context in which we shall use it. Let us just say here that our use of classes can be thought of as a decomposition of a set of objects into a collection of subordinate groups which will be called classes. According to *Encyclopoedia Britannica*, classification is "the arrangement of things in classes according to the characteristics that they have in common." It is not sufficient to think of classification as placing those objects in a class adjacent to one another, as is done in most library classification schemes, for we must admit the possibility that the objects are considered to belong in the same class even though they may be quite widely separated.

We may consider two types of classification—hierarchical and nonhierarchical—the former admitting the possibility that a class may be subordinate to a class other than the entire collection of objects, while the latter does not admit this possibility. It is unfortunate that some individuals interpret classification to always mean hierarchical classification.

INFORMATION HANDLING

For a better understanding of the following discussion it is convenient to give a diagrammatic description of information handling. To my knowledge, it represents all information-handling systems—including those which are purely manual, those with a man-machine intermix and those which are completely automatic. Since the diagram is representative of all systems, it is clear that the functions represented by the blocks take on different meanings depending on the particular system under consideration. In fact, for some systems one or more of the functional blocks may not be present. However, the end product of any information-handling system is the same—the presentation of information for decision-making.



The objects with which our information-handling system is concerned shall be referred to as items. A common information-handling system is one where the items are textual in nature. Our discussion will not be limited to this, however. We shall assume that an item may be a document in the usual sense, a book, a section, paragraph or sentence of a document or book; or the item may refer to an aerial photograph, a structural diagram of a chemical compound, a radar return, a sonar signal, and so forth. A decision must be made to determine those items which are to be included in the system. This decision may be an a priori one, or the decision may be made for each item individually at the time of accession.

Many different representations of the items are possible for a given collection. For example, if the items are textual in nature the representation adopted for the items might be: full text, full text with common words omitted, abstract, extract, keywords in context, title, index terms, first and last paragraph, etc. If the items are chemical compounds the representation might be: structural diagram, chemical name, one of several linear notations, a connection matrix, etc. If the items are signals the representation might consist of an explicit function of time, power spectral density, amplitude and phase spectrum, sampled data representation, etc. Of course, in every case an item may be used to represent itself. Again, the representation criteria may be established a priori so that the representation may be obtained either routinely or for each individual item on a judgmental basis, subject to general criteria established beforehand. Part of the function of obtaining the representation-of-item file is that of recording the results on a searchable medium.

If the item collection reaches any substantial magnitude (the collection is assumed to be dynamic), then consideration must be given to how the file should be organized. This is intended to include the establishment of format, search strategy, and classification of the recorded representation. At this point, updating of the file is complete and ready for searching.

Upon the formulation of a query an analysis must be performed in order to (1) make the representation of the query compatible with the item representation, and (2) establish appropriate permissible search strategy. Following this the file is searched and results of the search are delivered.

Within this framework, we may now describe a series of information-handling systems in which each system is more complex than the previous one with the ultimate being a system requiring no human intervention which operates in real time.

(a) *Natural System*

First of all let us describe what may be called a "natural system." An example of this is the individual researcher's personal file. This generally consists of a collection of items relevant to his particular field of endeavor. He is the user to the extent that he decides what items will be added to the collection; he formulates his own query, searches the collection, obtains those items which are responsive to the query, and makes a judgmental evaluation as to their relevance to the query. Dissatisfaction with the items retrieved may lead him to refine or modify his query and iterate the process. Note here that the researcher is using the items to represent themselves and in retrieving he actually retrieves the physical items from the files. Growth of the collection may require the researcher to develop and organize an auxiliary file—the representation-of-item file.

Libraries, either public, university or specialized have developed a system duplicating in large measure the information system of our individual researcher.

(b) *Machine-Aided System*

Because of one or more of the following reasons, one may bring in a machine to assist in the information-handling system. These reasons are: (1) increased speed; (2) magnitude of representation-of-item file; (3) reduced costs in processing; or (4) avoidance of errors in processing. Machines to assist in processing consist generally of three types: (1) tabulating equipment such as sorters, collators, and printers; (2) peek-a-boo devices; and (3) computers. The most common utilization of machines in information-handling systems is in performing the function of searching the file.

(c) *Automatic System*

Because of increased complexity of information-handling systems, it is frequently desirable to have a machine system called an "automatic system," behaviorally equivalent to the functions included in the solid rectangle (Fig. 1); this includes all those functions which can be mechanized.

(d) *Real-Time System*

For a real-time system four "times" appear to be of significance: (1) μ, the average time for updating; (2) ν, the average time for gaining information; (3) δ, the average rate of accession of new items; and (4) ξ, the average rate of accession of queries. Obvious conditions on these variables are $\mu \leq \delta$ and $\nu \leq \xi$.

Most information-handling systems in existence today fall into either those of types (a) or (b). For many systems it would be desirable that they either be of types (c) or (d).

The use of machines to perform each of the functions in the solid rectangle are in various degrees of development. As was indicated previously, the most highly mechanized function is that of searching the file. Displays, so far as printed or microform output are concerned, are fairly well mechanized. Much remains to be done for other types of displays. The other three functions represented are perhaps less well developed, but experiments are going on in each of these areas. For example, in obtaining the representation-of-item file, experiments in auto-abstracting and auto-indexing have been performed. The function of inquiry analysis may be avoided almost completely. An example of such a system is that in which the representation-of-item is full text. Once the file-organization characteristics have been established, a machine may assist in performing this function. However, little has been done in the way of machine classification.

CLASSIFICATION IN INFORMATION HANDLING

Restricting the concept of classification to information handling it is clear that the fundamental problem is that of deciding in what sense the items should be considered associated or similar. It is also clear that a classification scheme cannot be universal but will be specialized to the particular collection of items under consideration. For example, the criteria for association of two items will be quite different if the items are, on the one hand, documents, and on the other hand, signals. In fact, we can go further: the classification scheme of the same collection of items will be quite different depending upon the viewpoint of the classifier. This can be handled theoretically, however, by means of the criteria adopted for association. There are three principal reasons for classification in an information-handling system: (1) size of file; (2) increased speed; and (3) recognition of the appearance of new classes. These reasons are not mutually exclusive. For the first, unless the file is classified, it is necessary to search the entire file, but this may be impractical depending upon the size and mechanism, if any, used in searching. For the second, urgency of gaining access to the information may dictate that the items be decomposed into classes. The third purpose of classification is in identifying new concepts or knowledge that finds its way into the item collection.

TRADITIONAL APPROACH TO CLASSIFICATION

The following is a common approach to classification: From personal knowledge of the item collection some classes are established a priori which are felt to be representative of the characteristics of the entire item collection. After this, each item, whether in the original collection or a new accession, is considered individually and evaluated to determine the classes to which the item belongs. This is a judgmental evaluation which must be made yet cannot be made precisely since initially the definition of the class is vague. New classes are added reluctantly. When the new classes are formed, almost without exception there is little or no review of items already in the file to determine whether or not they fit into the new class.

In the usual library situation, classification consists of two functions, that of establishing cross-references and that of classifying, each of these being accomplished within the guidelines of a set of rules. It seems to me the purpose of classifying in this context is to narrow the search resulting from an inquiry to a limited portion of the representation-of-item file, and the cross-referencing or association established increases the possibility of retrieving all pertinent information from the file that is either directly or peripherally relevant to the query. Cross-references are included to the best of the individual's ability to remember and recall.

In order to automate both processes, it is necessary to establish an analytic procedure for making associations and classifications, since classification is made on an intuitive and experience basis. Thus, it would be desirable to have a classification scheme which is objective; that is, it removes the judgmental element, and gives complete updating when a new class is formed.

MOTIVATION FOR MATHEMATICAL MODEL

Perhaps the first step away from the traditional approach to classification was included in a paper by Vannevar Bush[4] in the year 1945. In this paper he defined a theoretical machine called the "memex." The memex has massive storage capability, the capability of retrieving any item from storage and displaying it, the capability of inserting written comments into storage during the viewing process, and most important, the capability of tying two related items together. This last capability Dr. Bush referred to as "associative indexing," by which he meant a mechanism whereby any item will select immediately and automatically another associated item. Furthermore, the operator of this machine, in viewing items which he wishes to associate, links these together permanently by simply pressing a key and thereby successively builds a trail of association. What this amounts to, in effect, is to put items into a class, as if they were bound together in one volume, from widely separated locations. Notice that here emphasis is placed upon the association between concepts or ideas—each concept forming a class in the individual's mind.

The need for the associative concept is evident when one considers the selection processes that are available in searching a file. At present there are two types: (1) search the entire file; (2) use a tree structure for searching the file. These methods have been implemented on card equipment and conventional computers. The association concept would be particularly effective for avoiding backtracking if a search is being made in one branch of the tree and it is required to search in another branch of the tree.

Tying together the ideas presented by Dr. Bush and the traditional classification approach to the library, it seems reasonable to think, instead, of reversing the process, that of first establishing the association between the items and then through some logical process form the appropriate classes. There are two cases to consider: (1) the items are either associated or not, and (2) more generally, the items are associated to a degree. This paper will be concerned only with the first.

THE MATHEMATICAL MODEL

A review of the mathematical literature reveals that little has been done in the way of a mathematical approach to classification. Apparently one of the few concepts in mathematics relating to classification as such is the well-known idea of an equivalence relation.

The fundamental features of an equivalence relation are that a binary relation ρ is defined on a set S. The relation satisfies the reflexive, symmetric, and transitive properties. By the phrase "a binary relation ρ is defined on a set S" is meant that for any pair of elements a, b, of S a definite rule is prescribed by which it can be determined whether or not a and b are in the relation ρ. This may be denoted by $\rho(a,b) = 1$ or 0. The significant property of an equivalence relation, defined on a set S, is that the relation separates the set into mutually exclusive, exhaustive classes. That is, each element of S belongs to one and only one class. Because of this the partitioning of the set S may be thought of as a classification of the elements of S. To be precise, let S be a finite set with elements $s_1, s_2, \ldots, s_m$. Since it will be assumed that, in general, the number of elements varies with time it will be necessary to require that the set be well-defined—i.e., given a new object s^* there exists a definite rule by which it can be determined whether or not $s^* \epsilon S$. It will be further assumed that ρ is a binary relation defined on S by an explicit rule which determines whether s_i and s_j, s_i and s_j not necessarily distinct, are in the relation or not. This will be denoted by $\rho(s_i,s_j)$ (or $\rho(s_j,s_i)$ if the order is important) and agree that $\rho(s_i,s_j)$ has the value 1 if s_i and s_j stand in the relation ρ; otherwise, $\rho(s_i,s_j)$ has the value zero. An alternative way of thinking of this is that ρ is a mapping of the cartesian product space S^2 onto the set $\{0,1\}$. Moreover, it will be assumed that when the set S is augmented by the addition of s^*, to form the set S^*, the same rule is applicable for evaluating $\rho(s^*,s_j)$ for $j = 1, 2, \ldots, m$ and $\rho(s^*,s^*)$.

The relation ρ is an equivalence relation if ρ is reflexive, symmetric and transitive, i.e.,

(1) $\rho(s_i,s_i) = 1$, for $i = 1, 2, \ldots, m$.
(2) if $\rho(s_i,s_j) = 1$, then $\rho(s_j,s_i) = 1$ for all i and j.
(3) if $\rho(s_i,s_j) = 1$, and $\rho(s_j,s_k) = 1$, then $\rho(s_i,s_k) = 1$.

In the application of this to information handling such a classification is clearly unsatisfactory since in general an element may belong to more than one class. This motivates the search for a generalization of the classification induced by an equivalence relation.

A study of the equivalence relation postulates shows that it is the transitive property which decomposes S into mutually exclusive classes. However, the classes induced by an equivalence relation do have the characteristic that the classes are maximal[1] with respect to the property that any pair s_i and s_j belonging to a class implies that $\rho(s_i,s_j) = 1$. This suggests that the transitive property be dropped and the maximality condition, just referred to, be imposed on the classes. The collection of classes determined by such a relation ρ are called "coherence classes." This terminology is consistent with Ref. 6.

Suppose now that a new element s^* is adjoined to the set S to form the set S^*. Let C_k, $k = 1, 2, \ldots, n$ be the coherence classes of S and $R(s^*)$ be the set of elements in S^* related by ρ to s^*. A precise inductive algorithm was given in Ref. 2 for obtaining C^*, a coherence class in S^*. The algorithm is based upon $C_k^* = \{s^*\} \cup \{R(s^*) \cap C_k\}$. As k ranges over the values $1, 2, \ldots, n$, C_k^* forms a new coherence class if it is maximal. This yields all new classes; none of the classes C_k of S can disappear. (If an element is removed from S, then, of course, a class may disappear.)

It should be pointed out that the decomposition of the set S into classes, either in the case of equivalence classes or coherence classes, is unique. If a different classification is desired then the association criterion may be changed, i.e., the binary relation ρ defined on S is modified. Because of the well-known correspondence between graphs, relations, and matrices, it is clear that these ideas may be expressed in either of the other forms. Some interesting matric relationships were given in Ref. 8.

Since initiation of this investigation two papers have come to the author's attention. Hillman[5] has explored, from a philosophical-logical point of view, Carnap's idea of a "concept-class." Bonner[3] develops some computer algorithms for what he calls "clusters." The "concept-class" and Bonner's "tight cluster" appear to be identical to the notion of a coherence class. The present paper establishes a firm mathematical basis for classification in case (1) referred to above and simultaneously affords a simple updating algorithm.

EVALUATION OF RESULTS OF CLASSIFICATION

How can one evaluate the results of automatic classification? The natural approach seemed to be to compare the classes obtained objectively with those classes obtained by means of people making a judgemental classification. This has been done for other automatic classification schemes, and as was to be expected, there were classes in the objective scheme which did not appear in the subjective scheme and conversely.

Further consideration reveals that there seems to be no sound reasons why the classes of an objective classification should agree with those of a subjective classification scheme to any significant degree.

The appropriate type of experiment to conduct for evaluation of classification schemes seems to be to evaluate the output obtained by each and in this way measure the performance of one versus the other.

Experimental results will be reported in another paper.

SUMMARY

A classification scheme has been demonstrated in the case where the association between two items is a binary relation whose field is the set $\{0, 1\}$. Significant features of the procedure are (1) objectivity; (2) ease of updating; (3) automatic; (4) focuses attention on the association concept and (5) independent of item representation.

REFERENCES

1. Bednarek, A. R., "Reflexive and Symmetric Relations: A Set-Theoretic Characterization," submitted for publication.
2. Bednarek, A. R., and O. E. Taulbee, "On Maximal Chains," to appear in *Publicationes Mathematicae* (Debrecen).
3. Bonner, R. E., "On Some Clustering Techniques," *IBM Journal* (January 1964), pp. 22–32.
4. Bush, Vannevar, "As We May Think," *Atlantic Monthly* (July 1945). pp. 101–108.
5. Hillman, D. J., "The Notion of Relevance," *Amer. Doc.* (January 1964), pp. 26–34.

6. Rado, T., and P. Reichelderfer, "On Cyclic Transitivity," *Fundamenta Mathematicae*, vol 34 (1947), pp. 14–29.
7. Russell, Bertrand, "The Philosophy of Mathematics" (Allen and Unwin), 1929.
8. Taulbee, O. E., "On Properties of Classification Matrices," presented at the 69th Summer Meeting of the American Mathematical Society, University of Massachusetts, Amherst, Mass. (Aug. 25, 1964).

16

Leviathan, and Information Handling in Large Organizations*

BEATRICE AND SYDNEY ROME

System Development Corporation
Santa Monica, California

The authors have devoted the past few years to studying information handling in very large organizations. We do this by growing large organizations in our computer-based laboratory and performing experiments upon them. These laboratory organizations are combinations of live and artificial personnel. Today we shall focus on the information handling facets of our experimental method. We shall present some of our initial findings, or, more rigorously, initial interpretations of initial findings.

THE COMMUNICATION PROCESS IN LARGE ORGANIZATIONS

THE SCOPE OF THE COMMUNICATION PROCESS

First a word about the communication process as we view it. All of you will agree that information handling is more than issuing memoranda, disseminating documents, making telephone calls, filing papers. It is more than abstracting, indexing, digesting, card punching, photocopying, or shuffling electronic pulses through computers. All these modes of processing information are merely instrumentalities that serve a higher function. They serve as media to convey and develop meaning and intent between person and person, persons and groups, and groups and groups, in large organizations.

Communication, then, goes beyond mere data processing. It includes all formal and informal conversations. It is a succession of encounters and a continual stream of dialogue among multitudes of organisms. When these organisms communicate with one another, they buffet, challenge, sustain, cajole one another. They address to one another their hopes, anticipations, plans, schemes, knowledge, misinformation. They submit

*This paper is an enlargement of two talks presented at the Conference by the authors. Development of the theoretical aspects of the research described in this paper is being supported in part by the Air Force Office of Scientific Research (Information Sciences Directorate) of the Office of Aerospace Research, under contract No. AF 49(638)-1188.

or seek to dominate; they conceal and reveal, laugh and joke, impress and depress one another, persuade and threaten each other—in sum, they express, covertly or overtly, entire worlds of hopes, fears, tendencies, motives, attitudes, intentions.

THE SOCIALIZING FUNCTION OF THE COMMUNICATION PROCESS

Consequently, we can understand the communication process as that process through which individuals enter into social, value-laden relations with one another. This process fuses separate, often conflicting and antagonistic, individuals into the solidary groupings which in turn make up large organizations. Through communication, large organizations become real social beings. Communication assimilates the resources of a large organization into its organic social existence. By means of communication the organization, once born, continues to recreate itself and to sustain its own social existence. Through communication the organization acts, accomplishes its objectives, realizes its values, and exerts its power. And once a large organization comes into existence and continues to be, it provides, through its communication process, the internal social environments in which its members have status and roles, realize tactics, develop strategies, and cope with the larger environments in which the entire organization lives, moves and has its being.

When information handling is viewed in the present way as carrying out the life processes of large organizations, every document and every symbolic expression within it can have many levels of revealed and concealed meaning. We are not speaking of ambiguity. We are speaking of the fact that any significant piece of information is potentially a many-layered communication having values and consequences that impinge differently on different departments, levels and subsystems within large organizations. We are also speaking of the fact that every symbolic expression hides while it reveals. As the noted French sociologist, Georges Gurvitch, puts it: "Social symbols . . . characteristically reveal while veiling and veil while revealing, and while inspiring participation also restrain it."* And, we add, all this multifaceted impingement of informa-

**The Spectrum of Social Time (La multiplicité des temps sociaux)*, Dordrecht-Holland, The Netherlands, 1964, p. 2. On page 49 Gurvitch elaborates this thought as follows:

> Social symbols are signs which only partially express the contents toward which they are oriented. They serve as mediators between these contents and the collective and individual agents who formulate them and to whom they are addressed. This mediation consists of encouraging the participation of agents in the symbolized contents and these contents in the agents. Whether the symbols are mainly intellectual, emotional, or voluntary, whether they are tied to the mystic or the rational, one of their essential characteristics is that *they reveal while veiling and veil while revealing, and even while they encourage participation, they check it.* From this viewpoint, all the symbols, including the sexual symbols, constitute a way of overcoming and dealing with obstacles and impediments to expression and to participation. The symbols vary because of many

tion grows and unfolds in time, calling for constant reevaluation and reinterpretation by all participants at all levels of an organization.

In short: A large organization is a union of people, relating in myriad ways, grouping and regrouping ceaselessly, and constantly making and remaking its evolving history. Through its communication process, the organization creates and regenerates its ongoing power and sustains itself. At the same time, the communication process reveals and expresses the social vitality of the organization.

Thus, paradoxically, communication is the creative force that gives birth to and preserves the organization, and, in turn, it is the organization that gives birth to and sustains the communication. Communication expresses the organization and the organization is the expression of its communication.

A TAXONOMY OF THE COMMUNICATION PROCESS

Were one to construct a conceptual model or taxonomy of the communication process, one would have to take into account at least five essential elements:

(a) *Linguistic medium.* The linguistic or symbolic medium through which the members of the organization talk with one another.
(b) *Information feedback.* The information feedback that reports on system and subsystem performances.
(c) *Formal authority.* The structure of formal authority and its interrelation with the feedback system.
(d) *Charter.* The process through which the organization expresses and enforces its values, image, and mission—an active process that constantly renews, recreates and reaffirms the "organizational charter."*
(e) *Extraformal interaction:* The extraformal process of person-to-person interaction in an organization.

factors: particularly because of the character of the subject-broadcasters and the subject-receivers, because of the variable importance of the symbols and that [which] is symbolized; because of the various degrees of their crystallization and flexibility, etc. This is why the symbols constantly risk being overwhelmed, of being slower than that which they would symbolize. Only rarely are they adjusted for their task, so much so that at each turn we are tempted to speak of their "fatigue," if not of their "defeat."

*E. Wight Bakke, "Concept of the Social Organization," in Mason Haire (ed.), *Modern Organization Theory* (New York, 1959), chap. 2, pp. 37-39. *Cf.* Kenneth E. Boulding, *Image: Knowledge in Life and Society* (Ann Arbor, 1956). Bakke describes what he means by the term, organizational charter, as follows:

> In many relationships of participants and outsiders to a social organization, it is essential that those involved have an adequate image of the uniqueness and wholeness of the organization. It is essential that the organization as a whole mean something definite, that the name of the organization call to mind unique, identifying features. This image and its content we label the *Organizational Charter*. . . .

Clearly, the communication process in a large organization is a complex, fluidly developing, all-pervasive medium. Clearly, too, a medium of this magnitude cannot be completely observed in any actual organization that exists in the real world. No one information scientist, or group, or army of scientists, can fully survey and evaluate the full information flow in large organizations. Therefore, information scientists have used a variety of means for conducting such study. Good as these means are, they have all lacked one vital element in order to be truly scientific—the ability to conduct experiments and thereby to test hypotheses in a laboratory environment.

We have developed a unique and, we believe, a fairly comprehensive and fruitful method for performing such experiments. This method, which we call the Leviathan, is itself a complex instrument. We shall now describe some of the information-handling features of the Leviathan method. You will observe, as we proceed, how the five basic taxonomic elements of the communication process are incorporated in our method.

It is the conception held by participants in the organization of what the name of the organization stands for, together with their basic and shared values, which tend to justify and legitimize such identifying features. Efforts to maintain the integrity of the organization will be governed by what is necessary to actualize and perpetuate this image of unique wholeness. It is basically a set of ideas shared by the participants which may or may not be embodied in written documents....

Although it is the image of the unique *wholeness* of the organization, it is not by any means a *summation* of its parts. It is created by selecting, highlighting, and combining those elements which represent the *unique* whole character of the organization and to which uniqueness and wholeness all features of the organization and its operations tend to be oriented ....

The Organization Charter contains at least the following identifying features of the organization:

1. The name of the organization.
2. The function of the organization in relation to its environment and its participants.
3. The major goal or goals toward the realization of which the organization, through its system of activities, is expected by participants to employ its resources (including themselves).
4. The major policies related to the fulfilling of this function and the achievement of these major goals to which agents of the organization are committed.
5. The major characteristics of the reciprocal rights and obligations of the organization and its participants with respect to each other.
6. The major characteristics of the reciprocal rights and obligations with respect to each other of the organization, and people and organizations in the environment.
7. The significance of the organization for the self-realization of people and organizations inside and outside the organization in question.
8. The value premises legitimizing the function, goals, policies, rights and obligations, and significance for people inside and outside the organization.
9. The symbols used to clarify, focus attention upon, and reinforce the above, and to gain acceptance from people inside and outside the organization. These symbols are actually particular items of the several basic resources which serve as cues to bring to mind the content of the Organizational Charter and reinforce its hold upon the minds of both participants and outsiders.

COMMUNICATING THROUGH THE COMPUTER IN NATURAL ENGLISH: THE LINGUISTIC MEDIUM

COMPUTER-BASED LABORATORY

The Leviathan method, first of all, utilizes a large, computer-based laboratory (Fig. 1). An essential feature of this laboratory is its 24 separate stations at which individual subjects communicate independently and

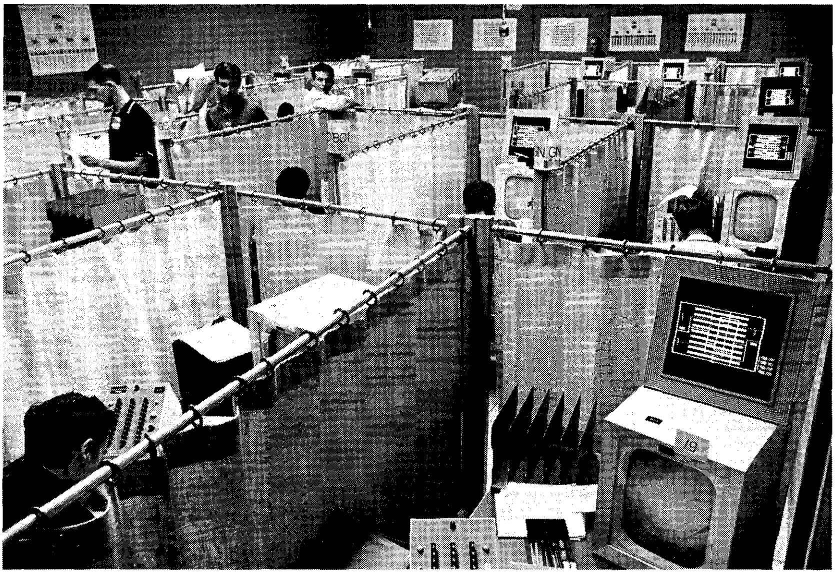


Figure 1. View of Leviathan Laboratory. Subjects in 21 booths enact roles of Officers in large information-handling organization.

directly with the computer in real time (Fig. 2). Each station contains a set of pushbuttons and a display scope. The pushbutton unit was especially designed for Leviathan experiments but has proved to have extremely wide practical and theoretical application. By means of these pushbutton units and displays, subjects communicate with each other through the computer. An example of a complete message follows: "Request approval to increase production rate to 999 at station A-1. Need maximum rate." (See Fig. 3.)

NATURAL LANGUAGE SETS

Note that the present message approximates natural English. It is one of a set of over three million well-formed sentences. This set of sentences

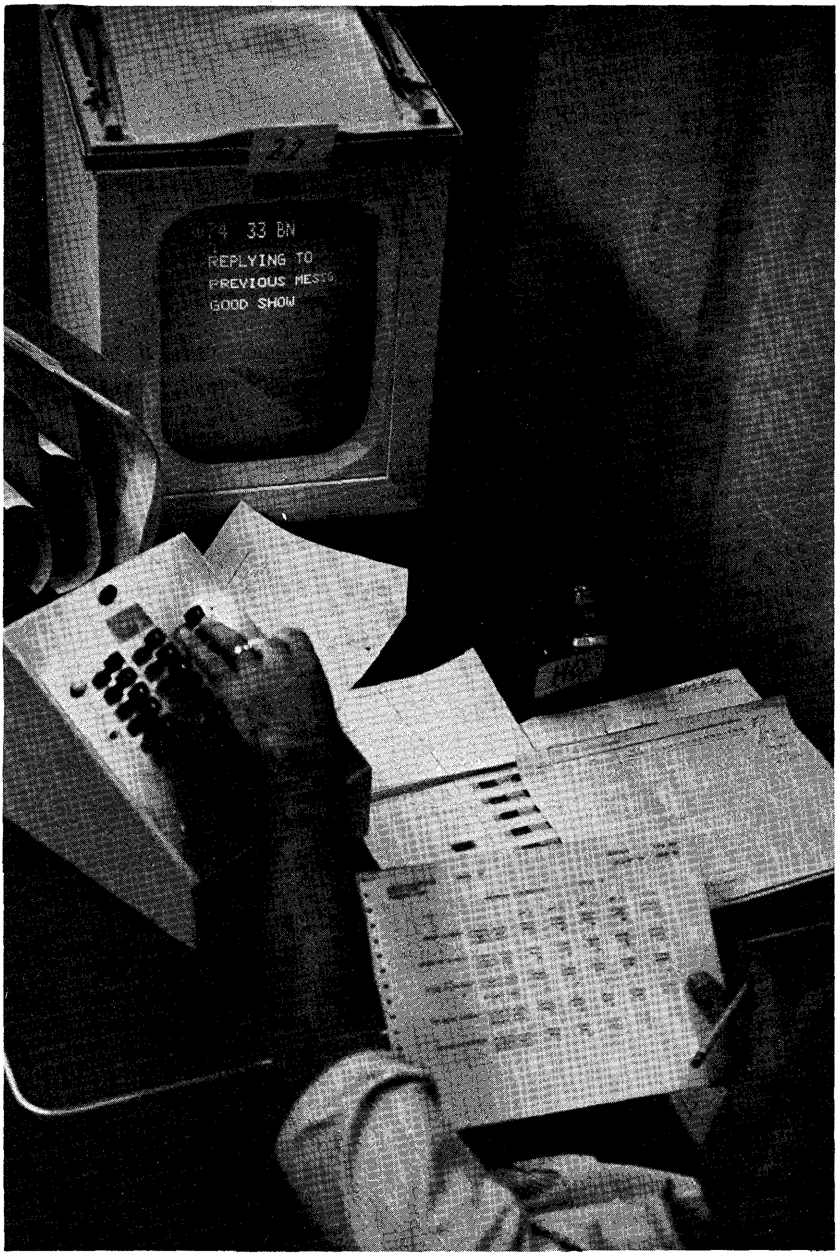


Figure 2. Subject in individual booth sending message over computer.

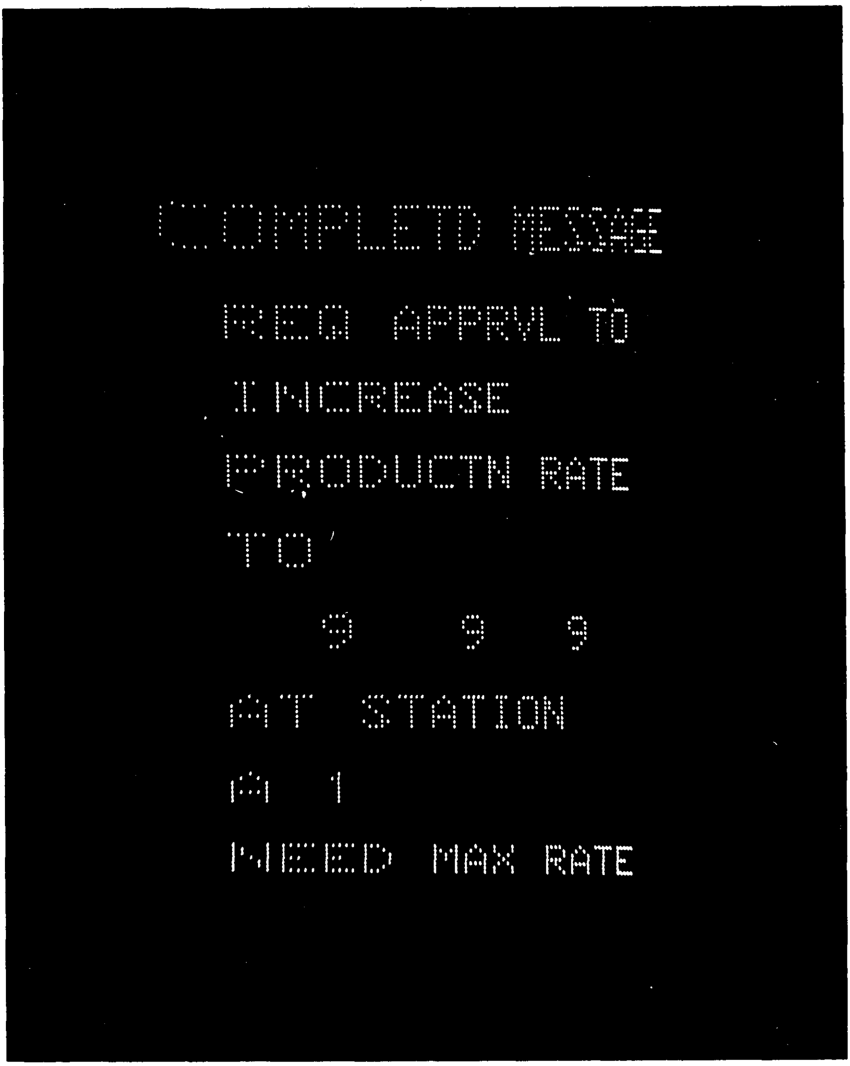


Figure 3. Example of completed message.

exists in the computer and is simultaneously and independently available to each individual subject. The entire set of sentences is a well-organized language. This language, moreover, can be varied from experiment to experiment without affecting the basic computer programs. In other words, the program system remains unchanged regardless of the variety of natural languages that can be imposed upon it. Any one language, or any version of a language, is supplied to the computer by the experimenters through a relatively small deck of several hundred IBM cards. And since new cards can be readily substituted, any one language can be grown and modified as the needs of the subjects become manifest to themselves or the experimenters.

COMPRESSED CODING

One great advantage of this language is its ease of mastery and use by the subjects. Another advantage is its extraordinarily compressed coding, for it is several times as efficient as any other existing means for communicating sentences over physical channels. The entire message shown above, for example, can be coded into and transmitted by less than two 48-bit words.* Furthermore, when our subjects compose these messages—which they can do faster than your eye can follow—they transmit at a rate which is equivalent to approximately three bits per second.† As a result of the extremely compressed coding, transmission of this language over physical channels can be very economical.

AUTOMATIC RECORDING

From the experimenters' perspective, the language has still another advantage, in that it provides an automatic record and analysis by the computer of the entire interactive communication process among the subjects. The computer records who talks to whom, at what levels of authority and domains of responsibility in the organization, the occasions and times when communications take place, the exact content of what is said, and the patterns in which the utterances succeed one another. Subjects use this computer-based language to manage and control a large-scale organization operated by hundreds of artificial employees. The language is also used by the live subjects to issue orders to the artificial personnel and to communicate to them the decision rules according to which the organization operates.‡

Using this language, the managers can also report information to one another over the computer. For example, a manager might compose the following message: "Reporting information on epoch 28. Value F is

*Except for special data such as "999" and "A-1."

†Actually the transmission takes place over parallel lines; the figure of approximately three bits per second is estimated for a single channel and optimal coding both of computer programs and hardware signals.

‡The language just described is a structured command or management language for directing, planning and operating a large organization in a laboratory. While its technical aspects have been perfected, its social elements are still being developed and refined. This is being accomplished by supplementing the computer-based language with handwritten messages and face-to-face debriefings. The latter are observed through one-way glass and recorded on sound tape, and subsequently transcribed.

being routed to line 3, to meet sender's demand." The message appears on the display scope in this way:

COMPLETED MESSAGE
REPORTING
INFO ON EPOCH
2 8
VALUE
F
ROUTED TO LINE
3
TO
MEET SNDR DMND

Simultaneously with displaying the message, the computer prints hard copy. The hard copy is delivered by courier to the sender of the message and to those to whom copies have been addressed. Any who wish can use the hard copy for their permanent records.

USE OF LANGUAGE TO REQUEST FEEDBACK

Finally, during a laboratory experiment, this same language enables the live managers to request various kinds of feedback information (which we call "indite"). This information is generated by the robots in the computer. An example of a request for feedback information is the following:

COMPLETED MESSAGE
REQUEST FOR
INFO ON EPOCH
4 2
SEND
STATION OUTAGE
INDITE TO
CO BL GM

This completed message contains a request made by an officer to the robots for information on operations that took place during the 42nd epoch or simulated day of laboratory operations. He is requesting that the information be sent to his commanding officer (CO), his branch leader (BL) and his fellow group head (GM). He is asking the robots to send these officers feedback information (indite) concerning station failures or outages.

The computer programs for the natural language that we have been describing are called the General Operator-Computer Interaction (GOCI)

Program System.* By means of this system of programs, the first prerequisite of a taxonomy of the communication process is realized: GOCI and the natural language superimposed on it represent the linguistic or symbolic medium through which the members of an organization talk with one another.†

THE LEVIATHAN INDITE SYSTEM: COMPUTER-GENERATED INFORMATION FEEDBACK

HIERARCHICALLY ORGANIZED

The information feedback that can be requested by means of the computer language is itself a major feature of our Leviathan method. It satisfies the second taxonomic prerequisite of the communication process —a feedback mechanism that reports on system and subsystem performance. An integrated system of computer programs, known as the Indite programs,‡ provides us with an extensive repertory of different kinds of feedback information. During the past two years of laboratory operations, we have given our subjects—on line and in real time—more than 20 different types of feedback. Each of these types is supplied in different forms to different subjects, according to their particular roles in a given experiment. We, the experimenters, specify which combinations are to be given to the subjects, depending on the design of the experiment. Almost all of the 20 types of feedback are aggregated to suit the various organizational levels of authority and responsibility.§ Each officer at each command level receives those abstracts of the total information store that are relevant to the particular offices which fall within his span of control.

EXPERIMENTALLY CONTROLLED

In a typical Leviathan experiment, the subjects simulated 21 distinct offices in a six-level hierarchy (Fig. 4). Each office had its own unique combination of authority level, functional specialty (or combination of specialties), and territorial domain. And each office received a distinct selection of appropriate feedback. More than 200 different reports were supplied to the subjects in a simulated day of operation, covered in 25 to 30 minutes of laboratory time. Thus our program system enables us to

*The GOCI programs were realized by Mildred Almquist.

†The handwritten messages and face-to-face debriefings complete the linguistic or symbolic medium in Leviathan experiments.

‡The Indite programs were realized by Robert E. Krouss.

§Ten of these different types are illustrated in Figs. 7–9 and 11–19. Figures 9 and 11 and 16–18 respectively show how two major types of feedback are aggregated at various levels of command.

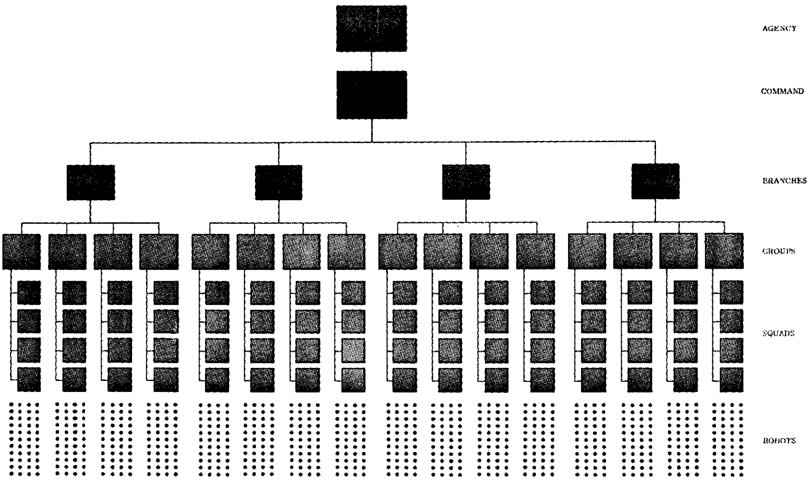


Figure 4. Six-level hierarchy in typical Leviathan experiment.

control the kinds and amounts of information that we supply our subjects. We also control by its means the rates, timing and patterns of information flow.

Clearly, as with large organizations in real life, our feedback system has been deliberately designed to *prohibit* any single officer in a command pyramid to form accurate, comprehensive and complete pictures of organizational performance on the basis of his own information alone. Each officer receives information relevant to the perspective of his office. If the officers as a group want systemic knowledge—if they want knowledge of organizational performance that is simultaneously relevant to all levels of authority, to all functional specialties, and to all theaters of operation—then they must work as a group to wrest this knowledge from the total corpus of feedback.

ABSTRACT AND GENERAL

One more point. The feedback programs, as all Leviathan programs, are very general and are independent of the particular interpretation that the experimenters choose to impose upon them. Hence the programs are amenable to staging any of a large variety of logistic simulations. We have been telling our subjects the myth or fable that they have been operating an intelligence communications control center embedded in a national intelligence agency. Equally feasible would be the myths of a military supply facility, a large public library, a personnel bureau, and so forth. Which myth we elect to use is a matter of convenience.

The only way that the myth affects the computer programs is to dictate some labels on some items. Changing the myth means only that, on the computer displays and the feedback printouts, some items are called by one name rather than some other. The computer programs are not otherwise affected by the change of myths. They stand above all myths.

This neutrality of the program system is of the first importance when we come to interpret our experimental results. We are not tied to the particular system we stage in the laboratory. Our subjects do operate a perfectly concrete system that has a proper name, has a setting in the real world, and operates in real time. But our experimental results do not have to be tied to that system in that setting at that time. Because the underlying computer program system operates according to abstract, general principles, experimental investigations that use the system can provide abstract and general results.

We have brought to your attention three basic features of our computer-based feedback system. The feedback flows in a setting that can be interpreted abstractly and generally. The feedback is elaborately hierarchical in character. We exercise complete experimental control over the composition of the feedback, its allocation, and its flow. These three features have enabled us to make information handling in large organizations a major area of investigation in the overall Leviathan program of experimental research.

THE INTERRELATION OF INFORMATION FEEDBACK AND FORMAL ORGANIZATIONAL STRUCTURE

One important problem that we have begun to investigate in our laboratory is the interrelation of formal organizational structure and the process of information feedback. This relationship constitutes our third taxonomic requirement for modelling the communication process. Thus far we have realized four basic varieties of formal organizations in our laboratory. Many other varieties can also be realized with the present system of Leviathan computer programs. Hence, formal organization is a parameter with respect to the program system. Of special interest here are the transients—what are the effects on feedback and system performance of radical shifts in formal organization?

The four types of formal organization that we have realized are shown in Fig. 5. Each circular symbol in each type of organization represents an office staffed by at least one live subject. In all four configurations, there are 16 live group heads (level III) reporting to four live branch

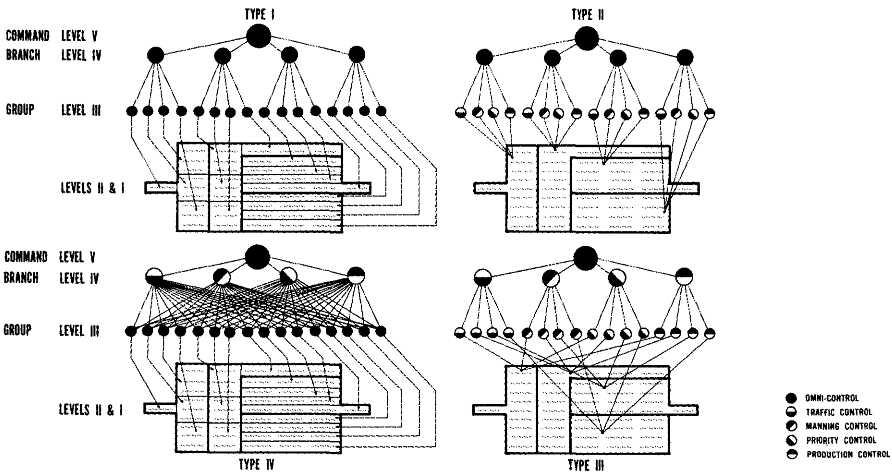


Figure 5. Four types of formal organization realized in Leviathan experiments.

leaders. The branch leaders (level IV) in turn report to a single commanding officer (level V). The commanding officer reports to his next-higher echelon, the superordinate embedding organization, not shown in this figure. In each of the four types of organization, underneath the live officers we show their territorial jurisdiction. In all four types of organization, 64 squads of robots are distributed over this territory (level II) and report to the group heads directly over them. Each squad of robots consists of artificial enlisted men (level I) who exist in the computer.

In 1963, three types of formal organization were realized by one group of subjects operating over a three-month period in two four-hour laboratory sessions per week. These types were, successively, II, I, and IV. In the spring of 1964, a different group of subjects operated type II organization twice a week over a three-month period. Approximately half of this group was then (summer of 1964) replaced by newcomers, and the new organism simulated types III and I successively, operating three times a week over an eight-week period.

NON-SPECIALIZED ORGANIZATIONAL STRUCTURE: TERRITORIAL DOMINION AND RANK DETERMINE FEEDBACK DISTRIBUTION

In all four configurations, four functional specialties are exercised. As shown in Fig. 5, these are traffic control, manpower control, priority control and production control.

In the organization labeled type I, all live officers on all three levels exercise all four functions. Each is an omnispecialist or, if you will, a nonspecialist. What criteria determine the distribution of feedback information to the various officers?

Clearly, in this type of organization, no distinction can be made on the basis of functional specialty alone. All officers are, or should be, concerned with the feedback reports for all the four functions. Territorial dominion, however, does decide who gets what part of the total information feedback. The theater of logistic operations is apportioned among the 16 group leaders. Each group leader has primary responsibility for his own proper and unique territorial domain, as shown in Fig. 5 for type I organization. He also has primary responsibility for the particular robots assigned to him and to his territory. Therefore, the interest of each group head in the total body of feedback is structured and circumscribed by, and centered on, his specific territory and on the artificial personnel assigned to him and to his territory.

Another basis for deciding who gets what information in the present type of organization is an officer's rank or level. While group heads have mainly disaggregated and localized interests, branch heads enjoy a larger perspective. Their theaters of operations and spans of authority overarch and combine those of their group heads. Their larger perspectives, however, do not necessarily imply that branch heads are interested in simply more of the information which group heads receive. Branch heads may have qualitatively different concerns and responsibilities and, therefore, different feedback interests. Hence, branch feedback may be far less detailed but far more integrated than group feedback.

The commanding officer, of course, has an all-inclusive and all-comprehensive interest concerning the entire organization. But this, perforce, places his interest on an even higher level of integration and necessitates a form of feedback commensurate with his all-inclusive commitment.

We can sum up feedback requirements for type I organization thus: Territory and rank determine who receives what feedback. Functional specialty does not count.

SPECIALIZED ORGANIZATIONAL STRUCTURE: PROFESSIONAL SPECIALIZATION AND RANK DETERMINE FEEDBACK DISTRIBUTION

Consider now the extreme opposite of type I organization, namely, type III. How is feedback affected by this type of formal organization? Here the commanding officer still retains the same breadth and scope of interest as in type I. But the branch heads and group heads no longer have exclusive territory. Each branch head now has exactly the same territorial dominion as every other branch head, and, indeed, as the commander himself. Now professional specialization dominates and differentiates branch interest. Each branch head is a specialist and shares his specialty with his four group heads. Necessarily, the entire system of feedback will feel and operate quite differently in this type of organization.

On the group level, territory is now divided four rather than 16 ways. But within each territorial quadrant, control is now no longer exclusive with a group. Four different group heads, each of whom represents a different branch, act in concert within each territorial quadrant. Consequently, in a quadrant, a group head need no longer have interest in the same kinds of feedback information as do his three colleagues in that quadrant. Each has radically distinct commitments, and his interests tend to follow his specialty.

As we ascend to the branch level, territory ceases to count, as we have seen. Thus the present type of organization looks like four autonomous empires, all trying simultaneously to command the same theater of logistic operations, but each employing means and information qualitatively different from all the others.

HYBRID ORGANIZATIONAL STRUCTURES: PROFESSIONAL SPECIALIZATION, TERRITORIAL DOMINION, AND RANK DETERMINE FEEDBACK DISTRIBUTION

Type II and type IV organizations are hybrid rather than pure types. In both, functional specialization and territory both play important roles. On the group level, type IV is like type I—nonspecialized; type II is like type III—specialized. On the branch level, type IV is like type III—specialized; and type II is like type I—nonspecialized.

In type IV organization, the line of command is broken at the branch level and flows differently in the different branches. Type IV organization calls, far more, for information relative to a leaderless or committee or bureaucratically decentralized operation. Type II organization, on the other hand, places territorial autonomy at the branch level. Feedback interests on the branch level are now integrative across the professional specialities while divisive geographically.

We have been focusing on the interrelationship between formal organizational structure and the feedback process. As we compared one type of organization with another, we saw feedback requirements can differ greatly in the different organizations. We shall now examine the kinds of feedback we supplied our subjects during our 1963 and just-completed 1964 experiments.

THE INDITE FEEDBACK FOR THE 1963 AND FIRST 1964 EXPERIMENTS

Throughout our 1963 and 1964 experiments, the subjects operated a logistic processing system consisting of a single initial receiving station, nine parallel traffic lines, and a common exit station. Along each processing line was a cascade of processing stations (see Fig. 6). The feedback which the subjects received in these experiments obviously reflected not

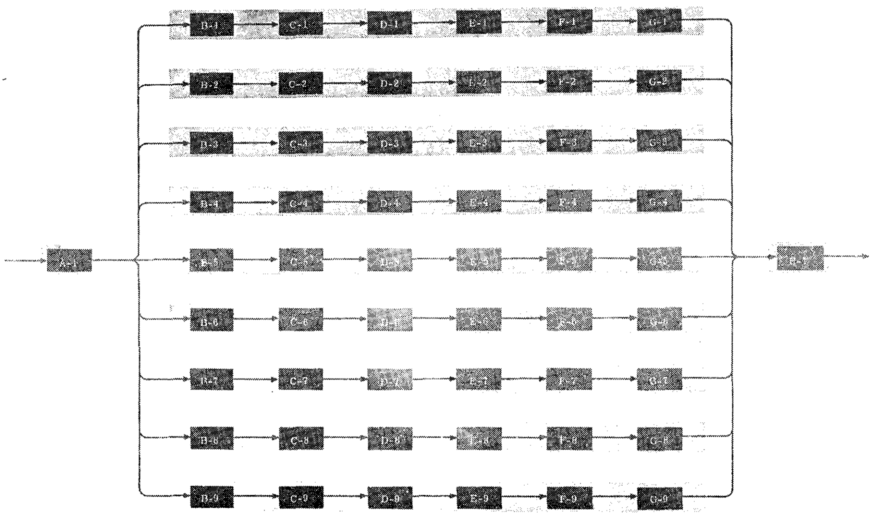


Figure 6. Leviathan logistic processing system.

only the structure of the processing system common to all the experiments, but also the particular type of formal organizational configuration that was enacted in a particular experiment. For convenience, we shall treat with just one of the four types in describing the Indite feedback system. This is type II which was used in our first 1963 and first 1964 experiments.

SYSTEM PERFORMANCE FEEDBACK

We gave the subjects 12 basic kinds of feedback information. The first related to the total productivity of the organization. As shown in the example of Fig. 7,* this feedback reports the total number of units (267) processed by the entire system in an epoch of time—in a simulated day

*Figures exhibiting Leviathan feedback (Figs. 7–9 and 11–19) are not confined to the first 1963 and first 1964 experiments; but,'for illustrative purposes, they are taken from a variety of experiments performed with the 1963 group and from one experiment performed with the 1964 subjects.

LEVIATHAN-INDITE EPOCH 25 08-29-63 PAGE 502
EXPERIMENTAL RUN - 305A
SESSION - 5A DELIVER TO CO

UNITS THROUGH SYSTEM COMMAND CO

LINE	PRTY 1	PRTY 2	PRTY 3	PRTY 4	TOTAL
1	7	7	13	0	27
2	3	11	15	0	29
3	12	6	9	0	27
4	18	9	4	0	31
5	3	8	15	0	26
6	2	12	14	0	26
7	6	4	12	0	22
8	6	4	26	0	36
9	0	10	31	0	41
TOTAL	57	71	139	0	267

Figure 7. System performance feedback: units through system.

(epoch 25). It breaks these down by the priority treatment accorded these units by the priority controlling group heads as the units passed over the traffic lines (57 were accorded highest priority, 71 priority 2, 139 priority 3, none was given the lowest priority). Notice that the information is also broken down more finely by individual traffic lines. This feedback report was distributed every epoch to the commanding officer. It was also received by the branch leader within whose territory fell the exit station and by the group head in charge of production in this exit branch.

The commanding officer and the same exit branch head also received another report that covered average transit time, that is, how long, on the average, it took the units of work to pass through the processing system (Fig. 8). This report was also supplied to the priority controller in the exit branch. Notice that this report is also broken down by priority history of the units and by the lines traversed.

FEEDBACK ON PERFORMANCE AT COMPONENT STATIONS —ACCORDING TO PRIORITY TREATMENT

These two kinds of feedback just presented—units through system, time through system—provide information on the system level. Finer-grain feedback, relating to specific locations within the processing system, was also supplied on the group level.

In Fig. 9 we see an example of information provided to the production controllers. It shows, for each station and each squad of robots, how many units were processed by the robots in a given epoch. Since our myth was that our subjects were managing an intelligence communications control center, the work units were construed as intelligence communiques. Group GN received information relative to squads R-1, R-2, R-3 and R-4. Figure 10 shows how the branches, groups and squads were located over the processing lines in this experiment. Group GN was responsible for production control in branch BL's territory which is lo-

LEVIATHAN-INDITE EPOCH 25 08-29-63 PAGE 494
EXPERIMENTAL RUN - 305A
SESSION - 5A DELIVER TO CO

TIME THROUGH SYSTEM COMMAND CO

LINE	PRTY 1	PRTY 2	PRTY 3	PRTY 4	TOTAL
1	17	17	17	0	17
2	11	11	11	0	11
3	13	13	13	0	13
4	15	15	15	0	15
5	11	12	11	0	11
6	20	20	19	0	20
7	14	15	14	0	14
8	19	19	19	0	19
9	0	17	18	0	17
TOTAL	15	15	16	0	15

Figure 8. System performance feedback: time through system.

LEVIATHAN-INDITE EPOCH 19 07-11-63 PAGE 539
EXPERIMENTAL RUN - 301A
SESSION - 7A DELIVER TO GN

NUMBER OF MESSAGES PROCESSED BY PRIORITY GROUP GN

SQUAD	STATION	PRTY 1	PRTY 2	PRTY 3	PRTY 4	TOTAL
R1	C1	0	0	14	0	14
	C2	0	0	18	0	18
SUBLEVEL TOTAL		0	0	32	0	32
R2	C3	0	0	12	0	12
	C4	0	0	17	0	17
	D3	0	0	12	0	12
SUBLEVEL TOTAL		0	0	41	0	41
R3	D1	0	0	14	0	14
	D2	0	0	34	0	34
	E1	0	0	14	0	14
SUBLEVEL TOTAL		0	0	62	0	62
R4	D4	0	0	17	0	17
	D5	0	0	8	0	8
	D6	0	0	25	0	25
SUBLEVEL TOTAL		0	0	50	0	50
GROUP TOTAL		0	0	185	0	185

Figure 9. Number of messages processed, by priority.

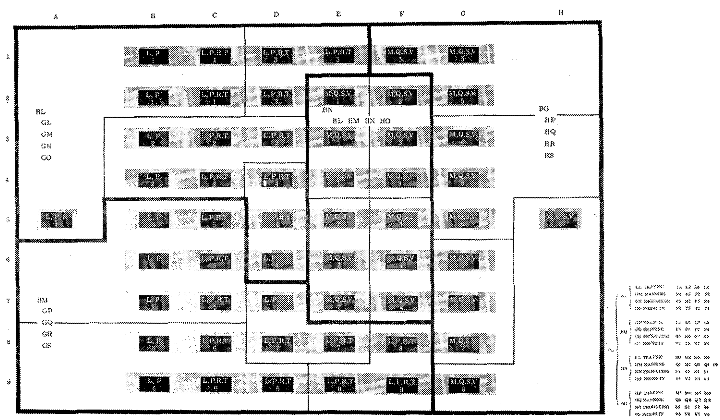


Figure 10. Branch, group, and squad Locations on processing lines.

cated in the northwest region of the processing system. In this territory, GN's squads are located as follows:

Squad	Line	Stations
R-1	1	C
	2	C
R-2	3	C, D
	4	C
R-3	1	D, E
	2	D
R-4	4	D
	5	D
	6	D

This type of information on the number of messages processed was supplied to the branch leaders and to the commander in aggregated forms suited to their respective territories and levels of command (see Fig. 11).

At each processing station along each of the traffic lines, there is a waiting queue, where the units of work (intelligence communiqués) wait until they can be processed at that station. Feedback reports were, accordingly,

LEVIATHAN-INDITE EPOCH 5 07-24-63 PAGE 106
EXPERIMENTAL RUN - 303A
SESSION - 1A DELIVER TO BN

NUMBER OF MESSAGES PROCESSED BY PRIORITY BRANCH BN

	PRTY 1	PRTY 2	PRTY 3	PRTY 4	TOTAL
GROUP HL	0	0	52	0	52
GROUP HM	0	0	46	0	46
GROUP HN	0	0	47	0	47
GROUP HO	0	0	47	0	47
BRANCH TOTAL	0	0	192	0	192

LEVIATHAN-INDITE EPOCH 5 07-24-63 PAGE 108
EXPERIMENTAL RUN - 303A
SESSION - 1A DELIVER TO CO

NUMBER OF MESSAGES PROCESSED BY PRIORITY COMMAND CO

	PRTY 1	PRTY 2	PRTY 3	PRTY 4	TOTAL
BRANCH BL	0	0	238	0	238
BRANCH BM	0	0	106	0	106
BRANCH BN	0	0	192	0	192
BRANCH BO	0	0	343	0	343
COMMAND TOTAL	0	0	879	0	879

Figure 11. Number of messages processed, aggregated at branch and command levels.

supplied to the priority and traffic controllers concerning how many units of work, on the average, stood waiting in the queue of each station (queue occupancy) and how long, on the average, these units had to wait before being processed at that station (delay time—see Fig. 12). These data were

LEVIATHAN-INDITE EPOCH 21 08-06-63 PAGE 46
EXPERIMENTAL RUN - 304A
SESSION - 4A DELIVER TO GL

AVERAGE DELAY TIME OF MESSAGES, BY PRIORITY GROUP GL

	QUEUE	PRTY 1	PRTY 2	PRTY 3	PRTY 4	TOTAL
SQUAD T1	Q-C1	0	0	1	0	1
	Q-C2	0	0	4	0	4
SUBLEVEL TOTAL		0	0	4	0	4
TOTAL		0	0	4	0	4

Figure 12. Message delay times, by priority.

broken down and presented in formats similar to the one shown in Fig. 9 for units processed. These queue occupancy and delay data were in turn aggregated at the branch and command levels.

FEEDBACK ON PERFORMANCE AT COMPONENT STATIONS —ACCORDING TO TYPES OF WORK UNITS

Thus far we have seen that the production, traffic and priority managers received detailed information for each station and squad within their respective territorial domains concerning the number of units processed, the number standing in queues, and the average delays at queues. All this information was broken down by priorities and aggregated for higher echelons. Besides these feedback reports, still other reports were given to the subjects in which the very same information was also broken down in another way, namely, by type (Fig. 13). At each station, the traffic managers stipulated one of eight different types according to which the robots would analyze the feedback information.* Because, in our myth, the units of work were said to be intelligence communiqués, their type classifications were accordingly interpreted to be the following: subject matter of the communiqués, source, area of origination, precedence treatment requested by sender, evaluations of source and of quality of infor-

*Type classification has fundamental importance in the Leviathan computer program system. It constitutes the device by which the live officers stipulate decision rules to the robots, and the robots, by using it, implement the decision rules on a contingent basis. This mechanism is described in our paper, *Communication and Large Organizations,* currently available from the System Development Corporation, Santa Monica, California, as SP-1690/000/00. The paper appears in somewhat compressed form in the December 1964 issue of *IEEE Spectrum,* published by the Institute of Electrical and Electronic Engineers.

LEVIATHAN-INDITE EPOCH 27 08-29-63 PAGE 714
EXPERIMENTAL RUN - 305A
SESSION - 5A DELIVER TO GL

NUMBER OF MESSAGES PROCESSED BY TYPE GROUP GL

SQUAD R1

STATION A1	W EUROPE	69	E EUROPE	63		
	N AMERICA	31	L AMERICA	20		
	AFRICA	11	NEAREAST	43		
	NCOMASIA	41	COM ASIA	20		
	POLAR	2			TOTAL	300
STATION C1	EXECUTVE	4	C I A	8		
	STATE	3	U S A F	2		
	US ARMY	4	US NAVY	5		
	N A S A	1	A E C	4		
	SPL COM	0			TOTAL	31
STATION C2	EXECUTVE	2	C I A	6		
	STATE	3	U S A F	5		
	US ARMY	6	US NAVY	6		
	N A S A	0	A E C	3		
	SPL COM	0			TOTAL	31
SUBLEVEL TOTAL	362					

Figure 13. Number of messages processed, by type.

mation contained in the communiqué, etc. In the example shown in Fig. 13, the manager had stipulated that at station A-1 the type classification was to be geographical area from which the communiqués, originated. At stations C-1 and C-2, the addressees of the communiqués were stipulated to be the basis of feedback analysis.

The feedback reports analyzed by type, like those previously described that were analyzed by priority, also covered each station and squad within the territory of each officer and again reported the number of units processed, the number standing in queues, and the average delays at queues. These reports by type likewise were aggregated at the branch and command levels.

FEEDBACK ON PERFORMANCE AT COMPONENT STATIONS—FAILURE REPORTS

When we stop to consider all these fine-grain feedback reports, namely, those on station productivity, queue occupancy and delay information, analyzed by priority and again by types of units, we see that they comprise a comprehensive but complex set of feedback reports on all levels of the management pyramid. Our subjects were receiving a great deal of information every 30 minutes. They could, however, turn to a different sort of report whenever they needed to form a more simplified feedback picture. This was the failure report.

In those cases where production requirements were so high that processing stations failed to meet the processing quotas set by the production managers, failure reports were provided, as shown in the example of Fig. 14. Stations that failed and time of failure (time unit is the "scan") were shown. These data were aggregated at the branch and command levels. Similar reports covered queue blockages at all the individual processing

```
LEVIATHAN-INDITE        EPOCH 27                                   08-29-63        PAGE 708
EXPERIMENTAL RUN - 305A
SESSION - 5A                                                                       DELIVER TO HO

     OUTAGE OF STATIONS              GROUP HO

     STATION              SCAN
       S-F5               160
       S-F6               160
       S-F7               160
```

Figure 14. Failure report: station outage.

```
LEVIATHAN-INDITE             EPOCH 27                              08-29-63        PAGE 705
EXPERIMENTAL RUN - 305A
SESSION - 5A                                                                       DELIVER TO HO

     QUEUE BLOCKAGE TIME                      GROUP HO

                           QUEUE     SCAN BLOCKED     SCAN UNBLOCKED
         SQUAD M4          Q-F6          156               158
                           Q-F6          158               159
                           Q-F6          159               160
                 AVERAGE TIME QUEUE BLOCKED FOR QUEUE        1
                           Q-F7          154               158
                           Q-F7          158               159
                           Q-F7          159               160
                 AVERAGE TIME QUEUE BLOCKED FOR QUEUE        2
                 AVERAGE TIME QUEUE BLOCKED FOR SQUAD        1
                 AVERAGE TIME QUEUE BLOCKED FOR GROUP        1
```

Figure 15. Failure report: queue blockage.

stations (see Fig. 15). These show which queues were blocked, when they were blocked and when they became unblocked.

FEEDBACK ON UTILIZATION OF MANPOWER RESOURCES

We have now covered ten kinds of feedback information supplied to the managers on-line and in real-time. These all served to measure the accomplishments of the processing system at its component and system levels. A totally different kind of information related to the resources available to the system. It measured the degree to which the managers were utilizing the productive energy supplied by the artificial personnel.

In Fig. 16 we see an example of the first page of the report supplied to branch BL's manpower officer, GM. Each unit of energy supplied by a robot is called a taylor, after Fred Taylor, who, together with his stop-watch, flourished in Pittsburgh half a century ago. Taylors come in four kinds, reflecting the fact that Leviathan robots have four kinds of aptitudes. (We tell our subjects that aptitude 1 is manual, 2 is linguistic, 3 is arithmetic, and 4 is logico-analytic.) Taylors available, taylors utilized, and per cent utilization are shown for each aptitude and for all four aptitudes, at the group level and at the level of each squad in BL's branch.

LEVIATHAN-INDITE EPOCH 5 07-24-63 PAGE 51
EXPERIMENTAL RUN - 303 A
SESSION - 1A DELIVER TO GL

MANPOWER UTILIZATION GROUP GL SUMMARY

	APT 1	APT 2	APT 3	APT 4	TOTAL
TAYLORS AVAIL	147000	124750	95750	79250	446750
TAYLORS USED	26781	26911	19271	34786	107749
PERCENT USED	18	21	20	43	24

SQUAD L1

	APT 1	APT 2	APT 3	APT 4	TOTAL
TAYLORS AVAIL	34000	34500	25000	24000	117500
TAYLORS USED	2548	9645	6101	11829	3012
PERCENT USED	07	27	24	49	

SQUAD P1

	APT 1	APT 2	APT 3
TAYLORS AVAIL	39750	27500	142
TAYLORS USED	12769	6622	
PERCENT USED	32		

Figure 16. Manpower utilization report.

Figure 17 shows these data aggregated at the branch and command levels. At these levels, fine-grained data are merged, and therefore concealed, while high-level data are revealed, just as Gurvitch describes to be the case for social symbols. Thus the group heads received very concrete detailed feedback not available on the branch and command levels. And branch heads received summaries not available in the commanding officer's feedback. When higher-level officers had need for the finer-grained data, they had to rely on their subordinates to furnish these, assuming the subordinates were willing to do so.

The husbandry of manpower resources was also monitored by a failure report (Fig. 18). This report simply showed, on the group level, which

LEVIATHAN-INDITE EPOCH 15 07-25-63 PAGE 446
EXPERIMENTAL RUN - 303A
SESSION - 2A DELIVER TO BL

MANPOWER UTILIZATION BRANCH BL SUMMARY

	APT 1	APT 2	APT 3	APT 4	TOTAL
TAYLORS AVAIL	376750	312500	224250	232750	1146250
TAYLORS USED	60420	60645	49042	88060	258167
PERCENT USED	16	19	21	37	22

LEVIATHAN-INDITE EPOCH 15 07-25-63 PAGE 450
EXPERIMENTAL RUN - 303A
SESSION - 2A DELIVER TO CO

MANPOWER UTILIZATION COMMAND CO SUMMARY

	APT 1	APT 2	APT 3	APT 4	TOTAL
TAYLORS AVAIL	1714750	1415500	1066250	1089000	5285500
TAYLORS USED	310352	352701	297283	413071	1373407
PERCENT USED	18	24	27	37	25

Figure 17. Manpower utilization, aggregated at branch and command levels.

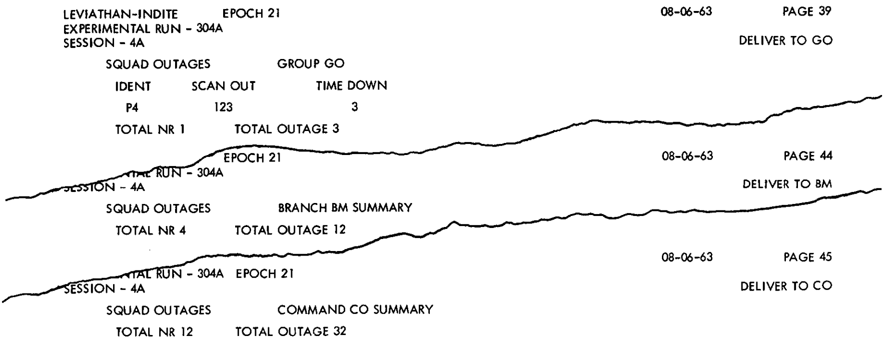


Figure 18. Failure report: squad outages, aggregated at group, branch, and command levels.

quads failed to provide sufficient robot energy to perform their assigned tasks and when in simulated time this failure occurred. As shown, these data were merged and aggregated at the branch and command levels.

INITIAL FINDINGS ON INFORMATION HANDLING IN LARGE ORGANIZATIONS

We shall now discuss some of our initial findings concerning the information-handling process in large organizations. To reiterate, these are preliminary results—initial interpretations of initial findings.*

INTERRELATION BETWEEN CHARTER COMMUNICATION AND FEEDBACK

Our earliest experiment in 1963 taught us that the feedback system, however extensive, accurate and well designed, is by itself not adequate for achieving effective and efficient management of a large organization. Our evidence seems to indicate that another process of communication is indispensable to the feedback process and may even be logically anterior to it. This other process is systemic communication or, better put, charter communication.

Charter communication is a telic and normative process—it orients the component individuals of an organization to their place in the organization's total systemic effort. It stipulates or presents what Bakke calls the

*For a more detailed discussion, see *Communication and Large Organizations*, previously cited, pp. 32ff.

organizational charter or image.* It sets the scene, identifies the unique wholeness of the organization, provides the system point of view, and defines the policies towards which an organization does and should aspire. Charter formation, development and renewal is the fourth of the taxonomic elements that we believe are essential to the communication process.

In our first 1963 experiment we gave our subjects the entire corpus of indite feedback previously described. Each officer received information fitted to his level of command, his functional responsibilities and his territory. This information was continually updated as the experiment proceeded. But the subjects were not antecedently instructed concerning how to use this information or to what use to put it. Nor were they instructed on their managerial roles in the intelligence center that they were to operate or on the goals, policies and missions of their center. In short, they were given almost no instruction or restriction concerning the organizational charter or image that they could accord to their center.

System performance was initially disastrously low and continued to fall until nearly extinct. We allowed the operation to continue long enough for clear trends to develop at all component offices and on the system level. Then we intervened in three ways: (a) The entire group was assembled together in a face-to-face debriefing, conducted by our debriefing officer. (b) We presented to the subjects exactly the same Indite feedback information they had heretofore been receiving epoch by epoch, but now it was shown in the form of trends over epochs and aggregated to the level of total system performance. (c) Finally, through our officer, the subjects were given a single instruction—to take the system point of view.

Following this debriefing meeting, the subjects resumed operation of their system. The computer was subsequently used to compare their actual performance following the debriefing with what it would have been, had the policies and decision rules in force just prior to the debriefing been frozen and preserved unchanged. Actual performance showed over 300 per cent improvement.

We infer that hierarchically structured feedback can be used simultaneously in different ways by a large organization. (a) It can be used to supply data upon the basis of which component problems can be solved on component levels of the organization. And it can even be used in such a way that these many alternative, and even conflicting, component problems can have fairly good solutions, yet these solutions can fail to contribute to the welfare of the total system. (b) An identical corpus of

*See footnote to fourth taxonomic element, p. 163.

hierarchically structured and distributed feedback can be used to achieve more efficient and more effective system performance.

Thus our first 1963 experiment indicates that an identical body of feedback can be used in two mutually inconsistent ways—it can be used to solve component problems at the expense of system performance, or it can be used simultaneously to help solve *both* component and system problems, with neither system nor component solutions completely excluding one another.

What permits an organization to use its feedback on both the component and system levels concurrently? The present experiment seems to show (a) that accurate, comprehensive and hierarchically structured and distributed feedback is not sufficient to guarantee good system performance in a large organization, (b) but when the organization adopts a community of interest and a sense of common system, mission or charter, then it can constructively use its hierarchically structured feedback simultaneously on its component, intermediate and overall system levels.

These conclusions are open to challenge as long as they are based on this experiment alone. One might argue, for example, that a major source of the variance in the subjects' performance before and after the debriefing might have been their learning to optimize on component levels. Was it not possible that, as a result of the briefing, each learned to perform his component task better but continued to ignore the system perspective? And did not the better system performance result simply from the better local performances? Clearly, to strengthen our interpretation of the present experiment, it was necessary to negate and exclude in turn alternative interpretations by conducting an ordered series of additional experiments.* Accordingly, four such experiments were subsequently performed, with the same group of subjects. These did help to rule out alternative hypotheses and thus helped to confirm the inferences concerning the interaction of charter and feedback that we drew from our first experiment.

VARIABLES TO WHICH LARGE ORGANIZATIONS ARE ESPECIALLY SENSITIVE

Thus far we have performed two series of experiments: one series of five experiments in 1963 with a fixed group of subjects and a series of five in 1964 with an evolving group. The two series seem to point to a class of variables to which large, information-handling organizations are especially sensitive. The variables of this class are intrinsically related to values, norms, orientation, mission and charter-development.

*This ordered series of experiments is reported in the document referenced previously: *Communication and Large Organizations*, pp. 61 ff.

Large organizations, we are finding, realign their extraformal coordinating communication processes when their members perceive themselves in system contexts and adopt system goals. They seem to do this no matter which formal structure is imposed upon them. To be sure, different formal configurations do proceed with different gaits: Different formal configurations do call forth different detailed procedures, different specific assignments and acceptances of responsibility, and different kinds and channels of reporting. But formal configurations and their characteristic gaitings seem less important to overall system performance than do the normative elements—the content of the organizational charter and its degree of acceptance by the members of the organization. The 1963 and the 1964 groups of subjects have, between them, operated all four of the types of formal structures shown above in Fig. 5. In every case, system performance improved as the organizations adapted their interpersonal communications and information-sharing procedures to system goals. When their leadership took this system point of view, performance improved; when leadership fought it or sought other, component objectives, performance declined or remained unimproved.

In sum, the charter or image (fourth taxonomic element) that an organization adopts seems intimately (a) to affect how it translates its formal structure of authority (third taxonomic element) into its extraformal processes of interaction (fifth taxonomic element), and (b) to affect its consequent record of system performance.

THE TEACHING MACHINE FOR INDOCTRINATING SUBJECTS IN HIGH-LEVEL MANAGERIAL ROLES

On the basis of our results concerning the interrelationships of the feedback process and the normative, systemic process of charter communication, we formally incorporated the telic process into an indoctrination or teaching machine. This machine we used to initiate our 1964 series of experiments. It consists of a sequence of briefings, presented to the subjects over the computer. The subjects enter into private dialogues with the computer which instructs them, each at his own pace and according to his special interests. The computer explains to the subjects their roles, the extent of their authority and power, the type and mission of the organization they are to manage, their resources, and the managerial controls at their disposal.

The results of employing this telic machine for indoctrinating subjects who are to enact the roles of high-level executives can be stated in a single sentence: Right from the outset, the 1964 organizations performed better than any of the organizations brought to life in 1963. This superior achievement, moreover, was manifest in every one of our measures of system performance.

POSITIVE AND NEGATIVE FEEDBACK—REPORTING BY EXCEPTION AND MANAGING BY EXCEPTION

Another set of interesting preliminary findings on information handling relates to positive and negative feedback in large information-handling systems. The subjects in the earlier 1963 experiments were essentially a close-tracking control group. They were content to exercise negative feedback, to attempt to prevent their organizations from deteriorating. These officers, accordingly, depended heavily on the three kinds of failure reports or reports by exception that we supplied them (see Figs. 14, 15, and 18). In several experiments we actually deprived them, most of the time, of all detailed quantitative feedback information on the component (station and squad) levels. Yet when they operated primarily with failure or exception reports, their performance improved.

We obtained very different results with the very first organization operated by the 1964 group. With this organization, we instituted reporting and managing by exception after the organization had built up a great backlog of unprocessed units of work and then gradually reduced this backlog almost completely. At this juncture, just when it had worked off a mountain of backlog, the group was experiencing almost no failures. By this time, moreover, it had a sustained history of developing long-range objectives and contingency plans for meeting these objectives. In short, this group was subordinating close tracking or negative control to positive, innovative behavior. When we deprived this group of positive detailed quantitative component feedback, they responded with resistance, protest, and expressions of discouragement. Our records tend to sustain the conclusion that when a group is planning positively and creatively, it needs information of another order of magnitude in amount than is provided by simple failure information. Control, on the other hand, requires far simpler and far less feedback.

EVALUATIVE FEEDBACK

In the second of the five 1964 experiments, we introduced a new category of feedback information. This information was presented by the computer in evaluated and interpreted form. In Fig. 19 we show an example of the system performance feedback given to the commanding officer. First is shown the total number of units processed. Next, five classifications of urgency or system importance are listed. Urgency 1 is the highest and 5 the lowest. These urgencies represent the assessments of various types of units of work made by the superordinate embedding agency to which the entire simulated organization reports. The urgencies are communicated to the managing officers in the form of quantified crisis scenarios. For each degree of urgency, this feedback report states the

LEVIATHAN-INDITE EPOCH 31 08-01-64 PAGE 309
RUN - 405A
SESSION - 6A DELIVER TO COMMAND CO

COMMUNIQUES THROUGH SYSTEM COMMAND CO

COMMAND	URG	NUMBER	NUMBER	AVERAGE PRIORITY				TIME THROUGH SYSTEM				
TOTAL		PROCSD	IN OFQ	1	2	3	4	0-13	14-24	25-48	49-96	OV 96
582	1	53	20	-	16	37	-	4	20	29	-	-
	2	62	0	-	1	45	16	1	28	33	-	-
	3	68	16	-	4	64	-	4	22	35	7	-
	4	36	8	-	-	28	8	-	12	24	-	-
	5	363	69	-	14	145	204	3	93	202	65	-
COMMAND TOTAL		582	113	0	35	319	228	12	175	323	72	0

Figure 19. System performance feedback: communiques through system. Evaluated according to urgency (URG) stipulated by higher-level embedding system.

number delayed in the system's overflow queue. Next, the priorities assigned by the officers are broken down by each class of urgency. Finally, the transit time through the system of each class of urgency is shown.

Similar evaluative feedback is supplied to all levels of command for numbers of units processed, numbers standing in waiting queues, and delays in queues.

We found that the 1964 subjects required a relatively long period of time to adjust to the new feedback and to learn to use its information effectively. Almost from their first exposure to the evaluative feedback, however, they began to develop, to a greater degree than before, very long-range contingency plans.

With the changeover to evaluative feedback, furthermore, the 1964 managers seemed to make a new use of exception or failure reports. These reports, it appeared, were being used to serve as prearranged triggers to call into play the previously formulated contingency plans. Thus the subjects were exploiting the exception or control information to subserve positive, innovative command objectives.

SUMMARY OF PERFORMANCE OF 1964 SUBJECTS

Whereas the 1963 subjects seemed throughout their five experimental runs to be trying to preserve a static equilibrium from declining, the 1964 group rejected any static equilibrium as its goal, in favor of a dynamic, progressive equilibrium. As the 1964 series of experiments ran its course, the group's performance steadily continued to rise, and it was still rising when we terminated. Eventually, at the end of the final experiment, the group's performance, using evaluative feedback, rose to levels twice those of any previous performance—an achievement equivalent to some large data-processing organization doubling its yearly output with no increase in manpower resources, facilities, or costs. In short, the group broke the bank.

What accounts for the high performance of this group? No doubt the format of the evaluative feedback contributed greatly to the subjects' achievement. Other variables also contributed:

- Learning, aided by the teaching machine and the hierarchically structured feedback, of how to operate the component offices.
- Development of functional specialists within the administrative hierarchy.
- Development of extraformal coordinating and reporting procedures.
- Formation of a repertory of contingency plans and imaginative envisagement of contingencies for which the plans might be invoked.
- Zealous motivation to realize an idealized and ambitious charter.

All these and other reasons account for the excellent performance of the group, especially during its final epochs of operation. It would not be proper experimental method, we believe, to view each of these possible reasons for high performance as though it were an isolable atomic element to which we could attribute just so much of the variance of the performance with just so much statistical confidence. As we review the reasons that account for the subjects' performance, we find what seems to us to be both an overlay and a mirror effect:

- The subjects were performing in a normative context surcharged with crises that waxed, waned and changed qualitatively at the will of the experimenters.
- Supported by the experimenters, who assumed the guise of the higher-level embedding agency, the subjects were imbued with a sense of the importance of their systemic effort.
- In the course of a series of experimenters' interventions, followed by subjects' reactions, the subjects were goaded and guided; and they responded with an evolving charter or image.
- The subjects were encouraged to develop specialties and staff appointments, and to unite the partial contributions of individual specialists in a systemic group effort.

All these elements and many more contributed to the formation and development of a normative, value-laden culture. Now, this structured, organized, many-layered, value-laden system, with its myriad subsystems of values and objectives, was mirrored in the evaluative feedback. This feedback provided direct, continuing and constantly updated assessment of the degree to which the subjects were realizing their image or charter. It was this solidary influence or overlay of evaluative, systemic feedback communication on the normative systemic context, we are confident, that resulted in the high performance of the 1964 group.

CONCLUDING REMARKS

This paper has dealt with the Leviathan method for laboratory experimentation on the information-handling process of large organizations. The method has focused on five essential elements in a taxonomy of the communication process. It has developed (a) a computer-based, dynamically evolving intercommunication language (GOCI). By this language, the interpersonal communication system of real-life organizations becomes simulated in the laboratory.

The Leviathan method has developed (b) an elaborate feedback system (indite), the distribution of which is governed by three criteria: professional specialty, territorial dominion, and hierarchical rank. (c) The functional specialization, territorial dominion, and hierarchical pyramid simulate the formal authority system of large organizations. The indite feedback reports, which flow according to the formal authority channels, simulate the data-handling and information-processing system of large organizations.

As part of the Leviathan method we have (d) a complex communication process that corresponds to the policy-formation-and-implementation system of large organizations (charter). This communication process consists of such features as the teaching machine, crisis scenarios, value-laden terms in the communication language, evaluative feedback, demands imposed on the subjects in the guise of consumer demands, demands laid on the subjects' organization by its embedding agency, and others. By these means, the experimenters cause to develop the ideals, goals, values and mission of a simulated organization.

From, and within, the interplay of these four basic elements of the total communication process emerges the fifth element. We have said that we view a large organization as a union of people relating in myriad ways, that creates and regenerates its on going power and sustains itself through its communication process. As a Leviathan simulation proceeds in the laboratory, (e) the extraformal, face-to-face interactions throughout the management pyramid come to life. How these develop—how this culture evolves—depends on how the experimenters manipulate the other four basic elements and how the group reacts to and assimilates them. Face-to-face interactive behavior is also reflected in and measured by the group's performance and accomplishments.

Finally, we have brought to your attention some of the first fruits gleaned from the use of the Leviathan method in the laboratory. These have been our initial interpretations of the interrelations between systemic communication and information feedback, of the normative, value-laden variables to which large organizations seem especially sensitive, and of the functions of positive, negative and evaluative feedback in large organizations.

VI. ELECTRONIC INFORMATION-HANDLING SYSTEMS—SHORTCOMINGS

17

Limitations of the Current Stock of Ideas about Problem Solving*

ALLEN NEWELL

Institute Professor
Systems and Communication Sciences
Carnegie Institute of Technology

Figure 1 shows a checkerboard with a domino beside it. The domino covers exactly two squares of the board. Suppose we are given an unlimited supply of dominoes and asked to cover the checkerboard exactly—i.e., with no dominoes extending over the boundary. This is a trivial problem. The dominoes can be laid down as in Fig. 2; and there are many other arrangements that would do the job equally well.

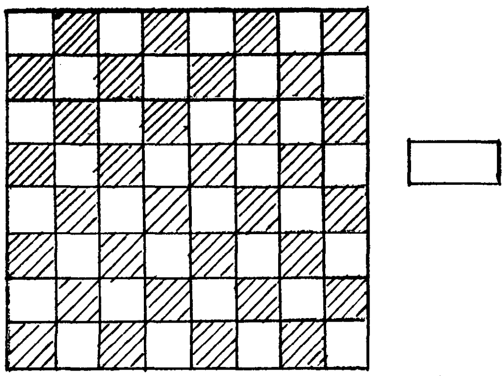


Figure 1. Checkerboard.

Now let us mutilate the board, as shown in Fig. 3, by removing the two corner squares. Again, the problem is to cover the board with dominoes. Only this time it is a hard problem. In fact, it is impossible. Therefore, the real problem is to prove that it is impossible. (Before reading further try to convince yourself of the impossibility and try to find a proof. You may already know the problem, of course, since it is a familiar chestnut.)

*The preparation of this paper was supported by Contract SD-146 from The Advanced Research Projects Agency of the Defense Department.

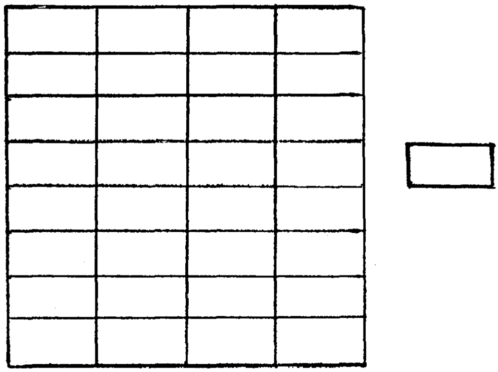


Figure 2. Covered checkerboard.

Most people find the proof difficult to discover, but transparent once found. Observe that the original checkerboard has thirty-two black squares and thirty-two white squares, and that a domino always covers one black square and one white square. With two white squares removed, the mutilated board has thirty-two black squares and only thirty white squares. Consequently, no matter how the dominoes are laid down eventually a position will occur with two black squares left and no white squares; and it will be impossible to cover these remaining two squares.

Our concern is with machines and not men. Hence, the ultimate problem is not to discover the proof, but to build a machine that can discover the proof to the domino problem. If is a fair statement, I believe, that no one today knows how to build such a machine—or equivalently, how to

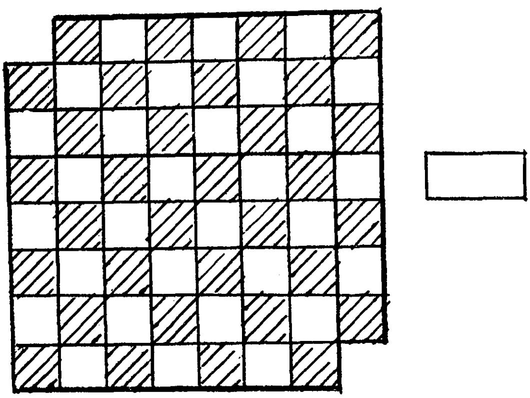


Figure 3. Mutilated checkerboard.

construct such a computer program. And this inability represents one of the limitations on the current stock of ideas about problem-solving by computers.

It may seem disturbing to have a limitation of ideas stated as the inability to solve a particular problem. It doesn't say what is missing. Even admitting that to say exactly what is missing is to say too much, one might still hope to describe classes of problems that could not be solved. Instead, the domino problem seems extremely particular.

In fact, proceeding by highly particular examples is characteristic of work in programming computers to solve problems. It is standard methodology—to write specific programs to do specific things—and in its own way represents a limitation on our current stock of ideas. Nevertheless, it is possible to use a single example as a tool to explore more generally our current knowledge about how to make computers into problem-solvers.

THE PROBLEM OF REPRESENTATION

The experience of many people with the domino problem is that they have no idea at all how to get started on finding a proof. When and if a proof is found, it occurs suddenly. This leaves them with a proof, but with no idea at all how a program might find it. Let me interpret this experience. Progress on a problem requires having some representation of the possible solutions to the problem that can be manipulated, searched, or explored in the process of determining the correct solution. With no representation, there is no possibility of manipulation and no way of making progress. Thus, the initial "lost" period is in fact devoted to finding a representation. The suddenness of solution arises from the extreme simplicity of the proof, so that once a representation is found, the "essential idea" of the proof is immediate, as is the verification of its soundness. Thus, there is little awareness of the representation of the possible proofs, which is what is needed to make a start on a computer program for finding the proof.

The proposition that a representation of possible solutions is necessary to finding a particular solution appears almost banal. However, the existing lines of attack in getting computers to problem-solve can be described in terms of the representations they have developed. And an important aspect of their limitations can be seen in what kinds of problems can be easily cast into these representations. We will put some flesh on this proposition by considering a number of these representations. As a common thread, we will ask whether each representation could help us in building a program that would find the proof of the domino problem.

HEURISTIC SEARCH

Perhaps the most notable approach in problem solving by computers is heuristic search. Almost all the successful theorem-proving, game-playing and puzzle-solving programs of the last several years belong in this class, as well as a number of programs for management problems of scheduling and allocation.[1] The basis of heuristic search is that I can look at any problem as if there are a set of situations (say S_1, S_2, ...) and a finite set of operators (say Q_1, Q_2, ... Q_n), such that given the situation S_i, the application of an operation, say Q, transforms the situation into another one, say S_j. As Fig. 4 shows, the situations can be viewed as the nodes of a tree, with the operations as the branches. The application of a sequence results in searching a part of the tree.

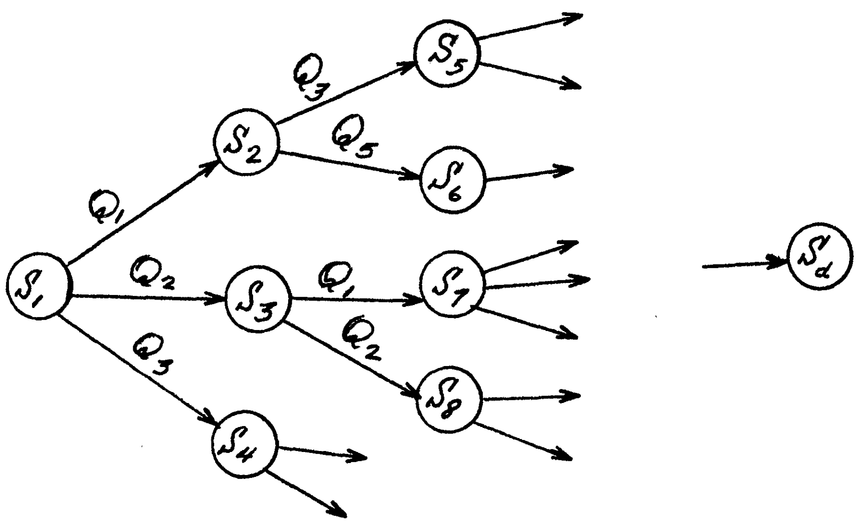


Figure 4. Tree representation of problem.

In this representation, a problem takes the following form: The initially given situation is the root of the tree, S_1; the desired situation is some S_d (possible a set of situations); the problem is to obtain S_d starting from S_1. Thus the problem is one of searching through the tree (as implied by applying different sequences of operators) until S_d is discovered. To pick one concrete example, if the game is checkers, the situations are checker positions, the operators are the legal checker moves, S_1 is the opening position, and the desired positions are those in which your side wins.

When problems of realistic difficulty (like chess and checkers) are cast into this representation, the trees turn out to be massively large and the problems cannot be solved simply by searching at high speed. Instead various rules (called heuristics) are used to narrow the search to the profitable part of the tree. These rules can be evaluation functions on the situations that approximate the final value, or rules that eliminate a branch, or rules that determine how much effort should be spent in searching a subpart of the tree. We are not interested here in the particular form of these heuristics. What is of interest is that having once represented the problem as search in a tree, there are a number of things we can do to bring the computer's problem-solving power (here, its capacity for sophisticated search) to where it solves significant problems. Indeed, the computer itself can modify and extend its own heuristics. For example, Samuel's checker program[4] modifies its evaluation function on the basis of its past experience.

Let us return to the domino problem and ask whether these ideas are of use. We can certainly represent the domino problem itself in this way: the situations are all the partially covered checker boards; the operators are the placing of a domino either vertically or horizontally so it covers two squares not yet covered; the initial situation is the empty checkerboard; and the desired situation is the completely covered board. But this doesn't lead anywhere. If coverings existed, a program could find them this way. But if coverings are impossible and the job is to prove it, then trying possible coverings, no matter how many, doesn't help a bit. Only if the program tried all possible coverings and knew it had exhausted them could it conclude that none were possible. But this implies searching the entire tree, and the tree is much too big (at least 10^{20} situations).

PREDICTING SEQUENCES

Let us turn to a different task, which has been solved by programs of a somewhat different kind.[6] The problem is to predict the next letter in the following sequences:

1. A B A B A B ___
2. A T B A T A A T B A T ___
3. D E F G E F G H F G H I ___

The answer to the first is clearly A; the answers to the others are not quite so clear, but are attained without difficulty by intelligent humans. However, for us the problem is not how humans can do it, but how to write a computer program that will do it.

This seems a difficult task—indeed, it involves a genuine induction—until one notes the absence of a representation of possible patterns, and takes steps to provide it. Consider the following scheme, which we can illustrate on the second task. A sequence will be generated by the iterated application of a set of rules; this set of rules, therefore, represents the pattern. There will also be some variables that maintain a memory of the current cycle, upon which the rules can act. For the second pattern, we start with one variable, m_1, which takes values in the alphabet (A,B) and initially has the value B. The rules are given by the expression:

$$\mathrm{A}, \mathrm{T}, m_1, n(m_1)$$

This is to be interpreted: Print A; then print T; then print the current value of m_1; then change the value of m_1 to be the next higher letter in the alphabet of m_1. Thus, on the initial run this prints ATB and changes m_1 to A (the alphabets are understood to be cyclical). The next run yields ATA and m_1 changes to B, and so on.

To give one more example, the third sequence above requires two variables, m_1 and m_2, both of which range over the standard alphabet (A, . . . , Z) and have initial values of D. The iterative rule is given by the expression:

$$m_1, n(m_1), m_1, n(m_1), m_1, n(m_1), m_1, n(m_2), e(m_1, m_2)$$

The first seven steps of this expression generate the four letters in a cycle; e.g., DEFG. Then m_2 is advanced one (e.g., from D to E) and m_1 is set equal to it. Thus the next cycle goes EFGH.

Once this language of patterns has been defined it is easy to write a program that will interpret it; that is, that will generate the sequence, given the expression. More important, it is also easy to construct a program that will discover whether any simple expression in this language agrees with a sample of a sequence. Given the language, it is clear that one must conjecture the cycle in the sample, and then discover the relations (expressed in terms of the operators n, e, and the various alphabets) between the letters both within the cycle and between corresponding members of successive cycles. Partial solutions can be tried (via the interpreter) and the discrepancies used to modify the expression. In short, once a representation is available for possible solutions, it is possible to construct programs that work on the problem in reasonable ways.

Returning to the domino problem, it hardly seems possible to apply the above language directly. Rather, we should look to the principle involved: "Build a language to express the possible solutions." Our problem is to find a language of proofs. We already have a way of talking about checkerboards and various coverings of dominoes; this clearly is not enough. Since proofs are normally given in a combination of natural language and notation about the task (this latter corresponding to our checkerboard and coverings) it is not easy to imagine what such a language of proofs might be like. However, there has been considerable work in constructing computer programs to find proofs, and we can look at these.

THEOREM-PROVING IN THE PREDICATE CALCULUS

Currently there are two distinct approaches to theorem proving. One of these considers the problem as one of heuristic search. The situations are theorems, the operators are the rules of inference, the initial situation is the collection of theorems that can be assumed true, and the object of search is the desired theorem. This approach has worked in areas where the rules of inference and the possible theorems are clearly set out, as in plane geometry or symbolic logic. But in the domino problem our difficulty is that we do not have any language for expressing possible theorems (other than the one given), nor are the rules of inference delineated. So we must solve our problem of representation prior to using heuristic search techniques for discovering the proof.

The second approach appears more hopeful. The development of mathematical logic has resulted in some formalized logical systems of great scope and power. One of these, called the first order predicate calculus, has received a great deal of attention from logicians interested in constructing programs to prove theorems. This calculus permits assertions involving the usual logical connectives (and, or, not, implies) and in addition, assertions of the form "There exists an x such that $A(x)$ is true" and "For all x, $A(X)$ is true," where $A(x)$ is any legal assertion in the calculus and x is a variable ranging over the basic objects that the calculus makes assertions about. The appeal of this system is not just that a great deal is understood about it mathematically, but that it appears to be rich enough in expressive power to cover most of the mathematics used in science and engineering. This gives rise to a vision in which all problems of proof are translated into the first order predicate calculus, and a single big theorem-proving engine is built for handling proofs in this calculus. Thus, the predicate calculus provides a universal means of representation. This vision has sufficient appeal that an entire subfield of artificial intelligence is devoted to its implementation, and numerous programs have been built to prove theorems in this system.[8]

Certainly we should apply this to the domino problem. First, we must translate the problem into the predicate calculus; then we can explore the possibility of current programs proving the theorem. Of course, there is more than one way to represent the domino problem in the predicate calculus—so the task of translation should not be passed over too lightly. However, analogously to the sequence-predicting problem already discussed, a representation already exists so the problem is quite tractable. We will not provide a translation here; it is too technical for this paper. Recently, however, John McCarthy has published a short memo, entitled, "A Tough Nut for Proof Procedures,"[3] in which he provides a translation of the domino problems into the predicate calculus and asserts that this theorem will be very difficult for present theorem-proving programs to handle. To quote him, "... I don't see how the parity and counting argument can be translated into a guide to the method of semantic tableaus, into a resolvant argument, or into a standard proof. Therefore, I offer the problem of proving the following sentences inconsistent as a challenge to the programmers of proof procedures and to the optimists who believe that by formulating number theory in predicate calculus and devising efficient proof procedures for predicate calculus, significant mathematical theorems can be proved." ["Semantic tableaus" and "resolvant arguments" are two special techniques developed in the field. "Proving the ... sentences inconsistent" refers to a standard approach in the field of conjoining the axioms and given theorems with the *negation* of the desired theorem to obtain a contradiction.]

PATTERN RECOGNITION

Let us consider just one more class of tasks, that of recognizing a pattern. Typical of such problems is recognizing the letters of the alphabet when printed, or when written by hand. Many computer programs (and hardware devices) have been constructed that do moderately well at these tasks; harder tasks are recognizing spoken words, or human faces. Now, an important superficial characteristic of human pattern recognition is that it appears to occur "all at once"—immediately, without protracted inferences. This is reminiscent of the suddenness with which most people discover the domino proof—"nothing" for a while, and then the proof is simply "there." Thus, we might look at pattern-recognition programs to see how they represent problems and whether this representation might be of use with the domino problem.

Enough pattern-recognition programs have been constructed, so we have a pretty good idea of the basic components. (At least, those that have been built have much in common; there might be other approaches which no one has discovered yet.) As Fig. 5 shows, there is an initial com-

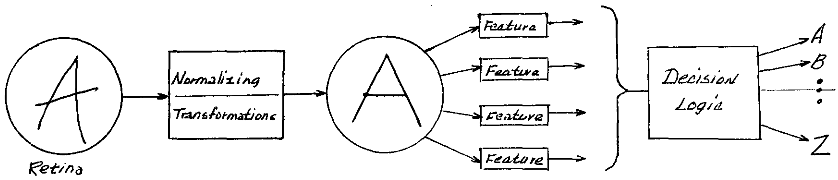


Figure 5. Schematic pattern recognizer.

ponent in which the item to be recognized is registered, often called the retina for obvious reasons. Then occurs a series of normalizing transformations, which get rid of variation by putting the input into standard form. In visual recognition these are such operations as centering, focusing, smoothing, enhancing contrast at edges, etc. Following this there are a set of feature detectors; each one reacts to some characteristic of the image. Taking vision, again, these might be "the existence of a vertical line segment," or "the number of corners," or "a marbled texture." Some of these features are themselves moderately complex, and may be thought of as involving the combination of other features. Finally, there is a component that combines all these features and arrives at a decision. This might be a "decision tree" in which discriminations on the various features finally lead to identifying the pattern; or it might involve measuring how close the input image is to templates of the possible patterns and choosing the closest.

The scheme of Fig. 5 can be taken as another general representation of how to make decisions or selections. Given a new task, the scheme directs attention to what pieces need to be defined and how they should then be related to produce a total system. It does not provide a representation of the possible solutions; rather, it is a representation of the problem-solving process. This is unfortunate, since if we try to apply the scheme to our domino problem, it provides us with little clue as to what should be made available at the retina (surely the checkerboard, but what else?), what features should be taken, or what the class of responses should be from which the right one (the proof) should be selected.

Although not appearing to help directly with the domino problem, the area of pattern recognition provides a good historical example of the difference between having a representation and not having one. In visual or auditory recognition, the representation on the retina and the set of responses are quite well defined; the real questions focus on the transformations, the features, and the decision logic. Of these, the features have seemed especially critical. A few years ago, it was an informal maxim in the field that one could undoubtedly design, ad hoc, a good set of features for any specific limited recognition task, but that the "real problem" was how to get new features for new tasks.[5] Up to this time, the features had always been thought up by the programmer on the basis of prior experience and investigation and simply programmed into the recognition program. The features that worked for one task did not necessarily work for another. The inability to construct recognition programs which built their own features was considered a significant limitation of the field.

In 1961 Leonard Uhr developed the first successful pattern-recognition program that obtained its own features.[7] The details of this program are not of interest here, but the essential idea is important. Since features had been anything a programmer could think up (as, for us, are the ideas for proving the domino theorem), there was no way of talking about the set of possible features (nor, for us, the possible proofs). Hence, there was no way of getting a program to manipulate features and develop new ones. Uhr's main contribution was to construct a space of possible features. The retina in his program was a rectangular grid of bits, 20 on a side, as in Fig. 6. The pattern to be recognized is written on the blank retina (consisting of all 0's) by putting 1's in the appropriate cells. A feature, said Uhr, is defined by a 5 × 5 subgrid having 0's, 1's and X's for entries (only the X's show in Fig. 6 to avoid confusion). The subgrid is swept over the entire matrix; at each position a measure of agreement be-

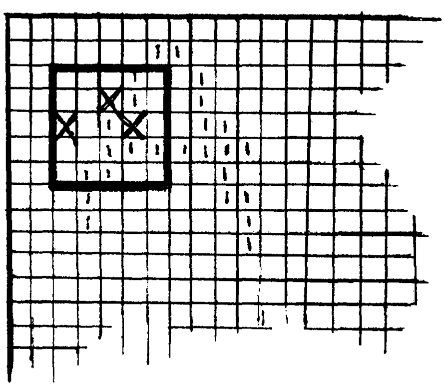


Figure 6. Retina and 5 × 5 feature.

tween it and the retina is taken by counting the 0's and 1's that match (and ignoring the X's). This distribution of measures is used to define the actual feature—e.g., the position where it is strongest, whether it ever exceeds a certain threshold, etc. The important thing for us is that there exists a set of possible features (all different subgrids), so that the program could introduce new ones. For instance, it could copy a part of a sample pattern and use it as a feature in recognizing other exemplars of the same pattern. This is an extremely simple scheme, almost naive; yet it was enough to permit his program to recognize a wide variety of different kinds of patterns, developing for each an appropriate set of features. It was enough to dispose of the maxim.

A FINAL LOOK AT THE DOMINO PROBLEM

Although the domino problem is not easily assimilated to any existing approaches, each of them has had something to say about how to represent a problem and how to proceed to solve it. Together they permit a slight reformulation of the domino problem. This is of interest in showing that, having represented the problem and surmounted one hurdle, the next hurdle we come to is again a matter of representation.

As noted earlier, we can formulate the task of covering the checkerboard as a tree of operations. Clearly, we can get the computer to try a series of domino placements, Q_1, Q_2, . . . , starting at S_1 to attempt to get a complete covering (see Fig. 4 again). Since the task is impossible, there is no path that leads to S_d, the final, perfect covering.

Now there must be something that prevents a path starting at S_1 from reaching S_d. That is, there must be some property of the initial situation that is true of all the situations (the S_i) reachable from S_1, is not true of S_d, and such that none of the operators, Q, changes it. Putting this more formally, let $P(S)$ be this property, determinable for any position. Then the conditions are:

1. $P(S_i)$ is true.
2. If $P(S_i)$ is true then $P(Q(S_i))$ is true for any legal Q.
3. $P(S_d)$ is false.

And the conclusion is:

There is no sequence Q_1, Q_2 . . . Q_m such that

$$Q_m Q_{m-1} \dots Q_2 Q_1(S_1) = S_d$$

Proposition (1) says that the property is true of the initial situation. Proposition (2) asserts that this property is hereditary; that is, if it holds for a situation, it holds for all those that immediately follow from it by legal moves—hence, for any that can be reached through any chain of legal moves. Finally, proposition (3) says the property does not hold for the desired position. The conclusion is that the final position can never be reached.

Note that the actual proof can be put in just this form. The property P is that the number of black and white squares uncovered are unequal. This is true of the initial board; and the placing of any domino, which covers one square of each color, leaves the property true of the resulting board. But the final position has equal numbers uncovered, namely, zero.

If the problem is reformulated as above, then the task shifts to the search for a property with the desired characteristics. But first it is necessary to ask whether a computer program could be expected to reformulate the task in this way. This seems reasonable to me, in support of which I offer the following plausibility argument. The formulation above is an example of the principle of mathematical induction, usually stated, "If $P(n)$ implies $P(n + 1)$, and $P(1)$ is true, then $P(n)$ is true for all positive n." Now there is only one such principle, just as there is only one concept of equality, one concept of a function, or one mathematics of the integers. Consequently, it is reasonable to assume that a problem-solving program would be given this principle. In fact, this is the way almost all humans get their basic intellectual tools. (That they are not easily discovered by the unaided human intellect is testified to by the long historical development of mathematics.) Therefore, the program does not have to discover the induction principle; it has only to evoke it and apply it. To evoke the principle does involve a recognition; however, there are relatively few basic ideas for proofs, so that this is not the difficult step. Likewise, transformation of the principle from its positive form into the essentially negative form used in the domino proof does not seem insurmountable. The machine has a representation of the principle and a representation of the final thing it wants to prove—i.e., proposition (4). Purely formal operations can be used to manipulate the principle to give (1), (2) and (3).

Despite the unfilled gaps—several programs have been built to use the principle of induction in sophisticated ways,[2] but none to adapt the principle to new situations—let us accept that the program can get as far as the formulation (1)–(4). Where does it go from here? Its task is now to find a feature. Again, the difficulty is that no space of features is given within which to search—i.e., a representation is missing. If we limit the features too severely—e.g., to relations among numbers of black and white squares, then in choosing the space of features we have already done most of the work. That is, it is we who have found the proof by selecting the feature space. If, on the other hand, we give it no representation at all, then the program can do nothing. It is not enough to give it the checkerboard; it must also have ways to measure aspects of the board and to combine and compare these in various ways. Even at this stage, for instance, it is clear that it makes a great deal of difference whether the program is given a checkerboard, with its squares alternating in color in the relevant way, or whether it is given a blank board. (Only the checkerboard's familiarity inhibits the checkering from immediately cluing the human.)

Actually matters are not quite so difficult, since the expressions (1)–(4) provide some good raw material to work with. However, in the interests of making the point we will not press the example to the limit. (For I believe, certainly, that given a modest amount of additional effort, a reasonable program can be constructed that finds the domino proof and does so fairly.) It is enough to observe the transformation of the original problem of representation into another (less severe) problem of representation.

CONCLUSION

Let me summarize the general argument, for which the domino problem has been only a means, although hopefully an entertaining one. We can look at the current field of problem solving by computers as consisting of a series of ideas about how to represent a problem. If a problem can be cast into one of these representations in a natural way, then it is possible to manipulate it and stand some chance of solving it. Different approaches, consisting of different global visions about representation, are not easily translatable, one into the other. Naturally, each of these visions turns out to have certain advantages and certain disadvantages, much of which can be summarized by describing the kinds of problems which can be easily so represented, and admitting that we can't yet stretch any one representation too far.

The natural response to this description of problem solving is to inquire where representations come from, and what is known about constructing new ones. Here we are on familiar, but unpleasant, ground. Currently, representations seem to arise in isolation—"out of nowhere." To put it in still more familiar terms, we do not yet have any useful representation of possible representations. This is possibly the biggest limitation on the current stock of ideas about problem solving.

REFERENCES

1. Feigenbaum, E. and J. Feldman (eds.), *Computers and Thought* (McGraw-Hill, 1963). Contains many examples, reprinted from the primary literature.

2. London, R., "A Computer Program for Discovering and Proving Recognition Rules for Backus Normal Form Grammars," *Proc. Assoc. for Computing Machinery*, A1.3-1–A1.3-7 (1964), p. 64.
3. McCarthy, J., "A Tough Nut for Proof Procedures," Stanford Artificial Intelligence Project Memo 16, July 17, 1964.
4. Samuel, A., "Some Studies in Machine Learning Using the Game of Checkers," *IBM J. Research and Development*, vol. 3 (July 1959), pp. 211–229. (Also reprinted in Feigenbaum and Feldman.)
5. Selfridge, O. G. and U. Neisser, "Pattern Recognition by Machine," *Scientific American* (August, 1960), pp. 60–68. See especially the last paragraph. (Also reprinted in Feigenbaum and Feldman.)
6. Simon, H. A. and K. Kotovsky, "Human Acquisition of Concepts for Sequential Patterns," *Psychol. Rev.*, vol. 70 (November 1963), pp. 534–546.
7. Uhr, L. and C. Vossler, "A Pattern Recognition Program That Generates, Evaluates, and Adjusts Its Own Operators," *Proceedings of the Western Joint Computer Conference*, vol. 19 (1961), pp. 555–570. (Also reprinted in Feigenbaum and Feldman).
8. Wos, L., D. Carson, and G. Robinson, "The Unit Preference Strategy in Theorem Proving," *Proceedings of the Fall Joint Computer Conference*, vol. 26 (Spartan, 1964), pp. 615–621.

18

Some Practical Aspects of Adaptive Systems Theory*

JOHN H. HOLLAND
University of Michigan

Al Newell started out this morning by putting down something that looked like a checkerboard and wasn't. I'm going to put down something that doesn't look like a checkerboard and is (Fig. 1). What I'd like to do in the time that I have is to relate information retrieval to what is perhaps the only really successful accomplishment in artificial intelligence as measured against the performance of a sophisticated human: Arthur Samuel's checker player.[1] I'd like to see if in fact some of the things that Samuel learned by writing his program have some bearing on problems in information retrieval.

This (the left side of Fig. 1) is really a tree representing successive legal moves in the game. Each vertex (node) stands for a possible board configuration. Each directed edge (arrow) represents a legal move; it points to the configuration (i.e., the corresponding vertex) that will result from that particular move. By way of simplification I will assume that my opponent's strategy—his reply to each possible move—is fixed. Thus, each move I take will elicit a specific reply from my opponent and hence the arrow in the reduced tree (the right side of Fig. 1) need only point to the set result of his reply to my move. The arrow then represents two successive legal moves: my choice, followed by my opponent's reply. The tree as a result shows only successive decisions or choices open to me in the face of my opponent's strategy. Each different strategy for the opponent will yield a different tree of decisions.†

The first thing I'd like to discuss is the way Samuel tackled this game. Samuel's approach is related to the "features" notion that Al Newell talked about. It involves pattern recognition in an essential way—the recognition of crucial situations (opportunities, pitfalls, etc.) as features

*The work discussed in the latter part of this talk was supported in part by the National Institutes of Health under grant GH-12236-01.

†For those familiar with automata theory this tree can be looked upon as a simple finite automaton—one with delays but no cycles, a generalized switch wherein successive inputs correspond to successive moves. If the opponent employs a mixed strategy the resulting automaton is a correspondingly simple probabilistic automaton. Hence corresponding to a game coupled with an opponent's strategy, there is a probabilistic automaton with a rather simple normal form.

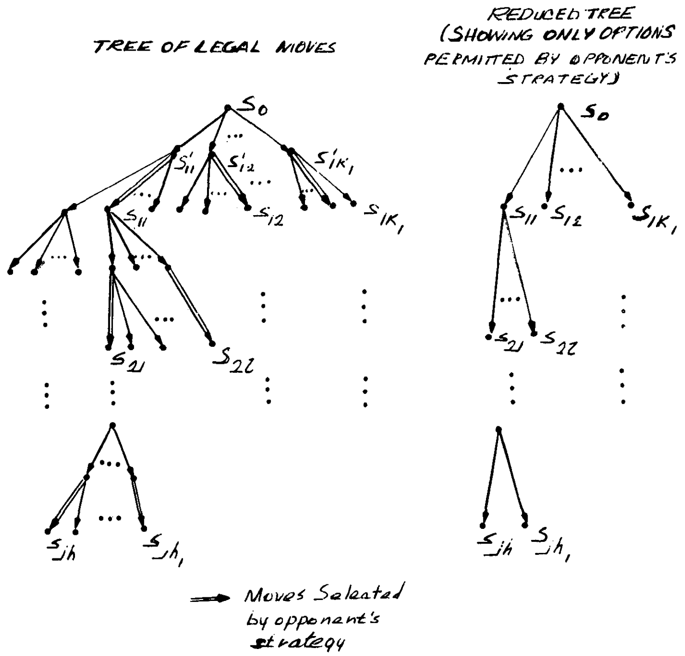


Figure 1

of the overall board configuration. Samuel started out by choosing a large number of features (he called them parameters) and programmed subroutines which were to detect these features in the various possible board configurations. Let's designate these different subroutines θ_1, θ_2, . . . , θ_j, . . . , θ_n. θ_1 might be the number of pieces I have on the board minus the number of pieces my opponent has on the board. Samuel has to have a subroutine in the computer that will scan the board and decide what this number is. Most of the time in a close game this property will have the value zero; that is, most of the time in a close game I'll have the same number of pieces as my opponent. But there can be more subtle properties which will often be nonzero. Thus, θ_2 might measure the average distance of penetration of my pieces into my opponent's territory minus the average distance of his penetration into my territory. What Samuel did was to select a large set of properties like this—actually not so terribly large—30 or so. The properties were so chosen that each of the related subroutines calculates a number when presented with a board configuration (piece count, distance, and so on). He then formed a polynomial by weighting the parameters and summing them:

$$V(s) = \Sigma_j a_j \theta_j(s)$$

where $s \epsilon S = \{s \ni s$ describes a broad configuration$\}$. Note that, formally, each parameter maps S into the rational numbers, $\theta_j: S \longrightarrow R$, as does the polynomial V. Having made an initial choice of weights, Samuel used the polynomial to make successive move selections. For example, to choose the first move in terms of the tree of Fig. 1, V is calculated for s_{11}, s_{12}, $\ldots$, s_{1k_1}. Then that move is chosen which leads to the vertex for which V is largest. Actually Samuel's program is more complex than this, but the description is sufficient for present purposes.

The problem in this simple situation is to decide what weights are appropriate. Some features will be worth striving for: if I can keep myself pieces ahead, ultimately I will win the game. Similarly, in the long run, if I manage to penetrate my opponent's territory more often and more deeply than he penetrates mine, I'll get more kings and eventually win. Positive weights seem appropriate for such parameters. On the other hand, there may be some θ_j which indicates double-jump traps by a positive value. A large negative weight here will assure that, whenever a situation s_{ih} occurs where θ_j is positive, the polynomial V will take a low value. As a result the situation s_{ih} will be passed over or avoided in favor of some alternative $s_{ih'}$ for which $V(s_{ih'})$ is larger. Note that θ_j can thus be very helpful even though it indicates situations to be avoided. There may be other properties which are redundant or irrelevant to which we would hope to assign the weight zero.

Briefly, and more formally, the problem is to make a linear combination of the basis functions, $\{\theta_j\}$, which will yield the best possible strategy in terms of this basis.* Moreover, Samuel wished to do this automatically through play of the game and not through his direct intervention. In other words, the overall program is to try various combinations of weights and then select the best set among those it has sampled.

One way this might be accomplished would be to generate and try n-tuples of weights at random. At each stage the best n-tuple up to that point is retained. Let us assume that $n = 30$ and that there are 10 possible weights (5 positive, 5 negative) for each a_j. A simple calculation shows that even if one could rate one n-tuple every millimicrosecond, it would take about 3×10^{12} centuries to try out all n-tuples. This makes it abundantly clear that, even for a relatively simple task like checkers, it is not feasible to enumerate and try possibilities (strategies, here) one-by-one, ignoring almost all the information returned by each trial. In other words, there is not, nor will there be, a computer large enough and fast enough to simply grind away and grind away until all possibilities are tried. As Al

*Cf. a truncated Fourier series as an approximation to an analytic function.

Newell remarked, a similar comment goes for any related approach to problems in the predicate calculus. I might say, "Alright, I'll start with a problem phrased in the predicate calculus and simply grind out proofs one by one. If a proof exists it will certainly be produced." And it will. But this guarantee is worthless, since the procedure which yields it can never under any stretch of the imagination have much bearing on how the answer might really be attained.*

Samuel's approach is demonstrably better. In fact, his scheme made enough use of the information it gained from playing the game to be able to beat him. This is already a good criterion. I guess the most recent piece of information I have is that about two years ago the program beat a tournament-level player. The player claimed to have made a mistake (since it involved a "look-ahead" of seven moves, it was not what an ordinary player would likely denote by that word) and in later rematches has beaten the program. He will readily admit that the program gives him a good game. Thus Samuel has given empirical proof that there is a way to design a checker player which adapts rapidly enough to play well by human standards—a way which is feasible on human time scales. This not only gives hope for success in similar programming endeavors (alas, there is little enough to date), but also indicates an area ripe for mathematical study. Surely we can gain a deeper understanding of what formal characteristics of checkers enable the success of Samuel's approach. We should be able to learn what generalizations of Samuel's approach will work in a broader context.

I do not have the time to go into details of Samuel's approach but I do want to discuss one aspect of it particularly relevant to information retrieval. Although Samuel treats the θ's as features of a checkerboard, they could as well be features of documents, i.e., descriptors. In other words, one could as well write a set of subroutines for detecting or extracting critical information from documents. Each subroutine could estimate, for example, the frequency of specific key words or phrases. Suppose now that I wish to extract documents on a particular subject from a system indexed by descriptors $\theta_1, \ldots, \theta_n$. Because it is desirable to keep the number of descriptors reasonably small in relation to the range of possible subjects, I will in general require a (weighted) combination of descriptors to access the documents of interest. Moreover, because the descriptor subroutines may be quite intricate, I will have only a general idea of their use or definition. Hence I may choose a very poor combination.

*It is worth noting that in the areas of adaptation, problem-solving, information retrieval, etc. such guarantees are almost always trivially available and have almost no bearing on the problem at hand.

Is there a way the system can adapt to my requirements, hopefully without modifying the descriptor subroutines $\theta_1, \ldots, \theta_n$ which after all have been very carefully conceived?

Samuel provides a very useful technique, his "book move" technique, which can be brought to bear. In conceiving his program, Samuel kept before himself the objective of having the program learn by playing against experts or, even better, against the recorded games of experts. For checkers, as for chess and go, there are many books which contain records of games between experts, often annotated to indicate the "best" move at each step. Let us assume now that we have followed a game to the j^{th} move and that N alternative board configurations $s_{\alpha 1}, \ldots, s_{\alpha N}$ are open to us (via legal moves). The weighted descriptor $a_i\theta_i$ will assign values $a_i\theta_i(s_{\alpha 1}), \ldots, a_i\theta_i(s_{\alpha N})$ to these alternatives. Suppose the book (or the expert) says that in fact $s_{\alpha k}$ was the "best" move. How can the program make use of this information?

Let us count those alternatives, s_α, for which $a_i\theta_i(s_\alpha)$ exceeds $a_i\theta_i(s_{\alpha k})$. Say there are N_1. Then there will be $N_2 = N - N_1$ alternatives with values less than or equal to $a_i\theta_i(s_{\alpha k})$. A little thought will show that if we modify a_i by an amount

$$\Delta_i = c \left[\frac{N_2 - N_1}{N}\right] = c\left[1 - \frac{2N_1}{N}\right]$$

where c is a small constant, $(a_i + \Delta_i)\,\theta_i$ will give the polynomial V a better chance of selecting the expert move when this situation presents itself again. That is the modified polynomial $V' = \Sigma_h(a_h + \Delta_h)\theta_h$ is more likely to select $s_{\alpha k}$ than the given polynomial $V = \Sigma_h a_h \theta_h$. Let us continue in this way to modify the weights of the polynomial on successive moves and plays whenever expert advice is available. Eventually we will obtain a polynomial V^* which is the best approximation, over the basis $\theta_1, \ldots, \theta_n$, to expert play.

Notice here that the expert (or book) need know nothing about the program or the subroutines for $\theta_1, \ldots, \theta_n$. He simply indicates what he would do at each move. The program takes over from there. In effect we get a kind of man-machine interaction where the man for once need know nothing about computers. There are in fact many problems where use can be made of advice via this technique. Here I will concentrate on the previously posed problem of document retrieval. The θ_i are once again descriptors. Each "move" is the selection and presentation of a document. I advise the system as to whether the document is acceptable or not. After a sequence of such trials, I can ask the system for a printout of the modified weights. The technique just described assures that next time I approach the system for information on the particular subject of interest, I will get better service simply by employing these weights. Note that this better service does not require any modification of the descriptors (a costly process both in terms of reprogramming and in terms of recataloging).

This technique is just one of several developed by Samuel; it is not the only one with applications outside of checkers (or game-playing for that matter). Many of Samuel's ideas are useful and important when translated to the context of information retrieval.

To repeat: we have empirical proof that Samuel's checker player plays a good game by human standards. Moreover it has reached this level in a relatively short time—certainly nothing like 10^{12} years. Why? And how much better could it be? These questions lead us immediately into deep waters. A careful answer would require at least a series of capacity or efficiency theorems for adaptive systems. At present we have no good way for comparing two adaptive strategies or techniques. Given two techniques for learning to play checkers, or for "adaptive" information retrieval, we are reduced to building and trying (or simulating) them. Even then, and even assuming we have satisfactory criteria for comparison, we will have little idea about the existence of still better strategies or about how much better they could be. We are at much the same stage as steam engine designers before the advent of Carnot. Or, more recently, the stage in information transmission technology preceding Shannon's famous capacity theorems. Here it was actually the case that a great deal of money was going into the development of a transmission system which simply could not be built because its existence would entail exceeding the capacity of the particular transmission technique involved. At the same time there was a transmission technique, receiving little development effort, which in fact was operating far from capacity. Shannon's abstract theorems had a real effect by directing attention to this latter technique—in a short time, and for a relatively small expenditure, large improvements were achieved—while preventing a large waste of effort on the former, an effort doomed ab initio.

Capacity theorems for adaptive systems would certainly effect similar reorganizations of research and development over a wide range of areas, including information retrieval. To see what some of these effects might be, let's take a closer look at the formal framework underlying Samuel's approach: as I mentioned earlier, Samuel's approach formally amounts to a search for the best strategy definable by a linear combination of the basis functions $\theta_1, \ldots, \theta_n$. Under what conditions will Samuel's weight modification technique yield rapid convergence to the best strategy over $\theta_1, \ldots, \theta_n$? Some thought shows that Samuel's technique will give rapid convergence only if the θ_i are independent or quasi-independent of one another with respect to the environment (the domain of the functions

$\theta_1, \ldots, \theta_n$). Interestingly enough, many—one is tempted to say almost all—schemes for adaptation proposed to date make the same requirements of the environment. To mention just three: Friedberg's learning machine,[2] the Bledsoe-Browning pattern recognition scheme,[3] including Uhr's modification,[4] and the work on adaptive threshold elements, for example, Mays' extension[5] of Widrow's work.

Before going further, it is important to note that, when we discuss the environment of an adaptive system, we are really discussing not a single environment but a set of environments. Why do I say that? Consider first the case of an information-retrieval system. Each user of the system is a different environment for that system—he puts different requirements on it. The system must distinguish different users and respond differently to each. More generally, and more precisely, if a system is faced with a problem of adaptation there must be some aspect of its environment unknown to it. Formally this can only mean that, from the system's point of view, the description of the environment involves a variable. This variable must have a set of substitution instances and each of these substitution instances yields a distinct environment. The set of environments so obtained is the set of environments the adaptive system must be prepared to face. We're really interested in how well the system can perform over this set.

And here we run into a real difficulty. Just how do we compare the performance of two systems over some set of environments ξ? One system may perform well on one subset of ξ, say ξ_1, and poorly on a subset ξ_2, while another system may do well on ξ_2 and poorly on ξ_1. One hope would be for the existence of a system which performs well over all of ξ, a kind of "universal" (w.r.t. ξ) adaptive system. Then we could at least compare various schemes of adaptation with the "universal" scheme, if not directly with one another. Even then we need a formal counterpart of the phrase "performs well over all of ξ." One possibility is to make use of a notion from probability: "gambler's ruin." Assume that we can measure performance in any given environment E of ξ in terms of some accumulated payoff (cf. von Neumann's theory of games.[6] Scheme T will be said to "perform well over E" with respect to scheme T' if T is not forced into "gambler's ruin" by T'. If this holds true for T for all T' and over all $E \epsilon \xi$, I'll call T "strictly near-optimal (sno)."*

Fortunately, over many interesting classes of environments and adaptive strategies, strictly near-optimal strategies exist.†

*For more details see the latter part of Ref. 3.

†In particular there exist sno strategies over broad classes of game-trees—these classes are probably most easily characterized in terms of the corresponding probabilistic automata. It is important that enumeration and rote learning schemes are *not* sno over any of these classes.

In game-playing terms, a sno strategy assures the inability of an opponent to bring about the ruin of the player. In biological terms, a biological adaptive system employing a sno strategy is assured of adapting rapidly enough to escape extinction.

Taking this into account, let's look once more at the adaptive strategy implicit in the work of Samuel, Bledsoe-Browning, et al. In effect, it is a particular scheme for sequential sampling of functions definable over the basis set $\theta_1, \ldots, \theta_n$, using the performance ratings of functions sampled to determine new samples. Hopefully, the sampling scheme (adaptive strategy) will be strictly near-optimal over the class of environments of interest. However, the previously noted requirement of independence of the θ's for rapid convergence—and this turns out to be a necessary condition for near optimality in this case—puts a very strong constraint on the basis set. In general this constraint will be satisfied only over very limited sets of environments.

There are, however, more general techniques than Samuel's for generating successive trials of functions over a basis set. Given any basis set, these techniques, closely related to the interacting phenomena of crossover, linkage, and dominance in genetic systems, yield strict near optimality over a much broader class of environments. Much remains to be done along this particular line and there remain many other definitions of "performs well over all of ξ" which merit examination.

To those of you extensively involved in information handling, I would urge the importance of doing some of this work. The invention of new heuristics and programming languages is important, and will continue to be so. At present there is no dearth of effort along these lines. But a concentration on invention without a parallel effort on theory—particularly theory relevant to efficiency or capacity—can lead to extensive development work along foredoomed lines coupled with ignorance of the potential of promising lines.

REFERENCES

1. Samuel, A. L., "Some Studies in Machine Learning. Using the Game of Checkers," *IBM J. Res. and Dev.*, vol. 3 (1959), pp. 210–229.
2. Friedberg, R. M., "A Learning Machine: Part 1," *IBM J. Res. and Dev.*, (1958), pp. 2–13.
3. Bledsoe, W. W., and I. Browning, "Pattern Recognition and Reading by Machine," *Proceedings of the Eastern Joint Computer Conference* (1959), pp. 225–232.
4. Uhr, L., and C. Vossler, "A Pattern Recognition Program That Generates, Evaluates, and Adjusts Its Own Operators," *Proceedings of the Western Joint Computer Conference* (1961), pp. 555–570.

5. Mays, C. H., "Effects of Adaptation Parameters on Convergence Time and Tolerance for Adaptive Threshold Elements," *IEEE Trans. Electron. Computers EC-13*, vol. 4 (1964), pp. 465–468.
6. von Neumann, J., and O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton, 1947).
7. Holland, J. H., "Universal Spaces: A Basis for Studies of Adaptation," *Automata Theory*, E. R. Caianiello, ed., Academic Press, 1965.

19

Information Processing and Bionics

JOHN E. KETO
Chief Scientist
Aeronautical Systems Division (AFSC)
Wright-Patterson Air Force Base, Ohio

INTRODUCTION: THE INFORMATION DELUGE

Much has been said about the problems of information handling brought about by the explosive growth of our advancing technology. For instance, the *Wall Street Journal* in the article "Fishing for Facts" (December 1960), pointed out that during the year technical papers around the globe had generated some 60 million pages of new material or the equivalent of about 465 man-years of steady around-the-clock reading.[1] The article highlighted the problem of industry in absorbing and reducing this information into relevant and significant data for application. Again, Dr. Milton S. Eisenhower, in an address at the 15th National Science Fair International (Baltimore, Maryland, May 6, 1964), added further statistics and comments on the problems of the "knowledge explosion."[2] To quote,

> The scientific revolution continues today with an incredible flow of new knowledge and new ideas. Though we stand at the center of the knowledge explosion, even we can hardly comprehend the scope and the impact of the scientific and technological information that pours from the world's universities and research laboratories.
>
> In the last year for which international statistics are available, it was reported that 1,250,000 technical papers were published in the fields of the life and physical sciences. And the production of knowledge is increasing exponentially. The number of technical journals has doubled from 50,000 to 100,000 in only the past 13 years. By 1980 it is estimated there will be a million such journals. In one field—the biological sciences, research findings have increased by 60 percent in the past five years. And the average biologist can now review only about five percent of the material published each year. The proliferation of articles, journals, and abstracts is so tremendous that we are now publishing abstracts of abstracts.

And so we have problems, and much is being done about those problems. Advanced information-storage and retrieval systems, reading machines, translation machines, automated library systems, documentation and data-processing centers, all directed at a solution to the problem of knowledge availability. This symposium is a critical indicator of the magnitude of the problem and the vigorous efforts towards its resolution.

It is the purpose of this paper to highlight advancing problems of process-related data handling—processes of scientific research, engineering design and analysis, biological and medical investigation and system mechanization as they relate to growth and advancing complexity of this type of information-processing problem. Again, the trends characterize another information explosion with all the earmarks of the "library" problem and questions of knowledge availability. It is my plea that this other "side-of-the-coin" of the information-handling problem receive like attention.

SOME SIGNIFICANT EXAMPLES

AIRCRAFT STRUCTURAL INTEGRITY

The modern airplane is indeed a complex machine and its operating envelope continues to advance in speed, altitude, performance and environment. Its structure is a maze of ribs, spars, and bulkhead frames to which the outer skin is attached by rivets, adhesives, welding or other means. The Air Force C-133 of Fig. 1 is a typical transport representative of the larger vehicle class. Such airplanes are much more aeroelastic in their structural charcter in contrast to the highly rigid body nature of the airplane of the early forties. Response and loads analysis induced by flight conditions of maneuver, speed, and atmospherics (altitude, temperature, and air mass motion, particularly turbulence) has become a very complex problem. When man is introduced into the control loop through the flight control system, the nonlinearities of the overall system provide further complication. Even under linear conditions, the equations of system motion are complex, requiring the energetic use of IBM-7090 computers to obtain quantitative criteria of system performance by mathematic modeling.[3] Considering the ever-increasing flight speeds and the severity of atmospheric turbulences being encountered, aircraft design to assure structural integrity for the required flight safety and operational life has become a priority problem. Structural design must accept not only the *dynamic* flight loads encountered in any given flight but must also concern itself with effects of wear-out due to *fatigue.*

Because of unknowns in the area of fatigue, it is current practice to use scatter factors (confidence or safety factors) of two to four in estimating operational life from the load cycling tests on the initial prototype. Other solutions are seriously sought to alleviate the problem, such as design approaches to provide more favorable gust-response characteristics,[4] and airborne detectors of atmospheric turbulence.[5]

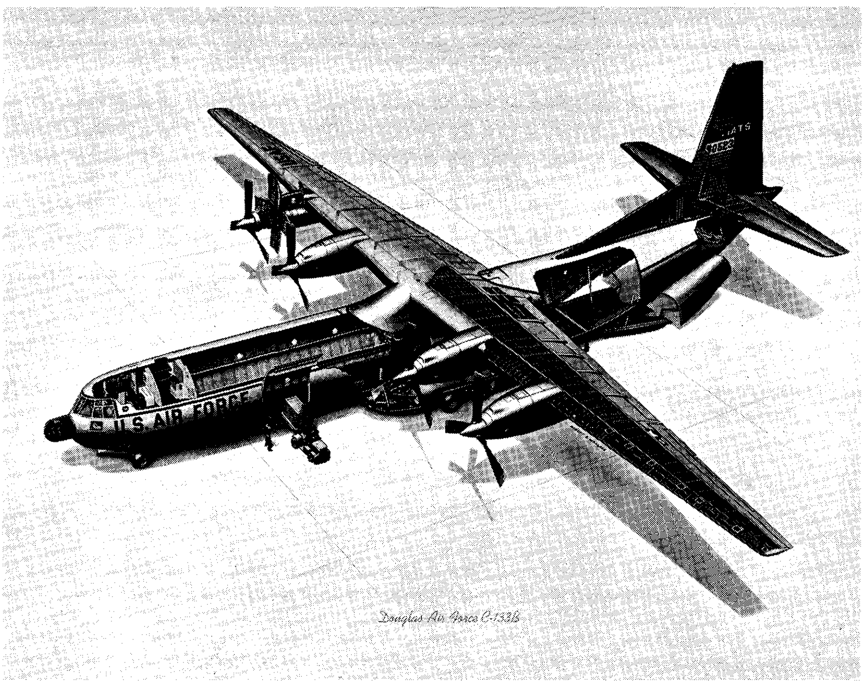


Figure 1. Air Force C-133 Transport.

In the past five to six years there has been a marked change in the engineering mathematics of structural design and loads analysis. Recognizing the random character of the atmospheric disturbances, approaches have been developed through the application of theory of random processes and techniques of general harmonic analysis. These methods have been much advanced by early investigators such as John C. Houbolt and Harry Press, and NACA Report 1272[6] published in 1956 is still a basic reference in this area. Atmospheric turbulence is represented as a continuous random disturbance characterized by power-spectral-density functions as plotted in Fig. 2. Data for such curves have been obtained by flight measurements with aircraft instrumented for the purpose, primarily through efforts of NACA (now NASA), the Air Force and Cornell Aeronautical Laboratory, and by observations from meteorological towers. The curves of Fig. 2 are plotted for different values of L, the scale of turbulence, and related to eddy size of the turbulence. The aircraft response to such disturbance in terms of acceleration and load spectra are obtained from the transfer functions of the airplane. Integrity of design is dependent upon failure-free response to the loads to be encountered in any given flight as well as the fatigue aspects of the stress-strain history

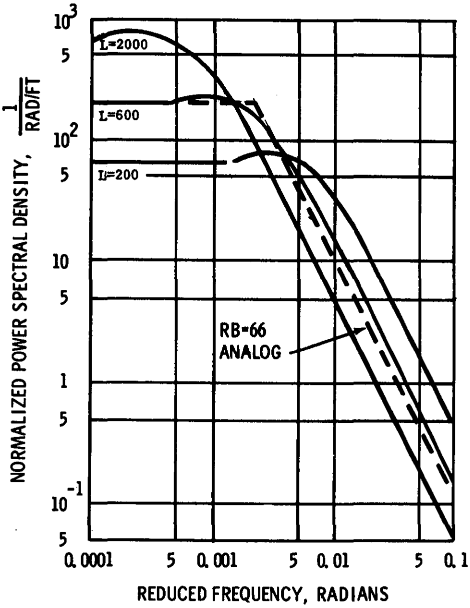


Figure 2. Analytic turbulence representation.

encountered in repeated flight. The probability character of the atmospheric disturbance therefore becomes the second aspect that must be considered. Typical probability data is shown in Fig. 3. Such curves are based upon direct observation as outlined before and from data derived from aircraft operations.[6] The currently available data is seriously restricted in applicability due to the limited scope of measurements and atmospheric conditions covered in the direct observations and assumptions and approximations involved in the derived data. An extended statistical model of true gust conditions is urgently needed.

This latter statement is borne out by recent experiences in the structural repair and improvement program of the Air Force B-52 bomber. In the course of this program, it was necessary to instrument and flight test representative B-52s to verify structural rework and to accumulate additional data on response to varying flight conditions. One such B-52 instrumented with sensors and recorders to obtain velocity, acceleration and altitude information (V, G, H), suffered severe lateral gusts in flying by the Spanish Peaks of the Sangre de Cristo mountains in southern Colorado at a clearance of approximately 1,000 feet, flight altitude 14,000 feet, flight direction south to north just east of the East Spanish Peak. Catastrophic loss of the rudder and 82 percent of the vertical fin occurred but, fortunately, due to outstanding performance by the pilot and crew,

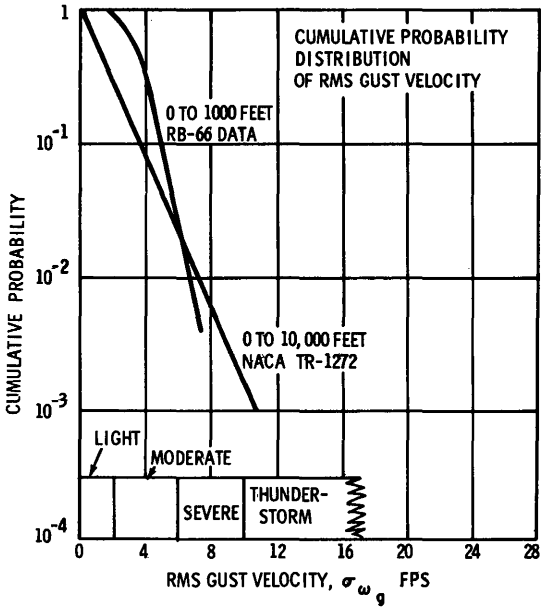


Figure 3. Cumulative probability.

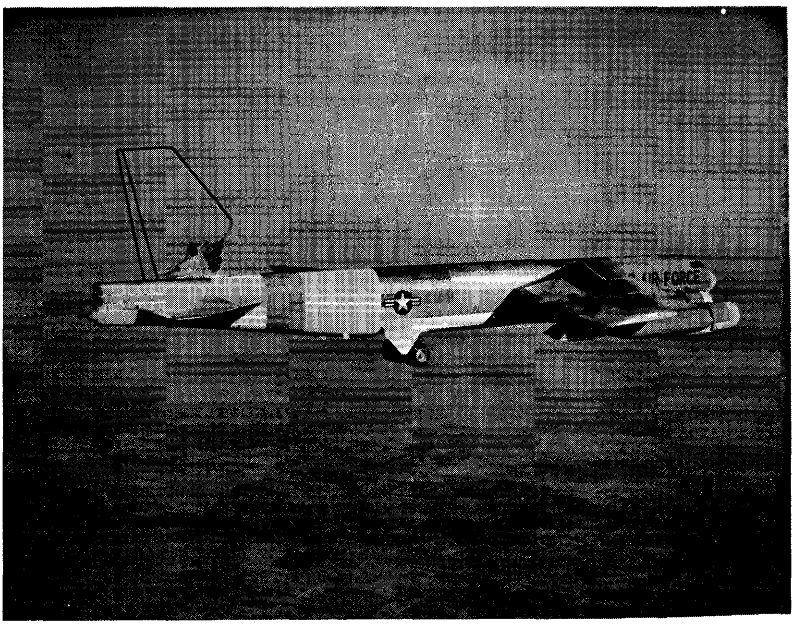


Figure 4. B-52 severe turbulence tests.

recovery was effected and the airplane returned to a safe landing, bringing home the test data accumulated from some 200 test points instrumented on the airplane. An inflight picture of the B-52 is shown in Fig. 4 with the portion of the vertical fin and rudder that sheared off outlined in black. The recorded acceleration and yaw response of the B-52 to the turbulence is given in Fig. 5 and the induced stresses in Fig. 6. Body station 1655 corresponds to the vertical fin location and fin station 135 was at the point

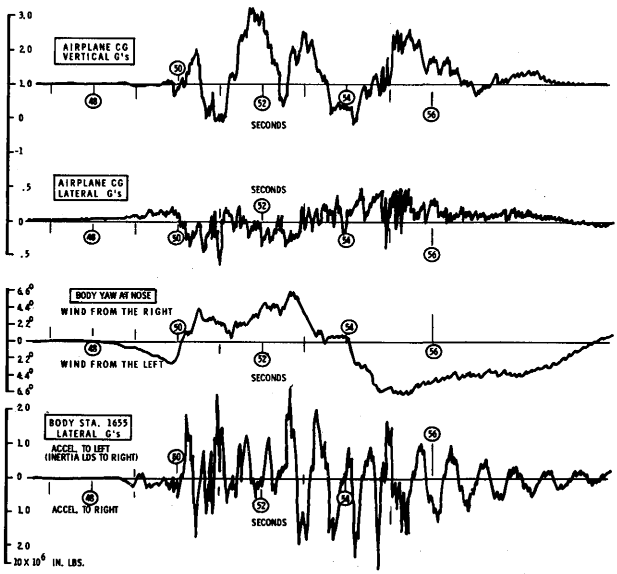


Figure 5. Severe turbulence effects, B-52 flight test.

of fin failure. Subsequent measurements by an instrumented Air Force F-106 interceptor in the same location of the B-52 incident recorded lateral gusts up to velocities of 120 feet per second providing new data at extended severity levels.

Because of the recognized limitations in the structural design of aircraft and the unknowns of the environment, program efforts are underway to install recorders in operational aircraft to obtain flight histories of accelerations, velocities and altitudes being encountered in operational flight. Such accumulated data will provide not only an advancing understanding of the flight environment and improved structural design but also provide a base for inspection, maintenance and operational procedures for increased flight safety. Because of stringent requirements, the development of a suitable VGH recorder has not yet been completed. The stringent requirements are imposed by the accuracy performance dictated by the

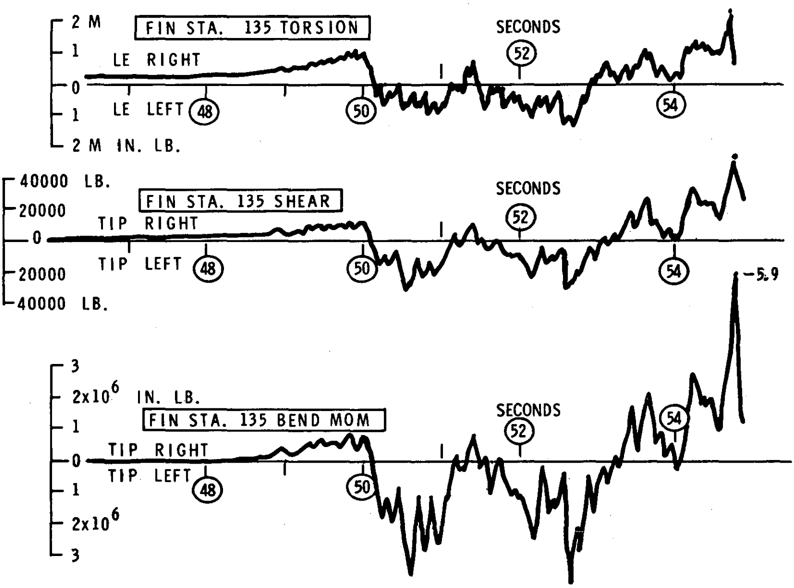


Figure 6. Severe turbulence effects, B-52 flight test.

problem, size and weight restrictions of installation and problems of obtaining necessary record time for allowable tape volume. Typical specifications are: 8 channels of recording plus time; maximum response, 12 cycles per second per channel; overall system accuracy including readout but not including sensor, 2 percent; record time 25 hours; size not to exceed 8 × 7¼ × 7¼ inches and weight shall be no more than 25 pounds. Several recorder developments have been supported by the Air Force over the past five years and although the requirements have been demonstrated to be a real challenge to the tape recording industry, good progress has been made. Concentrated effort is being applied to complete the development of the desired recorder as soon as possible. Meanwhile, statistical count recorders are being programmed for some aircraft installations, and oscillographs with manual readout are being used on a limited basis.

Advancements are needed toward improvement of sensors, recorders and other instrumentation through application of microelectronics and other advanced techniques. System logic and data processing innovations will be necessary to reduce the data processing load. A program analysis for the 8-channel VGH recorder indicates a need for 20 ground playbacks, 20 digital converters (analog to digital or digital to digital), 15 IBM 1401 computers and 2 IBM 7094 computers operated on a two-shift basis to machine process the estimated one and one-quarter million hours of data that would be acquired by the yearly operation of a fleet of 2,500 aircraft. Obviously, better signatures and mathematical modeling of the total problem are highly desirable to reduce this workload to more desirable proportions but its subtleties, particularly of the fatigue aspect, are highly complex. In particular, there is serious need for improvement in the understanding of the variant nature of the total environment and the treatment of aircraft or vehicle response to such environment. It is necessary that the response be considered not only in the light of the stress-strain integrity of the vehicle but also with respect to reduction of crew (and passenger) disturbances that would otherwise adversely affect mission success.

CORONARY-CARDIOVASCULAR RESEARCH

In a totally different field, that of coronary-cardiovascular research, one finds striking similarities to those problems just outlined under aircraft structural integrity. There is a system involved in each—one an airplane, the other a human being. Each is basically concerned with the welfare of a critical key of the system—structure in one, blood circulation in the second. Further, this welfare is directly related to the response of the total system to a complex environment not totally understood. In the research to obtain a better understanding of the problem toward improved welfare, there is a significant trend toward much more data accumulation and the processing of such data in an iterative process of knowledge acquisition.

The Cox Coronary Heart Institute is in the process of completion in Dayton, Ohio and expected to begin preliminary operation in April 1965. This new Institute for coronary-cardiovascular research is shown in Fig. 7. Director and principal investigator is Dr. G. Douglas Talbott, a pioneer in coronary research for a number of years. The Cox Coronary Heart Institute will be unique in the treatment and research on the coronary problem in its data processing approach. The Institute as a clinical research laboratory will provide 16 patient beds with each of the patients being "wired" in on-line for real-time monitoring by a data processing center. In addition to the 16 patient beds, there will be 10 research stations wired into the same data processing center to provide for off-line research analysis of coronary-cardiovascular data or for on-line processing of experimental research data generated at the station. The data process center therefore has the functions of generating alert and alarm signals for patient care, to provide a tool for medical diagnosis, for storing and retrieving information and finally, to facilitate analytical study to obtain a better understanding of the coronary-cardiovascular system and its functions or malfunctions. Another unique feature of the Institute program is the emphasis on a highly interdisciplinary approach.

Figure 7. Cox Coronary Heart Institute.

The data processing system is being designed and built around the GE/PAC 4000 Process Automation Computer by the General Electric Company. This is a sufficiently rapid and versatile computer for the job with its cycle time of 5 microseconds, add-and-subtract times of 16 microseconds, high-speed core memory with a storage capacity directly addressable up to 16,384 words, 24 bits per word, available on a modular basis, and other features to provide the required capabilities of both on-line data processing and off-line analysis and data correlation. Initially, five physiological parameters will be monitored—blood pressure (systolic and diastolic), heart rate, electrocardiogram (ECG), respiration rate, and body temperature. Later, this will be expanded to ten to include such further indicators as cardiac output, venous pressure (central and peripheral). Typical analog recordings are shown in Fig. 8, the top sawtooth being blood pressure, maxima of the sawtooth being systolic and minima diastolic; the center curve of high regularity being respiration rate and the bottom with its sharp peaks—the electrocardiogram. Digital and average data readout will be provided at the rate of 1 data point every 20 seconds. Such readout is illustrated in Fig. 9.

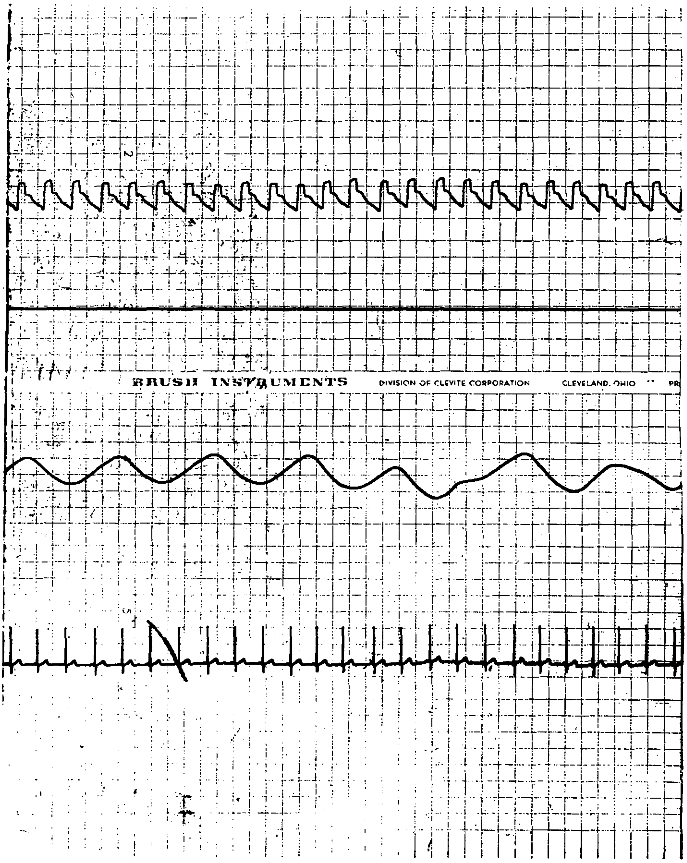


Figure 8. Typical blood pressure, respiration rate, and electrocardiogram recordings.

There are many new and interesting problems that are being encountered in the development of the instrumentation techniques for the Institute program. For instance, it becomes immediately apparent that new methods must be devised for coupling to the patient for long-time monitoring of such parameters as blook pressure, ECGs and the like, to prevent patient irritation either physiologically or psychologically. It's an entirely different situation to maintain patient comfort when he is "wired" to a data-processing center for days and weeks at a time in contrast to the usual observations that require only minutes. For electrocardiogram signals, the usual electrodes and skin contact methods are not satisfactory. A new conductive silicone with highly adherent properties was developed by Minnesota Mining and Manufacturing Company working in collaboration with the Institute to provide a suitable solution. The conductive silicone is simple to apply and connection is made by imbedding the bared end of the connecting wire in the silicone.

The problem of blood pressure monitoring has been more difficult of solution. The direct-pressure coupling by intravenous or arterial catheter, although accurate and positive in calibration, causes problems in application and is of obvious irritation to the patient. Several types of external pickups have been investigated but none has been found completely satisfactory as to required sensitivity and accuracy, and simplicity of application and calibration.

Again, inherent in this program is the basic problem of data processing. Vast amounts of data will be accumulated, processed, stored, retrieved, correlated, and otherwise analyzed for better understandings, signatures and models of system functions and malfunctions. System equations are obviously complex because of nonlinearities and numbers of variables involved, time variant in random and explicit combination. In reviewing the Institute program in connection with this paper, it was interesting to

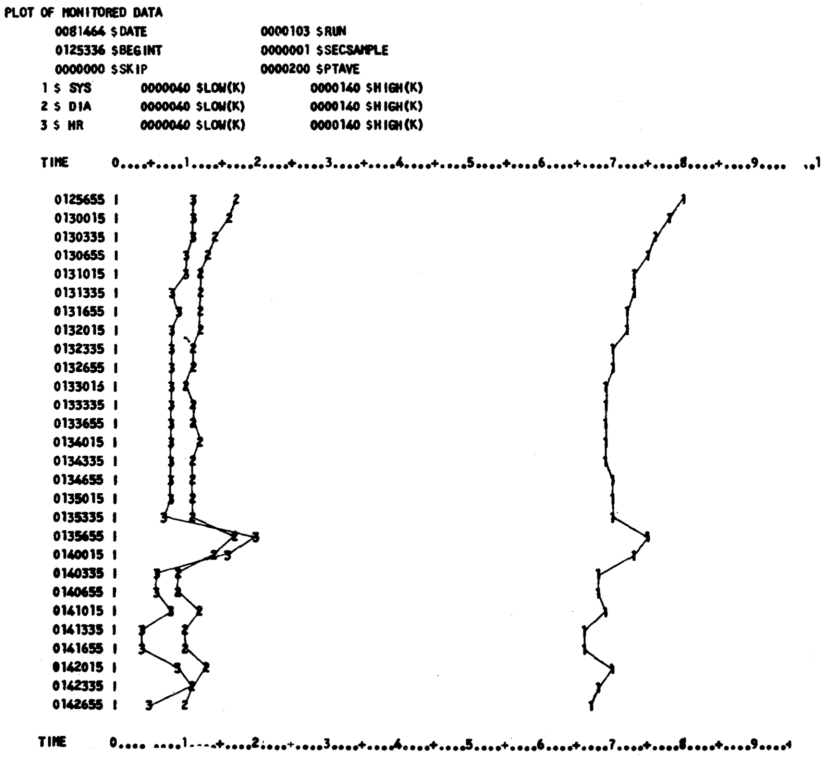


Figure 9. Computer readout of cardiac data.

note the use of advanced mathematics of variance developed in connection with the aircraft vibration and flutter problem under investigation for the Air Force.[7,8] An iterative program will provide improved signatures for medical diagnosis, alert and alarm criteria for patient care in the hospital and ultimately an understanding and model of the total coronary-cardiac system. Implicit as a possible trend is the use of recordings of selected diagnostic parameters for a period of typical individual activity to provide a more adequate data base for assessment of his physical well-being. Suitable sensors and miniature tape recorders of sufficient store capacity and high degree of portability can be anticipated from the advancing art. This approach could be used by the family physician in collaboration with established clinical centers to maintain a closer check of normal well-being.

THE TACTICAL WEAPON SYSTEM

A third and quite different type of process-related data-handling problem is involved in the typical military weapon system. With increasing demands of the military environment to perform against an increasing target complex with a wide arsenal of weapons under the extremes of combat and battlefield conditions, the performance of the advancing weapon system must be pushed to the limit that the state-of-the-art will permit. Quick reaction and alertness to rapid change, short time constants of maneuver, versatility of action, and quick turnaround has forced a high degree of sophistication with a maximum of instrumentation and automation to assist the crew in mission execution. At the same time, these demands must be traded off against the basic requirements of simplicity and minimized resource costs to provide operational and logistic practicability and effectiveness. Corresponding exponentially increasing demands have been placed on data processing for intelligence and communications, targetting, display and action. Throughout there is the interplay of manual, machine and man-machine approaches in the function to be performed.

The latest fighter-bomber of the Air Force, shown in Figs. 10 and 11, is the F-111. It's designed to provide high versatility and flexibility through use of variable geometry in its aerodynamic configuration—the wing sweep can be changed in flight. The extremes of full forward and rearward sweeps are shown in the figures. The required performance is thereby achieved for a variety of takeoff and landing conditions as well as speeds and altitudes of flight.

In a like fashion, the instrumentation and equipment must provide for a variety of functions if the airplane is to perform its job. Figure 12 lists all of the functions beginning with those required for flight-vehicle operation such as flight control, next the functions essential to the performance of

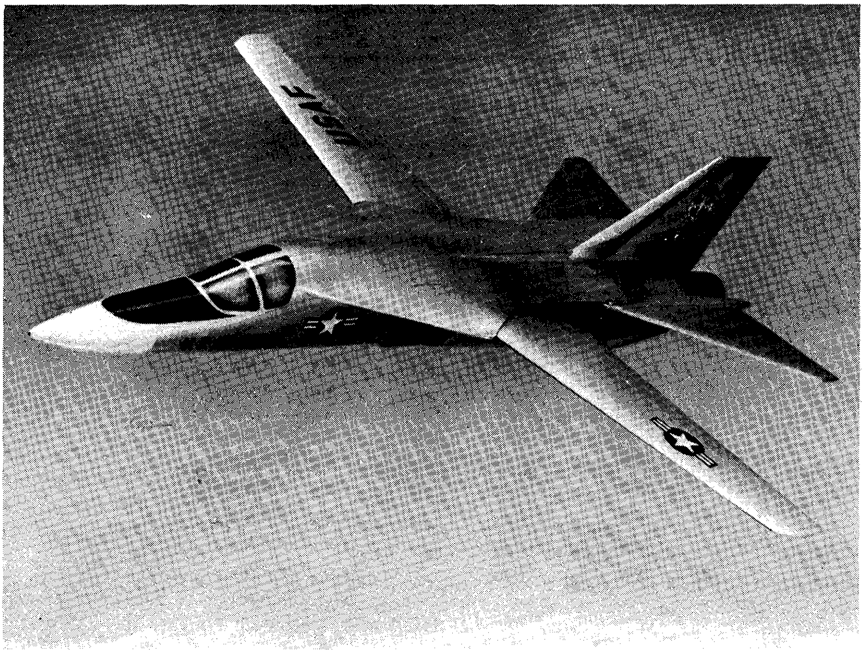


Figure 10. F-111 fighter bomber—minimum wing sweep.

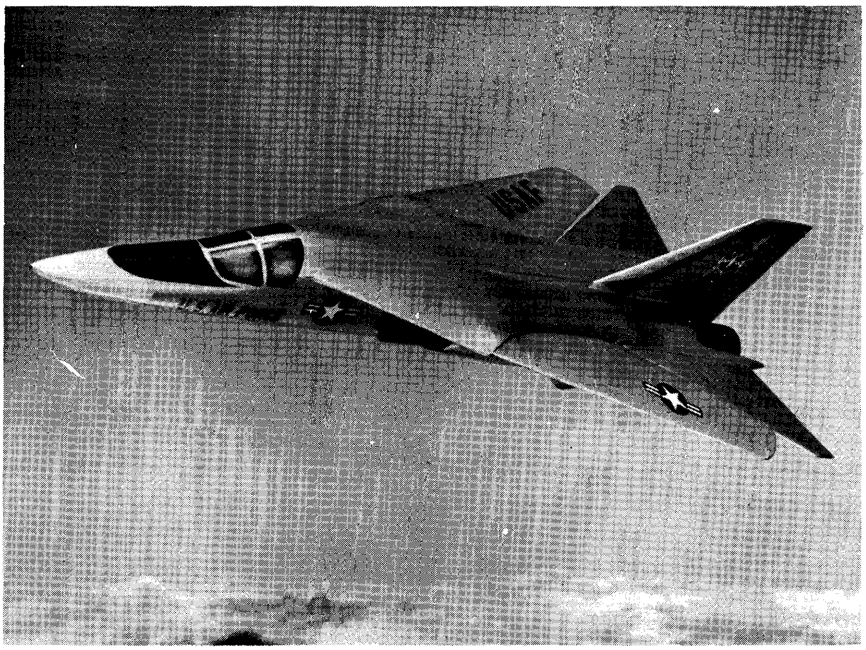


Figure 11. F-111 fighter bomber—maximum wing sweep.

- **FLIGHT VEHICLE OPERATION**
  - AIR DATA PROCESSING
  - FLIGHT CONTROL
  - FLIGHT INSTRUMENTATION
  - POWER CONTROL-PROPULSION
- **MISSION OPERATION**
  - TERRAIN AVOIDANCE
  - NAVIGATION
    - RADAR
    - DOPPLER
    - INERTIAL
    - RADIO
  - POSITION REPORTING
  - COMMUNICATIONS
  - DISPLAY AND CONTROL
  - TARGET ACQUISITION
  - BOMBING
  - AIR INTERCEPT
  - WEAPON CONTROL
  - ELECTRONIC WARFARE
- **CHECKOUT AND CALIBRATION**
  - SELF TEST
  - SYSTEM EVALUATION

Figure 12. System functions—tactical avionics.

the military mission such as target acquisition and weapon control, and finally, the functions of checkout and calibration to assure reliability and readiness-to-go. Every item listed requires a major subsystem. Space and weight conflicts of installing all of this equipment within the airframe are obvious. Equipment design must be modular since it is not possible and in many cases not even desirable to provide for all of the subsystem functions for every flight. Quick exchange is a necessary feature to adapt to the mission at hand.

The data-processing implications are clearly evident—each subsystem must handle large quantities, usually in real-time, and must interface together and with the crew for required performance of the total system. Shown in Fig. 13 (see Ref. 9) are typical block diagrams for the air-data sensing, flight instrumentation, navigation and flight-control functions of the total avionic system. There has been a significant advancing trend in the use and application of digital data processing for all of the subsystem functions and the total avionic system to facilitate the data-handling problem and requirements of system integration.

This trend has generated issues as to the proper logic of data processing for optimum system design. Should the system be highly centralized with a single general-purpose computer as the heart of the system, should it be highly decentralized with a separate computer for each function or is there a better approach somewhere in between these extremes? Fall-back modes of operation must be provided in case of equipment failure or battle damage. There is an obvious need of redundancy for reliability.

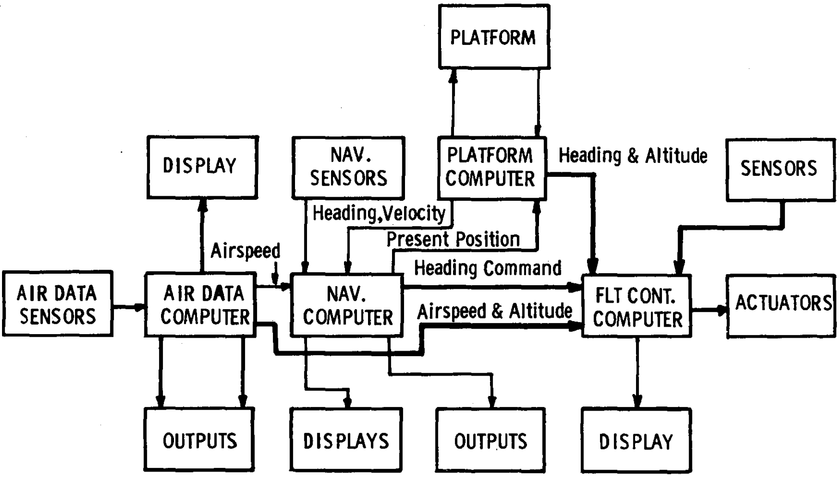


Figure 13. Integrated microelectronic avionic system.

Certain functions such as targetting, the selection and manipulation of tabular data, system mode control and self-test require or favor the general purpose computer. For other functions such as tracking and weapon control where only a simple updating of the problem is required, the Digital Differential Analyzer (DDA) is best suited. These and other aspects have led to proposals of hybrid approaches (David H. Blauvelt, Ref. 9) as well as a variety of other logic approaches to the data-processing problem. There is still another factor that the Air Force must consider—that of facilitating competitive procurement. Certain standardizations as to language, format, cycle times and the like become necessary considerations to permit subsystems supplied by different vendors to be integrated together into a totally operative avionic system. Further, it is highly desirable to update the performance of any subsystem function from the source that has achieved a significant advancement without required major changes in the rest of the system. The degree of standardization that will be required and the involvements connected therewith are currently under study.

PROGRESS IN BIONICS

RESEARCH NEEDS

Having examined some problem areas in need of research attention, it will be the further purpose of this paper to review promising avenues of investigation. In the previous discussion, needs have been identified for advanced sensors, high-density storage and retrieval, improved techniques and logic of data processing, advanced tools of analytical study and inquiry particularly in intimate symbiosis with man, more adequate treatment of complex problems of variance and general advancement of man-machine relations. In the material that follows, highlighting progress being made in bionics research, considerable correlation will be apparent to the needs just outlined. This is to be expected since it is one of the primary objectives of bionics to do research on living systems to gain insight and knowledge of their sensory and data-processing capabilities for application to our general technology. More advanced analytical tools will certainly come about by a better understanding of man's intelligence function and the tailoring of machines and equipment to assist that function. A deep probing of the living system will lead to a better understanding of complex problems of variance. The improved man-machine relationship will be not only a direct result of bionics research but more subtle payoffs can also be anticipated.

THE AIR FORCE PROGRAM

For this paper, progress trends will be extracted from the efforts of the 6570th Aerospace Medical Research Laboratories and the Air Force Avionics Laboratory at Wright-Patterson Air Force Base. Other Air Force research efforts are being carried out by Rome Air Development Center, Rome, New York, and the Air Force Cambridge Research Laboratories and the Air Force Office of Scientific Research of the Office of Aerospace Research.

The total program in the Wright-Patterson complex represents a current effort supported by contract funds of $1.7 million annually and a total laboratory staff of 31 people. These efforts are approximately equally divided between the research interests of the 6570th Aerospace Medical Research Laboratories and the applied research and applicational interests of the Air Force Avionics Laboratory. A good summary picture of the program efforts is provided by the following project breakdown.

6570th Aerospace Medical Research Laboratories

Two Projects

7232—Research on the Logical Structure and Function of the Nervous System

Objective—The objective of this project is the discovery and analysis of organizational and functional features of nervous systems which contribute to their ability to collect, store, and utilize information. Principles, methods, and techniques are sought, described, and developed. The methods are experimental and theoretical. Results will be new theories, more lucid descriptions and expanded understanding of communication, control, memory, pattern recognition, data selection, and data transfer and will expand the basis for engineering bionics and contribute to improved computer technology.

Subtasks
1. Functional Parameters Controlling Biological Reflexes.
2. Processing of Auditory Information.
3. Processing of Visual Information.
4. Neural Network Investigations.
5. Neurophysiology of the Central Nervous System.

7233—Biological Information-Handling Systems and Their Functional Analogs

Objective—The objective of the biological phase of the bionics research program is to select those features of living systems which excel present technological capabilities in one or more parameters; to discover and derive the biological principles and processes responsible for their superiority; and to develop mathematical and logical models, methods, and procedures appropriate for the description and theoretical understanding of highly complex biological systems in terms useful to design engineers. In essence, living organisms are studied as engineering prototypes and an attempt is made to bridge the gap between the biological and engineering disciplines.

Subtasks
1. Auditory Processing of Speech.
2. Neural Network Simulation.
3. Advanced Mathematical and Computer Methods in Biological Data Processing.
4. Theory of Pattern Recognition.
5. Research on Theory of Adaptive Processes.

Air Force Avionics Laboratory
One Project
4160—Engineering Bionics

Objective—It is the objective of this project to optimize, in a formal mathematical and physical sense, knowledge of the functional abilities of organic systems and to demonstrate the feasibility of translating this knowledge into dependable and efficient hardware to satisfy Air Force requirements.

Subtasks
1. Primary Elements and Techniques for Engineering Bionics.
2. Man-Machine Interface Phenomena.
3. Bionic Subsystem Techniques.
4. Bionic System Techniques.

5. Experimental Synthesis of Bionic Systems and Subsystems.
6. Experimental Analysis of Bionic Systems and Subsystems.
7. Growth, Form, Structure and Function in Bionics.

There are several other particular aspects of the Bionics effort at the Wright-Patterson complex that should be noted. First there has been a very deliberate interdisciplinary approach in the activity in recognition of the nature of the research and technology involved. This has been stressed by management in its policy, planning and direction. The emblem of Fig. 14 symbolizes this emphasis—the scalpel of the life sciences being joined with the soldering iron of engineering by the integral sign of mathematics. Group efforts exploit the multidiscipline attack with augmentation of applied disciplines as manpower ceilings permit. Individual interdisciplinary development is also highly encouraged by graduate training opportunities.

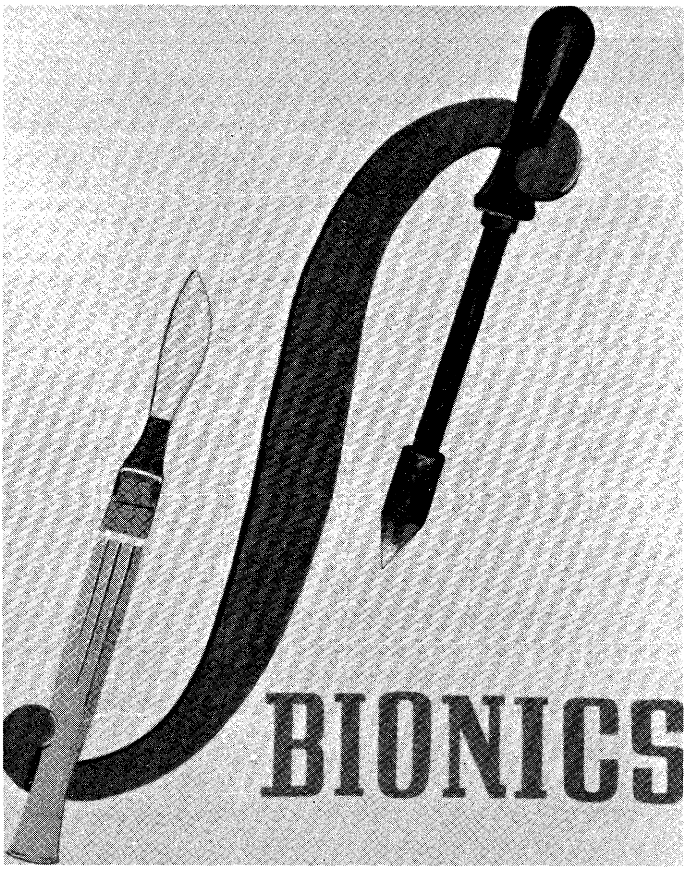


Figure 14. Bionics program emblem.

A further characteristic of the effort is the deliberate division of labor as to motivation. As the program breakdown indicates, the activity of the 6570th Aerospace Medical Research Laboratories is research directed whereas the interest of the Air Force Avionics Laboratory is applicationally oriented. Mathematical modeling is the common bond since it is the essential result of research and the beginning point of application. There is therefore a deliberate concentration on this bond in the development of mathematical models and signatures. There are of course other interface relations in the collaborative work relations between the two groups. These functional work relations are shown in Fig. 15.

As a final point, note should be made of the data-processing developments arising from the nature of the research and the emphasis on mathematical modeling. A very advanced real-time digital data-processing system for biological research[10] has been developed by the 6570th Aerospace

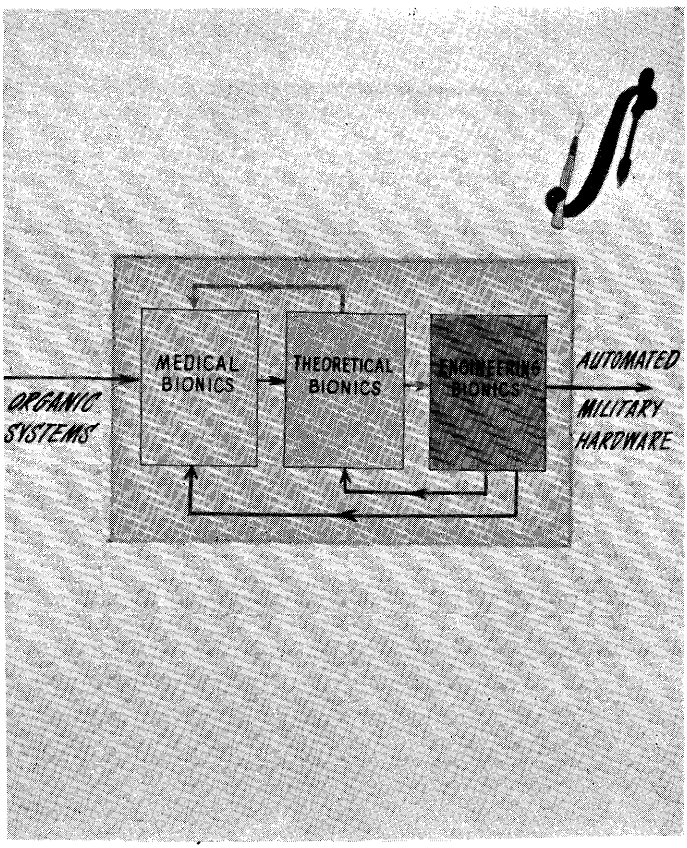


Figure 15. Organization of bionics effort.

Medical Research Laboratories and is shown in Fig. 16. The Central Processor, Digital Equipment Corporation, PDP-1, operates with a word length of 18 bits in fixed-point arithmetic. Core memory of 4,096 words is provided expandable to 65,536 words in units of 4,096. Cycle time is 5 microseconds and carries out arithmetic and logical operations in multiples of the memory cycle. Data can be entered directly into the core memory by bypassing the input-output register at rates up to 200,000 (9-to-18 bit) words per second. These speeds give virtually instantaneous

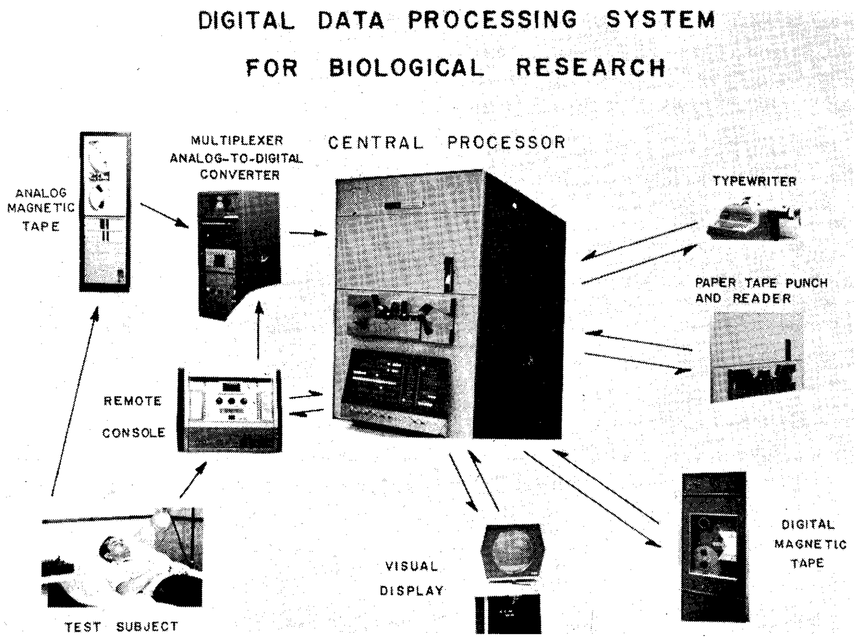


Figure 16. Digital data-processing system for biological research.

response to biological data where times are generally measured in milliseconds. The computer can carry out 100 additions or 50 multiplications during a single one-millisecond pulse from a nerve cell. By provision of flexible programming features and a wide range of arithmetic and logical machine operations coupled with the peripheral equipment shown in Fig. 16, a very versatile data-processing system is achieved. On-line, real-time operations are available for a wide variety of experimental approaches as well as an extensive list of off-line programs for data analysis. These include:

Statistical Analysis

Statistical package—mean, variance and standard deviation

Linear regression

Analog Signal Analysis
- Cross-autocorrelation
- Real-time cross-autocorrelation
- Correlation
- Fourier and Laplace transforms
- Function generator
- Transfer function
- Average response
- Data editing
- Zero crossing
- Vector magnitude
- Power spectra

Pulse-data analysis
- Occurrence histogram
- Moving average rate
- Average pulse interval

Full on-line and off-line readin and readout, control and display facilities are available at the experimental test stations by means of the Remote Control Console and Visual Display units, a feature most essential for experimental flexibility. This system development in conjunction with the program of the Cox Coronary Heart Institute provides a significant and interesting trend picture.

REPRESENTATIVE PROGRAM EFFORTS

It is convenient to consider the living system and the bionic program in terms of the functional breakdown listed in Fig. 17. At the input end of the sensor, the transducer transforms the stimulation, be it heat, light, sound, pressure or other, into the signal to be processed and transmitted along the nerve network. The property filter performs filtering and other selective modification to begin data reduction at the stimulation end of the system. Under the cognitive center, we include all of those data-processing functions which derive from the input information, the decision and action outputs. These in turn initiate actions, reaction and control functions which constitute the effector net of the system.

Sensors

Primary emphasis at the present is on the study of the "property filter" and signal processing characteristics of the visual, auditory and tactile perception functions of the living system—man and lower orders of animals. This is not to say that research on the transducer functions of rods, cones, cilia (cochlea) and the like are not worthwhile but at present, the property filter functions are lesser understood.

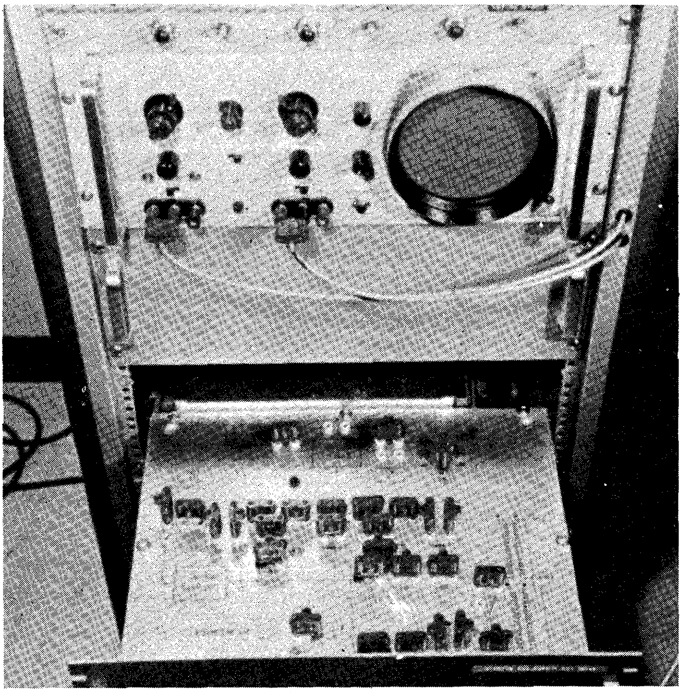


Figure 17. Classification of bionic functions.

Cochlea of the Ear

Shown in Fig. 18 is one of the first and most complete electronic analogs of the human ear. It is the result of contract efforts by the Aerospace Medical Research Laboratories with the Santa Rita Technology, Inc., Menlo Park, California, principal investigator Dr. J. L. Stewart and associates E. Glaesser and W. F. Caldwell.[11,12] The analog includes the external and middle ear, the cochlea, and part of the neural structure of the cochlea and the higher auditory centers of the central nervous system. Tests of the analog and functional components have established important similarities to the human ear. Certain psychoacoustic characteristics such as mutual inhibition and phasic-tonic neural behavior are not modeled, nor does the analog provide for middle-ear reflexes and fatigue. Subsequent study and modification is required to more closely approach complete simulation.

At present, the analog is being used to further the understanding of the functioning of the auditory system in speech recognition and in the analysis of communications program improvements. The analog is convenient to use and modify for experimental purposes because of its functional modular design.

Electronic Model of the Frog Retina

In the same manner, an electronic analog (Fig. 19) of the frog's retina has been fabricated based on research investigations of J. Y. Lettvin, H. R. Maturana, W. S. McCulloch and W. H. Pitts.[13] This was accomplished by M. B. Herscher and T. P. Kelley[14] of the Radio Corporation of

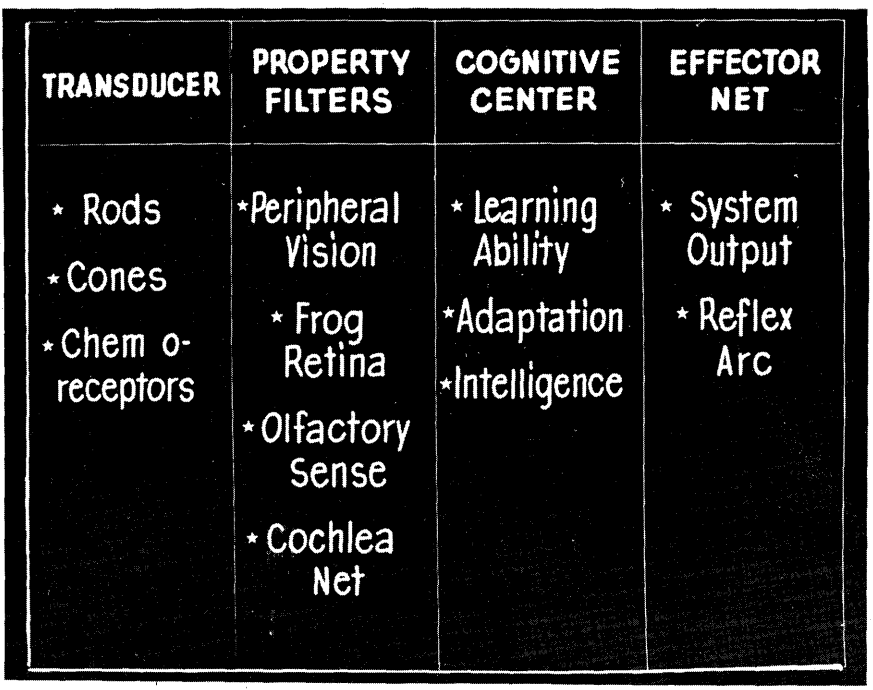


Figure 18. Electronic analog of the ear.

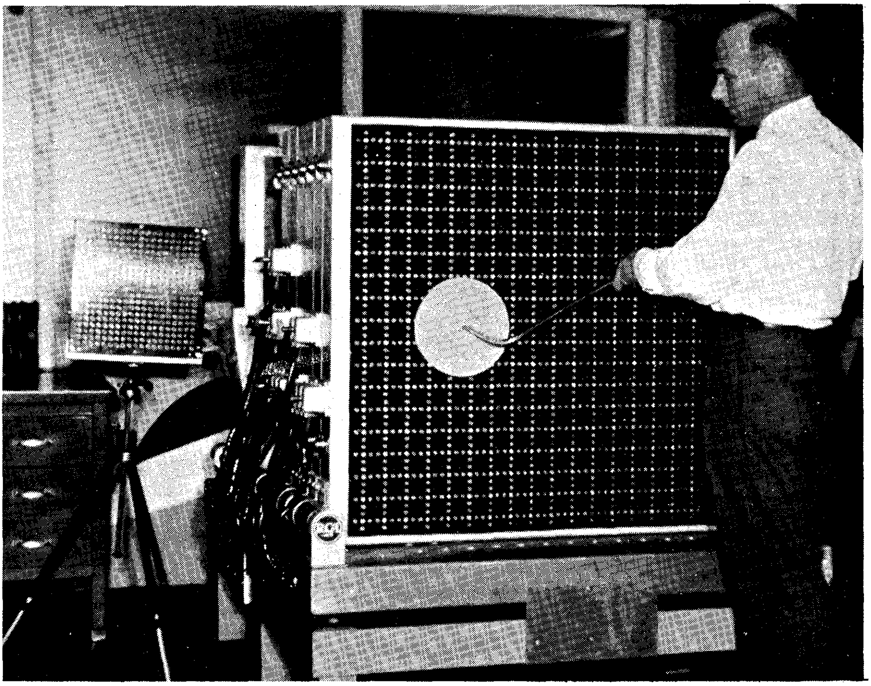


Figure 19. Electronic model of frog's retina.

America under contract with the Avionics Laboratory. It is being used as a simulator in further study of the property filter characteristics (discrimination, motion detection, resolution, etc.) of the retinal-optic nerve system with a potential payoff of improvements to surveillance and target tracking systems—optic, infrared, radar, and the like.

Tactual Perception

Another very interesting investigation in the sensory area is the work being done by J. C. Bliss and H. D. Crane of Stanford Research Institute on tactual perception.[15] This effort is supported by the Air Force (Avionics Laboratory), National Aeronautics and Space Administration (NASA) and National Institutes of Health. A 12 × 8 matrix of air jets as shown in Fig. 20 is controlled by a CDC-180A computer to provide a spatial and temporal pattern of air jet stimulation of the hand or other part of the body to communicate with the individual. The arrangements of instructor and control panel, tactile stimulator, display and control computer is shown in Fig. 21. The specially designed electromagnetic control valves for the air jet stimulator can operate at frequencies up to 200 CPS to produce an air jet having a rise and fall time of one millisecond and a duration of three milliseconds. The CDC-180A computer is

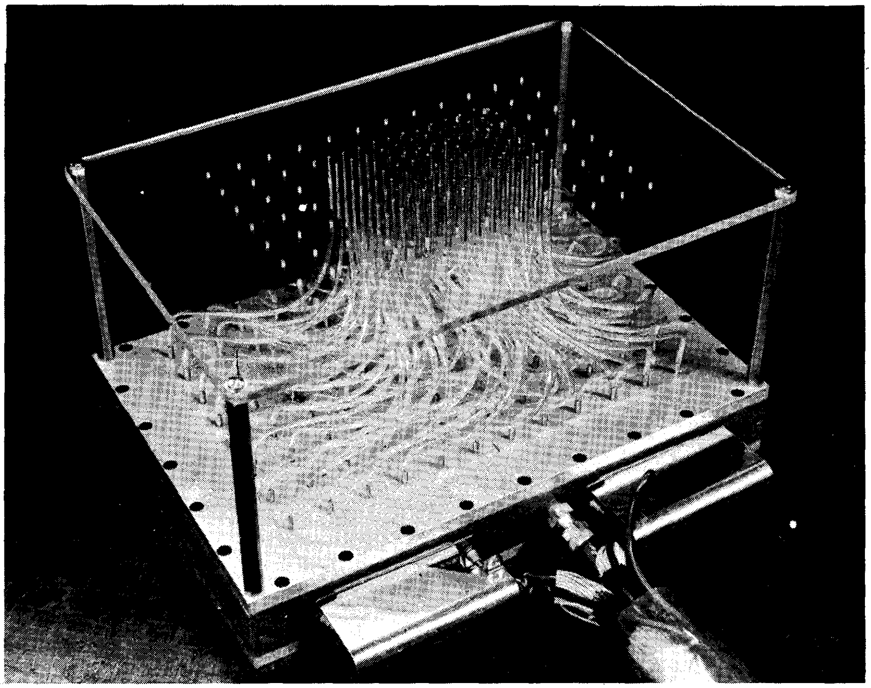


Figure 20. Tactile stimulator.

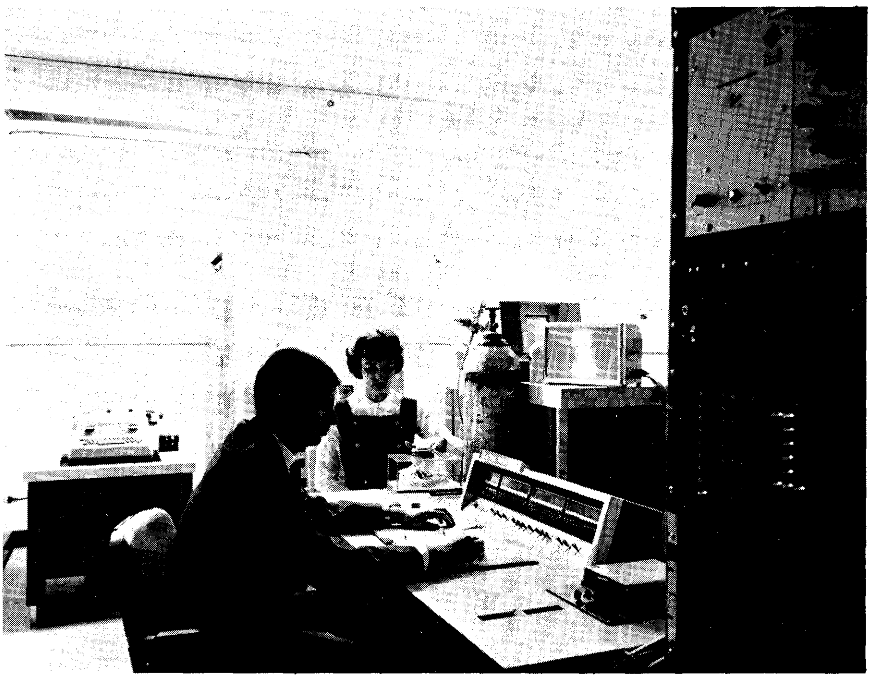


Figure 21. Equipment arrangement—tactual perception experiments.

used in real-time to store the stimulus patterns, scan them according to various temporal modes, output the scanned stimulus patterns, record and tabulate the subject's response and analyze the recorded data. The overall system provides a high degree of flexibility and facility for the conduct of many different kinds of experiments in psychophysical research and the development of tactual languages. Air Force interests are in providing additional channels of data input to the individual in communication, command and control functions and a more intimate relation between man and machine. An experimental set-up for investigating tracking functions is shown in Fig. 22 with the tactile stimulation being applied to the forehead.

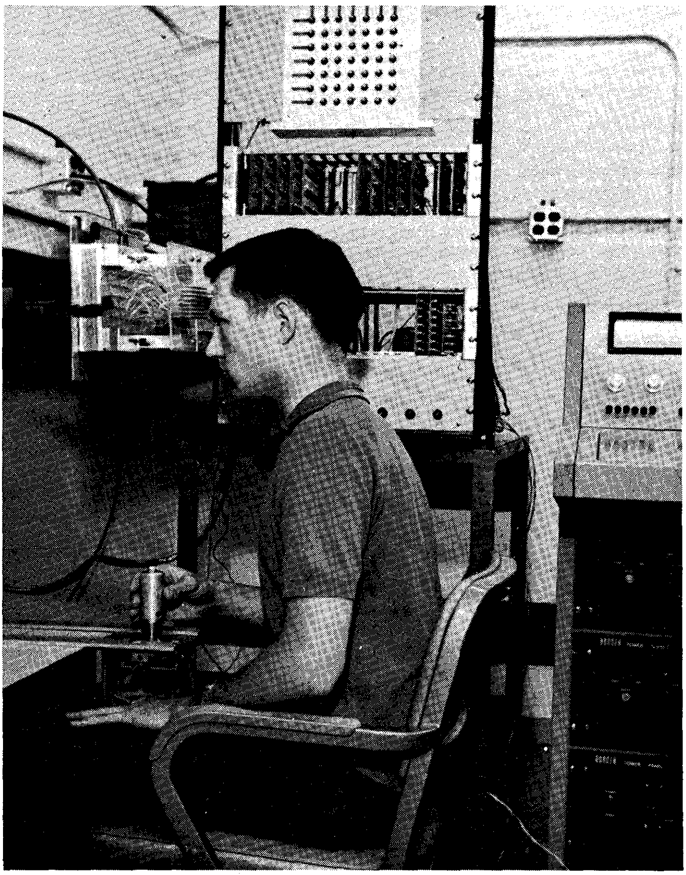


Figure 22. Tracking experiments with tactile inputs.

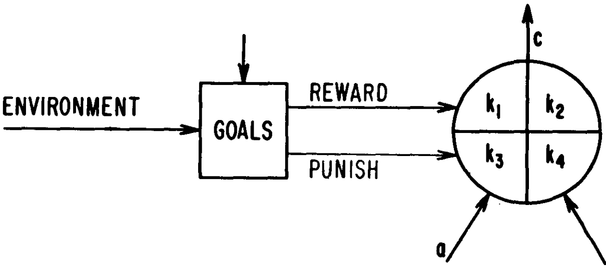
SELF-ORGANIZING MACHINES

More effort than in any other area of bionics is being applied to the intelligence or cognitive center function of the classification chart of Fig. 15. Pattern-recognition systems, perceptrons, adaptive, self-organizing, heuristically programmed learning machines, automata, artificial intelligence and thinking machines—the total list has indeed become staggering. The technical literature has mushroomed with large numbers of papers on a wide diversity of subjects representing the upsurge of activity in this area. One can very well ponder the question as to whether we haven't really gone overboard with program balance in serious jeopardy. However, when one considers the explosive growth of the data-processing problem across the spectrum of advancing technology of which only a small part has been highlighted in this paper and the magnitude of the programming load that has resulted, one concludes that the concentration of effort on self-organizing or adaptive logic machines is well justified. A further reaction is that there needs to be a "tightening" up of the program. We need to better define and classify the problem areas that urgently need and would benefit from advanced machine assist so that our research can be more responsive to the problem classes. We have in too many instances inventions in search of problems.

Air Force effort at the Wright-Patterson complex (Avionics Lab) in this area has been largely on statistically conditioned and self-organizing binary logical networks in learning systems using the reinforcement principle. Initiated by contract with Melpar, Inc., Falls Church, Virginia, in 1960, this work has been under the program direction of Dr. E. B. Carne of that organization. The network systems are based on the use of the artron (artificial neuron) shown in Fig. 23, where *a* and *b* are inputs (dendrites), *c* the output (axon) and *R* and *P*, biasing or conditioning signals.[16] Output *c* is some logical function of inputs *a* and *b*. There are sixteen possible states or gating functions which the artron may assume. Teaching the artron is essentially a process of changing the probability of existence of any state or states. This is done through the reinforcement and punish channels, *R* and *P*, an input at *R* increasing the probability of a given state recurring, an input signal at *P* decreasing the probability.

The generalized artron is capable of learning to perform Boolean function operations and of implementing decision processes. Although it is not intended to closely approximate the functioning of human neural cells, systems of generalized artrons are capable of simulating behavioral patterns.

An artron network controlling a maze runner has been designed and built[16] as shown in Fig. 24. As the maze runner proceeds from the starting to the finish point, it is required to make a decision at each intersection of



$$c = k_1\bar{a}\bar{b} + k_2\bar{a}b + k_3 a\bar{b} + k_4 ab$$

WHERE THE k_i ARE STATISTICAL VARIABLES (BETWEEN ZERO AND ONE) UNDER THE CONTROL OF THE GOAL.

Figure 23. Artron.

the maze as to a right or left turn or to proceed straight ahead. Subsequent experience is used to determine whether the right or wrong decision was made at a particular intersection and stored for future reference. The maze runner will make many mistakes in its initial attempt to learn the maze but will eventually find home. Fewer mistakes will be made on successive runs by success-and-failure conditioning of the artron control through coding of the reward and punishment signals on the basis of prior experience. A variety of simple learning experiments have been carried out to demonstrate the learning characteristics of the artron network.

Further theoretical and simulation studies of generalized machine learning have been carried out for two types of networks, the artron network and the self-organizing binary logical network.[17] General conclusion has been that machines can be designed and constructed that are capable of learning efficiently. Goal criteria have also been examined and computer simulation comparisons made of artron and self-organizing binary logical networks of varying complexity.

Another extension of this work has been the design and construction of a large Artificial Nerve Net (LANNET) by Melpar[18] (Fig. 25). The self-

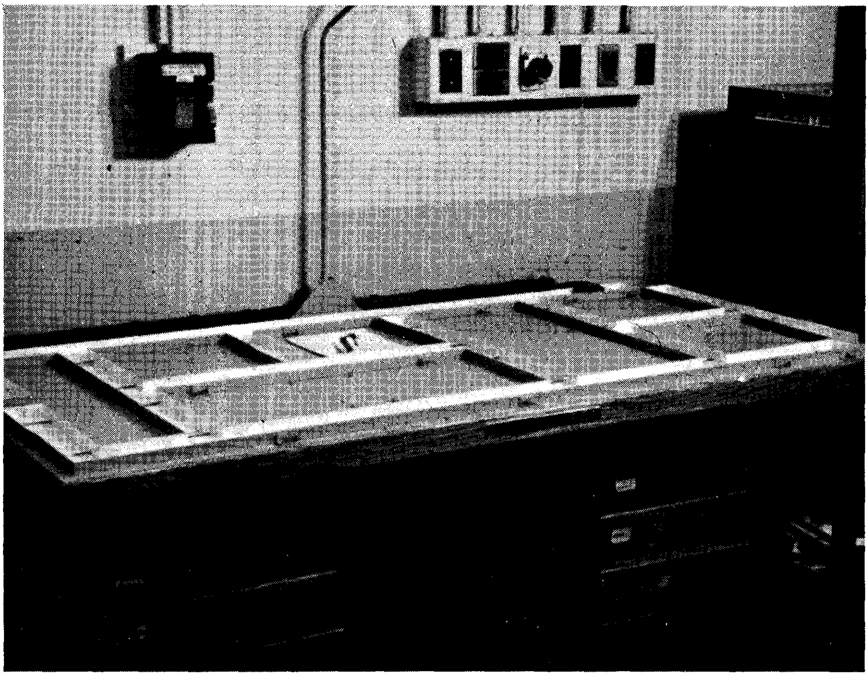


Figure 24. The maze runner.

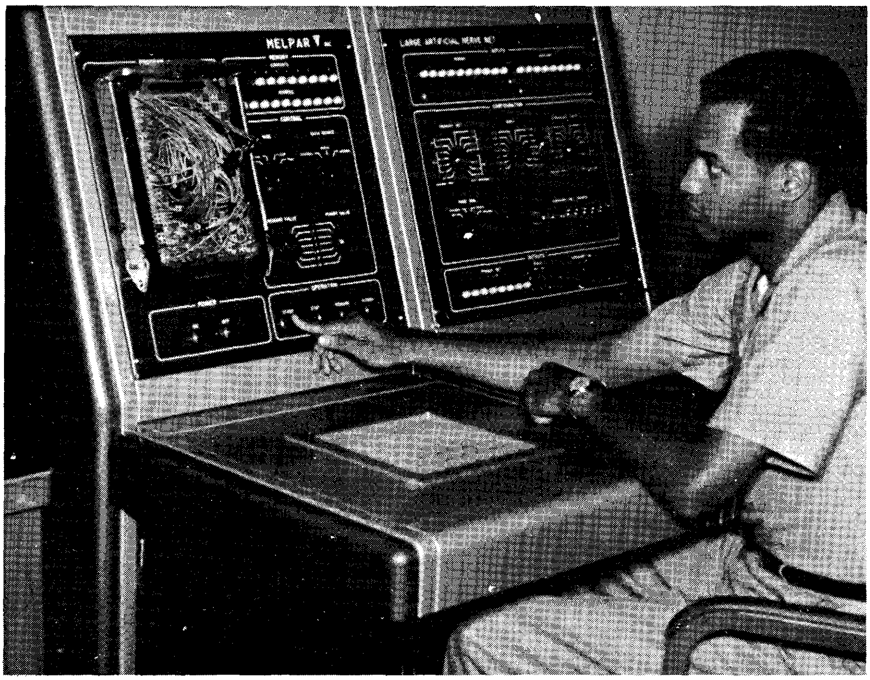


Figure 25. Large artificial nerve net (LANNET).

organizing binary logical network is used in this case as the primary component. The learning system is a 1,024 decision element network with a general purpose program to enable the operator to simulate a large number of problems to study machine learning. A variety of network combinations can be provided by the plugboard and switch arrangement. A 1,024 × 8-bit random-access memory is provided which can all be allocated to generating primary learning net outputs or divided between the primary learning net and subsidiary learning nets. Goal configurations available for training the primary learning net are as follows: Fixed, any one of the subsidiary learning nets, biased random, majority vote, priority and partitions.

LANNET can be used to study a number of different complex biological functions. Among these are the maze problems, classical (Pavlovian) conditioning, instrumental conditioning and depth perception. Air Force use is for continued study of machine learning and evaluation of possible problem application.

Control

In providing the action, reaction, motion, control, or other outputs at the effector-net end of the system, there are already available a wide host of electromechanical and servo control type devices to do the job. Living system capabilities offer further attractive features in areas of dexterity, versatility, accuracy, motion precision, or other performance qualifications. There are also situations where man requires machine assist to perform his normal function under conditions of environmental stress or physiological impairment. Two interesting developments have been achieved—the artificial muscle and the myoelectric servo control.

The Artificial Muscle

Study and development of muscle substitutes[19] has been carried out by the Laboratory for the Study of Sensory Systems, Tucson, Arizona, under contract with the Avionics Laboratory. Principal investigator in the research has been H. A. Baldwin. A composite structure membrane which analogs the functioning of the skeletal muscle is shown in Fig. 26. The required functioning of the membrane is obtained by the combination of two materials of widely different moduli of elasticity in its makeup—i.e., essentially inelastic fibers imbedded in an elastic base. Experimental prototypes were made of extremely fine fiber glass and natural rubber latex. When the composite membrane cylinder is inflated (Fig. 27) by low pressure air (2 to 10 psi), a contractive pull up to 100 lb is produced on the attachments to the ends of the cylinder. The force-distance properties of the device have been shown to analog those of the skeletal muscle. Further studies of the properties and applicational characteristics have been carried out in muscle engines such as shown in Fig. 28.

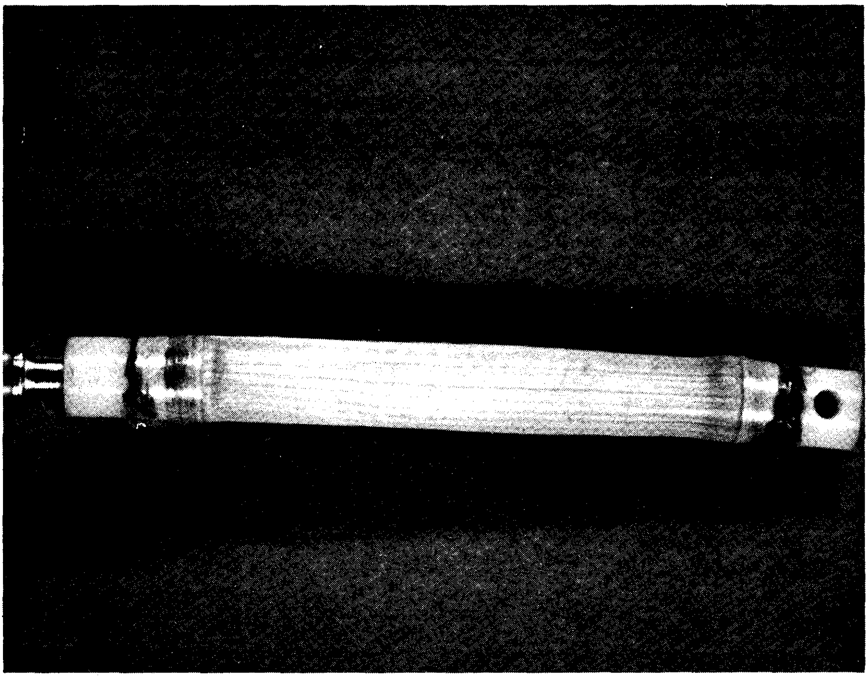


Figure 26. The artificial muscle—normal.

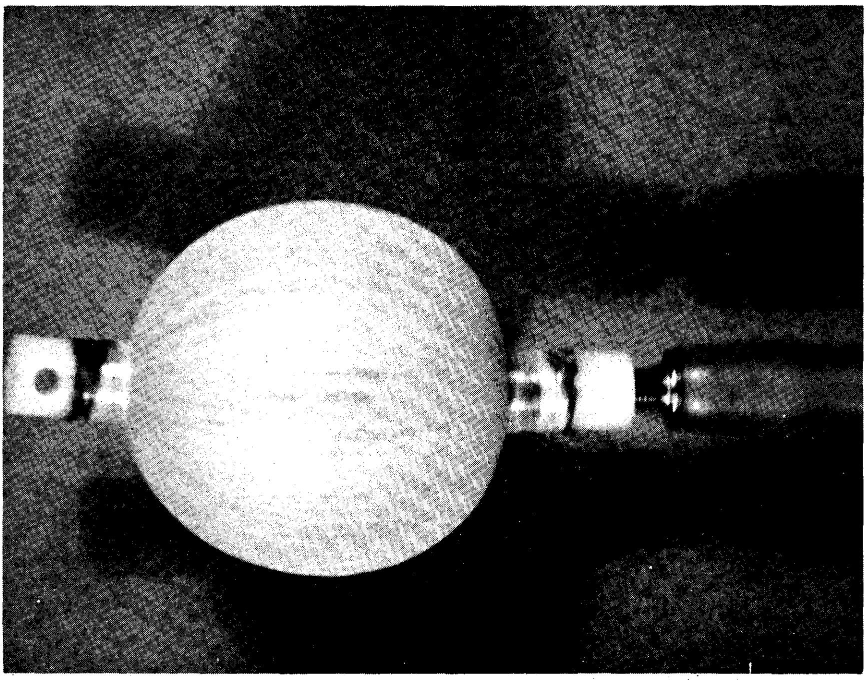


Figure 27. The artificial muscle—distended.

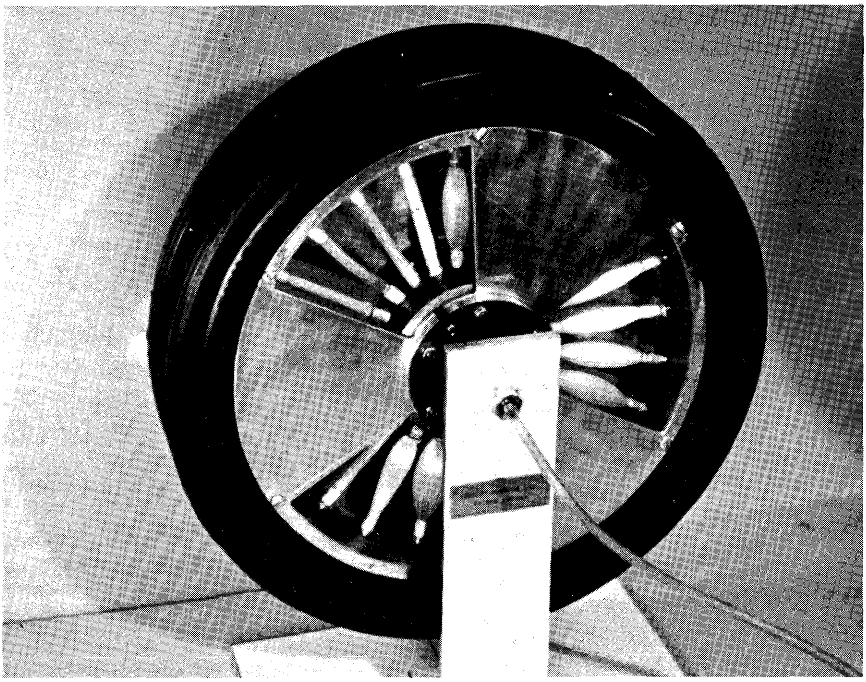


Figure 28. Muscle-powered wheel.

The sphincter muscle has also been analoged to provide a fluid amplifier sphincter valve. Pressure applied to a side tube operates on a nylon and rubber membrane concentric in a main tube to control flow in the main tube.

The artificial muscle and the sphincter valve in combination have a variety of possible applications in hydraulic and pneumatic control systems such as propulsion and flight control.

Myoelectric Servo Control

Of closely associated interest, studies on myoelectric servo control[20] are being carried out by Spacelabs, Inc., Van Nuys, California, again on contract with the Avionics Lab. Principal investigator is G. H. Sullivan, M.D., of Spacelabs, Inc. Important contributions have also been made by J. Lyman and F. C. DeBiasio, Biotechnology Laboratory, University of California, Los Angeles. In the experimental setup, myoelectric signals are picked up by small electrodes placed on the arm and shoulder muscles and after amplification are processed by a logic computer to control a servo-boosted arm-support sling. The arrangement of pickups and small amplifier worn as a chest pack are shown in Fig. 29. The test rig, arm-support sling and subject are shown in Fig. 30. Logic of the control com-

Figure 29. Myoelectric pickups and amplifier.

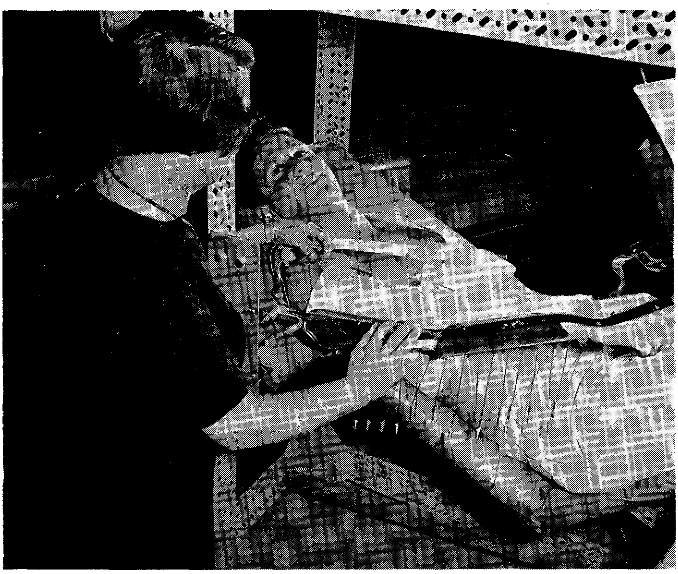


Figure 30. Myoelectric servo-boosted arm-support sling.

puter was designed to assist the subject by means of the arm-support sling in six arm movements—up, down, in, out and rotation (supination and pronation), in the manipulation of the controls on the control box (Fig. 30). *G* forces of vehicle motion can be simulated by the tension wires applied to the sling. A variety of experiments have been carried out to study the use of the myoelectric servo-boost system to assist an operator in carrying out control manipulations under high accelerative-decelerative motion conditions. Thus a significant advance has been made in the use of myoelectric potentials through a preprogrammed computer to control a servo-boost system. In conjunction with the muscle substitutes above, the application to prosthetic devices and control systems is obvious. Initial application to prosthetic aids have already been accomplished.

CONCLUSION

Significant problems and trends in process-related information handling have been outlined and discussed. It is patently clear that there will be a prolific growth in the quantities and kinds of data to be processed. Information "indigestion" will be a common complaint in more and more of our endeavors. Saturation barriers and knowhow limitations will generate an ever-increasing demand for relief. Based on the considerations brought out in the foregoing material, the following key points therefore warrant specific attention.

1. It is urgent that our information sciences give adequate attention to process-related information handling. There are equally fundamental and deep-rooted problems involved as in the library and knowledge availability issues. Current interests and efforts are predominantly on the "library" problem.
2. There is inadequate treatment of the fundamentals of process-related information handling as it relates to the total community of interests. Our technology growth is left essentially to free enterprise in limited interest areas. Our language barriers between machines has been a direct result. A community system approach is needed based on fundamentals derived from an information science attack.
3. As a direct corollary of the above, we need to apply more attention to machine-machine relations.
4. Bionics research has the promise of significant advance in our machine assist capabilities. Classification of needs and problem areas require emphasis to permit concentration of research attack for more selective progress.
5. Man-man relations are a serious problem in achieving required interdisciplinary relations. There is an urgent need to reduce and elimi-

nate technical language barriers. Knowledge availability must provide for the flow of information across disciplines so that findings in one can be effectively correlated and applied in another. There are aspects of the overall problem that warrant social science research.

REFERENCES

1. Stanton, Ted, "Fishing for Facts," *Wall Street Journal*, Dec. 20, 1960.
2. Eisenhower, Milton S., "The Third Scientific Revolution," *Science News Letter*, May 23, 1964.
3. Rynaski, E. G., P. A. Reynolds, and W. H. Shed, "Design of Linear Flight Control Systems Using Optimal Control Theory," Cornell Aeronautical Laboratory, Inc., Buffalo, N.Y., Air Force Flight Dynamics Lab. Contract No. AF33(657)-7498, Project No. 8225, Technical Documentary Report ASD-TDR-63-376, April 1964.
4. Balducci, J. D., F. L. Adams, and M. A. Schwartzberg, "Effect of Gust Alleviation Systems on Dynamic Air Loads," North American Aviation/Autonetics Technical Paper.
5. Fusca, J. A., "Clear Air Turbulence," *Space/Aeronautics*, vol. 42, no. 2 (August 1964).
6. Press, H., M. T. Meadows, and I. Hadlock, "A Reevaluation of Data on Atmospheric Turbulence and Airplane Gust Loads for Application in Spectral Calculations," National Advisory Committee for Aeronautics (NACA) Report 1272, 1956.
7. Bendat, J. S., L. D. Enochson, G. H. Klein, and A. G. Piersol, "Application of Statistics to the Flight Vehicle Vibration Problem," Ramo-Wooldridge Div., Thompson Ramo Wooldridge, Inc., Canoga Park, Calif., Flight Dynamics Lab., Aeronautical Systems Div. Contract AF33(616)-7434, Proj. No. 1370, Technical Documentary Report ASD-TDR-61-123, Dec. 1961.
8. Bendat, J. S., L. D. Enochson, G. H. Klein, and A. G. Piersol, "Advanced Concepts of Stochastic Processes and Statistics for Flight Vehicle Vibration Estimation and Measurement," Ramo-Wooldridge Div., Thompson Ramo Wooldridge, Inc., Canoga Park, Calif., Flight Dynamics Lab., Aeronautical Systems Div. Contract AF33(657)-7459, Proj. No. 1370, Technical Documentary Report ASD-TDR-62-973, Dec. 1962.
9. Blauvelt, D. H., "Capabilities and Limitations of Digital Control Computers in Airborne Applications," The Bendix Corp., Eclipse-Pioneer Division, Teterboro, N.J., 1964 Joint Automatic Control Conference, Stanford University, Stanford, Calif., June 24–26, 1964.
10. Mundie, J. R., H. L. Oestreicher, and H. E. von Gierke, "Real Time Digital Analysis System for Biological Data," 6570th Aerospace Medical Research Labs., Wright-Patterson AFB, Ohio, Proceedings of the 16th Annual Conference on Engineering in Medicine and Biology, IEEE, November 1963.
11. Caldwell, W. F., "Recognition of Sounds by Cochlear Patterns," Santa Rita Technology, Inc., Menlo Park, Calif., Air Force Bionics Symposium 1963, Aeronautical Systems Division and Aerospace Medical Division.

12. Stewart, J. L., W. F. Caldwell, and E. Glaesser, "An Electronic Analog of the Ear," Santa Rita Technology, Inc., Menlo Park, Calif., 6570th Aerospace Medical Research Labs. Contract AF33(657)-11331, Proj. No. 7233, Technical Documentary Report AMRL-TDR-63-60, June 1963.
13. Lettvin, J. Y., H. R. Maturna, W. S. McCulloch, and W. H. Pitts, "What the Frog's Eye Tells the Frog's Brain," *Proc. IRE*, vol. 47 (November 1959), pp. 1940–1951.
14. Herscher, M. B., and T. P. Kelley, "Functional Electronic Model of the Frog Retina," *IEEE Trans. Mil. Electron.*, vol. MIL-7 (April-July 1963).
15. Bliss, J. C., and H. D. Crane, "A Computer-Aided Instrumentation System for Studies in Tactual Perception," Stanford Research Inst., Menlo Park, Calif., 1964 National Aerospace Electronics Conference Proceedings, PGANE, IEEE.
16. Carne, E. B., "Electronic Realization of Functional Nerve Nets," Melpar, Inc., Falls Church, Va., Electronic Technology Laboratory, Aeronautical Systems Division Contract AF33(616)-7834, Proj. No. 4160, Technical Documentary Report ASD-TDR-62-266, June 1962.
17. Carne, E. B., et al., "A Study of Generalized Machine Learning," Melpar, Inc., Falls Church, Va., Electronic Technology Lab., Aeronautical Systems Div. Contract AF33(616)-7682, Proj. No. 4160, Technical Documentary Report ASD-TDR-62-166, April 1962.
18. Guinn, D. F., "Large Artificial Nerve Net (LANNET)," Melpar, Inc., Falls Church, Va., *IEEE Trans. Mil. Electron.*, vol. MIL-7 (April-July 1963).
19. Baldwin, H. A., J. V. Wait, R. L. Brown, and C. W. Wieske, "Study and Development of Muscle Substitutes," Laboratory for the Study of Sensory Systems, Tucson, Arizona, Air Force Avionics Lab. Contract AF33(657)-8875, Proj. No. 4160, Technical Documentary Report No. RTD-TDR-63-4181.
20. Sullivan, G. H., C. J. Martell, G. Weltman, and D. Pierce, "Myoelectric Servo Control," Spacelabs, Inc., Van Nuys, Calif., Electronics Technology Division, Aeronautical Systems Division, Contract AF33(657)-7771, Proj. No. 4160, Technical Documentary Report ASD-TDR-63-70, May 1963.

20

Artificial Intelligence Applications to Military Problems

RUTH M. DAVIS
Department of Defense

INTRODUCTION

There are two answers that one would be likely to receive upon asking a member of the military community if artificial intelligence was applicable to military problems. The first answer would probably be that, no, it was not applicable and that the most advanced techniques possible were being applied to military problems. The second answer would probably be that he didn't know what artificial intelligence really was and that it was too esoteric to be useful to the military. Neither answer reflects any discredit to the military. They merely reflect the nebulous aura that surrounds the use of the phrase "artificial intelligence" as well as the fact that those techniques of artificial intelligence which are applicable to military problems are known by other titles.

It is worthwhile at this point to emphasize that, rather than attempting to give a generalized definition of artificial intelligence, it seems more practical and more constructive to consider artificial intelligence to be a summation of definable techniques and subject areas. This type of definition is elastic in that as techniques or areas are added or deleted the definition of artificial intelligence varies accordingly. Its sole advantage is that it enables the user of the phrase to be understood by his listeners and thus eliminates a great deal of confusion. Accordingly, for the purposes of this paper, artificial intelligence will be assumed to have a minimum coverage where the addition of any other areas by those interested will therefore not detract from the statements made in subsequent paragraphs. The minimum coverage of artificial intelligence is stated to be:

1. Intelligent automata
2. Pattern recognition
3. Learning machines and theory
4. Adaptive and self-organizing systems
5. Nonnumerical data processing
   (a) Mechanical translation
   (b) Symbol recognition and relationships
   (c) Processing of visual, acoustic, electronic and textual data

6. Problem-solving and theorem-proving
7. Process control
8. Heuristic programming
9. Decision theory
10. Selected data processing system organization theory (as applicable to nonnumeric data processing), and
11. Man-machine interactions

In the context of this definition, there are many applications of artificial intelligence to military problems, and there is in being a great deal of research and development in the field of artificial intelligence. One should not be surprised, however, at not having such R&D activities neatly organized into a coherent self-contained package. It is evident to anyone who has watched the emergence of automation and of other techniques for simulation of the human intellect that the entire effort concerned with artificial intelligence or equivalently with methods of simulation of selected portions of the human intellectual process is young, erratic and in a state of flux. There is no acknowledged set of leaders or spokesmen, there is no established theoretical background and there is no agreement as to its current degree of success or its potential. It is against this background, then, that applications of artificial intelligence to military problems will be discussed.

THE MOTIVATION FOR INTEREST BY THE MILITARY COMMUNITY IN ARTIFICIAL INTELLIGENCE

It is interesting to consider some of the reasons motivating the application of artificial intelligence techniques to military functions. It must be remembered, first, that, as has been stated previously, the techniques being considered permit functions normally demanding application of human intellect to be performed instead by artificial means simulating the human intellect. The reasons include:

1. The need to conduct operations in remote areas. The operations are of the type traditionally assigned to humans to perform but in this case the existence of man in the desired area is impossible. Remote areas are either those where it is difficult or impossible for humans to exist such as space, subocean areas or deep underground sites, or those man-made hostile environments currently incapable of penetration by our personnel. The latter are, of course, denied areas and countries.
2. The need to perform operations at a rate not attainable by the number of individuals available for assignment to the function. One

meets this difficulty primarily when the data-collection process yields data so voluminous that it is simply impossible to process it manually; also, the same difficulty arises when the response time for decision making is so short as to preclude manually processing of all the relevant, related data.

3. The chronic lack of personnel trained or educated to perform the function in question. Such a shortage of trained personnel may occur for a variety of reasons. Typical instances are emergence of a new technology, traditional distaste for the job, inadequate compensation for the expense of education required and, particularly in the military, administratively imposed restrictions on the number of personnel allowed in a given field or on the length of tenure in the field for a given individual.
4. The need for mass training and for education of individuals in a defined field or profession where the speeding up of the educational or training process to pace each individual's capability will effect a marked improvement in present procedures. This problem area is related but not identical with the preceding in that an increase of teachers would, of course, materially improve the educational picture. It is considered separately, however, because self-teaching or self-educational processes seem to lend themselves to separate treatment.
5. The need to mass-produce or to control the production of materials or of material components. This area encompasses a wide spectrum of activities from the control of nuclear power plants to the production of material sections of an airframe.

All of these above reasons, and presumably many more, are currently motivating the application of techniques of artificial intelligence to military functions. To highlight the issue it is worthwhile to discuss briefly specific examples illustrative of the generalized picture presented above.

REMOTE OPERATIONS

Remote operations demanding immediate attention include the capability of repairing equipment in spacecraft through remote controls effected from the earth and the capability of decision-making in remote spacecraft or on remote surfaces on the basis of environmental data available to the remote equipment in question. Such decisions could run the gamut from deciding whether to take more detailed photographs based on the appearance of an object of interest in the panorama under view of remote optical equipment to a determination by a remote movable device as to whether to proceed in a given direction based on the terrain characteristics available to the device for analysis. Other remote operations receiving attention but requiring still more include underwater operations such as mapping, locating and acquiring objects of specific shape or having specific characteristics, and collecting data on the underwater environment. The ability to conduct remote operations in denied surface areas such as the collection of information of certain types is an extremely desirable goal. In the latter case, the capability of determining what data to collect and the capability of preprocessing and collating it to conserve communications is essential. Here, techniques of problem solving, theorem-proving and inductive reasoning are essential tools to be possessed by the remote information-gathering device. This practical problem area is just beginning to be tackled in the military research and development community. Another interesting example, which has become one of renewed interest, is that of either a self-operating or a remotely controlled polygraph device.

TIME-LIMITED OPERATIONS

Time-limited operations are probably those which come most frequently to mind. Here, it is worthwhile to stress again that these operations include those where the timing factor enters simply because manual processing cannot match the volume of data involved or the amount of detail to be generated from the data. Examples are rampant enough and have been considered by the scientific community to such an extent as to be readily understood and to require, therefore, only a simple listing as follows:

Automated Photographic Interpretation

This includes the functions of target recognition, detection of movement or of change in a given environment, area discrimination such as the determination as to whether a wooded area is being viewed as opposed to a suburban area and the recognition of specific atmospheric conditions such as the presence of water vapor clouds, nuclear clouds, and the like. Techniques of pattern recognition and inductive reasoning appear to be required for the attainment of this function. It should be obvious that the volumes of data, i.e., photographs, to be processed as well as the short response times often needed for decision making are the factors making the attainment of automatic photo interpretation so urgent.

Symbol Recognition

Symbol recognition includes character recognition as a special case. This function will be assumed here to also involve the determination of relationships between symbols. It is obvious that automatic photo interpretation is dependent upon the development of techniques in this area. Other military problem areas also are involved such as textual processing, the input of large volumes of formatted hand-printed data to computer systems, the generation of computer-driven displays and the automatic production of printed material containing different type fonts, mathematical symbols, drawings, photographs and the like.

Nonnumerical Data Processing

This is much too general an area to discuss in any detail here and indeed it is not difficult to generate a controversy as to what should be included in its domain. Certainly photographic processing, although considered separately in this paper for reasons of emphasis, is in the domain of nonnumerical data processing. For the sake of brevity and with the understanding that each of the types of data listed below could be considered as a topic in itself, it will be stated that the military community is deeply interested in the application of techniques of artificial intelligence to:

(a) The analysis of medical records.
(b) The processing and analysis of acoustical data, including voice.
(c) The processing and analysis of electronic signal data.
(d) The processing and analysis of optical data.
(e) The processing and analysis of textual material including indexing, abstracting, extracting, dissemination and the like.

Attainment of any real facility for automatically processing nonnumerical data in a manner simulating that of a human will certainly demand an improvement in adaptive processes, in self-organizing processes, in associative processes, in the organization of automatic data-processing systems, in heuristic programming techniques and in problem-solving and theorem-proving procedures.

AREAS CONSTRAINED BY THE LACK OF TRAINED OR EDUCATED PERSONNEL

Certainly the rise of mechanical translation and of computational linguistics as pseudosciences can be attributed to the seemingly chronic lack of linguists, especially in specialized scientific disciplines. The history of mechanical translation is an interesting one for all interested in artificial intelligence to understand because it appears indicative of what is coming to be a general trend in the development of the various techniques comprising artificial intelligence. A look at the history of mechanical translation reveals that there were five periods and/or factors characterizing its growth which are also recognizable in various degrees in other areas of artificial intelligence. These are:

1. The initial period of development where most proponents will state that the human procedures being simulated can be completely reduced to algorithmic-like steps for automatic accomplishment.
2. A second period of complete disillusionment following unsuccessful

attempts to simulate the required human techniques. During this period a vociferous group will insist that mechanical translation, for example, is impossible of attainment and should be abandoned as a goal.

3. A period of retrenchment and reeducation where goals are modified or made more specific and where a determination is made of what human procedures can now be simulated and of what research is needed where simulation is not now possible.
4. A final period of slower, steadier progress towards both short-term and long-term goals which are realistic in nature. The initiation of this period is dependent upon the recognition of the field and its placement in the proper scientific discipline by university faculties and by the initiation of formal education to train researchers and managers.
5. There was in mechanical translation a general tendency to underestimate the length of time to achieve desired goals as well as to state realistic goals. This should be recognized as characteristic of most efforts in artificial intelligence and should be compensated for by those responsible for the promotion and management of these efforts.

EDUCATIONAL FUNCTIONS

Although it seems somewhat contradictory in concept, one of the most useful techniques of artificial intelligence should turn out to be the improvement of the human in terms of better training and educational procedures. Fortunately, this is also one of the most popular fields among scientists today, invoking the interests of educators, psychologists, engineers, mathematicians, and the general layman. It is an area of utmost importance to the military community where there is a continuous need for training on new equipment and for education in new disciplines and where there is never adequate time available for conventional schooling procedures to be effective. Techniques which need to be advanced, improved, applied and evaluated, include:

1. The use of the learning machine principle.
2. Problem-solving aids.
3. Theorem-proving.
4. Decision-making aids.
5. Question-asking and answering, and
6. Evaluation procedures.

MASS PRODUCTION AND CONTROL PROCESSES

Automatic process-control techniques are being developed and advanced with great urgency and evidently with a fair amount of success by the Russians. We have not yet given the same recognition to them although the situation in this country appears to be changing rapidly. Automatic control of processes in major industries such as the transportation industry, the oil-refining industry and the power-producing industry will greatly benefit the military profession in times of crisis. Automatic control of logistics and of lines of supply for military needs should be hastened. The need for automatic control of nuclear power plants is obvious. Also, of course, the benefits to be derived from automatically controlled mass fabrication procedures cannot be overemphasized. Many such automatic-control procedures have been implemented, but a sound scientific basis for the field is lacking and must be developed before its real potential can ever be fulfilled. It must be recognized that progress in many cases will be painfully slow. Techniques for automatically generating ship lines to replace manual lofting procedures have been under development for at least twelve years and have still not in any measure replaced the tedious manual work required. It is essential that more interest in this field be generated in the scientific community and particularly in universities.

SPECIFIC EXAMPLES OF POTENTIAL APPPLICATIONS IN THE MILITARY COMMUNITY

Certainly, the various subject areas of artificial intelligence find many applications—and in fact are essential for success—in the large military information data-handling systems. These systems, known generally as command and control systems, intelligence systems and reconnaissance systems, are all characterized by the fact that most of the data processed by the system is nonnumerical and therefore requires the application of many of the techniques discussed in earlier paragraphs.

In addition, certain other specific potential applications will now be considered.

MECHANICAL MANIPULATORS FOR REMOTE SPACE OPERATIONS*

It has been suggested that many of the purposes for which it has been proposed to place human operators in space vehicles can be accomplished more effectively by placing in the space vehicle remote-control apparatus which is operated in real-time (except for limitations arising from the round-trip transit time of signals) by one or more human operators on the ground. The art of remotely controlling and operating a space vehicle is capable of enormous expansion in comparison with its present realization.

*As discussed by W. E. Bradley in an IDA Memorandum of January 1964.

Simple control from the ground of specialized operations in a space vehicle has been a common feature of many past programs. Telemetering back to earth of the responses of devices in the space vehicle to such ground control operations has also been customary, but only for a rather small number of critical degrees of freedom of the system.

In particular, the telemetering and control concept can be extended in scope until a truly general-purpose "telecontrol" system is the result. Typically, such a general-purpose system can perform most of the operations which could be performed by a human operator in the vehicle, but a great deal more conveniently and, in some cases, more effectively. Briefly, the goal would be to place in space the operator's hands and eyes while leaving the rest of him on the ground.

There would be in the vehicle one or more small television cameras, which may be mounted on jointed and articulated arms in such a way that they can be moved in both translation and rotation with at least six degrees of freedom within the translation limits imposed by the space within the vehicle. In addition, it is possible for one or more cameras to be operated outside the vehicle by extending the arms through an aperture in the vehicle wall, permitting scrutiny of the surrounding environment or of devices such as antennas or solar-battery arrays located outside of the vehicle.

While such television cameras can be moved about from the ground by direct control means common in controlling ordinary television cameras in a broadcast studio, it is preferable and perfectly feasible to control such cameras by the angle and location of the head of a human operator who is located in a control station on the ground. Such a head-controlled television camera system was constructed in 1958 and operated *very successfully*.

The hands of the human operator in this vehicle can be simulated by remote-controlled manipulators having the necessary large number of degrees of freedom. Remote-controlled manual manipulators have been built for performance of laboratory operations in a radioactive environment and have been used so extensively in AEC operations that a considerable body of data is presumed to be available regarding their design. In any case, there is nothing difficult in principle in reproducing the motions of a man's hand and arm at a distance, and relatively little bandwidth in the electromagnetic spectrum would be required to transmit the necessary information, since the degrees of freedom involved, although numerous, change only slowly.

The result of providing in the space vehicle, in effect, both the hands and the eyes of one or more human operators is very similar to the provision of the human operator himself, except that less weight and a much simpler supporting system is involved. Such a general-purpose, remote-control system can replace many of the special-purpose systems which have been used or proposed in the past in somewhat the same way that a general-purpose computer can perform the function of specialized computers.

INTELLIGENT AUTOMATA APPLICATIONS TO RECONNAISSANCE SYSTEMS

A potential use of intelligent automata is to assist the operation of sensor systems designed to provide indications of hostile intent. A categorization of sensor systems by functional breakdown of components is as follows:

Typical Sensor System Components

1. Input Subsystem
   (a) Input stimulus
   (b) Noise (interference) perturbation
2. Detection Subsystems
   (a) Sensor (sensory field)
   (b) Translation connections—used to translate incoming sensory patterns into forms convenient for recognition.
3. Processing Subsystems with the functions of
   (a) Abstraction—reduction in dimensionality of input field.
   (b) Recognition—predetermined response to each of many varying sensory patterns.
   (c) Generalization—similar response to two or more varying sensory patterns.
   (d) Synthesis (association)—combination—either linear or non-linear—convolution, etc. of responses for purposes of decision.
4. Decision Subsystem† with the functions of
   (a) Estimation
   (b) Prediction
   (c) Extrapolation
   (d) Complex decision procedure
5. Output Subsystem
   (a) Transmission links
   (b) Noise perturbation

It is believed that all known or envisaged sensor systems can be described in terms of the above subsystem representation. In particular, such a description is useful for the purpose of this paper, which is to discuss the potential role of logical automata for improving sensor systems.

†Included with Processing Subsystems in later sections because of interrelatedness of functions.

Now, logical automata can be thought of as any artificial (nonliving) device which can be made to simulate any combination of the logical processes performed by human beings. It follows then that logical automata may perform as any component of a sensor system other than that of input or output. In practice, logical automata are split into two broad classes—automata which would seem to possess intelligence and those which in the most rigorous sense do only what they have been instructed to do. This report emphasizes potential applications of the first class. The latter class are of course quite important and currently are the *workhorses* of automated systems taking the form of standard programmed computers, both analog and digital, guidance systems, photomeasuration devices, etc. Both classes of automata have a role in remote sensor systems and in local sensor systems. Fully to exploit logical automata, their capabilities should be applied to those functions of sensor systems that are most complex.

Detailed investigation yields the conclusion that there are three primary ways in which intelligent automata may best be utilized to improve the capabilities of sensor systems in the near future.

(a) They may be used to *design sensor system components.* In this mode intelligent automata in a laboratory environment "learn" procedures of pattern-recognition, synthesis, search, etc., that are oriented towards the analysis and interpretation of particular sensor inputs. Once an adequate level of intelligence has been attained, the process evolved is "frozen" either into hardware or fixed computer programs and the resultant device becomes a component of the sensor system. In this manner, adaptive logic machines will serve the planner of sensor systems somewhat as the analog computer served the aircraft designer. With such a device the planner may adjust his variables and observe directly the effect of the changes upon the learning process.

(b) They may be used themselves as a component of sensor system. This is particularly feasible in remote sensor systems where all components are located in friendly territory. The capabilities of intelligent automata must be applied to the functions of recognition abstraction, generalization, and synthesis in the processing subsystem and to the problem of pattern recognition in the translational connections. The word "must" is used intentionally because of a strong conviction that any real advances in the data-processing tasks of sensor systems will evolve from uses of intelligent automata. The use of intelligent automata in denied areas is limited by the current lack of knowledge on how to control such devices remotely. Therefore, their use as components of local sensor systems will probably not be realized until after 1975.

(c) Finally, they will provide a means for analyzing large amounts of data that would otherwise be discarded for lack of manpower. It may take a great deal of data and a long period of analysis to fully develop means of discriminating a real threat from a distraction where the distraction itself may be either man-made or a natural phenomenon. This utilization could begin to be realized by 1965 if proper direction were applied to existing research projects. Results from such research would benefit both remote and local sensor systems.

PROMOTION OF RESEARCH AND DEVELOPMENT AIMED AT MILITARY APPLICATIONS

This section is a very brief statement of what appear to be essential principles underlying a program of research and development which would stimulate, encourage, and bring to fruition many successful applications of artificial intelligence to military problems.

First of all, each of the many constituent subject areas of artificial intelligence can and should be individually developed. There will be some known and recognized duplication of effort which will not be harmful or wasteful of funds expended.

There should, on the other hand, be an overall goal which will knit together as many constituent subject areas as possible. Such a goal would have to require the successful application of all these constituent techniques with all the attendant problems of interaction and feedback among techniques. The goal should be difficult of realization but by the same token it should result in the intermediate solution of an existing military problem. This goal has been tentatively defined as the development of a mobile device capable of nontrivial activity and possessing goal-seeking ing device with as large a memory as technology and cost permit. The overall goal should be approached through the successful attainment of a set of predetermined subgoals, each nontrivial in itself and each resulting in an advance in the state-of-the-art of some subject field. It is hoped that the projects which will represent the first step towards achieving the overall goal can be started within the year, and we are currently engaged in formulating the appropriate program.

Another essential ingredient to a successful R&D program is the education of the potential user, of those management personnel within the government responsible for the program, and of the scientists contributing to the program who must understand the practical importance of the goal. Talks, reports and conferences such as this form the means for so doing.

21

Computer Augmentation of Human Reasoning

RICHARD H WILCOX
Head, Information Systems Branch
Office of Naval Research

This paper is concerned with some of the limitations represented by the current state-of-the-art in electronic information handling. There are at least two approaches to an orderly examination of limitations, the first of which involves consideration of "better" ways to do what is already being done. The word "better" usually implies economics in one way or another —lower cost, increased efficiency, simplified operation, etc. Use of this approach would tend to emphasize current limitations of memory and logic devices and subsystems, fabrication techniques, file organization, programming languages, and display methods.

These are all significant problem areas whose improvement is important to widespread economic use of electronic information-handling systems. However, in this paper a different approach to the subject is used— namely, the consideration of barriers to performance of additional functions by information-handling systems over and above what is permitted by the current state-of-the-art. While this will undoubtedly lead to examination of some of the same limitations brought out by the other approach, the general emphasis will tend to be on research to develop new capabilities rather than on engineering to improve existing ones.

The following section will define and discuss a relatively new area of possible interest, to serve as a vehicle for looking ahead. Following that, the potential utility of this area will be considered (although it is maintained in some quarters that contributions to the information sciences come from the establishing of new procedures rather than the solving of problems, it is argued here that the incentive for devising new procedures stems from a desire to solve real problems). Finally, the requirements for realizing practical implementation in this new area will be examined in order to illuminate the limitations of the current state-of-the-art, together with possible avenues for alleviating such limitations.

AUGMENTATION OF HUMAN INTELLECT

The title of this paper very carefully avoids the term "artificial intelligence." Although the contents will certainly sound familiar to members of the artificial intelligentsia, the concern here is not directly with machine accomplishment of humanlike activities, but rather with machine help for the human who is himself performing intellectual tasks.

There are many ways in which computers might provide such help. The current myriad applications of machines—both digital and analog—to solution of mathematical problems arising in scientific research and engineering design certainly comprise a major augmentation of man's intellect. Automation of libraries is another important class, but a library, whether mechanized or humanized, is not an end in itself; it exists solely to help men solve problems. Therefore, this paper, without intending to minimize either the importance or the difficulty of mechanizing information retrieval, attempts to look beyond the process of making archival knowledge available and considers two processes involving interaction between a computerized data base and a human problem solver.

The first process—for which many examples are now available—is as a glorified "scratch pad": a mechanism for obtaining quick and accurate calculation of incidental numerical problems, and a temporary memory to retain intermediate results for later use. But the second potential way for computers to help in sophisticated tasks is actual participation in the intellectual processes themselves, much as a human assistant or colleague would contribute. This is much more than simple performance of more difficult or sophisticated tasks by machine; in particular, the computer must be able to analyze the man's input and to criticize or correct it, at least to some extent. Ideally, the machine might even represent or act in consonance with an alternative point of view, so that out of continual interaction between the man and the machine would grow a problem solution which exceeded anything the man might produce based upon his own ideas alone. Perhaps the highest example of this process in human society lies in the American court system, wherein a plaintiff and a defendant argue their opposing views in detail so that a judge or a jury has the best chance of deducing the true situation. A similar but less formal example, one upon which the progress of science depends, is the discussion and debate which takes place in technical journals and at scientific meetings. The corresponding man-machine process—that is, interaction at an intellectual level to permit synthesis of a more valuable solution than either man or machine could produce independently—might properly be called "dialectic programming."

Note that this process of hypothesis, antithesis, and synthesis carries with it the necessity of *on-line* or *real-time* interaction between man and machine; that is, the man must get a response at least as rapidly as he would in a conversation with another human. That, in turn, requires elimination of the programmer or any other intermediary between the problem solver and his mechanical aide. This provision for direct access to the machine by scientists and managers lacking specific programming skills has been called "implicit programming" by the Air Force; in itself (i.e., without inclusion of dialectic capability for the machine) it carries several important implications for reduced noise and distortion in the communication link, for direct knowledge of assumptions on the part of the user, and for ensuring that decisions are made by the proper decision maker and not by an intermediate programmer.

UTILITY OF COMPUTER AUGMENTATION

These two approaches to computer augmentation of human reasoning, the scratch pad and dialectic programming, are quite intriguing ideas. As such, they are certainly valid subjects of research within the university. But before private industry can risk its capital in pursuit of such ideas, and before the government can invest a significant portion of the funds entrusted to federal care by the taxpayers, it is necessary to ask whether any form of computer augmentation is justified either by economics or by some other benefit to society. In other words, what practical reason is there for devising machines to perform pseudo-intellectual tasks if the tasks can be performed by other humans themselves?

There are at least three answers to this question which are pertinent to any use of computers. First, if machines can do an equivalent job more economically, their use is generally justified. Second, there are some tasks in hostile environments, such as space exploration, for which it is desirable to minimize the number of humans involved. In addition, machines offer greater speed, memory capacity, and accuracy than is available from humans. These are the capabilities which justify the use of computers in most present day applications, and such capabilities are no less important in augmentation of human reasoning—particularly by the scratch-pad method.

But there are other reasons for pursuing dialectic programming, potential advantages to utilization of machines which have seldom been advanced for other types of applications. It may be simpler to train (or program) a machine than a technician or research assistant; or, more properly, it may be simpler to train a whole cadre of machines than the requisite number of humans, because once the effectiveness of training has been demonstrated on one machine, its subsequent copies may be depended upon to exhibit equal capabilities—a consistency which is not very evident in our present selection and training methods for people. Machines have highly predictable requirements for power, maintenance, and environment, and they ask no special management considerations. Finally, and perhaps most important, machines can be made available in desired numbers at planned times, and they can be stored or destroyed when not needed (this may not be economically desirable, but neither is it a social problem). These arguments do not imply that computers should replace humans wherever and whenever the state of technology makes it possible; rather, in those situations for which human individuality is not an advantage, then the use of machines may be preferable provided it is economically sound. The objective of research and development in computer augmentation of human reasoning, then, is to increase the variety of cooperative intellectual tasks for which machines are economically feasible, in order to permit consideration of computers as one practical alternative in as many situations as possible for which a human problem solver is going to need additional help.

Typical examples of situations for which dialectic programming of computers might be valuable may be found in both military and civilian contexts. Senior military commanders, for example, must consider and test a wide variety of alternative courses of action to meet the challenges presented by potential or actual opponents. To prepare a human "sounding board" for effective discussion with such a commander requires many years of training and experience, and even well-qualified staff members find it wise to temper their comments in view of the superior-subordinate relationship; further, a good staff man in one command is obviously unavailable elsewhere. Now of course there is no foreseeable prospect of being able to replace senior military staffs with computers. But a computer capable of pseudo-intellectual discourse, even within limited spheres of subject matter, could be an extremely valuable augmentation of the human staff—one that was unbiased by the presence of rank. And if it worked in one command, copies might well work in others with relatively slight modification. Even if no reduction in total staff size resulted from introduction of dialectically programmable machines, the increased effectiveness in decision-making could justify their presence. If in addition the size of the larger staffs could be reduced, the benefit would be compounded, for really good senior staff men are a precious commodity not widely found or easily generated.

This military example has its parallel in the executive world of private industry, where analogous uncertainties exist and good staff men are also scarce. Another example may be found in the processes performed by an intelligence analyst, and of course scientific research has always produced pioneers in new uses of computers. One must be careful, of course, not to eliminate useful apprenticeships held by management trainees, graduate students, etc., where inefficiency may be justified in terms of investing in the future.

REQUIREMENTS FOR PRACTICAL COMPUTER AUGMENTATION

It would appear that the potential utility of machines capable of participation in intellectual tasks with humans is sufficiently great to justify their development. Is such development possible now, or is additional research required? How much can be done within the limits of present technology? In what directions should appropriate research go? To explore such questions it is first necessary to examine the probable characteristics of computers capable of being dialectically programmed.

The first and most obvious requirement is for simple and direct communications between man and machine. In particular, if machines are to be of real value to human problem solvers then the language of communications must be one which is natural to the human—the same language he uses to solve problems by hand. This means English (with all its ambiguities), algebra, formal logic, block diagrams, and two-dimensional curves. It means charts and special terminology (both of disciplines and of the problem solver himself); conversely, it means access to data bases without restriction to narrow, previously established nomenclatures. And it means vocal and handwritten inputs, not necessarily typewriters. Printed character recognition would also be useful when utilizing previously prepared material. Can these things be done now? The answer is "yes, partly—at least in the laboratory." But these capabilities have never been combined in one system, and in general, they are far from operational, for reasons that will become clearer below.

Closely related to simple communications is physically convenient access. Experience with conventional computer facilities indicates that their use varies roughly inversely with the distance from the user. A problem solver needs the console near his own desk, where he has all his reference material and other familiar paraphernalia. Thus mechanized scratch pads and dialectic programming call either for separate machines scattered around near individual users, or else remote consoles tied to a large central facility through telephone or telegraph lines. However, when a man attacks a problem in collaboration with a helper (human or machine), he often stops to think—particularly when a thorny or unexpected response has been presented to him. But the helper might just as well be aiding someone else while the problem solver thinks, and this argues for having many remote consoles tied to a single processor on a time-sharing basis. In addition, such an arrangement introduces the possibility of two or more human problem solvers simultaneously attacking the same difficult problem, with interaction taking place through the computer. One possible application of this approach—which incidentally is well within the state-of-the-art—is to war gaming, wherein two opponents could be on-line simultaneously with each affecting the other's actions; this may be much more realistic than are current simulation methods.

The capabilities called for above imply the use of very large central processing facilities and rather sophisticated remote consoles. To handle a large number of users the central facility must have vast memory resources which are quickly accessible. To achieve speed and efficiency it must perform several tasks simultaneously through multiprocessing, and must be able to capitalize on peculiarities of the problem. Reliability of such a large system implies judicious use of redundancy techniques. The sophisticated nature of tasks to be performed calls for extensive programming (note that this refers to original programming of the system to provide its general capabilities, not the dialectic programming subsequently performed on-line by a problem solver); in particular, the capability to perform heuristic processes must be provided. It may be necessary for the machine to carry out some of its own basic programming in a learning, or self-organizing, mode of operation. This, incidentally, introduces the possibility of the machine adapting to individual users' specific requirements. Again, these capabilities are at least partially within the state-of-the-laboratory-art, but much remains to be done before they are operationally useful.

Operational usefulness involves economic feasibility. Most of the characteristics described above are achievable today only through the investment of large amounts of time and money in the hand assembly and programming of special pieces of equipment—both central processors and individual users' consoles. Wide application under such circumstances is completely out of the question. In other words, a major stumbling block to operational introduction of dialectic programming is lack of cheap mass-production techniques for fabricating and programming the large systems required. In large memories, for example (10^7 words or more), current technology permits cheap storage at unacceptably slow access speeds (magnetic tape) or rapid access at unacceptably high cost (magnetic cores). The hope here lies in the batch fabrication processes brought about by thin-film technology, microelectronics, cryogenics, and optical methods; in content-addressed (search) memories and iterative logic organizations; and (for software) in list processing techniques and self-organization with its potential for partially eliminating conventional programming through substituting of an example-showing process.

This author also believes that one other factor is needed before sophisticated computer augmentation of human reasoning becomes an operational reality—and this last item definitely is not within the current state-of-the-art. It would seem to be necessary that there be some quantitative techniques for measuring the effectiveness—and the shortcomings—of current and proposed computer systems. Undoubtedly, a few experimental systems will be built and operated on a pilot basis, just to see what happens. At least one would hope so. But general introduction of highly novel approaches such as dialectic programming will not follow until it can be clearly demonstrated that real benefits accrue as a result. With conventional computers in applications such as payroll preparation and inventory, it was possible to collect numerical data concerning relative processing times, frequency of errors, man hours reduced, etc. But when the process involved is that of helping a man solve a difficult problem, it is not obvious just what accessible measures are appropriate (or vice versa). Much more remains to be done in this area, and current emphasis on cost-effectiveness justification within the federal government implies that it must be done if sophisticated computer augmentation of human reasoning —such as dialectic programming—is ever to appear outside of the laboratory.

CONCLUSIONS

The picture painted above is one of exciting and useful capabilities, stringent requirements to provide them, and many remaining technological difficulties in meeting the requirements—at least outside of the laboratory. Man-machine interaction at the intellectual level has much to offer for improved decision making, more effective problem solving, and release of highly skilled humans from participation in some tasks so that their scarce skills may be utilized elsewhere. But achievement of this calls for much-improved (and simplified) man-machine communications; better understanding of very large computer systems (to provide improved synthesis); advances in microelectronics, optical techniques, iterative logic, associative addressing, and self-organization to permit economic hardware and software for these large systems; and—least available—methods for measuring the effectiveness of what is available and for predicting the characteristics of what is proposed. Possible techniques and approaches are in sight to alleviate all the current shortcomings, but much additional research and development—expensive research and development—will be required to realize economically feasible solutions. Can we afford to neglect the investment?

VII. PLANNING FOR THE FUTURE

22

Information Technology and the Information Sciences—"With Forks and Hope"*

HAROLD WOOSTER

Director, Information Sciences
Office of Aerospace Research
United States Air Force

EXPLANATION AND APOLOGIA

This preface is written with a profound and humble apology to those five or ten readers of this paper who already understand both the source and aptness of the major title. A Random Serial Search of 40 Documentalists in Philadelphia found an incidence of ½/40 (the numerator representing the, obviously better, half of a Documentalist); a Simultaneous Parallel Search of an audience of Electronic Information Handlers in Pittsburgh, employing the accepted "Is there a Carrollite in the House?" technique found 2/400 (and one of those cheated, as I'd explained it to him the night before) who recognized the source. My estimate of the logical intersection of the two classes is probably wildly optimistic. Hence this explanation.

The phrase occurs, as all students of the writings of Charles Lutwidge Dodgson (Cantabriggian mathematician, 1832–1898) know, *passim* in "The Hunting of the Snark." A proper KPIC (Key Phrase in Context) system would show the following:

> You may seek it with thimbles and seek it with care,
> You may hunt it *with forks and hope*
> You may threaten its life with a railway share
> You may charm it with smiles and soap.

If one takes advantage of the ambiguity of "it," and substitutes the "Long Range Goals of Basic Research" for "Snark" (carefully and deliberately ignoring the problem of the Boojum), the need for "Hope" becomes obvious.

"Forks," in this context, cannot be clarified without resort to the ikons.

*AFOSR 64-1897.

In the illustration accompanying the 1914 edition, it becomes clear that at least three separate sorts of forks are implied. One is a trident, standard Retiarius Mk 1(a) mode for pinning the prey. Another is a two-pronged agricultural implement, suitable for short-range prey transport and termination. The third, representing the using commands, is a smaller, also two-tined, carving fork. Other necessary implements, illustrated in the ikon although not in the text, are a microscope and telescope.

In summary, then, if one is purusing basic research one should do so with both hope and forks.

WITH FORKS AND HOPE

It seems appropriate in a university ambience to begin with a historical anecdote—one of the very earliest instances I have been able to find of the relations between academic research, military applications, and the government—the story of Galileo and the telescope. I am indebted to the Oxford *History of Technology* and to Arthur Koestler in *The Sleepwalkers* for this information.

Galileo did *not* invent the telescope, but he probably made more money from it than the man who did. According to a reliable record of 1634, Johannes Janssen or Jansen, son of the Dutch spectacle maker who probably did, declared that his father "made the first telescope amongst us in 1604, after the model of an Italian one, on which was written *anno* 1590." Giambattista della Porta of Naples (1536–1605) describes in the second edition of his *Magiae Naturalis* (1589) various ways of improving vision at a distance, including the use of a convex and concave lens.

Galileo may or may not have seen one of the Dutch telescopes. He claimed (in *The Messenger from the Stars*) that he had merely read reports (from DDC—the Dutch Documentation Center?) of the invention, and that these reports had stimulated him to construct an instrument on the same principle, which he had only succeeded in doing through extensive basic research in "the principle of refraction." This may or may not have been a snow job—it certainly didn't take the mind of Galileo to put a concave and a convex spectacle lens together once it was known that it could be done.

Be that as it may. Galileo proceeded to make a presentation and demonstration to the Venetian Senate on the tower of Saint Marco on August 8, 1609. Three days later he gave the instrument to the Senate, together with a Technical Manual cum brochure explaining that this instrument, which magnified nine times, would prove of utmost importance in war, since it made it possible to "see sails and shipping that were so far off that it was two hours before they were seen with the naked eye, steering full sail into the harbour," thus being invaluable against invasion by sea.

Koestler adds, in a sentence I tend to use in my more paranoid Pentagon briefings:

> It was not the first nor the last time that pure research, that starved cur, snapped up a bone from the warlords' rich banquet.

The story does not end there. Galileo gave the telescope to the Senate; the grateful Senate in return doubled his salary to a thousand scudi a year, and gave him tenure in his professorship at the University of Padua, which belonged to the Republic of Venice.

I am not entirely sure what the moral or morals of this story are. If the Senate had issued RFP's to meet their Military Requirement for an improved Command and Control System, their proposal evaluation might have reflected the needs of the service which opened the proposals. I can imagine that aerial types would have put in for a fire tower on top of the Tower of San Marco; that aquatic types might have preferred a fleet of picket boats; and that those with more terrestrial proclivities would have asked for a double appropriation for coast artillery, on the theory that more and bigger guns could take care of any problem.

Like all good stories, this has a happy ending. The military got a solution to their problem that would never have turned up through normal development channels. And Galileo, rewarded for Keeping Up With The Technical Literature and seeing an Immediate Practical Application, went on to build better telescopes and actually do good basic research in astronomy.

The ostensive, if not ostentatious, point of beginning with a hidden passage in the history of Galileo and the telescope, may become clearer with the following definition:

> Electronic information handling, the subject of this meeting, is a rapidly developing technology. It is parasitic upon, symbiotic with, and host to all other technologies. Like all other technologies, it is dependent upon a body of fundamental scientific disciplines and knowledge. Advances in information technology can only come in three ways; by specific research and development efforts aimed at information handling per se; by exploiting the fortuitous advances in ancillary technologies; and, by improvements in fundamental scientific knowledge and understanding.

The invention, or continued reinvention, of coordinate indexing is an example of the first; the continuing improvements in computers designed for either business or mathematics of the second; and, perhaps, the epistemological battle now being waged between syntax and semantics of the third.

More than most technologies, with the possible exception of medicine which it curiously resembles, information handling is involved with people as producers, processors, and consumers of information.

Most technologies can get along very nicely without people; in fact, much of their engineering effort is devoted to protecting their systems from people. A little old lady in tennis shoes can do more damage to a car in a hundred miles driving to and fro through the Liberty Tubes than a lead-footed test driver will do in 1,000 miles on the proving ground; whether rightly or wrongly, most aircraft accidents are attributed to pilot error, and the majority of automobile accidents happen to cars in excellent mechanical condition. One can build foolproof machinery, but there is no such thing as a people-proof information system.

Let me talk about the problems of people as producers of information. Last February in Bangalore I met a young British engineer who had been sent out to India to manage a Horlick's malted milk factory. After the third gin and tonic (the first two were spent in discussing, *seriatim*, King George III and the relative merits of the European four-wheel drift versus the American power broadslide as a way of getting around corners), he began to speak enviously of the American milkshed system where the manager of a factory like his could count on tank trucks of pure milk pulling up to the loading bay on regular schedules.

In India, it turns out, each cow is owned by an individual who gets up before dawn, milks it into a little tin pail with a lid, ties the pail on the back of his tall black bicycle, and wobbles precariously down the middle of the road for 10 miles to the factory. There he exchanges his full pail for a sterilized empty one, rides 10 miles back to his village and promptly washes out the pail under the village pump.

Most of us who run information systems would like to be in the position of the American dairy manager, with large amounts of pure reliable material arriving promptly. We actually find ourselves in the position of the Indian dairy manager, with milk that may never get in the pails and/or be consumed in the village (I am reminded somehow, of Mark Twain's village that lived by taking in each other's washing), or gets spilled or turns sour en route to our factory, dealing with producers far more anarchic than the Indian cow owner, with far feebler incentive to encourage delivery at the factory docks.

We need people to run our systems—trained, skilled, intelligent, creative people who will neither be bored by routine nor become too inventive in their indexing, much as we would like to automate them out of our stacks, our accessions departments, our cataloging rooms and our reference desks.

Most of all we need people as customers. We cannot live solely by talking to other information centers and to our Federal sponsors. There comes a time when people must use our products.

Ranganathan can talk of "Every reader his book"; *Time* can talk of "Every non-reader his non-book." We must deal with carnivores, who want only small amounts of highly concentrated information and turn savage if not cannibalistic when they don't get it; with placid herbivores, who are willing to munch vast heaps of cellulose to extract a minimum of nutrition; and, with the vast run of omnivores, who, in spite of their innate ability to digest almost everything, have developed sophisticated, jaded or even perverted appetites.

I will now return to the specific and implied subject of this talk—research needed for the improvement of information technology. You will remember that I said that this improvement could come in only three ways:

1. By specific research and development in information handling per se.
2. By exploiting the fortuitous advances in ancillary technologies.
3. By improvements in fundamental scientific knowledge and understanding.

Let me speak of the easiest part first—by exploiting the fortuitous advances in ancillary technologies.

Information handling, at least in the very strict sense as it applies to the handling of scientific and technical information, is not likely to be a major customer for many large new equipments. A certain inherent reluctance to talk about rope in the house of one who lost an ancestor when the platform gave way while he was attending a public function keeps me from mentioning the fate of the last computer to be designed specifically for information retrieval. Nevertheless, computers have been getting bigger and better, faster, and cheaper every year. We might well be using the Indian pattern of Leicas for microfilming and studio enlargers for making photocopies if there were not a major business market for microfilming checks and industrial records.

I am not at all sure that equipment manufacturers always understand this aspect of the information retrieval market. People do occasionally buy Rolls Royces, Pegasos, Ferraris, and Walnuts, but most of us are in the position of borrowing time on someone else's Chevrolet.

Perhaps an analogy from another field, that of mechanical translation, will make my attitude clearer. I was visited recently by a representative from a small software firm which had sunk (I refuse to use the word invested) $500,000 of corporate funds into a mechanical translation program.

I said, "How do you justify this to your stockholders?"

"What do you mean?"

"Look, DOD has said somewhere that they need about sixty million words of Russian text translated a year. You know damn well that we can buy fair-to-middling human translation at twenty bucks a thousand words, and probably wouldn't be interested in machine translation unless we could get it considerably cheaper—say ten bucks a thousand. Assuming that a contract was let for this, and assuming that you were the successful bidder, this would give you a gross of $600,000 a year and, at ten percent profit, a net of $60,000. Are you sure that you want to be in this game?"

Or, to switch to another field, a recent report on the mechanization of the Library of Congress set a price tag of $30 million for the minimum automation of the central bibliographic system. John Walsh, in one of his quasi-editorials in *Science* [vol. 143 (1964), pp. 452–455] doubted seriously that the Congress would ever appropriate the money to do this job.

Yet, *Missiles and Rockets*, in a recent survey of display systems for command and control (Oct. 5, 1964) estimates in a matter-of-fact way that:

> Command and control system displays, on the order of $1 million each, are expected to continue at the rate of 25–30 a year for at least 5–10 years.

It is a lot cheaper to make a Bookmobile out of a commercial bus than to start from scratch. Most of us when it comes to major capital equipment are going to find ourselves on the winning end of the game that the Government Printing Office plays with me every time I send a book over for printing—they let me pay for the costs of setting and printing the first 4,000 copies and then charge themselves only the incremental costs for any additional copies they want. We can let the equipment be developed and paid for by someone else, and then modify and/or borrow it for our own purposes, rather than pay all the research and development costs for the first prototype.

Much of research and development in information handling per se seems to me to be deficient in at least three aspects:

1. The absence of exciting new ideas.
2. The test of the market place.
3. Clear-cut proof to the complete satisfaction of the shirt sleeve scientist, the grey eminences of the invisible colleges, and those concerned with the disbursement of public funds, in both the legislative and executive branches of the government, that the job we are trying to do is socially beneficial rather than socially harmless. (I refuse, even for the sake of symmetry, to admit the third possibility.)

It is difficult, at least in serial speech, to discuss these three separately. (1) must be closely linked with (2) lest we wind up with handset letterpress Selective-Dissemination-of-Information systems, or nationwide microwave color television links between laboratories, turning on automatically with the laboratory lights, with all messages going automatically on videotape into a central file dwarfing anything that any dreamer of national information systems have yet conceived.

(2) and (3) have equally close links, against the day when the full national expenditures on scientific and technical information are finally dragged out from under all their ingenious covers and some cold-eyed gentleman says, "O.K. This is what you're spending. What are you getting for it?"

To return to my first point. Six months ago I spoke in this same hotel on the problems of scientific creativity under the title "The Scientist, The Engineer, The Inventor—One World or Three?" We are slowly training a competent body of information engineers—people who can apply known principles cleverly and skillfully to the solution of specified problems. Scientists, as I shall point out later in my talk, are being attracted to the field in growing numbers even though under my operating slogan of *Sic vos non vobis mellificatis, apes*—"Thus you bees make honey, but not for yourselves alone," they may not realize that that is what is happening. But we're running short of inventors.

This Wednesday, at the banquet of the American Documentation Institute, a moving tribute was paid to the memory of a gentleman whom I would hope considered me a friend—Hans Peter Luhn. I have never made an exhaustive search of all of Pete's contributions to our field, but let me just mention three which have crossed my rather high threshold—Selective Dissemination of Information; Keyword in Context Indexing and Auto-abstracting. For years now much of the traffic in my office has been with people who would say, "Yes, I know Pete invented this technique, but I can improve on it." It is not difficult to improve on someone else's invention—Steve Juhasz, Ed Rippberger, and I have been, we hope, guilty of it with WADEX—but it is difficult, and for most people impossible, to make an invention of one's own. It is even more difficult for an invention to meet, as have at least two of Pete's—SDI and KWIC—the test of the market place. I do not know where we will ever find more people like Pete Luhn, but the field certainly needs them.

I am not sure that my job description calls for me to be either inventive or creative; one of the prices of becoming an administrator is to decline the fame and envy of original composition, but there are two notions that I've been gnawing on for a while.

One is the need for a scaling factor for information systems. I hinted at this in my most unrequested reprint—*Journal of Chemical Documentation*, volume 3, number 216, 1963—where I voiced my suspicion that the square-cube law—that as an organism grows, its surface increases as the square of the diameter, while the internal volume, and mass, increase as the cube—that affects all living organisms also applies to information systems. I feel intuitively, but lack both the evidence and the mathematics to prove, that the surface area of an information system available for radiation—the transfer of information outside the system—increases at a slower rate than the complexities of interaction between the items in the store, and that both of these tend to grow far more rapidly than does the nutrient supply of people and money needed to operate the system.

An interesting consequence of the square-cube law in nature is that it sets both a lower limit—something the size of a shrew has to spend all its time eating lest it starve to death—and an upper limit to the size of organisms. You just don't build a land-based animal much larger than the elephant.

I wonder if this square-cube law may not also set up an upper limit to the size of information systems; if the internal complexities are growing at a much faster rate than the public contact area, the manager inevitably becomes more concerned with the internal management than with the public service and, inevitably, gets a key to the dinosaur club.

I wonder also if we have not been remiss in forgetting that there are, after all, *four* laws of thermodynamics in our concentration on the second. I can't do anything constructive with the *first*. I started thinking about the *third* when I started thinking about the entropy of *knowledge*—that subset of information which gets inside the skull and stays there long enough to do some good—and think that I could do something about that in relation to Boring's minimum set of dissonant paradigms by which we actually operate.

I do think, though, that we need something like the *zero*th law of thermodynamics. Thermodynamics operates on the assumption, amply corroborated by experimental evidence, that heat flows from hot bodies to colder ones, and never in the reverse direction; that heat flows from heat sources to heat sinks. It was many years before that they realized that they needed one more law, the *zero*th law—that when two bodies are in thermal equilibrium no heat flows from the one to the other—to provide a logical axiomatic basis for the other three.

We operate, I submit, on the assumption that information invariably flows from information sources to information sinks. Is this a safe assumption? Has anyone ever proven it, either theoretically or empirically?

Let me return to my points two and three. We are not practicing a branch of aesthetics where we can concern ourselves with art for art's sake. We are dealing with the engineering of systems to do a variety of jobs, not least of which is satisfying both our customers and our sponsors. We think we know, although we probably do not, a great deal about our *milieu internale*. What do we know about our *milieu externale*?

What do we know about how scientists and engineers now communicate and use information?

What do we know about the relation of information to the actual processes of scientific research, of engineering development, of invention?

Just what is it that information and information services actually do?

What sort of accepted (and acceptable) methods and criteria can be used for evaluating objectively the design and operation of information systems and, perhaps most important of all, their actual and potential utilities?

Or, to use a phrase which some of you must have heard before, how do you do a cost-effectiveness study on an information system?

I would be less than gracious if I did not call the attention of those seeking problems on which to do research to the prospectus of the Knowledge Availability Systems Center which, at least in the draft I have (dated August 1, 1963), outlines some 29 more or less separate problems under such general headings as:

Criteria for systems design
Comparative anatomy of systems
Language manipulation
Behavioral studies
Hardware studies
Media studies

At least a third of these studies fall into the third and last area I wish to discuss today, basic research in the underlying scientific disciplines—the third way in which I said improvements in information technology could come about. This is not a field for one who expects quick results, nor immediate applications, nor, for that matter, is it a field for crash programs. I am rather amused by the plaint of a former principal investigator of mine, who once did good basic research for me and now finds himself operating a multimillion-dollar information center, that there is little coming out of any of the three major basic research programs in this field (the classification is by sponsoring agency) that helps him with his practical operating problems.

Of course not. Those of us who have been administering basic research programs in this field would be derelict in our duty if we yielded to our chronic temptation and cooked our seed corn—sought the approbation of our bosses by buying research on the basis of its immediate applications.

Our job in managing basic research is to bet on long shots at the $2 window. We try to do this on a little more rational basis than the horses' names or the color of the jockeys' eyes—although I must admit that we do pay a little attention to the color of the jockeys' silks, especially if they are those of a major stable. A horse-playing former chief scientist of ours once said that our job was looking for overlays—cases where the true odds are better than the apparent odds. Other agencies have much larger sums to bet on favorites to win, place or show, at correspondingly lower odds. Favorites do drop dead in the stretch; long shots do come from behind to win. This, together with the traditional difference in opinion, is what makes horse playing, and the administration of a basic research program, a sporting game.

Where does one go looking for research workers who might be able to take solid steps towards solving this problem? [In much that follows, I might quite properly be accused of exercising the *droite du seigneur* on a report, "Information Processing Relevant to Military Command: Survey, Recommendations and Bibliography," prepared by A. E. Murray and H. R. Leland of Cornell Aeronautical Laboratory under Contract AF 19(628)-1625 for the System Design Laboratory, Electronic Systems Division, Air Force Systems Command. ESD-TDR-63-349.] Sometimes, but only sometimes, in schools of documentation and/or library and/or information science. They are likely to be scattered all over the university campus, not infrequently in the electrical engineering department (which has become the liberal arts college of engineering), but also in such departments as biophysics, philosophy, psychology or mathematics. Some are not even on university campuses at all, but hidden away in remote corners of great industrial research laboratories or in small R&D firms in deserted shopping centers.

If you ask them what they are working on, they are unlikely to answer, unless they have been corrupted by the thought of government funding, by such phrases as "Information storage and retrieval" or "Electronic information handling." They are far more likely to answer with such phrases (or descriptors) as:

Automata, especially logical or computing automata
Pattern recognition
Signal detection
Artificial intelligence, mechanization of thought processes, brain mechanisms, artificial organisms, cognitive processes
Bionics
Self-organizing systems
Cybernetics
Nerve (or neural) nets

Perception mechanisms and logics
Discriminating functions
Decision-making
Problem-solving, game-playing, heuristic programming, hill-climbing, optimization, linear programming, dynamic programming
Linguistics
Logic, especially multivalued and modal logics
Information theory, channel capacity, entropy and uncertainty, coding theory
General aspects of correlation, prediction and filtering
Control theory, servomechanisms, theoretical and experimental dynamics of feedback systems
Signals and noise
Psychology of value judgments
Statistical prediction theory
Vision, speech and hearing
Concept and percept formation
Network and switching theory
Speech analysis, synthesis, and recognition
Existential and analytical philosophy
Epistemology
Combinatorial mathematics
Random processes
Probability theory
Circuit theory
Cryptology
Statistical communications theory
Programming languages

Use of these terms as descriptors in querying several very large document collections produced some 7,000 different citations to documents!

The odds that one or more of these 50 fields or 7,000 documents may yield results relevant to the problems of electronic information handling may seem staggering, but I submit that they are far less than the odds that out of the tens of thousands of young men and women in our colleges and universities will come another Hans Peter Luhn.

The names of the possible fields given were deliberately randomized. A rough classification—remembering that all classifications are personal to the point of being solipsistic—might yield the following five areas which seem in especial need of encouragement and acceleration.

1. The link between language and epistemology defines the single most important front for an advance in information processing technology. Linguistics occupies a uniquely pivotal position in relation to various aspects of intelligence and automata. Natural language

breaches the interface between conscious reasoning and the underlying mechanisms and serves as the medium for the conscious organization, transmission, storage and retrieval of information.

Formal versions link machines to man's will and, within the machines, primitive formal languages govern and are represented by the states, transitions and interactions of the active parts. To understand the nature and basis of intelligence so as to exploit this understanding in the use and development of automata, we need to know much more about language. Similarly, to understand more fully the techniques of symbolizing and systematizing meaning or concepts in order to exploit this understanding in analysis, storage, cross-linking, searching and retrieval of information, we, again, need to know much more about language.

2. Well conceived, firmly based and definitely, purposefully, and theoretically oriented, as opposed to vague, exploratory or empirical, research is needed to discover, at approximately the "neural" level, plausible fundamental mechanisms for the development of intelligence in information processing organisms and automata.

The problem of discovering the basis of intelligence appears to be essentially the problem of elucidating how any brainlike system can, through contact or interaction with its environment, become functionally organized in that special way we call "intelligent."

By referring this investigation to the "neural" level, one seeks the ultimate mechanismic basis of intelligence by taking explicit account of the importance of the nature, characteristics and interaction of relatively simple components in those special aggregates capable of acquiring and exhibiting intelligence.

3. Both philosophical and experimental evidence indicate that a satisfactory explanation or mechanization of *visual* pattern perception must incorporate both analytic and holistic concepts. Analytic pattern recognition, without regard for the problems of segmentation of a complex visual field, and suitable only for clean, separated figures, is receiving most of the attention devoted by physical scientists for all too obvious reasons.

What is needed more is much more difficult to supply; that is, information and understanding on the interrelation between the analytic and Gestalt aspects of pattern recognition; how and what subsets of point stimuli are perceived as unitary entities; figure-figure and figure-background separation mechanisms; and the meaning of the direction and limitation of attention.

This example has been set in the field of visual pattern perception. Similar and probably more complex problems face us in the field of speech perception, which may serve as an orbital stage before we tackle the vastly more difficult problem of semantic perception. It is becoming increasingly clear that speech recognition cannot be done on the basis of the acoustic properties of the speech signal alone; that general solutions will rely upon the interplay of linguistics and semantics.

The most exciting step of all will come when we are able to study pattern recognition in text. How does a reader, for example, recognize that novel A has the same plot as novel B? How does a scientist realize that a piece of work in, say, psychoacoustics contains the clue to solving his problems in cloud cover analysis? And one wonders how long will it be before a computer will actually be able to take a document and:

Make a true abstract.
Recognize that it is related to work not cited in the bibliography.
Describe it as brilliant, pedestrian, or unsound.
Tell the plot of a novel.

4. Self-organization appears to be a basic phenomenon manifested in the greatest variety of systems which can be described and understood in terms independent of the particular system in which it is observed. One of our needs is for research which studies self-organization as the central phenomenon of any system or systems, and attempts to describe it in the most basic and general of terms. In this regard, two facts are noticeable:

   (a) While learning may be regarded as a certain kind of self-organizing capacity, the bulk of the work by nonbiologists in systems which "learn" is not directed to the central issue, which is the epistemological problem for automata.
   (b) The principles of self-organization in fields outside of cognitive systems research are all but neglected by interdisciplinarians.

Some attention must be directed to self-organization as manifested in the most central phenomena underlying intelligence, and to the possibility of generalizing on the principles of self-organization over fields as remote as morphogenesis and socioeconomics.

5. It has become apparent in recent years that the major breakthroughs in computer capability in the future will come from improvements in the logical organization of computers and in new programming techniques. The organization of the digital computer as conceived by von Neumann seems increasingly inadequate for the types of

> problems people actually wish to solve. Concepts such as associative memory, built-in stacks, multiprocessing, multiprogramming and parallel organization, represent a radical departure from traditional ways of building computers, quite apart from the hardware used. At the same time, the difficulties that people have in communicating problems to computers have become more and more pressing as the complexity of these problems has grown.

Areas of effort most likely to extend the capability of the digital computer include *machine organization*, *programming techniques* and *information-handling techniques.*

The problems of *machine organization* are concerned with ways of constructing deterministic, programmable devices that can be used to solve problems. Continuing success in the study of relatively large complexes of relatively simple components, as in distributed element computers, will require, either for its own prosecuting or its exploitation in useful automata, a solution to the problems of space consumption, power requirements and the costs of layout, assembly and interconnection of the components. While microminiaturization itself probably needs no further encouragement, attention to the comprehensive solution of space, power and *interconnection* problems is especially recommended.

Computers, at least from the programmer's view, are mathematically well-defined structures in which random events are virtually nonexistent, or so he hopes. Nevertheless, although a number of abstract modeling devices for machines, such as finite-state machines and other constructs of automata theory, do exist, the general description of these structures has never been fully formulated. Such a formalism could provide a basis for a complete yet uniform mode of machine description or, more pragmatically, could also serve as a device to permit automatic generation of programs for many different machines.

Programming techniques are concerned with ways of applying a computing engine to solve many different unrelated problems. Very early in the computer game it became recognized that machine language was not a particularly efficient way of posing problems to a computer. An increasing number of programming demands are being met by problem-oriented languages.

Conversation at a recent Association for Computing Machinery convention:

"Hi, Joe. What's new?"
Joe (proudly): "I've invented a new programming language."
"So? So what else is new?"

One question concerns the way in which such languages are described—a crucial question because of the increasing need for translators for these languages. Each new language generates a requirement for a translator for many existing machines. Formal, and hence machine manipulatable, descriptions of programming languages are therefore increasingly in demand.

Another question concerns bridging the gap between human languages and programming languages. There are significant structural differences between the two. Human languages, at least when talking to inferior beings like children, wives and computers, are constructed mainly of imperatives. Most of the work in developing new programming languages has been concerned with their local structure rather than with their global structure—i.e., with the way that things are said rather than with the kinds of things that are said. Better impedance matching between human and programming languages could improve materially the ability of people, even trained programmers, to communicate with computers.

Computer programs with learning ability are needed—some way to use the computer in the process of finding problem-solving algorithms as well as in the process of executing these algorithms. Human beings can deal with complex problems only if they have a means of organizing them; computers can deal with complexity through brute force. Problems that people often think of as ill-defined are really problems for which the solution algorithm is too complex for human comprehension.

In such circumstances, a man-machine dialog, at a slightly more complex level than "Me Tarzan—you IBM" must be created, with the machine playing a more active role. The machine must learn about the problem, and in order to learn, it must be able to ask questions.

L'ENVOI

This talk has covered a span of some four centuries, from the *Magiae Naturalis* of Giambattista della Porta, *ca.* 1584, to an Orwellian world of dialectics with intelligent computers in 1984.

There are two things that I hope you will take away with you from this talk.

One is the moral (or immoral) of the story of Galileo Galilei and the telescope—that apart from moral, legal and ethical considerations, it doesn't really matter where an idea comes from if you can figure out a better use for it.

The other is the following set of premises on which this talk is based.

Electronic information handling is a rapidly developing technology. It is parasitic upon, symbiotic with, and host to all other technologies. Like all other technologies, it is dependent upon a body of fundamental scientific disciplines. Advances in information technology can come only in three ways:

By specific research and development efforts aimed at information handling per se;

By exploiting the fortuitous advances in ancillary technologies; and

By improvements in fundamental scientific knowledge and understanding.

For, after all, the motto of my organization, the Air Force Office of Scientific Research, is taken from Ecclesiastes: *Primum acquirere cognitionem*—"First, get thee understanding."

23

Future Hardware for Electronic Information-Handling Systems

DONALD L. ROHRBACHER
Goodyear Aerospace Corporation
Akron, Ohio

INTRODUCTION

The purpose of this paper is to examine hardware in the light of requirements for electronic information-handling systems. Currently available hardware as well as some of the approaches still in the laboratories will be considered. Within the scope of this conference it is impossible to present an exhaustive listing of the many techniques currently under development. However, some of the more promising ones are discussed and from these an indication of what the future holds can be obtained.

Furthermore, the electronic information-handling field is too broad to analyze the requirements for the multitude of different systems. However, there are certain general areas of consideration which are applicable to many of these systems. One of these is storage and the other is the need for processing of the stored data. These are certainly not the only system considerations, but for purposes of restricting the scope of this paper to a reasonable size only these two will be considered.

The major portion of this paper is written in the context of a large-scale information-retrieval problem which requires an electronic information-handling system. The problems found in information-retrieval systems are very similar to those found in the larger class of electronic information-handling systems. This is particularly true in the bulk-storage and file-processing areas. These are the areas that have been chosen for consideration.

BULK STORAGE

INTRODUCTION

Much has been written concerning the tremendous amount of world literature, but nothing demonstrates the size more vividly than the consideration of the required bulk-storage capacity for an information-retrieval system. For example, consider a system for 2×10^6 documents in the biomedical field. Each of these documents contains approximately

2,500 words. If 20 bits are used to encode each word, then the bulk storage will require a capacity of 10^{11} bits! Furthermore, approximately 250,000 documents are being added each year to this particular body of literature. Therefore, the bulk storage must have the capacity to accept a growth of 1.2×10^{10} bits per year.

The type of information-retrieval system (statistical, syntactical, etc.) will define some of the other characteristics required of the storage medium. A system[16] considered at Goodyear Aerospace Corporation (GAC) for this biomedical literature was essentially statistical. It required random-access storage for a large matrix ($10^4 \times 10^4$ for even a small pilot study of 100 documents), in addition to the bulk storage for the main file of documents which did not require the random-access capability.

Obviously, since many requirements for bulk storage are dependent on the type of system to be implemented, it will be impossible to consider them all. However, the more general requirement of large capacity is common to all systems and will be the major consideration of this section of the paper.

CURRENTLY AVAILABLE

Magnetic tape is still one of the least expensive forms of storage for large quantities of data. In this field the IBM 7340 Hypertape Drive[14] is one of the most advanced systems currently available. It uses an 1,800-ft reel of one-inch magnetic tape and has the capability of reading in either direction. It has a high-character density of 1,511 8-bit alphanumeric characters per inch. This high density plus a smaller record gap of 0.45 inch permits reel capacities of up to 30×10^6 characters per reel or 240×10^6 bits per reel. It has a rate of 170,000 alphanumeric characters per second.

In systems requiring a random-access memory, there are several choices available. Bryants Series 4000 Disk Files[8] feature up to 24 disks, each 30 inches in diameter, rotating at speeds up to 1,200 rpm. There are six magnetic heads with 768 concentric recording tracks for each disk surface. A hydraulic positioning system moves all heads simultaneously and can select any track within 100 milliseconds. This system has a maximum capacity of about 1.6×10^9 bits.

Another type of random-access storage is the IBM 2321 Data Cell Drive.[13] This system stores the information on strips of magnetic tape ($2\frac{1}{2} \times 13$ in.). Ten of these strips are contained in a subcell, twenty subcells forming a cell. Ten of these cells are then arranged in a circular array. A hydraulic system is used to position the selected subcell beneath the access station. A pneumatic mechanism is then used to select one of the ten strips. This strip is placed on a revolving drum and rotated past the read/write head. This system has a maximum capacity of 400×10^6

8-bit characters or 3.2×10^9 bits. The worst-case access time is approximately 600 milliseconds.

Another random-access storage system is RCA's RACE (*R*andom *A*ccess *C*omputer *E*quipment).[5] This system uses flexible magnetic cards ($4\frac{1}{2} \times 16$ in.). There are 166,400 characters on each card, which are divided into blocks of 650 characters. Up to 256 cards fit into a magazine. For every 16 magazines there is a read/write head and selection mechanism. Solenoïd-actuated bars select the card. It is then moved by pinch rollers and friction belts onto a spinning drum where the data is read or written. Using two control units a maximum of 128 magazines can be used. This gives the system a capacity of 5.4×10^9 characters.

FUTURE SYSTEMS

There exists a definite gap between the systems which are currently available and the needs of some users for larger systems. Fortunately, there are a large number of techniques under development which may go a long way toward closing the gap.

For some time superconductive memories have been expected to provide large, fast, inexpensive memories.[18] The reason for this expectation is the fact that they offer the possibility of batch fabrication of not only the storage elements, but also the addressing switches and all other connections. Conventional transistor drivers and sense circuits can be used; their number increases only moderately with capacity as they need not be partitioned. It has been expected that even large memories could have cycle times of about one microsecond. Unfortunately, the technological problems of operating at cryogenic temperatures (approximately 4° Kelvin) have greatly slowed progress. It is quite possible that other approaches, currently being developed, may become available more quickly.

In the development of large-capacity magnetic memories, batch fabrication techniques will be necessary, if for no other reason than that it is simply impossible to wire the billions of conventional magnetic elements in a reasonable amount of time.

One of the more promising approaches to batch fabrication of ferrite memories is IBM's Flute.[4,15] The basic element (Fig. 1) is a tubular ferrite structure with a conductor which runs axially through the tube serving as a word line. Bit lines intersect the tube at right angles to and displaced from the word line. A memory plane is composed of a number of such parallel tubes, with the same bit lines intersecting each tube. The fabrication of this complete prewired memory plane is accomplished by sandwiching a rectangular grid of wires between matched dies. The grooves of these dies are filled with a ferrite material, and after appropriate curing and sintering the complete prewired plane is ready for testing. It is antici-

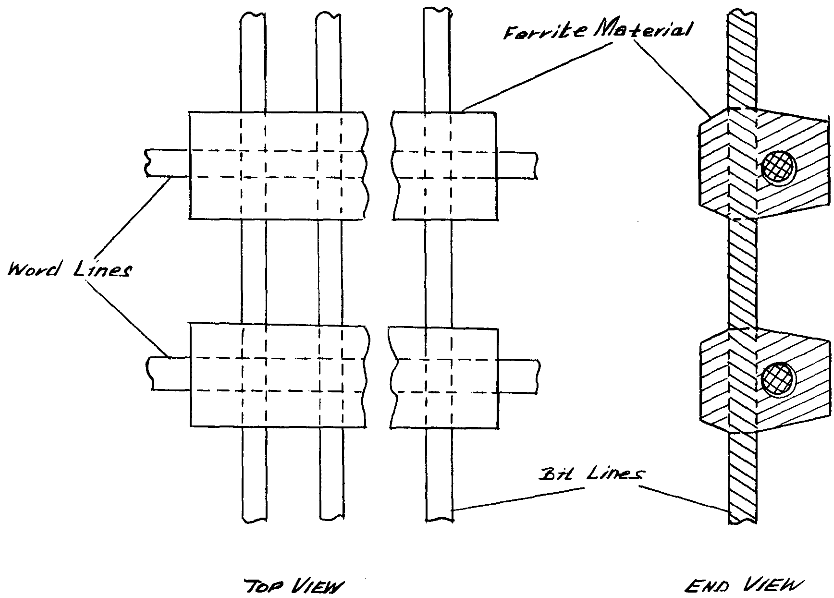


Figure 1. Flute memory elements.

pated that word and bit line spacing of up to 100 per inch are possible. This will result in a high packing density of 10^4 bits per square inch. It is also expected that cycle times of 250 nanoseconds will be practical.

A different approach which appears promising is the "*D*ove *D*ata *D*evice" (3D)[9] which is being developed by Rome Air Development Center (RADC). The recording is done on a 3-micron-thick (Fig. 2) nickel film. An electron beam is used to put 2-micron-diameter holes in the film, spaced about 1.5 microns apart. The reading is performed by aiming the electron beam at the bit location and sensing the existence or nonexistence of a hole. The sensing is performed by the use of a metal plate beneath the film which is used to collect the electrons passing through the holes. A flow of current indicates the existence of a hole. This system has a capability of storing up to 10^9 bits per square inch and a maximum system capacity of about 10^{11} bits. A feasibility model (10^7 bits/in.^{2}) is expected to be fully operative in about 12 months.

Stanford Research Institute (SRI) is developing an approach[19] with tremendous potential capability, but still some time in the future. This technique utilizes micromachining to record by etching holes in a metal film which has been properly covered with antireflection material. One side of this film will be illuminated and an array of light detectors will be used on the other side. Each light detector will be 0.2 micron in diameter

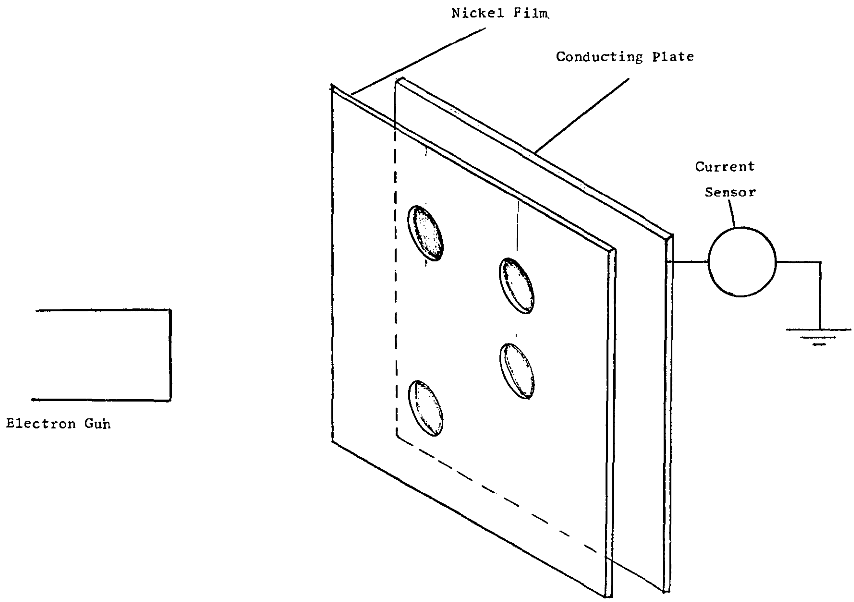


Figure 2. Dove data device.

and therefore covers an area which corresponds to 100 bit positions. This makes possible the detection of many light levels depending on the number of holes. It is expected that practical considerations will limit the number of light levels to 10, which in turn results in a system capacity of about 3.3×10^{10} bits per square inch. The system will also use 0.2-micron-diameter light sources and amplifiers. This will permit the data to be read in any desired series-parallel manner. The entire process for recording the data will take about four minutes for 10^{12} bits of information. The microminiaturized electronic circuits being developed in this program will be useful not only for the data-storage system, but for a large number of electronic information-handling problems. Using the techniques being developed in this study, it is believed possible to build 10^{11} electronically active components in a volume of one to several cubic inches. This small size may permit one to hand-carry a complete data-processing system!

FILE PROCESSING

INTRODUCTION

All large electronic information-handling systems require some type of high-speed processing and very often on a large amount of data. As was mentioned earlier, even the pilot information-retrieval study conducted at GAC (100 documents) required the determination of a $10^4 \times 10^4$ matrix. Larger systems will require matrices which are orders of magnitude larger, and hence the required amount of processing will be extremely large. This ability to process a large data base at a high speed will be the major consideration of this section of the paper.

CURRENTLY AVAILABLE

Historically, data-processing systems have been built faster and faster—each one with increased capability over its predecessor. Today we have such systems as the IBM System/360 Model 70.[17] This system has a main storage capacity of up to 512,000 8-bit characters. It has a 1-microsecond memory cycle time and also the capability of overlapping parts of consecutive core cycles to obtain effective access times less than one microsecond. It has six available I/O channels which can be overlapped with the processing to permit simultaneous read, write and compute. The Hypertape Drive and the Data Cell Drive, described earlier, can be used with this system.

Another large-scale data-processing system is the CDC 6600.[6] This is comprised of 10 peripheral and control processors plus one central processor, which is a high-speed arithmetic device. The peripheral and control processors can execute programs independently of each other or the central processor. Each has its own 4096 12-bit words of storage. The central processor has 131,072 60-bit words of storage with a cycle time of one microsecond. Available to use with the system are the CDC 626 tape units which handle binary data recording at 800 bits per inch on one-inch tapes up to 2,400 feet long. A card reader which reads at a 1,200-cards-per-minute rate is also available.

FUTURE SYSTEMS

Despite the increases in computer speed which have taken place in the period since 1947, the computer has remained basically sequential and inadequate for many of today's problems. This is particularly evident in the field of information retrieval and other areas which have very large data bases or require real-time computation. Furthermore, additional significant increases in speed are not likely, since current techniques are approaching performance limits imposed by the speed of light. If major increases in capability are to be obtained in the future, they will need to come about as the result of devices and organizations which permit the parallel execution of many operations.

A relatively new entry into the data-processing field which promises some of this parallel operation is the associative memory,[7,10] sometimes called the Content Addressable Memory. This is a memory that has the basic capability of addressing by content rather than location. It is capable of simultaneously interrogating the content of every word location to find all those locations which contain the same information as that stored in a special register known as the comparand. This is called the Exact Match Instruction (Fig. 3).

Knowledge exists to show how to greatly extend the basic capability of the associative memory. With the addition of some control logic it is possible to perform more complex searches such as Less-Than, Greater-Than and Between-Limits. With some additional control it is possible simultaneously to search the entire memory (or any chosen subset) for the maximum value. If the capability of modifying the contents of the memory at the word level, as a function of the response to a previous search, is added, then arithmetic computations can also be performed in parallel. Two fields in memory can be chosen and the memory can add, subtract, multiply, divide, etc., the corresponding numbers in each field and simultaneously store the results in a third field. The following is a partial list of

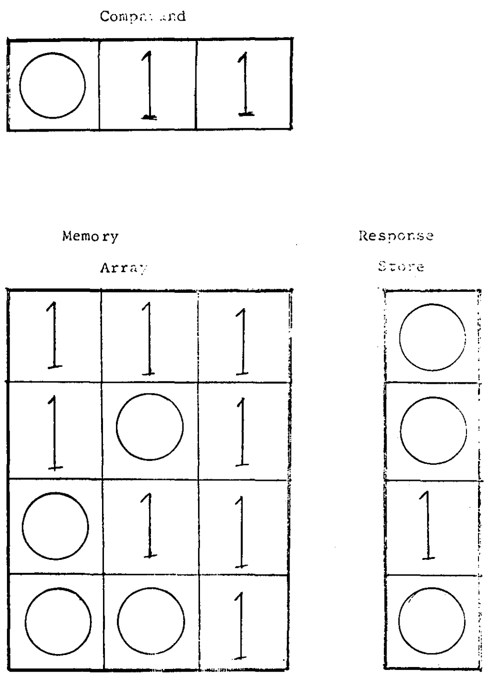


Figure 3. The associative memory executes an exact match search. Any location containing the same information as the comparand register is indicated by a "1" in the response store.

the instructions which could be implemented in an associative memory using present technology.[1,2,3]

ASSOCIATIVE ALGORITHMS

LOGICAL INSTRUCTIONS

- Exact Match of Comparand
- Mismatch of Comparand
- Less Than Comparand
- Greater Than Comparand
- Less Than or Equal to Comparand
- Greater Than or Equal to Comparand
- Between Limiting Comparands
- Minimum Value
- Maximum Value
- Next Lower Than Comparand
- Next Higher Than Comparand
- Long Left Shift
- Long Right Shift
- AND To Storage
- OR To Storage
- Exclusive OR To Storage
- Masked Store
- Store
- Masked Read
- Read
- Set Bits Plus
- Set Bits Minus
- Complement Bits

ARITHMETIC INSTRUCTIONS

- Add One
- Add Comparand
- Add Comparand, Save (Augends)
- Add Fields
- Add Fields, Save (Augends)
- Subtract One
- Subtract Comparand
- Subtract Comparand, Save (Minuends)
- Subtract From Comparand
- Subtract From Comparand, Save (Subtrahends)

Subtract Fields
Subtract Fields, Save (Minuends)
One's Complement
Two's Complement
Multiply by Comparand
Multiply by Comparand, Round
Multiply Fields, Round
Multiply Fields, Save (Multipliers)
Multiply Fields, Save (Multipliers), Round
Square
Square, Round
Round
Divide by Comparand
Divide Into Comparand
Divide Fields
Square Root

Associative memories are not available with all these capabilities, but more capability can be expected with each succeeding hardware generation. One of the most advanced memories currently on order is one being procured by RADC. It will have a capacity of 2,048 48-bit words. It will be able to perform Exact Match, Less Than, Greater Than, Between Limits, Next Higher, Next Lower, Maximum Value, Minimum Value instructions and also have variable-word-length capability. The Exact Match operation will be performed in 10 microseconds.

Future generations of associative memories will undoubtedly increase in both capacity and capability. Therefore, the associative memory may hold the key to increasing data-processing capability.

One step further into the future, beyond the associative memory, is the parallel processor. A parallel processor can be thought of as a machine which is capable of executing an arbitrary number of subprograms simultaneously. The first machine organization with this type of capability was proposed by John Holland[11,12] in 1959. In his paper he described a two-dimensional example which was essentially a rectangular grid of identical modules, each containing arithmetic capability, storage, path-building, and a certain amount of control logic. It was then possible for different groups of these modules to work together to execute a subprogram. This machine organization was not intended to be practical to implement with hardware. Currently, studies are being performed in an attempt to find feasible implementations of this basic capability. The associative memory exhibits many of the desired characteristics and may, in fact, be the building block that is needed.

CONCLUSION

A definite discrepancy exists between the bulk-storage and file-processing requirements of some of the larger electronic information-handling systems and currently available hardware. However, techniques for batch fabrication of ferrites, currently being developed, promise much larger memory systems. In addition, such memories as the Dove Data Device promise read-only capacities of 10^{11} bits. When these memory improvements are coupled with the studies in machine organizations, such as the parallel processor studies, then the result will go a long way toward satisfying large-system requirements. The work in the micro-miniaturized electronic circuits field promises a large-scale systems package in an amazingly small volume.

REFERENCES

1. Falkoff, A. D., "Algorithms for Parallel Search," *ACM J.* (October 1962), pp. 488–511.
2. Estrin, G., and R. H. Fuller, "Algorithms for Content-Addressable Memories," *Proceedings of 1963 Pacific Computer Conference*, pp. 118–130.
3. GER-11434, "Associative Memory Response Store Comparisons," Akron, Ohio, Goodyear Aerospace Corporation (September 1964).
4. Brownlow, J. M., E. A. Bartkus, and O. A. Gutwin, "An Approach Toward Batch Fabricated Ferrite Memory Planes: Part II, Fabrication and Yield," IBM, Watson Research Center, Yorktown Heights, New York.
5. "Cards Form Biggest Memory," *Electronics* (Feb. 14, 1964).
6. No. 60045000 (Rev. 6-64), Control Data 6600 *Computer System Reference Manual.*
7. Davies, P. M., "Design for an Associative Computer," *Proceedings of 1963 Pacific Computer Conference*, pp. 109–117.
8. McLaughlin, "Disc File Memories," in Milton H. Aronson (ed.), *Data Storage Handbook* (Instruments Publishing Company, Inc.).
9. RADC-TDR-64-307, *Dove Data Storage and Retrieval System,* AF30(602) 3192 (May 1964).
10. Gall, R. G., "A Hardware-Integrated Search Memory," Fall Joint Computer Conference, 1964 (Spartan, 1964).
11. Holland, John, "A Universal Computer Capable of Executing an Arbitrary Number of Subprograms Simultaneously," Eastern Joint Computer Conference (1959).
12. Holland, John, "Iterative Circuit Computers," Western Joint Computer Conference (1960).
13. *IBM 2321 Data Cell Drive* (Form A26-5851-0).
14. *IBM 7340 Hypertape Drive* (Form A22-6616).
15. Elfant, R. F., K. R. Grebe, and W. A. Crapo, "An Approach Toward Batch Fabricated Ferrite Memory Planes: Part I, Device Performance and Array

Characteristics," IBM, Watson Research Center, Yorktown Heights, New York.
16. GER-10936 S-7: *Information Storage and Retrieval Study for Diabetes*, Goodyear Aerospace Corporation (1962–63).
17. New System/360 IBM Data Processor, vol. VII, no. 10 (Apr. 16, 1964).
18. Rajchman, Jan A., "Magnetic Memories—Capabilities and Limitations," *J. Appl. Phys.* (April 1963).
19. Shoulders, K. R., "Research in Microelectronics Using Electron-Beam—Activated Machining Techniques."

24

Education Needed

WILLIAM F. ATCHISON
Rich Electronic Computer Center
and School of Information Science
Georgia Institute of Technology

The approach I will take in making this presentation on the education needed is to discuss what the fields of information science and computer science are and, in so doing, try to point out what I feel is needed in these areas. In so doing, I believe the area of Electronic Information Handling will at least be broadly covered. These are new fields just appearing on the horizon. They typify the fact that education is on the march. New approaches are being taken. Particularly in the new fields, new ideas will have to be used.

Let's put the computer in the classroom and let it help us in our teaching and learning processes. After all, it was made to serve us so let's use it in education too. A number of schools, in fact, already have remote input-output units to their computers. Many others are moving in that direction. Most of this use is in the application area, but I believe we will see this diversified further to get closer to actual teaching processes such as we see at the University of Illinois, Systems Development Corporation, and others.

I like the new trends I see in education. I approve of the new mathematics in the school systems, provided it is well done. I like the push in education to get away from rote learning and to teach the student to discover and learn for himself. It is not enough to have the student learn facts—he must get knowledge, understanding, and wisdom.

Three years ago, when Georgia Tech made its National Science Foundation-sponsored study on the training of science information specialists,* several of us interviewed a number of thoughtful people concerned about these problems. Several of these people, as well as many since, indicated the transient state of the field of information science. It was and is felt that student educational time should not be spent in teaching things that will disappear from the scene in a short period of time. As one of my friends in the computer science field put it, "Don't train a man in college in a computer technology that five years from now will be obsolete. Be

*See *Proceedings of the Conference in Training Science Information Specialists*, October 12–13, 1961 and April 12–13, 1962, Georgia Institute of Technology, Atlanta, Ga.

sure that you give the student basic knowledge and understanding on which he can build. He must have an understanding and an ability to adapt his learning to new situations."

It is with these things in mind that Georgia Tech recommended the development of a graduate program in Information Science aimed at educating personnel and doing research in the underlying principles on which Information Science is based. Our program admits only students who have a basic background in science or engineering. We want students who are grounded in the scientific method and thus, hopefully, will be in a better position to attack the problems confronting the field. The problems are not trivial and will require research workers of the highest caliber.

My interest in the field of information science stems from the fact that I feel there is a large overlap between Information Science (I.S.) and Computer Science (C.S.), i.e., to employ the terminology of the new math, the intersection of the set of I.S. knowledge with the set of C.S. knowledge is not the null set. The truth of the matter is that it is a large subset. Even further, I would not be scornful of anyone who would claim either one as a subset of the other. Each point of view could be defended.

The hierarchy of these new fields, such as Computer Science, Information Science, and Communication Sciences, have received considerable attention. Recently, John Hamblen and I published† a conjectured relationship among a number of these fields. Neither of us felt strongly dedicated to this table but, rather, we did it to invite discussion and provoke thought. No one has demanded a change in the table of relations, but many have expressed interest in it.

Keenan's recent article on "Computers and Education"‡ gives a very excellent discussion of computer science. His article discusses what computer science is at some length, but essentially he states that computer science is what computer scientists do, and this is largely covered by the following four topics:

1. Organization and interaction of equipment constituting an information processing system. The system can include both machinery and people, and its organization will be influenced by the environment in which it is embedded.
2. Development of software systems with which to control and communicate with equipment. Here is included, for example, mechanical languages, executive systems, systems to facilitate the reception and display of visual or aural information, etc.

† *Communications of the ACM*, vol. 7, no. 4 (April 1964), pp. 225–227.
‡ *Ibid.*, pp. 205–209.

3. Derivation and study of procedures and basic theories for the specification of processes. Specific topics included would be numerical analysis, list-processing procedures, heuristics and a theoretic basis for information retrieval.
4. Application of systems, software, procedures, and theories of computer science to other disciplines. A continuing awareness of potential applications is a stimulus to the computer scientist as it is in other disciplines.

Now, on the other hand, let me state the definitions for information science that came out of the Georgia Tech study. It states that I.S. is the science that investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability. The process includes the origination, dissemination, collection, organization, storage, retrieval, interpretation, and use of information. The field is derived from or related to mathematics, logic, linguistics, psychology, computer technology, operations research, the graphic arts, communications, library science, management, and some other fields.

Let us now take a look at each of these definitions in turn and, in so doing, try to point out the education needed. I assert that if you reverse the terms I.S. and C.S. in the two definitions you do not get a bad definition for the other field. Or, perhaps said more fairly, neither field can deny the pertinence of the subject matter of the other.

Looking first at the definition for computer science, let's start with *Topic 1* in the C.S. definition. This gives a reasonable picture of what our libraries and information centers do right now. The definition may intuitively imply more machinery than one currently finds in the conventional library. Looking ahead, however, to the automated library, it's not a bad description. It is, in fact, a very good description of what goes on at Documentation, Inc., or the Defense Documentation Center at Washington, D.C. Centers such as these, or their future replacements, are the kind of thing that we have to slant our education program toward. These are indeed the concern of both the computer scientist and the information scientist. If there is a difference, it is probably more a matter of viewpoint than it is of fact.

Topic 2 of the computer science definition is a point of major interest to me as an information scientist as well as a computer scientist. One of my fond hopes is that some day we will come up with a good computer language (or languages) for information storage and retrieval. I go along with the current trend for development of special computer languages to do special jobs. I have been trying to move Georgia Tech toward development in this area.

This past summer we ran an experimental course in I.S. which was a survey of computer languages. We are repeating it again this fall. This time each man already has a fairly thorough knowledge of at least one problem-oriented language and some have good knowledge of several. Our emphasis in this course is on list-processing languages such as IPLV, LISP, and COMIT. We will run this course again this winter and then in the spring, put these students in a course on how to construct a compiler language. The hope then is that one or more of these students will catch fire and help build one or more special languages for information storage and retrieval.

This effort to develop a special language points up one of our major problems in information science which is at the same time an education problem. People will need a lot of education to be brought around to using the new systems. The problem of being afraid of the computer is not unique to the librarian. We all had to face it with every kind of engineer and scientist at Georgia Tech. I know also that this is a problem shared by all of my colleagues who are directors of computer centers.

Fortunately, this problem has been diminishing in size, thanks to the new languages for computers. It is now much easier to talk to a computer and tell it your problem. My hope now is that we can soon have a computer language that makes it easy for the special librarian, the science information specialist, or the information scientist to communicate with the computer. This is why we are developing the above sequence of courses at Tech. A better language—a special one—might help. The availability of ALGOL on the Tech campus made a big difference. We need a good Information Processing Language for information science. Perhaps we can call it ISARL (*I*nformation *S*torage *A*nd *R*etrieval *L*anguage).

If you remove the mention of numerical analysis from *Topic 3* you sound explicitly like you are talking about information science. Clearly the need for a theoretic base for information storage and retrieval is something that our educational processes must move toward. This is a clearly stated aim at Georgia Tech. Clearly, also, heuristics and artificial intelligence, when they are sufficiently developed, will greatly contribute to information retrieval contrasted with fact retrieval.

Included under this third topic is the study and development of major systems. This covers the area frequently referred to as systems analysis or design. A number of people in the computer area say the system is the important thing and would let this be the framework on which they would hang everything covered by computer science.

Here the education needed has to be broken apart into at least two major areas which I will mention for illustrative purposes. Certainly there are many shades of these both between and beyond.

The first of these is the man who sets up the system and is probably responsible for maintaining and updating it to best meet the needs of the outside user. This man clearly needs a solid background of knowledge and experience. I hope that special programs such as ours at Georgia Tech, and others in various stages of development across the country, can help meet this need.

A second is the user who has little contact with the functioning system itself but for which the system is only a valuable tool to help him get the information he wants. Educationwise, this involves the wide spectrum of training the scientist, the science librarian or the science information specialist in how best to exploit the facilities of the large scale system. It involves such things as knowing how to frame your inquiries to knowing how to interpret the answers you get. It involves knowing not to be upset if your answers come back all in capitals or even back in a coded form which might be all numerical. Education for things like this is our responsibility, though, of course, we are going to have to obtain help from many others than just the information or computer scientist.

Topic 4 as mentioned by Keenan was the application area. Certainly this phrasing would read just as well with information science replacing computer science as it does now. The fact is that it probably has more meaning using information science in it than it has with computer science. You have an even broader area of application with information science if you are sufficiently broad in what you mean by I.S. It is from the applications people that you can expect to get a lot of help in spreading the necessary education in the development and use of information systems. Certainly this has happened in the computer field. It is not a joke to report that the professors at schools have learned to use computers because their students did first.

Finally, let us look at the definition of information science given. We will have to look at each of the three sentences one by one. The first one is applicable to computer science, but is meant to be broader than one normally considers computer science. This is particularly true if you think of computer science as data processing rather than information processing. The interest of computer science now, however, is going far beyond data processing and is truly information processing. As evidence of this, at the ACM meeting this past summer our biggest crowds showed up at the list-processing sessions. There appeared to be a much bigger interest here than in the numerical analysis sessions to draw a direct contrast. This gives some justification for our need to emphasize list-processing languages in our education programs.

The second sentence is again probably too broad for computer science as normally conceived. The fact of the matter is that this definition and, in particular, this second sentence, is so broad that by proper interpretation you can include almost anything, including computer science.

It is the words, origination and interpretation, that cause the most trouble, but even these, to a partial extent, are applicable. For example, much information now does originate inside of a computer, and in some instances we can ask the computer to make some "interpretations" for us. Finally, there is no question but what the last sentence is applicable to both fields. Careful study might, in fact, reveal that additional fields could be added.

In conclusion, I would say that it behooves each one of us to push the educational aspects of information science, computer science—or, if you like, electronic information handling—in every way possible. We should encourage the development of separate programs in universities if possible, or if this is not possible, amplify existing programs. One illustration of this is the fact that many computer-science programs are developing within mathematic departments and, similarly, efforts to develop information science programs are progressing in library schools across the country. It will take all of these efforts to get the job done and we should do everything we can to insist on the high caliber of each of these programs.

25

The Information Retrieval Game*

ALLEN KENT
Director, Knowledge Availability Systems Center
University of Pittsburgh

PURPOSE

Those who design, operate, use and/or evaluate information retrieval systems are forced to make assumptions concerning the objectives, functions, performance requirements, and environmental variables of these systems. Some of these assumptions are explicit, some are implicit, and some are buried deep in the subconscious.

The purposes of this paper are:

(a) To identify and question the validity of some of these assumptions;
(b) To suggest basic problems that have not been investigated to date because of the interference of invalid assumptions;
(c) To describe an approach to investigating several of these problems;
(d) To present preliminary results of investigating one methodology developed in order to elucidate these problems.

INTRODUCTION

The problem of designing, operating, using, and evaluating an information-retrieval system would be a trivial one (a) if each event impinging on the consciousness of any human beings would result in identical streams of observations, (b) if each observer would use the identical words in identical configurations to describe each such single event, and (c) if each human being interested in learning of the event would phrase questions using identical terminology.

However, each individual has his own paradigms, or ways of perceiving nature. These paradigms are fundamental hypotheses or models in respect to which thinking occurs. As in all perception, a shift from one hypothesis to another may occur at any moment, and unpredictably.†

If this premise is accepted, then it follows that requests for service from an information-retrieval system will be based on clues which are verbali-

*Supported by National Institutes of Health Grant FR00202-01.
†E. G. Boring, *Science*, vol. 145 (1964), pp. 680–685.

zations of subjects based upon the *requestor's* hypotheses or models in respect to which their own thinking occurs.

How, then, can we design a system to react effectively to the paradigms of the requestors rather than those of (a) the authors of source materials included in the system, or (b) the interpreters of these materials when the system is designed or operated?

SOME ASSUMPTIONS MADE EXPLICIT

Information-retrieval systems have as a common goal the provision, on demand, with maximum precision and at minimum expense, information relevant to reasonable questions posed by persons who have socially important reasons for desiring responses.

Assumption 1

*Since information seekers approach information retrieval systems for service, they have been unwilling or unable to perform the service for themselves. Accordingly, they have made a conscious decision to delegate to others one or more of the unit operations involved in obtaining information.**

Some of the major reasons why individuals delegate information retrieval tasks to others relate to their inability (or unwillingness) personally to acquire, analyze, and/or store all of the information that may eventually be useful to them. Since no individual can predict, with absolute certainty at the time of acquisition, which source material will be useful at a later time, those who delegate information-retrieval tasks to others expect to receive, at the time that they make a request for information, only that subset of source materials from the entire store that is most closely relevant to their current interest.

Assumption 2

Some state or level of processing of original source materials will be a "best" level to permit identification of subsets, which are relevant to requestors' interests.

Common experience in operating information retrieval systems makes it quite clear that neither the system operator nor user considers all responses to questions as relevant. Accordingly, one or more of the following conditions may prevail:

(a) The system user has not stated his problem with sufficient precision.
(b) The system operator has not comprehended the problems as presented.

*See, for example, A. Kent, *Textbook on Mechanized Information Retrieval* (Wiley, 1962), pp. 9–10, 109.

(c) The system has not been designed properly.
(d) The system has not been operated properly.
(e) There may be no relevant responses in the file.

Assumption 3

Some level of analysis of user problems (as verbalized), can lead to effective operation of the system.

The three assumptions listed above may be valid or not, may be made consciously or not; nevertheless, they influence design, operation, use, and evaluation of systems.

WHERE THESE ASSUMPTIONS LEAD

SYSTEMS EVALUATION

When the number of source materials being collected exceeds the ability of a potential "inquirer" to read and remember the contents of every document, the rationale for the delegation of tasks to designers and operators of information retrieval systems becomes apparent. Obviously, it is precisely at this point that the designer and operator can no longer assume that a potential user of the system will have previously read the text of source materials that may be of interest to him. Nevertheless, the designer and operator must select (index, classify), from the text, clues that will be useful in organizing the materials for ready identification *even though questions directed to the system will not come from the text of the documents on file but rather based on the users' paradims.*

Here, then, is the basis for much of the uncertainty in predetermining the effectiveness with which a system will operate in providing responses that meet the users' criteria for excellence.

And compounded upon this uncertainty has been much of the recent work directed to evaluating competing information storage and retrieval systems. These approaches have involved the processing of identified collections of source materials in parallel. The collections are then searched in response to questions using each of the systems, in an attempt to determine the effectiveness of each system to produce relevant material and suppress irrelevant material.

One such test method was based on the formulation of test questions by scientists and engineers. Each scientist and engineer participating in the experiment was provided with a set of source materials and asked to frame questions each of which could be satisfactorily answered by one of the source materials. The systems under test were operated and the quality of results analyzed.

The test results exhibited less significant differences in the performance of the systems compared than the systems operators would have led one to believe.

This investigation has buried in it an assumption that threatens the validity of all the results reported. Questions were "framed" by test participants when the "answers" were in their hands, a situation that is so unlike the real reference problem that one is tempted to examine the questions in order to discern whether they are indeed realistic.*

An example of one was: "Impedance testing of aircraft power control units" and its proper answer was provided by a technical periodical article entitled: "A possible method of impedance testing aircraft power control units." Although it is obvious that any system which restricted its indexing or cataloging to titles of source materials might have performed well, this is not the fundamental danger signal that is raised by this evaluation approach. A question formulated as in this investigation mirrors or attempts to mirror the problem faced by a person who has seen a desired report or article before, and who now frames a question based on his best recollection of its title or contents.

However, since it cannot be assumed that a potential user will have previously read the text of source materials that will be of interest, the systems must be evaluated in their performance in responses to *real questions which reflect users' paradigms, and not influenced in advance by exposure to source materials.*

RESPONSE PRODUCTS OF SYSTEMS

One of the consequences of uncertainty in the performance of systems has been to permit the user to evaluate intermediate response products before being exposed to the source materials themselves. It is expected that these intermediate response products will be useful to the users as predictors of the actual relevance of the final response products. Systems designers and operators have traditionally assumed that titles, abstracts, and/or extracts will be useful *intermediate* response products. However, these products are prepared by authors of source materials or by operators of systems, and again there is *no quantitative evidence available as to how accurately they may reflect the users' paradigms.*

In considering which final products will be most effective in providing service to users, it has been observed that many source materials contain more information than is apparently desired by the user, *as reflected by the formal statement of his question.* Accordingly, some systems designers and operators have chosen to provide information or data derived from source materials as final products. In so doing the final product is re-

*D. Swonson, *Library Quarterly*, vol. 35, no. 1, pp. 1–20 (Jan. 1965).

moved from the author's paradigm as represented by the full source material. The tacit assumption is thus made that the operator understands sufficiently the users' paradigms, *an assumption not generally borne out even by qualitative evaluation of systems responses by users.*

PROVISION OF SYSTEMS PRODUCTS IN PARALLEL

So deeply embedded are the implicit assumptions with regard to the ability of systems operators to reflect accurately the paradigms of potential users during initial processing of source materials that there results a basic criterion engineered into systems which is highly questionable. That criterion is that the operation of a system in response to a question shall result in the provision of all materials which meet search specifications prepared as a result of analysis of the formal statement of a user's requirement.

The number of responses resulting from a single search may be large or small; however, all of them are provided, in parallel, to the user. The user, on the other hand, can only review responses one at a time, with learning possibly taking place as information is assimilated during the review process. It can be assumed that at least in some cases this learning results in reformulation of the user requirements, and loss of interest in those responses still to be reviewed.

Since requirements for speed of operation of systems have been formulated on the basis of parallel responses, it is therefore prudent to reexamine the basic criterion in terms of more limited responses, with ability to reformulate questions in real-time.

TERMINOLOGY CONTROL DURING INPUT AND OUTPUT PROCESSING

In recognizing that significant differences may exist between the "language" of information retrieval systems and that of questions directed to the systems, various terminology-control approaches are used to assure effective service by providing a bridge between the two languages. The approaches involve:

1. Establishment of a "standard" indexing language by the system designer or operator which is used to express essential characteristics of source materials processed for inclusion in the system; and analysis of questions in terms of this "standard" language.
2. Use of terminology of authors of source materials for processing of source materials, and use of:
   (a) A thesaurus of related terms which is available to operators and users of the system for review during analysis of questions.

(b) Weighting of the terminology with regard to probable usefulness in identifying desired information for specific users, in terms of experience and feedback derived from operation of the system.

Both of these approaches are based on the paradigms of authors of source materials and operators of systems, with feedback from users serving to adjust search strategies. Although, empirically, satisfaction in use of systems may be obtained, there is no basic information derived which throws light on user paradigms without reference to the contamination of author or operator paradigms. Also, these approaches involve redelegation by the systems operators to the users of tasks that the users wished to delegate to others.

CURRENT RESEARCH INVESTIGATIONS IN THE FIELD

The assumptions discussed earlier have also influenced significantly much of the research that is now being conducted throughout the country.

Based on the assumption that some level of processing of original source materials will yield an optimum system for retrieving relevant information on demand, attempts are being made to:

1. Identify "key" words of titles, abstracts, extracts, or full texts in order to index, classify, abstract, or extract automatically.
2. Seek regularities in structure of language in order to normalize abstract or full texts as a basis for indexing, abstracting, or extracting automatically.
3. Analyze terminology from source materials used for indexing in order to discern inherent concepts which would serve as reference points for searches.
4. Select and assign indicators which would display the role played by words selected for indexing purposes, in an attempt to limit nonrelevant responses from the system.
5. Assign linkages among words selected for indexing purposes, also in an attempt to limit nonrelevant responses from the system.
6. Weight usefulness of words selected for indexing purposes on the basis of (a) frequency of occurrence in natural text, or (b) qualitative value judgments by system operators.

Each of these approaches concerns itself with author and system operator paradigms, without consideration of pure user paradigms, uncontaminated by prejudgments made by others.

It is in an attempt to isolate and examine user paradigms that the investigations described below have been designed.

*THE HEURISTIC INFORMATION RETRIEVAL GAME**

INTRODUCTION

As discussed above, there have been many hypotheses made about users of information during the development of information storage and retrieval systems which have not been examined experimentally. Some of the questions which will be investigated in the program described below are:

1. Are there any individual or common patterns exhibited by users in making decisions regarding the relevance of materials provided in response to questions that can be discerned experimentally?
2. What is the effect, if any, on relevance patterns of:
   (a) Subject field of user?
   (b) Organizational level of user?
   (c) Nature of question?
3. What is the effect, if any, on relevance decisions made by users, of the order in which materials are provided in response to questions?
4. What is the effect, if any, of the type of evidence of contents of source materials provided to a user in response to questions (e.g., titles, abstracts, extracts), on the ability of the user to predict accurately the relevance of the actual source materials?
5. To what extent do the words or expressions found in user questions correlate with words or expressions found in the evidences of contents of source materials which users find relevant?
6. To what extent can associations among words found in questions, with words found in evidences of contents of relevant source material, be predicted by word association tests?

In designing an experimental program to throw light on these questions, there are two fundamental assumptions that have been made:

1. The user of an information retrieval (IR) system is the ultimate judge of which information provided to him is relevant to questions that he wishes to have answered, regardless of how he has verbalized these to the system operator. Thus, *there can be no expert opinion which rules a question to be inappropriate, or a response relevant or not. Only the user's paradigms are to be served by a system rather than some consensus by others who may feel they know what is really wanted by the user, or who claim to know what he should want.*

*See A. Kent, *Amer. Doc.*, vol. 15, no. 2 (1964), pp. 150–151.

2. In order to measure the effectiveness of an IR system in providing relevant information to users, *the questions posed to the system must be derived from real needs of users who are motivated in some real way to have responses.*

DEVELOPMENT OF THE GAME

The human thinking process seems to follow a procedure in which we create in our minds a map or model of the real world. An individual uses several aids both in constructing these models and in communicating their salient features to others; one of these aids is the simulation, in which an attempt is made to recreate the basic functions, processes, and their interrelationships that most accurately depict the situation under study. The game is one of the forms of simulation. The traditional business or war game consists of a controlled situation in which an individual or a team competes against intelligent adversaries and against an environment in order to attain predetermined objectives. In the game, the players contend with several interacting variables, some of which are under their control. The heuristic IR game is developed in analogous fashion, except that the only "opponent" will be the entropy of the IR systems environment.

The IR game has as its chief purpose the investigation of the behavior of the three human components of the game: the players—IR system users; the instructor—IR systems operator; and the referee—the information scientist. The game is being developed heuristically with intermediate objectives emerging as the game proceeds. The ultimate objective is to gain insight into what constitutes relevance in an IR system, so that quantitative systems design criteria may be developed on the basis of user paradigms.

The primary players of the game are controlled groups of IR systems users who are attempting to derive maximum benefit from a collection of source materials by locating information relevant to a problem or question that interests them.

The instructor in the traditional game is responsible (1) for teaching the game in order that the players may know what rules to use in developing their strategies (in this case, the strategies of search), and (2) for indicating to the players what constitutes success. In the IR game the players have joint responsibility with the instructor in defining success, at least initially, so that the game may develop heuristically. However, the player reverts to his traditional status once he has helped define success (by his reactions) and then is scored on his consistency in applying rules that he has helped to establish.

In the traditional game the referee is a person (or computer) who scores responses and monitors the play. However, as stated earlier, since the IR game is developed heuristically, the referee, an information scientist, is observing the behavior of the players and the instructor and is developing tentative rules, scoring on the basis of these rules, and modifying them as appears appropriate. The referee is also responsible for modifying input stimuli to the players as appears appropriate.

A set of questions or problems of interest to the players is elicited in advance of the play. A set of source material documents is selected, some of which are of probable relevance to the questions, some of which are probably only of partial relevance, and some of which are tacitly irrelevant. Each document is prepared in a variety of levels and forms of processing for presentation to the players.

Responses to questions in a variety of states and forms, and in a variety of probable relevances or irrelevances, are presented to the players:

1. At random.
2. Structured according to probable relevance.
3. Structured according to state of processing.
4. Structured according to probable desired form.

The players are asked to rate the relevance of material presented in the response to their questions:

1. On the basis of a yes-no decision.
2. On the basis of a tentative scale of values.

After a pattern of response may be discerned for each player, further presentations are programmed by the referee to investigate the consistency of response. Cross correlations among players' responses in similar and dissimilar groups are also investigated by programming derived presentation patterns of one individual for response by another.

DEBUGGING THE PLAY OF THE GAME

A new experimental procedure to be used for studying the nature of a complex behavioral phenomenon usually must be perfected by successive approximations. Various segments of the heuristic IR game for studying the nature of relevance have been, and will continue to be, subjected to various debugging trials before the full game is attempted, and before plays will be expected to yield reliable data. Some of these trials are described below.

Trials to Debug Procedures

A class of thirty-four students* in the Information Sciences curriculum of the University of Pittsburgh was chosen as the first group to be subjected to the play of the IR game.

*Class entitled "Mechanized Information Retrieval," taught by A. Kent in the Master's program of the Graduate School of Library and Information Sciences, University of Pittsburgh.

A question was prepared which would be understood by all players and where a general educational background would be sufficient to permit evaluation of the relevance of responses provided. The administration of the game proceeded as follows:

1. *Explanation of purposes:* The general objectives of the entire experiment and of the specific trial were described.
2. *Mechanics of the play:*
   (a) Students were exposed to the question (Fig. 1) which they were to adopt as their own, and against which they would be asked to judge the relevance of responses provided to them.

> I would like to have all the information available on the amendments to the national constitution now pending in various state legislatures.

Figure 1. Question Chosen for First Play of IR Game.

   (b) Stimulus evaluation forms were distributed (Fig. 2) to each student and instructions were given on how to complete them.

RESPONSE SHEET—UNIVERSITY OF PITTSBURGH INFORMATION RETRIEVAL GAME

Major Field of Interest ______________________

Doc.	Pertinent	Maybe			Not
1.					
2.					
3.					
4.					
5.					
6.					
7.					
8.					
9.					
10.					

Figure 2. Stimulus Evaluation Form Used. One of three evaluations was permitted for each response submitted to players: pertinent; may be pertinent, and nonpertinent.

(c) Play was commenced by presenting to the students stimuli consisting, successively, of segments of source materials which might be pertinent or not to the question shown in Fig. 1. The contents of the stimuli were abstracts or excerpts from the source materials. The excerpts were one of the following: title; first paragraph; last paragraph; or one or more sentences, or a paragraph, selected anywhere from the total text. Seventy-four such stimuli recorded on transparencies were presented on a screen, using an overhead projector. Stimuli were exposed to the students for varying periods of time. In several instances, identical stimuli were repeated without warning to the students.

Results of Trials

Examples of the texts of stimuli presented to the students, as well as their evaluation, are given in Table 1. A complete tabulation of the results of the trial play is given in Table 2. These same data are rearranged in Table 3 to bring together the evaluations of the same source material for each level of processing, so that the predictive value of each level of processing in assessing relevance of the full source material (as determined by the referee) may be compared. In Table 4 the number of agreements and disagreements on relevance of source materials between referee and players is tabulated for each of the levels of processing; these data are summarized in Table 5.

For those stimuli which were exposed to the players twice, the evaluations provided for each of the stimulus pairs are given in Table 6.

Discussion of Results of Debugging Trials

A number of impressions were obtained from the initial trials which are to be taken into account in planning for subsequent plays:

(a) *Time of exposure of stimuli.* Each player was permitted to view the stimulus for a set period of time as shown in Table 2. Student comments following the trials suggest that a control group be permitted as much time as it requires to make the decisions required in the game. It then might be instructive to determine the effect of the amount of time taken on relevance decisions for the control group as well as for the time-restricted group of players.

(b) *Method of presentation of stimulus.* Given the physical shape of a classroom and the large number of students engaged in the trials simultaneously, it was evident that some players were not able to read the projected stimuli as well as others. Accordingly it is believed that future trials will be designed so that each player may have an individual viewing screen or individual notebook with stimuli more readily readable.

TABLE 1. EXAMPLES OF STIMULI, TEST CONDITIONS AND RESULTS

Test of stimulus	Level of processing	Period of exposure, seconds	Evaluation of source material			
			By Referee	By Students		
			(P—Pertinent) (N—Nonpertinent)	Pertinent	May be pertinent	Non-pertinent
Silent amendments	Title	10	P	23	1	10
The current fate of some proposed constitutional amendments	Abstract (annotation)	15	P	33	0	1
Report on the latest attempt at amending the United States Constitution	Abstract (annotation)	15	P	30	1	3
The assault on the Union	Title	10	P	9	6	19
The book includes a reprint of the Constitution, but the body of the book takes up the Constitution point by point, emphasizing and clarifying. Specific examples of modern day legislation are given to show how the Constitution and its principles are affecting legislation.	Abstract	20	N	14	3	17
There is underway a movement to change radically the American form of government. Indeed, there has been nothing like it for a hundred years. There are three proposals to amend the Constitution now being pressed in various state legislatures which strike at the very foundation of the American Union as deeply as anything which has been agitated seriously since nullification and secession.	First paragraph	30	P	32	0	2

Tortoise vs. Hare	Title	5	P	5	0	29
There is only one hope—an informed electorate, an intelligent electorate, a community wherein the majority of citizens endeavor to hear both sides. This majority can and will reaffirm in our free republic that only by the votes of the majority of the national legislature—not by the executive edicts of a few—shall laws be passed, and only by a three-fourths vote of our states and two-thirds vote of Congress shall the Constitution be considered to be properly amended.	Last paragraph	30	P	27	0	7

TABLE 2. COMPLETE RESULTS OF ONE PLAY

STIMULUS No.	Level of processing	Exposure time	Evaluation of source material: By Referee (P—Pertinent) (N—Nonpertinent)	By Student: Pertinent	By Student: May be pertinent	By Student: Non-pertinent	Source material identification letter
1	Title	10	P	23	1	10	A
2	Abstract	15	P	33	0	1	E
3	Abstract	15	P	30	1	3	G
4	Last paragraph	30	N	14	5	15	K
5	Title	10	P	9	6	19	E
6	Abstract	20	N	14	3	17	N
7	Title	10	N	8	7	19	Q
8	Last paragraph	20	P	10	4	20	C
9	Extract	15	N	16	0	18	Q
10	Extract	30	P	28	5	1	C
11	Title	10	N	16	1	17	L
12	Extract	15	P	8	1	25	J
13	First paragraph	30	P	32	0	2	E
14	Extract	25	P	25	3	6	M
15	Abstract	15	P	30	2	2	F
16	Title	5	P	5	0	29	D
17	Last paragraph	30	P	27	0	7	H

18	Title	10	P	31	0	3	J
19	Abstract	15	P	13	6	15	A
20	Last paragraph	20	N	18	6	10	I
21	First paragraph	30	N	6	7	21	L
22	First paragraph	15	P	13	2	19	D
23	Last paragraph	15	P	30	1	3	A
24	Last paragraph	20	N	15	9	10	L
25	Abstract	20	P	24	4	6	J
26	First paragraph	20	P	34	0	0	C
27	Title	5	P	3	4	27	C
28	Last paragraph	10	P	4	1	29	D
29	Title	10	P	5	1	28	H
30	Abstract	15	N	3	10	21	K
31	Extract	15	P	29	1	4	F
32	First paragraph	20	P	30	0	4	J
33	Extract	15	P	29	3	2	D
34	Last paragraph	30	P	27	4	3	E
35	First paragraph	15	P	18	3	13	D
36	Extract	10	N	6	9	19	I
37	Title	10	N	23	0	11	N
38	Extract	20	P	25	2	7	J
39	First paragraph	20	P	22	3	9	G
40	Abstract	15	P	27	4	3	D
41	Extract	20	P	30	2	2	B
42	Title	10	N	6	8	20	K
43	First paragraph	30	N	7	7	20	K
44	First paragraph	20	P	33	1	0	F
45	First paragraph	25	N	9	6	19	I

(Continued)

TABLE 2. COMPLETE RESULTS OF ONE PLAY (Continued)

STIMULUS No.	Level of processing	Exposure time	Evaluation of source material				Source material identification letter
			By Referee (P—Pertinent) (N—Nonpertinent)	By Student			
				Pertinent	May be pertinent	Non-pertinent	
46	First paragraph	10	P	10	0	24	H
47	Last paragraph	15	P	28	2	4	B
48	Last paragraph	15	P	26	5	3	F
49	Abstract	15	N	17	6	11	Q
50	Abstract	30	P	9	7	18	M
51	Title	5	P	22	5	7	F
52	Title	5	P	32	0	2	G
53	Abstract	10	N	28	1	5	I
54	Abstract	15	N	5	10	19	L
55	First paragraph	20	P	32	1	1	A
56	Title	5	P	9	5	20	M
57	Extract	15	P	30	2	2	G
58	Title	5	P	10	2	22	B
59	Extract	10	P	6	5	23	H
60	First paragraph	20	P	20	2	12	M
61	Last paragraph	20	P	31	1	2	G
62	Extract	20	P	27	4	3	J

63	Title	5	N	19	3	12	I
64	Extract	30	P	33	1	0	E
65	Abstract	15	P	34	0	0	E
66	Extract	15	N	13	5	16	Q
67	Extract	15	P	11	3	20	J
68	Title	10	P	31	3	0	J
69	Last paragraph	15	P	34	0	0	A
70	Abstract	15	P	26	5	3	D
71	Title	10	N	0	2	32	K
72	Extract	20	P	23	6	5	J
73	Last paragraph	10	P	0	2	32	D
74	Extract	15	P	34	0	0	F

TABLE 3. COMPARATIVE RESULTS
Same Source, Different Processing Levels

Source material identification letter	Evaluation of source material				Level of processing
	By Referee	By Student			
		Pertinent	May be pertinent	Non-pertinent	
A	P	23	1	10	Title
		13	6	15	Abstract
		30	1	3	Last paragraph
		32	1	1	First paragraph
B	P	10	2	22	Title
		30	2	2	Extract
		28	2	4	Last paragraph
C	P	3	4	27	Title
		10	4	20	Last paragraph
		34	0	0	First paragraph
		28	5	1	Extract
D	P	5	0	29	Title
		27	4	3	Abstract
		4	1	29	Last paragraph
		13	2	19	First paragraph
		29	3	2	Extract
E	P	9	6	19	Title
		33	0	1	Abstract
		32	0	2	First paragraph
		27	4	3	Last paragraph
		33	1	0	Extract
F	P	22	5	7	Title
		30	2	2	Abstract
		26	5	3	Last paragraph
		33	1	0	First paragraph
		29	1	4	Extract
G	P	32	0	9	Title
		30	1	3	Abstract
		22	3	9	First paragraph
		31	1	2	Last paragraph
		30	2	2	Extract
H	P	5	1	28	Title
		27	0	7	Last paragraph
		10	0	24	First paragraph
		6	5	23	Extract

(Continued)

TABLE 3. COMPARATIVE RESULTS (Continued)
Same Source, Different Processing Levels

Source material identification letter	Evaluation of source material				Level of processing
	By Referee	By Student			
		Pertinent	May be pertinent	Non-pertinent	
I	N	19	3	12	Title
		28	1	5	Abstract
		18	6	10	Last paragraph
		9	6	19	First paragraph
		6	9	19	Extract
J	P	31	0	3	Title
		24	4	6	Abstract
		30	0	4	First paragraph
		8	1	25	Extract
		25	2	7	Extract
K	N	6	8	20	Title
		3	10	21	Abstract
		14	5	15	Last paragraph
		7	7	20	First paragraph
L	N	16	1	17	Title
		5	10	19	Abstract
		15	9	10	Last paragraph
		6	7	21	First paragraph
M	P	9	5	24	Title
		9	7	18	Abstract
		20	2	12	First paragraph
		25	3	6	Extract
N	N	23	0	11	Title
		14	3	17	Abstract
O	N	8	7	19	Title
		17	6	11	Abstract
		16	0	18	Extract

(c) *Selection of question.* Although an artificial choice of question was made to permit debugging trials to be performed on a large number of students, it is believed to be the sine qua non of this experimental procedure that there be considerable motivation on the part of players to view the stimuli, and eventually to read the full source materials.

However, since there was a great deal of student interest in participating in a new experimental procedure, there may have been an unconscious adoption of a favorable attitude toward the question imposed on them. In

TABLE 4. AGREEMENTS (DISAGREEMENTS) WITH REFEREE RATINGS AT VARIOUS PROCESSING LEVELS OF SOURCE MATERIALS

Title		Abstract		First paragraph		Last paragraph		Extract	
Reference rating		Reference rating		Reference rating		Reference rating		Reference rating	
Pertinent	Non-pertinent	Pertinent	Non-pertinent	Pertinent	Non-pertinent	Pertinent	Non-pertinent	Pertinent	Non-pertinent
23(10)	19(8)	33(1)	17(14)	32(2)	21(6)	10(20)	15(14)	28(1)	18(16)
9(19)	17(16)	30(3)	21(3)	13(19)	20(7)	27(7)	10(18)	8(25)	19(6)
5(29)	11(23)	30(2)	11(17)	34(0)	19(9)	30(3)	10(15)	25(6)	16(13)
31(3)	20(6)	13(15)	5(28)	30(4)		4(29)		29(4)	
3(27)	12(19)	24(6)	19(5)	18(13)		27(3)		29(2)	
5(28)	32(0)	27(3)		22(9)		28(4)		25(7)	
22(7)		9(18)		33(0)		26(3)		30(2)	
32(2)		34(0)		10(24)		31(2)		30(2)	
9(20)		26(3)		32(1)		34(0)		6(23)	
10(22)				20(12)		0(32)		27(3)	
31(0)								33(0)	
								4(20)	
								23(5)	
								34(0)	

TABLE 5. REFEREE RATINGS
Agreements and Disagreements
Summary

Processing level	Correlation with referee ratings: Agreements (when pertinent; when nonpertinent)	Disagreements (when pertinent; when nonpertinent)
Title	291 (180; 111)	239 (167; 72)
Abstract	299 (226; 73)	117 (50; 67)
First paragraph	304 (244; 60)	106 (84; 22)
Last paragraph	252 (103; 47)	150 (103; 47)
Extract	384 (331; 53)	134 (99; 35)

any case, the atmosphere during the trials was cooperative and reflected a desire on the part of the students to be helpful, even though they knew that their grades did not depend upon their participation.

(d) *Repeating identical stimuli.* The mechanism of repeating identical stimuli unexpectedly during a long series of plays seems worthwhile, since this might throw some light on:

1. Consistency of players in making relevance decisions; and/or
2. Influence of learning on player decisions.

(e) *Predictive value of various levels of processing.* In the instructions, players were asked to rate the probable relevance of the *full source material to the question in terms of the stimulus presented* (reflecting various levels of processing of the full source materials). It is obvious from the data presented in Table 3 that with only some minor exceptions the level of processing had a very strong influence on the players' decisions regarding relevance of the source materials to the question.

It would be extremely interesting to determine whether plays of the game involving questions for which better motivation for procuring results may be assured, would lead to results as interesting as these. As will be noted from Table 5, percentage ability of the various levels of processing to predict relevance ratings by the referee was:

Title	55 %
Abstract	72 %
First paragraph	74 %
Last paragraph	63 %
Extract	74 %

The very significant observation that, if validated, would be extremely important, is that the first and/or last paragraph (which can be selected

TABLE 6. RELEVANCE
Decisions Made on Identical Stimuli Pairs

Source material identification letter	Stimulus number	Processing level	Referee rating	Evaluation by Students		
				Pertinent	Don't know	Nonpertinent
E	2	Abstract	Pertinent	33	0	1
	65	Abstract		34	0	0
O	9	Extract	Nonpertinent	16	0	18
	66	Extract		13	5	16
J	12	Extract	Pertinent	8	1	25
	67	Extract		11	3	20
J	18	Title	Pertinent	31	0	3
	68	Title		31	3	0
A	23	Last paragraph	Pertinent	30	1	3
	69	Last paragraph		34	0	0
D	40	Abstract	Pertinent	27	4	3
	70	Abstract		26	5	3
K	42	Title	Nonpertinent	6	8	20
	71	Title		0	2	32
J	62	Extract	Pertinent	27	4	3
	72	Extract		23	6	5
D	28	Last paragraph	Pertinent	4	1	29
	73	Last paragraph		0	2	32
F	31	Extract	Pertinent	29	1	4
	74	Extract		34	0	0

clerically at minimal processing expense) may be better predictors of the relevance of the full source documents than:

1. The Title, which is being used so much for "automatic" indexing using "Keyword in Context" procedures.
2. The Abstract, which is being prepared at considerable expense in many information storage and retrieval activities.
3. The Extract, which must be selected from the text of source materials by competent subject specialists.

Although it is recognized that it is totally premature to extrapolate at all from such invalidated debugging trials, the above observations are made only in order to stimulate additional investigations.

ANOTHER TRIAL TO DEBUG THE GAME

Another opportunity for investigating segments of the play of the IR game presented itself with another class* of 34 students in the information sciences curriculum at the University of Pittsburgh. The subject matter is taught in terms of:

1. A major, national survey of specialized information centers conducted by the instructor in 1963–1964.
2. Analysis of fundamental unit operations conducted at such centers.

The analysis of unit operations† (acquisition of source materials, analysis, terminology control, recording results of analysis in searchable medium, storage of source materials, question receipt and analysis, conducting of search, and delivery of search results), reveals that confidence limits claimed for performance by systems operators may be overly optimistic.

Accordingly, several of the unit operations were selected (acquisition, analysis, searching), and an attempt made to investigate the IR game as a tool for estimating confidence limits of performance of each operation.

The acquisition operation was one which lent itself best to this approach, and accordingly is described here.

Acquisition Policy

A policy for acquiring source materials for a specialized information center was presented, in writing, to each student (Fig. 3). In addition, a list of questions considered to be typical by the center involved was presented (Fig. 4).

*Class entitled "Specialized Information Centers," taught by A. Kent in the Master's program of the Graduate School of Library and Information Sciences.

†A. Kent, *Specialized Information Centers* (Spartan, 1965).

Ideally, of course, everything that has been written about a culture or area should be included in the file. For some cultures or areas, however, the material is so extensive that only a sample of the literature can be processed. This is the case with the Soviet Union. On the other hand, the bibliography on some cultures may be limited, as it is with the Burusho, and in those instances it is likely that all the available material will be processed.

Figure 3. Acquisition Policy of Human Relations Area File, Inc.

Explanation of Purpose

Each of the students participating in this play had also been a member of the class that was involved in the first debugging trials, so that it was necessary only to review the general objectives, and to specify the purpose of the current play.

Mechanics of the Play

As before, students were provided with stimulus evaluation forms. Play was commenced by presenting to the students stimuli consisting, successively, of segments of source materials which they might consider pertinent or not to the acquisition policy shown in Fig. 3. As before, the contents of the stimuli were abstracts or excerpts from the source material (see, for example, Table 1). Again, stimuli were exposed for varying

1. Do the Iroquois have the institution of blood brotherhood?
2. Where can one find information on the cultivating and processing of sugar?
3. Soil conditions, climate, and topography of Korea and Formosa.
4. Were smoke signals used by the Senecas or the Creeks, and what other communication methods were employed?
5. Facilities and methods for water transportation in Finland, but not kinds of craft used.
6. If poultry or dairy cattle are raised in Iraq, what methods are used?

Figure 4. Typical Questions Representing Range of Service Provided by Human Relations Area File, Inc.

periods of time (depending upon length of stimulus), and "yes-no" decisions as to whether to acquire or not were recorded by each student.

Results of the Play

A complete tabulation of the results is given in Table 7.

TABLE 7. RESULTS OF TRIALS ON ACQUISITIONS

Document identification	Order of presentation of stimuli	Level of processing of source document	Student evaluation		
			Yes	No	Can't tell
A	16	Title	12	18	4
	19	Abstract	14	12	8
	1	First paragraph	21	11	2
	6	Last paragraph	18	11	5
	10	Extract	25	4	5
	4	Extract (map)	15	18	1
	13	Extract (table)	11	14	9
	18	Extract (picture)	11	16	7
B	5	Title	18	14	2
	12	First paragraph	14	12	8
	2	Last paragraph	25	5	4
	20	Extract	21	8	5
	17	Extract (picture)	9	19	6
	9	Extract (picture caption)	20	8	6
C	11	Title	14	18	2
	7	Abstract	17	9	8
	8	Last paragraph	25	5	4
	3	Extract	29	4	1
	14	Extract (preface)	10	16	8
	15	Extract (table)	13	16	5
	21	Extract (map)	15	18	1

It will be noted that, almost regardless of the stimulus used, responses were widely scattered. The only pattern that might be seen is that relating to first paragraphs, last paragraphs, and carefully chosen extracts, which led to more agreement on acquisition decisions than any of the other stimuli.

Further experiments will be conducted later to correlate results with decisions made by subject specialists who would make their decisions based on exposure to the entire source material.

TRIALS WITH MOTIVATED PLAYERS

An attempt was next made to introduce into the game development the element of player motivation. The local Veterans Administration Hospital identified two physicians engaged in research who had current problems which they believed would require literature searches. The questions were identified, literature searches conducted at a local library, and game materials prepared. The play was then conducted with each individual. A report on one of the plays is presented in the following:

The Question

After verbal discussion, the problem facing one of the players (physician) was recorded and checked by him to ascertain accuracy of expres-

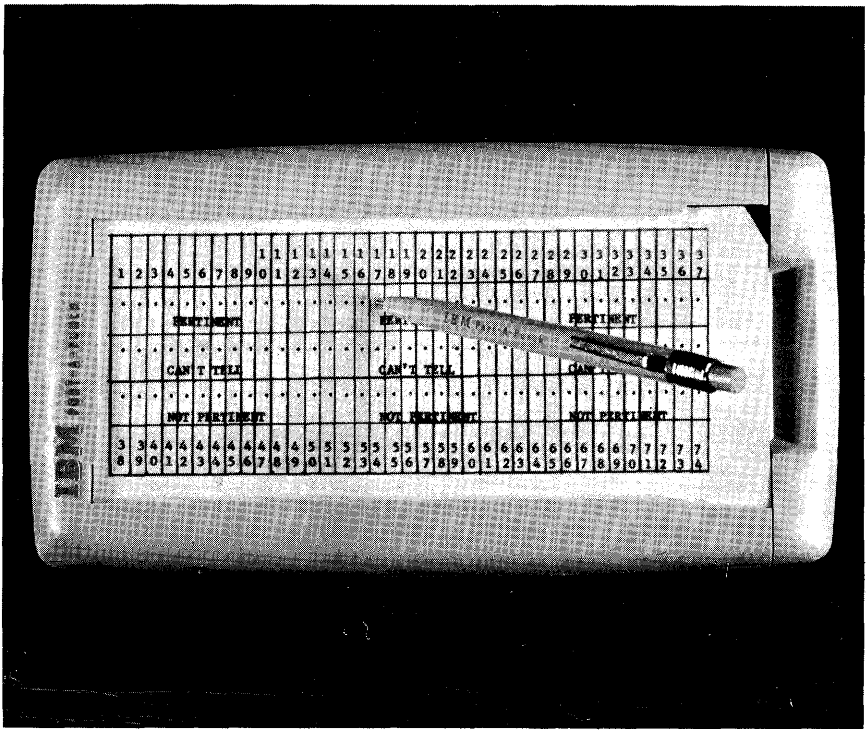


Figure 5. The IBM Port-A-Punch—is used by players of the IR Game to record their responses into special data processing cards which have been perforated for ease of answer recording. The players indicate their selections by pushing the Port-A-Punch stylus through the appropriate hole (Pertinent, Not Pertinent, Can't Tell) in the Port-A-Punch template which corresponds to the document fragment under consideration. This action automatically punches out the appropriate hole in the IBM Port-A-Punch card which is contained immediately behind the template.

sion. The resulting question was:

I would like to have the available information on the clinical deficiency of vitamin E in humans and in other mammalian animals as it relates to pancreatic insufficiency resulting in muscular dystrophy.

Mechanics of the Play

The play was conducted somewhat as before, with four differences, however:

(a) Two sets of stimuli were used instead of one; the first stimulus consisted of abstracts and extracts as before; the second stimulus consisted of full source materials (journal articles) which were predicted as being relevant by the player when responding to the first stimulus.
(b) The first set of stimuli was presented in looseleaf booklet form considered more suitable for review by a single subject.
(c) No limit was imposed on time to be spent with each stimulus—the player was asked to proceed at his own best speed.
(d) Responses to stimuli were recorded by the player using a port-a-punch device (Fig. 5).

Examples of the texts of the first set of stimuli presented to the player are given in Fig. 6.

Stimulus 1 (First paragraph of document):
There has been a recent surge of interest, which is reflected by a growing literature, in diseases of muscle. It is impossible to review the major developments by giving appropriate citations of the literature without a bibliography of major proportions. An attempt has therefore been made to present the present status of thought in this field in general terms.
Stimulus 2 (Extract):
Apart from the reduction of serum tocopherol concentration and resultant increased susceptibility of erythrocytes to peroxide hemolysis, no ill effects are known to result from reduction of dietary tocopherol content in normal infants.
Stimulus 3 (Title):
Biochemical Abnormalities of Primary Diseases of Muscle
—Marvin Smoller, M.D., Chicago, Illinois.

Figure 6. Examples of First Set of Stimuli Presented to VA Physician.

Results of the Play

A complete tabulation of the results of the play is given in Table 8. In Table 9 are given the number of agreements (and disagreements) on relevance of source materials, based on first (abstracts and extracts) and second (full source materials) sets of stimuli.

TABLE 8. RESULTS OF PLAY OF IR GAME WITH VA PHYSICIAN

Source document number	Responses to stimuli (P—Pertinent; N—Nonpertinent; C—Can't tell): First stimulus (fragment)	Responses to stimuli: Second stimulus (full source document)	Level of processing of first stimulus	Sequence of presentation of first stimulus
1	P	N	Extract	10
	P		First paragraph	25
	P		Extract	53
	N		Last paragraph	56
	P		Title	60
2	P	N	Title	9
	P		Last paragraph	12
	P		First paragraph	19
	P		Abstract	27
	P		Extract	34
	P		Extract	62
3	P	P	Bibliography	14
	P		Title	33
	P		Last paragraph	57
	P		Extract	66
	P		First paragraph	68
4	P	P	Title	28
	P		Last paragraph	35
	P		First paragraph	40
5	P	P	Extract	18
	P		Title	20
	P		Last paragraph	21
	P		Graph	24
	P		Abstract	26
	P		First paragraph	47
6	P	P	First paragraph	1
	P		Title	32

(Continued)

TABLE 8. RESULTS OF PLAY OF IR GAME WITH VA PHYSICIAN (Continued)

Source document number	Responses to stimuli (P—Pertinent; N—Nonpertinent; C—Can't tell)		Level of processing of first stimulus	Sequence of presentation of first stimulus
	First stimulus (fragment)	Second stimulus (full source document)		
	P		Last paragraph	39
	P		Extract	45
	P		Extract	48
7	P	P	Last paragraph	5
	P		Title	43
	P		Extract	55
	P		First paragraph	61
8	P	P	Title	3
	P		Last paragraph	17
	P		First paragraph	30
9	P	P	Extract	2
	P		Abstract	31
	P		Extract	41
	C		First paragraph	49
	N		Last paragraph	65
	C		Title	71
10	P	P	Abstract	37
	P		Title	38
	P		Extract	44
	P		First paragraph	54
11	P	P	Last paragraph	8
	P		Abstract	15
	P		Title	51
	N		List	58
	P		First paragraph	70
	N		Extract	74
12	P	P	First paragraph	22
	P		Last paragraph	46
	P		Title	50
	P		Abstract	69
	P		Extract	72
13	P	N	First paragraph	7
	P		Abstract	13
	P		Title	16

(Continued)

TABLE 8. RESULTS OF PLAY OF IR GAME WITH VA PHYSICIAN (Continued)

Source document number	Responses to stimuli (P—Pertinent; N—Nonpertinent; C—Can't tell): First stimulus (fragment)	Second stimulus (full source document)	Level of processing of first stimulus	Sequence of presentation of first stimulus
	P		Extract	23
	P		Extract	36
	P		Last paragraph	52
14	P	P	Extract	4
	P		Title	6
	P		Abstract	42
	P		Abstract	64
	P		Last paragraph	67
15	P	P	First paragraph	11
	P		Extract	29
	P		Abstract	59
	C		Title	63
	C		Last paragraph	73

TABLE 9. AGREEMENTS (AND DISAGREEMENTS) ON RELEVANCE BETWEEN STIMULUS SETS

Level of processing on which predictions are based	Agreements	Disagreements
Title	12	3
First paragraph	10	4
Last paragraph	11	3
Extract	15	6
Abstract	8	2
First and/or last paragraphs	12	3

Discussion of Results of Play

A number of impressions were obtained from the play of the game, which attempted to more realistically simulate an IR situation where the player is sufficiently well motivated to receive useful information so that the play of the game may seem like a positive step in the direction of satisfying his needs.

Since this was the first play attempted with a "live" user, no attempt was made to contaminate the stimuli with materials which were tacitly irrelevant to the question posed. Accordingly, all source materials considered "nonpertinent" by the player, were selected as being fully pertinent by the referee.

If the criteria for relevance posed earlier in this paper are to be continued, then those source materials, and only those, judged to be pertinent to the question by the user (player) are indeed relevant. Accordingly, we may have some initial, possibly valid information, regarding the relevance-predictive value of various levels of processing. As noted from Table 9, the percentage ability of the various levels of processing to predict relevance ratings by the user was:

Title	80 %
First paragraph	78 %
Last paragraph	78 %
First and/or last paragraph	80 %
Extract	71 %
Abstract	80 %

The significance of these results, if results of tests with valid samplings of users bears them out, is that first and/or last paragraphs of documents (which can be selected clerically) are no worse predictors of the relevance of the full source documents than the other levels of processing (some of which require the use of talent with suitable subject background).

The results of this play are still too sparse to permit even first attempts at deriving response patterns, especially due to the lack of contamination of induced nonpertinence in the stimuli. However, one interesting pattern emerged which seems worth discussing.

As seen from Table 10, the responses to the first 48 stimuli were all "Pertinent," before more discrimination in decisions became evident. In contemplating the reasons for this unusual skew in responses, it was considered that this pattern was analogous to that exhibited by any individual seriously seeking information; that is, those stimuli seen first are viewed more hopefully with regard to relevance; more discriminatory patterns emerge as the user gains confidence that some really relevant information will be provided. Until this confidence is attained in viewing the products of an IR system, the threshold of relevance would tend to be lower than might be the case later.

If the data of Table 9 are adjusted to include only responses to stimuli given after apparent confidence has been achieved by the user (stimuli No. 49 to end), then Table 11 results. These data would lead to the fol-

TABLE 10. RELEVANCE RATINGS AS A FUNCTION OF SEQUENCE OF PRESENTATION OF FIRST SET OF STIMULI

Stimulus Number	Player Rating
1–48	Pertinent
49	Can't tell
50–55	Pertinent
56	Nonpertinent
57	Pertinent
58	Nonpertinent
59–62	Pertinent
63	Can't tell
64	Pertinent
65	Nonpertinent
66–70	Pertinent
71	Can't tell
72	Pertinent
73	Can't tell
74	Nonpertinent

TABLE 11. AGREEMENTS (AND DISAGREEMENTS) ON RELEVANCE BETWEEN STIMULUS SETS (Stimuli 49–74)

Level of processing on which predictions are based	Agreements	Disagreements
Title	4	2
First paragraph	5	1
Last paragraph	4	3
Extract	3	4
Abstract	2	0
First and/or last paragraphs	7	3

lowing values for ability of the various levels of processing to predict relevance ratings by the user:

Title	67 %
First paragraph	83 %
Last paragraph	57 %
First and/or last paragraph	70* %
Extract	43 %
Abstract	100 %

*This rating would jump to 88 percent if two disagreements are neglected (for one, the first paragraph was not in the 49–74 stimulus sample; for the others, all levels failed to predict relevance, but only this level fell into the 49–74 stimulus sample).

This approach toward eliminating a first set of stimulus-responses as contaminated, will be investigated in later experiments. Of course, the reasons for response pattern changes during the play of the game may be caused by a learning experience relating to the contents of the stimuli rather than to a change in confidence in the responses.

EXPERIMENTAL PROGRAMS PLANNED

INTRODUCTION

From experience gained during the trials described in the previous section of this paper, a series of experimental programs relating to the heuristic information retrieval game are being designed. These will be discussed below under the following headings:

1. General play of game at Veterans Administration Hospital (players: physicians).
2. Special plays to determine relevance patterns, when level of processing is constant.
   (a) Patrons of university medical library as players.
   (b) Patrons of public library as players.
   (c) Patrons of special technical library as players.
   (d) Clients of NASA specialized information center as patrons.
3. Special plays to determine effect of learning on relevance.
   (a) Information sciences students as players.
   (b) Medical students as players.
4. Relationship between association test results and relevance of source materials.

Each of these programs is being pursued in order to develop a validated series of procedures which may be employed in various gaming situations relating to the information storage and retrieval field.

EXPERIMENTAL PROCEDURES

General Play with Physicians

The Director of the Veterans Administration Hospital (Pittsburgh) has agreed to address a memorandum to professional staff members encouraging them to participate in the experimental program with the Knowledge Availability Systems (KAS) Center of the University of Pittsburgh.

This memorandum will suggest that interested staff participate in discussions with KAS Center staff whenever they wish to obtain information from the literature, from clinical records, or from other sources, which relate to problems or questions in any area of the health sciences.

When contacted by a VA staff member, one of the KAS Center staff will interview the subject in order to obtain a statement of the problem or question. The subject will be considered suitable for involvement in the play of the IR game when the following conditions are met:

1. Response to the question is required no less than three days from the time that the statement of the problem is negotiated.
2. The subject is able to spend approximately two hours reviewing materials selected by the KAS Center.
3. The subject is willing to participate in an interview and to complete a questionnaire relating to:
   (a) Professional background.
   (b) Reasons for need for information relating to the question.
   (c) Evaluation of relevance of materials provided.

When agreement on the above procedure is reached with a subject, a search of appropriate resources in the Pittsburgh area will be conducted, leading to the selection of 5–25 source materials relating, in the opinion of KAS Center staff, directly, peripherally, or tenuously to the question statement.

Source materials selected will be processed in preparation for the play of the game, as follows:

1. Abstracts and extracts (title, first paragraph, last paragraph) of each source material will be prepared and placed on separate sheets, randomly arranged, and placed in a looseleaf notebook.
2. Source materials will be photocopied and rated for relevance to the question by a KAS staff member (referee).
3. Two relevance rating forms will be prepared, one for evaluation of stimuli (abstracts and extracts), the second for evaluation of source materials.

The subject will be asked to review the stimuli and to complete the evaluation form, with the understanding that he will immediately review and evaluate the source materials identified as probably relevant.

As a control, every other subject will be asked to review the entire set of source materials, regardless of relevance ratings.

Special Plays

Medical Library Patrons. Patrons of the Falk Medical Library who approach the reference desk, either in person or by telephone, will be screened for suitability as subjects for the play of the IR game. The criteria for selection of subjects will be as follows:

1. Response to question required in no less than three hours.
2. Willingness to spend approximately one hour reviewing materials selected.
3. Willingness to participate in an interview and to complete a questionnaire, as in the section above.

The play of the game and collection of data will then proceed substantially the same as for the VA physicians above.

In addition, in order to determine the extent to which questions used in the play represent a valid sample of all of the questions submitted to the reference desk, reference staff will be asked to collect the following information relating to patrons:

Ten half days during the trimester will be selected at random, and all questions submitted will be recorded, together with information relating to background of patron and reason for question.

Public Library Patrons. The same procedure as discussed above will be used at the reference desk of the Science-Technology Division of the Carnegie Library of Pittsburgh.

Special Library Patrons. A special library of an industrial organization in the Pittsburgh area will be selected for play of the game as discussed above.

Specialized Information Center Patrons. The KAS Center operates a regional facility for spinoff of technical information under contract to the National Aeronautics and Space Administration. At present, eleven companies participate in the program. Approximately 400 questions have been submitted and are searched monthly, with abstracts of appropriate documents provided.

A sample of these questions will be taken, and the game will be played with this group as discussed above.

Effect of Learning on Relevance Patterns

Subjects. In order to investigate the effect of learning on relevance patterns, an experiment is planned to derive data from the body of medical students at the University of Pittsburgh. However, in order to debug procedures for the investigation, an experimental group of students in the information sciences curriculum will be chosen. These will be masters and doctoral candidates taking courses in "Mechanized Information Retrieval" (Instructor: Prof. Allen Kent) and "Computers in Information Retrieval" (Instructor: Prof. Jack Belzer).

Source Materials. A file of 80 documents is being selected from the book, periodical, and report literature. These documents (ranging in size from 1 to 20 pages in length) may be whole documents, or self-contained segments of documents relating to the topics covered in the classes. Each "document" will either have or be provided with, a title and an abstract, so that each will have a reasonably "standard" format. These "documents" will be filmed and the entire "library" of 100 "documents" on film will be replicated in sufficient quantity so that students may have access to them at time of need. Suitable readers will be provided to facilitate review of individual documents.

Stimuli. As in other plays of the IR game, each document will be processed, and a set of stimuli prepared, consisting of the following:

1. Title
2. Abstract
3. First paragraph
4. Last paragraph
5. Extract

The stimuli will be recorded on sheets of paper, one stimulus per sheet, and notebooks containing them will be prepared, with stimuli arranged in different configurations, both random and structured.

Questions. Questions relating to skills which the students will be expected to acquire during the school term are being formulated. These questions will be presented to the students about mid-term, and again at the end of the term, and they will be expected to provide responses which can be rated objectively.

An attempt will be made to motivate students to desire responses to the questions, and to wish to use the file of documents by causing the question responses to have a bearing on the grade the students receive for the course.

Relevance Ratings. Students will be exposed to the stimuli, and asked to rate them for probable relevance to the questions. The documents identified as relevant from responses to stimuli will be examined by the students, who will, in turn, rate the documents as to relevance.

Reduction of Data. Individual relevance responses will be examined in terms of student progress at mid-term and end of term as measured by ratings derived from responses to the questions.

Individual student as well as class patterns will be sought which may throw some light on relevance decisions as a function of learning. As a minimum, information relating to searching strategies of individuals in a controlled situation will be developed.

Association Testing and Relevance

Introduction. A body of information will be derived during the course of this program which will consist of:

1. Actual questions presented to various information agencies.
2. Ratings of documents with regard to relevance to the questions.

All of the documents used in this program will have been derived from existing collections which have been organized by a library or information center in terms of an indexing or classifying system. There will then be an opportunity to operate these systems in retrospect in order to determine their effectiveness in providing relevant materials and withholding non-relevant material.

Hypothesis. It is hypothesized that the effectiveness of operation of IR systems may be improved, or their ability to perform may be evaluated, by finding some procedure which will permit user paradigms to be related to system operator paradigms (as evidenced by reference points made explicit for search purposes).

Association Tests. In order to investigate this hypothesis, apparent key words derived from user questions will be exposed to users in a test situation, involving the use of association tests.

Three association tests will be devised: One will involve free associations, in which the user will be asked to provide, for each key word derived from his question, a word that comes to mind during a set time period. The second and third tests will involve controlled associations, in which the user will be conditioned by instructions to provide controlled associations as follows:

1. Synonyms to key words derived from questions.
2. Generic terms relating to the key words derived from questions.

Reduction of Data. The responses to the association tests will be compared with the reference points (and cross-references) made explicit by IR systems operators for search purposes in order to discern the level of correlation between them.

The ability of a system to produce relevant documents in terms of correlation level will be investigated.

CONCLUSIONS

No conclusions will be presented in this paper, since the discussions involve mainly the experimental design for the initial stages of the program.

It is hypothesized that as the experiments progress, procedures which may be useful for evaluating and predicting relevance may emerge.

In any case, it is hoped that some quantitative information relating to user paradigms may be developed which may throw light on the nature of the information retrieval field—from the user's point of view.

ACKNOWLEDGMENTS

The development of the program described in this paper, as well as the debugging trials, have been the result of effort on the part of a number of members of the KAS Center staff, including John Canter, Irene Kreimer, and Jack Belzer.

Index