AFOSR 64-1897

Information Technology and The Information Sciences

"WITH FORKS AND HOPE"

HAROLD WOOSTER
Director, Information Sciences

OFFICE OF AEROSPACE RESEARCH
Air Force Office of Scientific Research
United States Air Force

AFOSR 64-1897

INFORMATION TECHNOLOGY AND THE INFORMATION SCIENCES

"WITH FORKS AND HOPE"

Harold Wooster
Director, Information Sciences
Air Force Office of Scientific Research
Office of Aerospace Research
United States Air Force
Washington, D. C. 20333

(Talk given at the Conference on Electronic Information Handling, Pittsburgh, Pennsylvania on 9 October 1964. To be published in the Proceedings of that Conference.)

WITH FORKS AND HOPE

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, can not be clarified without resort to the ikons. 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 Retarius Mk l(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 pursuing basic research one should do so with both hope and forks.

WITH FORKS AND HOPE

INFORMATION TECHNOLOGY AND THE INFORMATION SCIENCES

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 you knew 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 8 August 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 is/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 doubled 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 re-invention, 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 vs. 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 incentives 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 can not 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 photo copies 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 60 million words of Russian text translated a year. You know damn well that we can buy fair-to-middling human translation at 20 bucks a thousand words, and probably wouldn't be interested in machine translation unless we could get it considerably cheaper -- say 10 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 10% 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 (143, 452-455, 1964) 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 (5 October, 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. One must be closely linked with two, lest we wind up with hand-set letter press Selective Dissemination of Information systems, or nation-wide 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 has yet conceived.

Two and three 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; Key Word 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 your 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 -- J. Chemical Documentation, 3, 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 sub-set 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 zeroth 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 zeroth 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 1 August 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 multi-million 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 <u>droite du seigneur</u> 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 multi-valued 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 brain-like 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 can not 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, 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 socio-economics.

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 to 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 modelling 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 language 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 circa 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 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 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.

For, after all, the motto of my organization, the Air Force Office of Scientific Research, is taken from Ecclesiasteses: *Primum acquirere cognitionem* -- "First, get thee understanding."

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Air Force Office of Scientific Research Office of Aerospace Research Directorate of Information Sciences	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

Informati Technology and the Information Sciences--With Forks and Hope

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Symposium paper--Pittsburgh, Pa., Conf on Electronic Info Handling, 9 Oct 64

5. AUTHOR(S) (Last name, first name, initial)

Wooster, Harold

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
1 Nov 64		3

8a. CONTRACT OR GRANT NO. In-house b. PROJECT NO. 9769 c d.	9a. ORIGINATOR'S REPORT NUMBER(S) AFOSR 64-1897 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. AVAILABILITY/LIMITATION NOTICES

Qualified requestors may obtain copies of this report from DDC. Copies have been furnished to the Office of Technical Services thru DDC for sale to the general public.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Directorate of Information Sciences Air Force Office of Scientific Research

13. ABSTRACT

Electronic information handling is a rapidly developing technology. It is parasitic upon, host to, and symbiotic with many other technologies. Like all other technologies, it is presumably dependent upon a body of fundamental scientific disciplines and knowledge. Advances in information technology can only come in three ways: by specific R&D efforts in information handling *per se*; by exploiting fortuitous advances in ancillary technologies; and, by improvements in fundamental understanding. The emphasis of the paper is on the last of these, with emphasis on epistemology, intelligent automata, pattern recognition from visual to semantic, self-organization, and computer organization and programming.

DD FORM 1 JAN 64 1473

Unclassified

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Epistemology						
Intelligent automata						
Pattern recognition						
Computer programming						
Basic research						
Information retrieval						
Computers						
Research & research equipment						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ______."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ______."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ______."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.