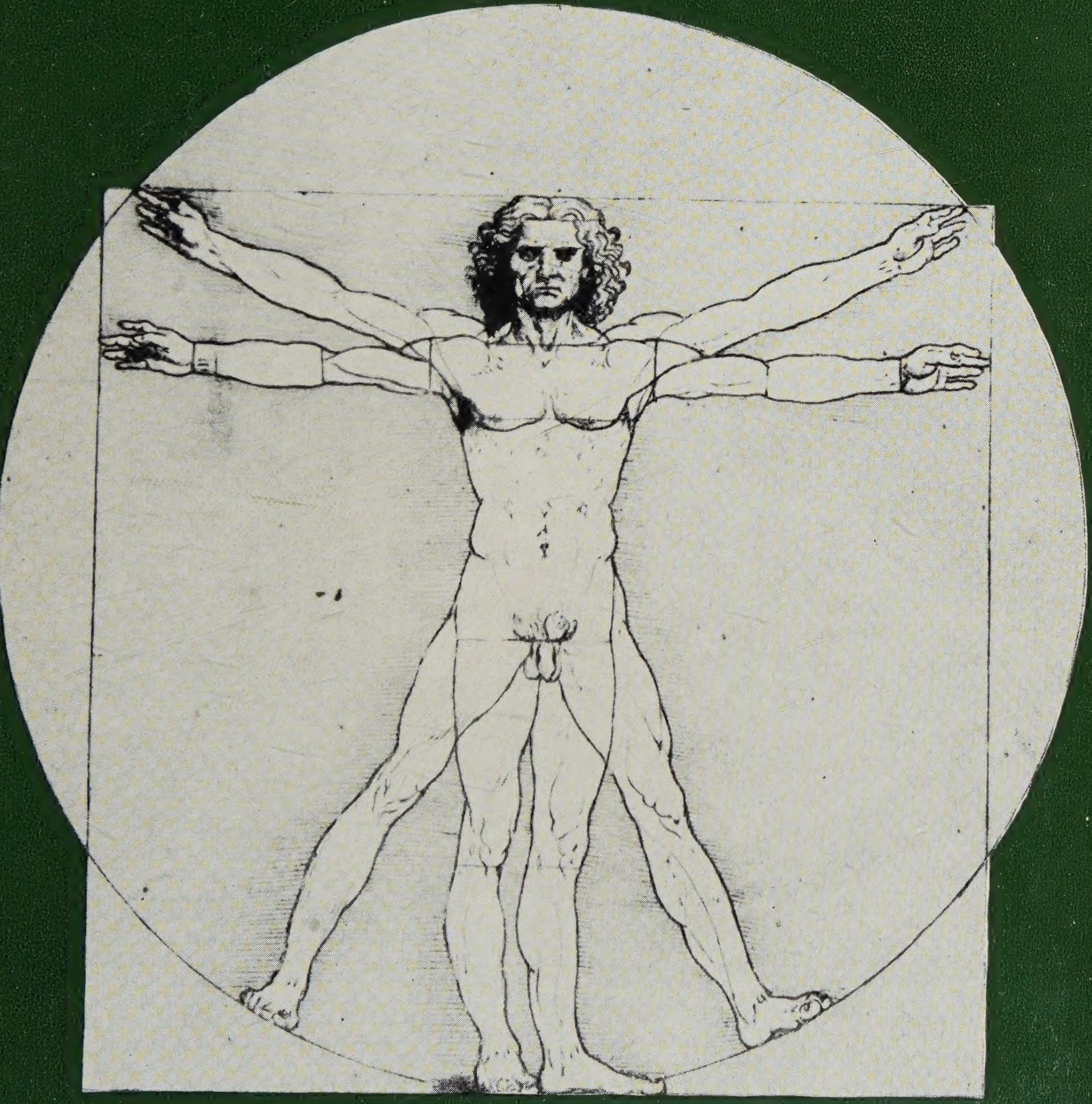
Progress of CYBERNETICS

VOLUME 1

Edited by J. ROSE



GORDON AND BREACH

Progress of CYBERNETICS

Proceedings of the First International Congress of Cybernetics, London, 1969

Edited by J. ROSE

Principal of Blackburn College of Technology and Design, Hon. Secretary International Cybernetics Congress Committee

Volume 1

Main Papers

Section I
The Meaning of Cybernetics (Definition and Philosophy)
Section II
Neuro- and Biocybernetics

Volume 2

Section III
Cybernetics and Industry (Automation)
Section IV
Social and Economic Consequences of Cybernetics (Including Management)
Section V
Cybernetics and Artifacts

Volume 3

Section VI
Cybernetics and Natural Sciences
Section VII
Cybernetics and the Social Sciences

Eight areas of work are covered in this book. The collection of main papers – contributed by renowned experts from five countries – deals with all aspects of cybernetics. This is followed by seven sections, each containing many papers and dealing with different disciplines. There are over 100 papers contributed by authors from many countries. Two indexes complete the work.

This book marks the coming of age of cybernetics and an advance into maturity of this new science that will have a tremendous effect upon society.

Progress of Cybernetics

Volume 1

Progress of Cybernetics

Volume 1

Main Papers
The Meaning of Cybernetics
Neuro- and Biocybernetics

Edited by

J. ROSE

Blackburn College of Technology and Design
Hon. Secretary International Cybernetics Congress Committee

Proceedings of the First International Congress of Cybernetics
London, 1969

GORDON AND BREACH SCIENCE PUBLISHERS
London New York Paris

Copyright © 1970 by

Gordon and Breach, Science Publishers Ltd.
12 Bloomsbury Way
London W.C.1

Editorial office for the United States of America

Gordon and Breach, Science Publishers, Inc.
150 Fifth Avenue
New York, N.Y. 10011

Editorial office for France

Gordon & Breach
7-9 rue Emile Dubois
Paris 14e

Library of Congress catalog card number 73-111944. ISBN 0 677 14310 9. Printed in east Germany.

Preface

An International Congress of Cybernetics was held in London in September, 1969, the date coinciding with the twenty-first formal birthday of the new interdisciplinary science of cybernetics. This event was held under the aegis of an International Committee (I.C.C.C.) composed of eminent academics and cyberneticians from eighteen countries, and was supported by many international bodies concerned with management, labour, cybernetics, the sciences and technologies, including U.N.E.S.C.O., I.L.O., etc. The aims of the congress, which marked a milestone in the history of cybernetics and may well become the basis of a world organization, were as follows:

(1) To establish cybernetics as an interdisciplinary science on solid foundations without the spurious accretions of the last two decades.

(2) To exchange up-to-date information and meet as an international academic community.

(3) To develop more efficient liaison between various scientists on an international scale.

In accordance with the above the congress has decided to explore the possibility of establishing a World Organization of Cybernetics, under the aegis of an international agency, and a Cybernetic Foundation; the latter to finance research, publications, establishment of institutions, etc.*

The proceedings of the congress are grouped in eight parts, viz. the main papers, followed by seven sections dealing with various aspects of cybernetics; authors from eighteen countries are represented. The main section comprises eight papers contributed by the most eminent cyberneticians of our times, the subjects treated covering the whole range of the science.

Section I is concerned with the philosophy and meaning of cybernetics,

---

* *Note added in proof* As a result of a world-wide enquiry, a World Organisation of General Systems and Cybernetics has been established. The Chairman of Council is Professor W. Ross Ashby (U.S.A.), the Vice-Chairman is Professor Stafford Beer (U.K.), and the Director-General is Dr. J. Rose (U.K.).

including its relation to education, religion, general systems theory and information. The second section—the largest in these proceedings—deals with a wide variety of topics relating to neuro- and biocybernetics. Among the subjects discussed are pathological conditions, the brain, vision and hearing, bioholography, language, medical diagnosis, etc.

Topics in Section III show a different aspect of the science, that of its relation to industry, e.g. automation, robots, man–machine systems, system dynamics, and cybernetic control. The economic consequences of cybernetics are considered in Section IV, where special emphasis is placed on management problems, and on economic and simulation models. These are also among the subjects of Section V, where the whole field of artifacts is treated from the point of view of information, perception and pattern recognition. This section also includes two papers on art.

The impact of cybernetics on the natural sciences is discussed in Section VI of the proceedings, including such diverse considerations as evolution, biology, chemistry, town planning, laws of nature, logic and mathematics. The final section (Section VII) examines the relation between cybernetics and the social sciences, emphasis being placed on effects of technology on society, cybernetics of systems, impact on political science, human values, languages and communication, social and economic stability and others.

It is evident that the range of subjects concerned is wide. Some of the papers are truly pioneering in their philosophy and sound in their scientific approach; others examine the fringes of the subject because of limitations on the size of the papers. Indeed, a number of papers had to be greatly abridged in order to cope with the sheer weight of the contributions. Above all, it is intended to demonstrate the scope and maturity of cybernetics, though a few papers bear the stamp of a rather exotic approach. These somewhat fatuous contributions were included in order to bring to the surface certain undesirable accretions. A mature science has to be able to live and cope with those who are trying to jump on the band-wagon and use it as a vehicle for their exuberant claims. In the editor's view the twenty-first anniversary of the formal birth of cybernetics was the proper time to lay the ghost of the futile accretions and thus fulfil one of the main aims of the congress. Indeed, discussions during the congress reinforced the sound structure of cybernetics, while exposing some of the fatuous hypotheses and concepts contributed by pseudo-cyberneticians. A sensible reader of the proceedings will be able to spot the few spurious accretions and identify the sound main body of the interdisciplinary science of cybernetics, which is considered to be that of the 21st century, as brought out by some speakers at the congress and at the Lord Mayor's Banquet at Guildhall, London, the latter held to celebrate the coming of age of cybernetics. Above all, the programme of lectures was supposed to give a broad prospective. In the words of one of the authors (Dr. E.E. Green), every facet of interest, from pure humanism to pure mechanism had its place in the congress with the accompanying integration of concepts and a hierarchical arrangements of functional levels, mechanical, philosophical, physiological, psychological, economic, political, and social. Communications received from various quarters after the congress clearly emphasize this point.

I wish to thank the International Congress Committee, and its chairman Professor Stafford Beer, for their encouragement in the arduous task of organizing the Congress, the Guildhall Banquet and the editing of the Proceedings. My thanks are due to the financial sponsors of the Congress, I.B.M. Corporation, both U.S.A. and U.K.; the Research Centre for Mathematics, Morphology and Psychology (U.S.A.); and Technology Forecasting Institution Inc. (U.S.A.). I am particularly indebted to Mr. J.G. Maisonrouge (President of the I.B.M. World Trade Corporation), Mr. E.R. Nixon (Managing Director, I.B.M., U.K.), Mr. Martin B. Gordon (President, Gordon and Breach Inc.), Mr. A.M. Young and Dr. C. Muses for their help and goodwill. Thanks are also due to my assistants, including Mr. J. Oldcorn, and Miss R.R. Black (Blackburn College of Technology and Design), my wife and sons, Professor T.C. Helvey, Miss Moya G. Head (outside bookings of Imperial College) and her staff. I am most grateful to Lord Penney, O.M., F.R.S., Rector of Imperial College, London for his kindness in offering to the congress the facilities of this great institution and associated halls of residence, and to Lord Jackson of Burnley, F.R.S., for his acceptance of the invitation to open the congress. My thanks are also due to Professor H. Marshall McLuhan for his kindness in gracing the Guildhall Banquet with his presence. I wish to express my gratitude to all the authors and the congress delegates for their cooperation, and to the representatives of the various international agencies for their support, particularly Dr. S. Chamecki of U.N.E.S.C.O. and Mr. R.H. Bergmann of I.L.O. It is hoped that this cooperation and goodwill of all concerned will continue and extend to the World Organization and also to the next international congress to be held in Great Britain, possibly in 1972.

J. Rose
Editor

Committee members of the International Cybernetics Congress

Professor W. Ross Ashby	University of Illinois, Cybernetician (U.S.A.)
Professor Stafford Beer	Development Director, I.P.C., Author and Cybernetician (U.K.)
Professor G. R. Boulanger	President, International Association of Cybernetics, Namur (BELGIUM)
Professor Carlos Chagas	President, Brazilian Academy of Sciences (BRAZIL)
Professor P. B. Fellgett	Professor of Cybernetics, Reading University (U.K.)
Professor A. Ghizzetti	University of Rome (ITALY) Member of Council I.F.I.P.
Acad. Professor V. M. Glushkov	Director Ukr. U.S.S.R., Institute of Cybernetics, Kiev (U.S.S.R.)
Dr. G. K. Goransky	Director, Institute of Engineering Cybernetics B.S.S.R., Academy of Sciences (U.S.S.R.)
Professor A. D. C. Holden	Chairman, International Joint Conference on Artificial Intelligence, University of Washington, Seattle (U.S.A.)
Acad. J. Kozesnik	Vice-President, Czechoslovak Academy of Sciences; Director Institute of Information Theory and Automation (CZECHOSLOVAKIA)

Professor Börge Langefors	Royal Institute of Technology; Stockholm (SWEDEN), Member of Council, I.F.I.P.
Professor Leon Lortie	President, The Royal Society of Canada (CANADA)
Professor Manea Manescu	Director, Rumanian Academy of Economic Computing and Economic Cybernetics (RUMANIA)
Professor Kassa Wolde Mariam	President, Haile Sellassi University (ETHIOPIA)
Professor Aldo Masturzo	President, International Society of Cybernetic Medicine (ITALY)
Mr. P. M. Murton	Chairman, Australian Computer Society (AUSTRALIA)
Dr. C. Muses	Director, Research Centre for Mathematics, Morphology and Psychology (U.S.A. and CANADA)
Professor Dr. Ing. H. Naplatanoff	Director, Institute of Engineering Cybernetics; Chairman, National Council of Automatic Control (BULGARIA)
Professor Miron Nicolescu	President, Rumanian Academy of Sciences (RUMANIA)
Professor Josef Panasewicz	Head, Polish National Centre S.I.M.C., Head, Biocybernetic Committee (POLAND)
Dr. Talcott Parsons	Harvard University; President, American Academy of Arts and Sciences (U.S.A.)
Professor Gordon Pask	Brunel University; Director, Systems Research Ltd. (U.K.)
Professor J. G. Santesmases	University of Madrid (SPAIN), Faculty of Sciences, Member of Council I.F.I.P.
Mr. Kiichiro Satoh	President, International Council for Scientific Management (JAPAN)

Dr. J.P. Schadé	Director, Netherlands Central Institute for Brain Research (THE NETHERLANDS)
Professor J.F. Schouten	Delegate, Royal Netherlands Academy of Sciences (THE NETHERLANDS)
Professor Benjamino Segre	President, National Academy of Italy (Lincean) (ITALY)
Lord C.P. Snow	U.K.
Acad. I. Tarjan	Delegate of the Hungarian Academy of Sciences (HUNGARY)
Dr. W. Grey Walter	Director, Burden Neurological Institute, Bristol (U.K.)
Hon. Secretary	Dr. J. Rose, Blackburn (U.K.)
Provisional Secretariat	c/o College of Technology and Design, Blackburn BB2 1 LH, Lancs, England

Contents

On the first page of each section is given a full contents list of papers for that section.

Volume 3

Main papers

CHAPTER M–1

Artificial intelligence (General introduction)

GEORGES R. BOULANGER

International Association for Cybernetics, Namur, Belgium

Summary

Intelligence has always been considered as an attribute solely present in man, though it is sometimes admitted that it also exists in higher animals (ape, dog, etc.). However, a machine is not and could not be intelligent. Cybernetics throws a new light on this problem. Cybernetic machines behave in a more advanced fashion than conventional machines; they simulate life, thus being unique. By means of a remarkable synthesis cybernetics progressively demolishes the partition between the living and non-living matter and enhances the scope of action of the machine. Its ultimate objective is to construct intelligent machines and explain the mechanism of life. Cybernetic machines show at present a certain degree of intelligence. It is fairly certain that in the future these machines may even overtake human intelligence. An artificial brain built by man on the basis of the human brain could even discard in the future the latter, and new methods of thought could thus be created. These could open up new intellectual fields and possibly supply solutions to important problems of great interest to mankind.

Professor Boulanger began his lecture by conveying greetings from the International Association of Cybernetics of Namur in Belgium to the assembled delegates of the Congress in London organised so successfully by the I.C.C.C. The speaker referred to cyberneticians as pioneers, who are still in the vanguard of science at present. In his view, cyberneticians are trying to span the bridge between living and inanimate matter, including in the same class animals and machines (one should say "cybernetic machines"), as far as they may be considered as goal-seeking systems.

Norbert Wiener opened the way some twenty years ago and originated the word "cybernetics" to cover the field. It might be of historical interest to know that when the word came into France, the people there discovered that Ampère had used it, in its French version, more than one century before—in a very restricted sense indeed.

The ultimate goal of cybernetics as we see it today, is to build intelligent machines and explain the "mystery of life". This goal includes many others, because once the mechanism of life is explained and the technology of brain-like systems is made available, this will have far reaching effects in engineering, economic, social, medical, psychological and many other sciences.

This, then, is the justification for choosing "Artificial intelligence" as the subject of a general introduction to this Congress.

One cannot resist the temptation in introducing this lecture to read here a few lines which were written by Marvin Minsky[1] about ten years ago and are still remarkably significant and of value today.

Minsky's text is as follows: "A visitor to our planet might be puzzled about the role of computers in our technology. On the one hand he would read and hear all about wonderful 'mechanical brains' baffling their creators with prodigious intellectual performances. And he (or it) would be warned that these machines must be restrained, lest they overwhelm us by might, persuasion, or even by the revelation of truths too terrible to be borne. On the other hand, our visitor would find the machines being denounced, on all sides, for their slavish obedience, unimaginative literal interpretations or initiative; in short, for their inhuman dullness."

"Our visitor might remain puzzled if he sets out to find, and judge for himself, these monsters. For he would find only a few machines (mostly 'general-purpose' computers, programmed for the moment to behave according to some specification) doing things that might claim any real intellectual status. Some would be proving mathematical theorems of rather distinguished character. A few machines might be playing certain games, occasionally defeating their designers. Some might be distinguishing between handprinted letters. Is this enough to justify so much interest, let alone deep concern?"

"The speaker believes that it is; that we are on the threshold of an era that will be strongly influenced, and quite possibly dominated by intelligent problem-solving machines."

This 1961 statement is still true today. It cannot be denied that more work is devoted to the field of problem-solving processes nowadays than in 1961, but the field is paved with confusion, contradiction, overenthusiasm, and deception. No general trend can be traced in the mass of theoretical and technological work on brain-like solving systems which is carried on all over the world, and we have not yet succeeded, in Namur, to develop a consistent philosophy, although major contributions to the subject must be credited to such scientists as Gordon Pask (Great Britain), Jacques Sauvan (France), and others.

If we go back to the earliest times of man's life on this planet, we must take for granted that the first human beings made a strong differentiation between man himself, and the other living creatures, vegetables and animals. They were certainly well aware of a kind of superiority they possessed over the other animals and—although they probably had no name for it—they felt that they had some intelligence, some cleverness that was of another order than that of the animals.

And, as a matter of fact, they made weapons to kill living creatures, they elaborated strategies to catch them alive, and they even domesticated some of them. Later they mastered fire, they constructed shelters, they invented the wheel, they built machines and, steadily increasing their power, engaged in the extraordinary adventure of the twentieth century which sees the species *homo sapiens* exploring the universe and submitting the natural elements to his own volition.

Nowadays, scientists make clear, with modern language, what man has always realized in the past but never tried to explain scientifically (it used to be considered as a philosophical problem) and they use, to cover the field, the famous word, "intelligence".

It is a famous word indeed, famous for the confusion it throws in the field. And so we must first clarify it.

"Man is intelligent, machine is not, even if in some special cases it doesn't seem to be so." This is the most commonly heard statement today, not only at the laymen level, but in most intellectual spheres, scientific or not. As far as animals are concerned, the ideas may diverge, depending on the degree of organization of the species. Some people might claim intelligence for a dog or an ape, but nobody will do so for an oyster or a butterfly.

Let's consider man and machines. When the situation is clarified there, the case of the animals will be solved automatically.

It is a surprising fact that almost everyone is prepared to attribute intelligence to man, and to refuse it to machines, but that nobody says what he means by intelligence and, moreover, that nobody knows what intelligence really is.

This "prise de position" is evidently a convenient one. We are proud to hear about the supremacy of man in the intellectual field and therefore we do not ask for proofs, ignoring voluntarily that this is today impossible to prove.

Let us, on the other hand, introduce the thesis that a machine can be built which may be intelligent. At once everyone strongly denies the fact, and requests proof without questioning himself on the subject, and the one who alleges such a thing is rapidly accused of being an impostor, a charlatan.

So, let us go a little closer to our subject.

If we want to discuss intelligence in man and machines, it seems reasonable to say that we first have to agree on a given definition of intelligence. And here we have to stop. There is no such definition.

One may be astonished when hearing such an affirmation. As a matter of fact, tests exist which permit us to measure intelligence. Moreover, there are units to extrapolate the results of these tests. We all know about the famous I.Q. used by pedagogists to evaluate and compare the intelligence of students, or by sociologists to study the intellectual structure of populations. This is all true. But we must be well aware that all those tests involve only certain aspects of what is called intelligence. No attempt has ever succeeded in approaching the notion of intelligence *per se*, the notion of general intelligence.

Therefore, our problem is poorly posed from the beginning. If we want to speak about intelligence of machinery in terms of human intelligence, we should first investigate and determine exactly what human intelligence really is.

Has this lecture then, reached an impasse with no further means to be carried on? On the contrary. And this for three reasons, namely because (1) artificial intelligence may be developed without reference to human intelligence, (2) many of us do have our own provisory definition of intelligence, and (3) human intelligence may well, one day, appear like a particular case of machine intelligence.

To make this point clear, let us go back into the world of energy and consider machines which have been built by man to perform human tasks. In most cases, these machines do their jobs without imitating life. Automobiles use wheels and not legs. Airplanes do not move their wings to fly. And so on.

Why, then, should a mechanical brain imitate the living brain? In the same line of thinking, one could reverse the situation and argue that if we want to succeed in building an intelligent machine, we should not try to imitate the functioning of the living brain, but should try to discover the basic principles of intelligence and build a machine along those principles.

Bearing this in mind, we must, however, turn to the human brain to get an initial approximation of what we want to do. But—I insist—the human brain, with all its weaknesses and shortcomings, should never be considered as a kind of ideal model for an intelligent machine. Only the most common approach to intelligent machines goes through the human brain which will surely be considered in the future as a starting point, not as an ultimate goal.

After having stated this firmly, it seems to me impossible to go further on this subject without quoting at length W. Ross Ashby who, in my opinion, is the man who has penetrated with the deepest insight into the arcanum of the human brain, and whose contribution of introducing rationality into a field "infested with myths"—these are the proper Ashby words, as you will hear in a moment—has proven to be invaluable.

In the FIER Distinguished Lecture—FIER stands for Foundation for Instrumentation Education and Research—held in New York in 1961, W. Ross Ashby says, about the attempt or the hope to develop machine-behaviors that are really brain-like: "To get the matter clear, it is essential that we first understand what the human brain really is and what it actually does." And when Ashby says "what it actually does", what he has in mind is much more the negative side of the question, i.e. what it actually does *not* do.

Ashby is particularly severe on the scientists of the past. "The fact is," he says, "the subject, ever since the days of the Greeks, has been infested with myths and superstitions."

One example of such myths, according to Ashby, is the common idea that the brain can predict the future. This idea, says Ashby, certainly has a superficial plausibility, but it is fundamentally wrong. Norbert Wiener put the truth succinctly when he said that to predict the future is to carry out an operation on the past. If the past shows repetitions, and if the future sustains the regularity, then the brain will score a hit when the future arrives. But the process is based entirely on past events and on the degree to which the real world has consistent regularities. Let the world produce something really new, and the brain, human or other, is helpless.

Such a point of view is, quite evidently, of primordial importance for the engineer who wants to build a mechanical brain, because it lessens what we must expect from artificial intelligence (at least at its first stage of development), if we want to stick to a realistic position.

Another myth—I quote Ashby again—stands exposed if we ask: "How does the brain generalize?" For experimentation shows that the answer is, it does not generalize. And Ashby presents, as a proof of that, the experiments of Papert, with his prismatic spectacles interchanging left and right, so giving him a reversed image of the world. Though Dr. Papert was by no means deficient in intelligence, at no stage of re-education has a generalization of "reversal" occurred.

The third example—this is Ashby's favorite one—is what I would call the "genius" example. Ashby enjoys claiming that it is a myth to believe certain people have the special property of being a "genius"—of having some quality that makes them, in some absolute sense, more effective than other people even if all other factors are equalized.

The myth, Ashby explains, is based on bad statistics. Many workers tackle a problem, using all sorts of preconceived ideas of how to solve it. One is successful, and then we pick him out and assume that he must have something that is good for all problems and for all time. What remains after this fallacy has been corrected, however, is that the persistently productive so-called geniuses (Mozart for instance) were mainly people who were obsessed by their subject, and who devoted to it a fraction of the day's twenty-four hours, a fraction far greater than the amount given by the average man.

Now you can probably see more clearly how the game is. The trend is no longer to place the human brain at the summit of a hierarchy and to try to make a machine capable of climbing there, but rather to discover the features that would enable us to claim that the brain is nothing other than a cybernetic machine, thus bound to see its performances not only imitated, but most certainly surpassed by the ones of the other machines which will one day be built by man himself.

The approach to artificial intelligence poses many problems which are of three fundamentally different natures. First there are the theoretical (scientific) problems involved in the projects, then the technical problems which arise when we try to build the actual machines, and finally the philosophical problems which appear as a consequence of our incomplete knowledge of the phenomenon called "intelligence" and especially of the lack of a general theory of information.

Let us first have a look at the theoretical aspects of artificial intelligence.

Proceeding along the line which says that the first step toward artificial intelligence should normally be the imitation (with a different technology) of the functioning of the brain itself, we shall turn to Ashby again. He told us, a few minutes ago, what the brain is not; let us ask him now what the brain is.

Ashby points out that every skill or power that the human brain possesses has been developed, either by natural selection acting on the species, or by the processes of experience in childhood and education as it affects the individual. After a billion years or so of shaping by the environment, the living brain is now suited to that environment with quite remarkable precision, but only to that particular environment. Only there can it be efficient.

And here comes the question: efficient for what, efficient to solve what kind of problems, efficient towards what kind of goals?

Here is Ashby's answer. The living brain has been shaped for survival. And survival means the persistent achieving of certain goals—food must be won, injury avoided, thirst satisfied, heat or cold tempered, and so on. "Intelligence" is shown in retrospect of these goals. Thus a first principle of any "brain-like" or intelligent system, living or mechanical, is that a system can be brain-like or intelligent only in relation to a defined goal. In other words, the construction of a brain-like system cannot be begun until the question has been answered: what do you want?

A second principle is that control is the touchstone of any brain-like behavior. This, says Ashby, is the lesson of twenty years of cybernetics. If one adheres to this view, one remains in the world of the scientific, practical, and modern. If one forgets it, one may drift back into the mythical, the superstitious, and the downright impossible.

As far as the theory of intelligent machines is concerned, I think we should stick to Ashby's views. This means that, as the heart has previously been found to be simply a pump (notwithstanding all its mythical attributes), so has the brain been found to be a regulator and therefore a mere processor of information.

This simplifies the problem as far as the basic principles are concerned. But the lack of a satisfactory general theory of information, to which I referred a few moments ago, exerts here a profound negative impact. This theory is urgently needed. I hope that the present Congress will unveil new and significant contributions to this topic.

Let us go, at present, to the technical problems involved with building intelligent machines.

On the present state of the art we will get first-hand information during this Congress. It would therefore be nonsense for me to present to you a list of achievements which would show itself obsolete a few hours from now.

It is, however, impossible for me not to mention here at least the brilliant work that is carried on in France by Dr. Jacques Sauvan, developing what he calls an active memory, and the provocative contribution to cybernetics of my friend and colleague, Prof. Gordon Pask.

It is, for me, a very great pleasure, while in Great Britain, to pay tribute to Gordon Pask, not only for his invaluable contribution to cybernetics, in the theoretical field as well as in actual achievements, but also for the firm support he has always given to the International Association for Cybernetics.

And now, on to philosophy. First of all, we find there the problem of "creation" of information.

Strangely enough the fact exists that many people, and among them a good many scientists, claim that a machine cannot and will never be able to "create" information, but that man—on the contrary—can do so.

One could quote hundreds of statements along this line. I shall give you only one, because I do not want to be tiresome. Here it is.

Roger De Staercke, the President of the Fédération des Industries Belges (Federation of Belgian Industries) wrote, fairly recently, in an Editorial published in BP Review (Antwerp, Belgium, 1967): "L'intelligence de l'homme garde le pas sur celle de la machine parceque, et avant tout, la pensée humaine est pensée créatrice." This means: "The intelligence of man keeps ahead of that of machines because, and above all, human thinking is creative."

To sustain such attitude no proofs are given. No proofs are asked for, because people generally like to hear such things. Moreover, no proofs can be given in the present state of science, and this is what makes this a philosophical problem.

Yet this problem is of fundamental importance, because it may emerge, from the tissue of the controversy, that it would be theoretically an absolute impossibility ever to build an intelligent machine.

No proofs can be given, I just said. But there is strong evidence favoring the following thesis: neither man nor machine are able to create information. Both can only process (i.e. transform, combine, and so on) information, with the consequence that there is no theoretical obstacle to the approach of artificial intelligence.

Let us see how one can arrive at such a conclusion.

There is no satisfactory general definition of information at the present time. Yet I shall speak about information in its commonly held, but rather vague meaning. And I hope that those who make claims about non-intelligence of machines without any proof or even any definition of intelligence, will allow me to speak for a few minutes with some degree of confusion.

So, let us proceed. The argument will develop in three points which sounds like this.

As a first statement one recalls that, whatever the definition of information might be, there is no transmission of information without the support of energy. If we hear something, it is due to vibration of air (or another sound-conducting material), if we see something it is the energy of light waves that allows us to do so, and so on.

On the other hand, and this is the second statement, it is a well established fact that any machine or machine component (including under this heading all living systems) that we can imagine, is only capable of transforming energy, and cannot create even the smallest portion of it (except by destroying matter, and this case is irrelevant here). This is the law of conservation of energy.

A combination of those two statements leads to the following reasoning. The human brain, like any other system, living or not, can just transform energy (this comes from the second statement). But there is no information without energy (this is the first statement). It seems, therefore, impossible that the brain, which is only able to *transform* energy and not to create it, would be capable of creating information.

This looks sound, yet cannot be proved today because of the lack of a satisfactory theory of information.

Therefore one concludes in the following way.

If what has been said is true—and it seems it absolutely is—then the brain is only a processor of information. This permits us to affirm that the certainty exists that we will someday be able to build intelligent machines, because we know how to build machines which transform information. Data-processing machines are well known, even to the general public now.

But if, on the contrary, and against all probabilities, it would once appear that the human brain creates information, then the intelligent machine would not be in view. And that, because we have not the slightest idea on how to proceed to build a machine that could create information.

Another philosophical problem related to artificial intelligence is connected with the notion of consciousness. It may be formulated as follows. If we succeed one day in building intelligent machines, would consciousness then emerge in such machines?

Clearly, we must see that there is no scientific answer to such a question now. The reason is that we have no experiment available which would enable us to determine if a given system—alive or not—is conscious. Even for man we have no way of proving the existence of consciousness.

This way seem nonsense. As a matter of fact, we probably are all aware of being conscious. As far as I am concerned, I am sure I am so. But it is only by analogy that I am authorized to admit that other people are conscious also.

However the matter may be, the consensus is absolute. We all agree in accepting the fact that every human being is conscious.

Divergencies arise when we go over to the animals, but unanimity forms again when we enter the field of machines. Notwithstanding certain analogies between the behavior of cybernetic machines and that of living systems, nearly all of us are reluctant to attribute to such machines, even the smallest degree of consciousness.

There is no scientific issue here. Only philosophers may conduct endless discussions on that problem, but they stand before the absolute impossibility to prove any of their conclusions.

From a practical point of view this has no importance. If a machine shows all the external characteristics of consciousness, the cybernetician must admit that such a machine *is* conscious, or what is practically the equivalent, that such a machine acts as if it were conscious.

In conclusion, the following point may be considered, which is related to the possibilities and limitations of natural and artificial brains.

The human brain—contrary to popular opinion—has very severe limitations. And this will also be true for any artificial brain that could be constructed.

Ashby has shown that, if a set of disturbances can be met by a set of responses, the smallest variety that can be achieved in the set of outcomes cannot be less than the quotient of the number of rows (corresponding to the disturbances) divided by the number of columns (corresponding to the responses) in the matrix of the outcomes.

This is Ashby's law of requisite variety. In a more direct language it states that the amount of appropriate selection that a control device can achieve is bounded by the amount of information that it has received and processed. And this is evidently applicable to brains.

The discovery of the limitation of brains (including the human one) is for many people a disillusion. It should not be.

"In fact," said Ashby after discovering the law, "the situation is here very similar to what it was a hundred years ago, when the power engineers were inventing new engines almost daily, and every imaginative engineer confidently expected that the machine with perpetual motion would be invented shortly. Then emerged an unpleasant surprise. Evidence began to grow that all machines were fundamentally limited, that energy was conserved, and that every erg of energy that came out first had to be put in. For a time many engineers felt profoundly disappointed. Nevertheless, the acceptance of this limitation was the first step to a new and altogether more realistic grasp of the principles of machinery. In the long run, the engineers who accepted the limitation outdistanced those who clung to the hope of the perpetual motor, precisely because the ideas of the acceptors were more in accordance with actual realities."

"Thus I feel sure," says Ashby in conclusion, "that those who accept the modern idea of the brain, and accept the limitation of it and all brain-like mechanisms, will eventually go far beyond those who want to hold on to the old idea that the living brain has some mysterious 'extra' that can do things in a flash."

The final statement will be this one. The discovery of the limitation of the power of the human brain is offset by the tremendous possibilities which will be opened up by artificial intelligence.

First of all, the limitation of a brain is a consequence of its structure. It is by no means an absolute limit that a brain could never overtake. It is thus possible, once the technology of brain-like systems is mastered, to build more and more sophisticated, artificial brains which may one day surpass the human brain.

Secondly, it was stated that the human brain has been shaped for survival in its terrestrial environment. It is not *a priori* a good thinking instrument—and, in fact, it is not. It is, therefore, not unreasonable to believe—and I do so—that in the far future, artificial thinking systems might be devised, which would enable the machine to think differently than the human brain does. Those thinking machines may well penetrate into intellectual spheres which are not accessible to the human brains. From there, if their conclusions prove to be accessible to our powers of comprehension, they would perhaps bring back to us answers to the cumulative and seemingly insoluble problems of modern science.

References

1. M. Minsky, *Proc. I.R.E. (Inst. Radio Engrs.)*, **49**, No. 1 (1961).

CHAPTER M-2

The meaning of cybernetics in the behavioural sciences (The cybernetics of behaviour and cognition; extending the meaning of "goal")

GORDON PASK

System Research Ltd. and Institute of Cybernetics, Brunel University, U.K.

Summary

The paper discusses the impact of cybernetic ideas upon behavioural and cognitive studies in general but the main thesis is developed in the context of human psychology. An effort is made to trace the influence of cybernetics upon the development of psychological theories, experimental techniques and methods for modelling mental and behavioural activity. Particular emphasis is placed upon the key concept of a "goal directed" system. It is argued that this concept becomes differentiated to yield two specialised forms of system, namely "taciturn systems" and "language oriented systems"; of these, the latter are peculiarly important in connection with studies of man or attempts to control, teach, or otherwise influence human beings. As it stands, the notion of "goal directed" system is unable to adumbrate the phenomena of evolutionary development (as in open ended concept learning) and conscious experience. Problems entailing both types of phenomena are ubiquitous in the human domain and the paper considers several ways in which the connotation of goal directedness can be enlarged sufficiently to render it useful in these areas.

INTRODUCTION

For many years, there has been a fruitful interplay between the interdisplinary pursuit of cybernetic ideas (bearing this label or not) and the special departments of the life sciences. Since the early 30s, for example, anthropologists have recognised that societal homeostasis depends upon symbolic regulatory programmes manifest as rituals, conventions, and traditions. Likewise, social change is commonly understood in terms of the competitive or co-operative interaction between subsystems characterised by these symbolic structures.* Similar comments apply at the level of animal populations, where the maintenance of density, dispersion and interspecific mutualism depend upon comparable processes (see, for example, Wynne Edwards[4]). The whole of ethology is, by definition, the study of behaviours mediating control and communication; hence, cybernetics is an essential part of this science.† Moving in one direction the area of cybernetic influence extends into studies of linguistics and kinship structures.‡ In another direction, it infiltrates biology (see, for example, Young[6]), embryology (see Waddington[7]), genetics, and developmental studies (for instance, Bonner[8]).

The crucial notion is that of a purposive or goal directed system. As the examples suggest, this concept has served very well to increase our understanding of natural processes. But the concept, as it stands, is not entirely satisfactory. The phenomena of evolution and of conscious experience are ubiquitous in all biological, social, or behavioural systems. It is far from clear that these phenomena can be explained (or even predicted and manipulated) within the existing cybernetic framework. A fundamental reappraisal of the concept "goal" is probably necessary.

Uneasiness over the adequacy of the existing framework has been expressed in various quarters; notably at the series of Wenner Gren symposia on conscious purpose and human adaptation, convened by Gregory Bateson. This is not just an academic matter. In order to control the social and ecological systems which nowadays show signs of instability or even destructive and autocatalytic degeneration, it does seem necessary to take the consciousness, self description and evaluation of these systems fully into account. Much the same theme will be developed by Stafford Beer in the

* The pioneering work is due to Bateson[1]. Recent developments along similar lines are documented in Rappaport[2] or Schwartz[3].

† This is especially obvious in the works of Lorenz, Tinbergen, and Mittelstadt.

‡ A representative selection of papers is contained in Garvin[5].

context of government and management. Hence, the present paper will examine the theoretical issues rather than dwelling upon their practical consequences.

Before embarking on this task, I must emphasize that these comments refer to behavioural and cognitive cybernetics. They are made from the viewpoint of some one concerned with natural systems and in no way contradict Prof. Boulanger's contention that issues of consciousness, etc., are often *irrelevant*. In the previous paper Boulanger adopted the attitude of an engineer who is anxious to make purposive or intelligent artifacts. From that point of view, of course, he is absolutely right. Wearing my engineering hat, I entirely agree with him.

CYBERNETICS IN RELATION TO HUMAN PSYCHOLOGY

To be specific, I shall trace the influence of cybernetic ideas upon a single discipline (human psychology). Here, as in the general domain, the key concept is "goal directed system" and it can be usefully refined in several ways. Once again, however, the concept of "goal" must be broadened in order to deal with outstanding issues of consciousness, conceptual development and the like, which a comprehensive psychology cannot afford to neglect. After showing that the requirement for a more liberal interpretation of goal directedness arises quite naturally from the application of the concept as it stands, I shall suggest several ways in which the connotation of "goal" can be usefully extended.

History

At the moment when the word "cybernetics" first made its appearance, there existed two classes of psychological theory, each carrying its own experimental trappings. On the one hand, there was behaviourism: either a brash, almost Watsonian, behaviourism or a mellowed "functionalism done with a behaviouristic bias" (chiefly represented in this country by the Cambridge School of Applied Psychology). On the other hand, there existed a sort of mentalism, born of the Gestalt psychologies amongst others, which was pursued in a thoroughly eclectic spirit, for example, by Bartlett.

Wiener's book[9] became widely known in the early fifties. It gave a name to an ongoing way of thinking and added mathematical stamina to a body of embryonic concepts. Of course, Wiener had spoken as a pioneer before he published. But his greatest innovation was philosophical and mathematical. The psychologists had been whittling away at broader Cybernetic notions for some years. Amongst them were McCulloch[10] and Pitts in the U.S.A. and Ashby[11,12], at that time in Great Britain, who laid the foundations of that peculiarly cybernetic edifice, "the brain as a communication and control system". Working at the behavioural level, Craik[13] saw the regulatory character of human performance with enormous clarity. Finally, there was a group of psychological information theorists, centred about Hick[14], and quite closely allied in their way of thinking to physical information theorists like Cherry[15] and Gabor (at Imperial College), Mackay and Shannon.

Thereafter, cybernetic ideas became increasingly popular. Their proliferation can be followed both in the psychological literature and in the relevant sections of various interdisciplinary forums (the Macy Foundation Symposia; the London Information Theory Symposia; the Congresses of the International Association of Cybernetics; the Conferences on self-organizing systems, sponsored by ONR; the Bionics Symposia, etc.). But, at the time in question (the early 50s), these concepts made a clear philosophical impression.

Philosophical Impact

The impact of cybernetics upon human psychology was many faceted.

1) Cybernetics drew attention to the form and dynamics, i.e. the *organization* of systems, which is often of greater relevance than their physical particulars. Usefully, but more superficially, it mustered a number of mathematical techniques for talking about organisation.

2) By establishing the basic concepts of feedback and stability, cybernetic thinking resolved those teleological dilemmas that had lingered on since the vitalist–mechanist controversy of the early years of this century and gave substance to the already ubiquitous notion of "goal directedness".

(3) Within the cybernetic framework, the constituents of organization, namely *information* and *control*, acquired a status just as respectable as that already accredited to "matter" or "energy".

(4) Conversely, it became evident that no system is completely specified by its physical description *alone*. The system's informational content and its control structure must *also* be described (for example, the system "gene" is not completely specified by talking about DNA molecules; in addition, a

gene entails the information encoded in the molecular configuration and the protein synthesising control loops in the context of which a gene is an hereditary unit).*

(5) As a result, the Cartesian Dualism, of which the distinction between behaviourism and mentalism is redolent, was replaced by a Systemic Monisim.

Systemic Monisim

The crux of systemic monisim is contained in the assertion that any system is a goal directed system which can be analysed into or (in context) synthesised from a collection of goal directed subsystems. The organisation of a

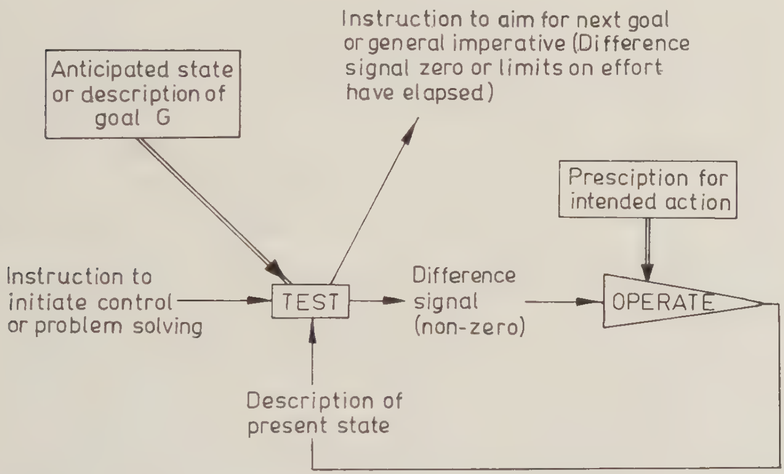


FIGURE 1 The basic goal directed system: a TOTE unit (modified)

goal directed subsystem, the basic building block, is the familiar *process* depicted in Figure 1. Notice that the organization is isomorphic with any of the following entities.

(1) A TOTE (or TEST, OPERATE, TEST, EXIT) unit, in the sense of Miller, Galanter, and Pribram[16].

* In view of later (essentially cybernetic) work in molecular biology, there is currently some doubt about hereditary units; the DNA configuration probably does not *uniquely* specify the organisation. However, the meaning of the example is clear enough.

(2) The interpretation and execution of an IF, THEN, ELSE programme segment.

(3) A properly interpreted control system.

(4) A problem solver.

(5) A game player (the equivalence of (3), (4), and (5) was mooted by Ashby and has recently been developed by Banerji)[17].

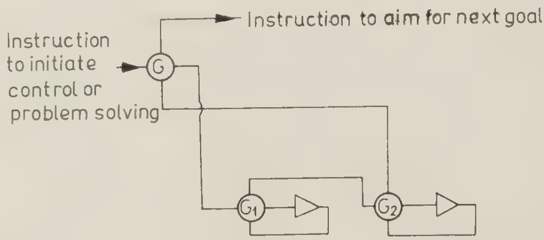


FIGURE 2 A typical decomposition of goal G: a TOTE hierarchy (modified)

(6) The execution of a state achievement command in the sense of Rescher[18].

(7) An elementary concept (regarded as a *procedure* for knowing or recognizing or doing)[19].

(8) Either the process of "abstraction" or of "completing an analogy"[20].

There are several ways of conjoining goal directed systems so that the entire system is goal directed (or, conversely, of dissecting a goal directed system into elementary units). One of them is shown in Figure 2. Here the entire system has a goal G and subgoals G_1 and G_2. In order to attain G, the uppermost unit calls for the execution of a G_1 subroutine and a G_2 subroutine, such a predictive sequence being a *plan*. Other types of composition and decomposition are discussed in a recent paper[21].

Holism and Atomism

Since it holds that composition and decomposition are universally possible, systemic monisim is a reductionist philosophy. But, in general, it is *holistic* rather than *atomistic* since, apart from a few trivial cases, the *whole* goal directed system is *more* than the sum of its goal directed parts. Further, in a sense that will be clear from the slightly unusual labelling of Figure 1, each subsystem can be said to *interpret*, to *intend*, and to *anticipate* or *expect*.

Hence, the reductionist explanations of human behaviour and mentation that feature in a cybernetic discussion are quite distinct from those (to my mind fallacious) mechanistic explanations in which man is reduced to a bag of associations and responses. To parody the position of naïve behaviourism, man is conceived *as something that reacts to stimuli.* In contrast, the cybernetic theories of psychology envisage man as *someone who interprets, intends, and anticipates.* To put it differently, a human being does not so much respond to stimuli as interpret certain states of his environment as posing *problems* which he makes an attempt to *solve.* Clearly, this point of view bridges the gap between behaviourism and cognitive psychology. A human being has the qualities ordinarily associated with mental activity; nevertheless, the human system is, in principle, reducible to elementary subsystems which have the same quality in a primitive form. Whether or not such a reductive explanation is generally *possible* is undecided at the moment; at any rate, useful explanations can be offered for certain aspects of human activity. Whether the goal directed unit is a rich enough construct to adumbrate "interpretation, intention, and anticipation" is a question to be taken up in a moment. Some disquiet on this score has been voiced already.

Cogency of the Cybernetic Approach

Since it resolves the arid conflict between behaviourism and mentalism, the cybernetic approach effects a salutory unification of the psychological field. Several theories of learning and cognition have been built up in overtly cybernetic terms; for example, in the U.S.A., the theories of Miller, Gallanter, and Pribram[16], McCulloch[10], and von Foerster[22] (or von Foerster *et al.*[23]); in the U.S.S.R., the theories of Anohkin, Amosov[24], Glushkov[25], and Napalkov and in Great Britain, my own theory[26,27]. To these should be added all theories involving cognitive and artificial intelligence models which are built up from elementary constituents with the status of information structures, for example, the models of Minsky[28], Reitman[29], Hunt *et al.*[30], Fogel *et al.*[31], George[32], Taylor[33], Uttley[34], and Young[6]. Many theories are primarily cybernetic in calibre; notably, Festinger's[35] "cognitive dissonance", Kelly's "personal constructs" and Laing's theory (see Laing *et al.*[37]) of interpersonal interaction.

Apart from this, the literature abounds with papers that are couched in behavioural terms but which are really talking about cybernetic constructs. Nowadays, when mental mechanisms are called "mediating processes" or (at a different level of discourse) are signified by "intervening variables" the author is nearly always referring to goal directed systems, programmes, and the like. The nomenclature of S.R. theory is cumbersome in this context; during a transition phase it is used by way of an apologia to classical behaviourism but is gradually being replaced by direct reference to the cybernetic entities.

Alongside this theoretical development a number of cybernetic methodologies have come into prominence. The most general of these, information theory and control theory, have already been mentioned. More specifically, the idea of a basic experimental situation is undergoing rapid change. The paradigm used to be a stimulus and response situation; now it is being equated with a game or a conversation between the subject and the experimeter (or his equipment, in some cases).

The consequences of this change in attitude extend far beyond the laboratory. They are particularly dramatic in connection with teaching, training, and computer assisted instruction, where sensible developments almost certainly rest upon the recognition that a tutorial *conversation* is the minimal, non trivial, transaction with a human being (here, incidentally, I mean conversation in the full-blooded *logical* sense; hence, "conversational interaction" with a simple computer terminal is generally insufficiently rich to qualify). These matters have been discussed extensively in other publications and I shall not dwell upon them[38–45].

Cybernetic theories and methods can be justified on psychological grounds; for example, the acquisition of either skills or concepts is most naturally described in cybernetic terms, as are the phenomena of selective attention. Empirically, cybernetic theories "work" quite well. However, the experiments of Dr. Grey Walter (which he will outline later this morning) lend a great deal of more specific support to the cybernetic contention. This work* has uncovered the physiological foundations for the goal directedness of man. Broadly, a complex of mechanisms involving the frontal cortex and certain lower regions such as the reticular formation, focus the attention upon relevant *evidence* and set up an anticipation or expectation with respect to its correlates and to properly equilibrating actions. In particular, the activity of this "expectancy" system depends upon either (1) a goal setting instruction (do so and so when something happens) or (2) an internal goal

* This is a refinement and extension of other work in the field. For particular theories in this matter, see Grey Walter[46] and Kilmer *et al.*[47] For the physiological background see Lynn[48].

orienting state. Generally speaking, a human being habituates against stimuli that are irrelevant to goals. Those which are relevant become *significant*, i.e. they pose *problems* or provide *evidence* (they do not simply "elicit responses").

It is right and reasonable to be impressed when physiology and psychology are happy bedfellows. In view of this work, there can be no serious doubt that human beings can be fruitfully represented as cybernetic systems.

DETAILED EXAMINATION OF CYBERNETIC THEORIES

Earlier in the paper I questioned the theoretical adequacy of the goal directed system as currently conceived and suggested that it is necessary to broaden our view of what such a system *is*. This calls for a more thoroughgoing appraisal of "goals" and of "cybernetic theories" in general.

Theory Building

Any theory starts off with an observer or experimenter. He has in mind a collection of abstract models with predictive capabilities. Using various criteria of relevance, he selects one of them. In order to actually make predictions, this model must be interpreted and identified with a real assembly to form a theory. The interpretation may be prescriptive and predictive, as when the model is used like a blueprint for *designing* a machine and predicting its states. On the other hand, it may be descriptive and predictive as it is when the model is used to explain and predict the behaviour of a *given* organism.

Now, in order to establish the identification and to form a predictive *theory* (indeed, in order to select one model from the set of possibilities), the observer needs to know the purpose *for* or the purpose *of* the system with which the model is identified. Lacking such a purpose, the observer would be at a loss to know what constitutes a sensible interpretation of the model or what properties of the world are relevant. In essence, of course, the purpose *for* or the purpose *of* the system is invented by the observer himself and it is stated in an observer's metalanguage for talking about the system. Thus, in the prescriptive mode, it is clear that people do not build purposeless machines. Equally, in the descriptive mode, an observer gets nowhere unless he has a systemic purpose in mind; for example, no headway was made with the explanation of amphibian vision until Lettvin, Matturana, McCulloch and Pitts* conceived the frog visual system as a machine for (with the purpose of) catching insects and avoiding predators. At that point, it became evident that the visual system consists in a set of attribute filters, evaluating properties relevant to this purpose.

In contrast, some systems have a purpose built into them; a "purpose *in*", i.e. a *goal*. Depending upon the type of observation we have in mind this may mean either (1) the models with which these systems are identified necessarily contain the mechanism of a goal directed system or (2) the system can state goals *to* the observer and accept some goals *from* him (or both). The principal cybernetic hypothesis can now be phrased as follows. *Any system with a purpose for it (any system for which a cybernetic theory can be constructed) also has a purpose in it, i.e. a goal; all systems are goal directed systems.*

Notice, in passing, the consequences of this definition. A cybernetic theory of adding machines is not just a theory of mechanical devices which have no goal. It refers directly to the *process* of addition and indirectly to the *user* of the adding machine, i.e. the mechanical device is necessarily embedded in the context which makes it meaningful.

Structural and Organizational Models

It follows from these comments that the truth of the cybernetic hypothesis cannot be decided (in respect to a particular system) at the level of the most fundamental and the simplest type of model: Ashby's "black box". However long a system identified with such a model is observed and however many experiments are carried out by varying the "black box" input, it will only be possible to say that the system behaves as *though* it is (or is not) a goal directed system. The whole concept of goal directedness depends upon the interpretation of a structural or organizational model for the system; something having enough detail to delineate the goal seeking process.†

Hence, in talking about goals, there is a tacit commitment to structural and organizational models containing a modicum of detail. At this level of

* "What the frog's eye tells the frog's brain" in *Embodiments of Mind*[10].

† This is quite clear in Ashby's work, of course. For example, in *Design for a Brain*[11] the concept of *essential variables* with limits upon their permissible values is employed to set up a goal directed system. The matter is generalized in Ashby's later work (see e.g. Ashby[48]). The present point is that structural notions, such as "essential variables" do not stem directly from the observation of a black box system. They are imported as a result of independent observations, e.g. data bearing on the nature of the animal.

discussion, still with human psychology in mind, it becomes useful to introduce a distinction between two types of goal directed system, namely *taciturn* and *language oriented* systems.* The former are systems in (roughly) the sense of general systems theory. The latter are based upon the concept of an *object language* (defined or described in the observer's metalanguage) in terms of which the system is able to accept goal statements (by programming or reprogramming) and to *describe* its current goals. The distinction "taciturn/language oriented" fundamentally entails the observer; we mean, to be strict "observed as taciturn/observed as language oriented". Nevertheless, these are features of the system, per se, which dispose us towards one mode of observation or the other.

Taciturn and Language Oriented Systems

Taciturn systems are those for which the observer asserts or discovers the goal (purpose *in*), which is thereafter equated with the purpose *for* the system in question. In contrast, language oriented systems can be asked or instructed to *adopt* goals by anyone who knows the object language and they may state and describe their own goals, using the same medium; in a very real sense these are "general purpose" systems. Ostensively, the distinction is determined by the following features.

(1) A special purpose, goal directed, computing machine (such as an autopilot) is a taciturn system. In contrast, a general purpose computer together with the compilers, interpreters, etc., required for processing statements in a programming language is a language oriented system. The programming language is the object language upon which the system is based.

Although this example is instructive, the peculiar character of general purpose computers must be kept firmly in mind throughout the discussion. In the case of a computer, an observer knows the programming language either because he has designed the machine or because he has a program-

---

* The basic distinction between taciturn and language oriented systems can be made in several ways of which this one is the most convenient for the present purpose. For example, Gregory[50] makes a similar spirited distinction between systems with a *deductive* capability (roughly, language oriented) and those without such a capability (roughly taciturn). Although Gregory's differentiation is elegant, and just as proper as my own, it does not fit the present framework quite so well. The caveat, *roughly*, must be taken seriously. As Figure 1 is labelled, any system with a goal (a purpose *in*) has a claim to deductive power.

ming manual written by someone who did so. In the case of a psychological system, an observer knows the programming language either because he speaks and understands it, or because of arguments involving inferences of similarity between the system and himself.

(2) A taciturn system can neither be given new goals nor can it state its goals (although an autopilot interacts with its environment, the legitimacy of calling the symbol system employed for this purpose, a "language" is suspect. Certainly it is not a language for stating new goals*). In contrast, the language oriented system is vacuous unless either it is *given* goals to aim for (by some sort of programming operation) or it already *has* goals which it is able to describe.

(3) Since a taciturn system cannot "speak" (i.e. communicate in a visual, auditory, or other symbolic modality), the notion of "mind" is irrelevant. On the other hand, within a language oriented system, it is usually possible to distinguish between a class of processes and procedures (for example, the class of programmes being executed in a general purpose computer) and the system in which these procedures are embodied (for example, the computer itself). The class of processes is an organization in the interpretative system and has the properties of *mind*, in contrast with the interpretative system itself (loosely, *brain*). Notice that the example of the general purpose computer though illuminating, is again misleading if taken too seriously. Computing systems are designed in such a way that the interpretative system, a box of logicians building bricks, is virtually independent of the organization. Brains are not like this.†

(4) In respect to a taciturn system, *information* has but one technical sense, which is developed in Prof. Ashby's paper at this congress. Briefly, information is a property of the relations existing between entries in the contingency tables which summarize the behaviours or possible behaviours of the system. It is crucial that the states so designated are defined in the observer's metalanguage and that the probability estimates, uncertainties, etc., are *observers* probability estimates, uncertainties, etc. (i.e. they are *objective*). Of course, the term information can be used in exactly the same sense with respect to the behaviour of a language oriented system. But here there is another pos-

* According to this argument, the course changing instructions delivered to an autopilot change the parameters of a given goal.

† There is, however, a fairly close relationship between brains and more complex computational systems with supervisory director programmes and resource allocation executives.

sibility as well, which is not open for the taciturn system. Clearly, the language oriented system can define a set of alternatives in terms of its own object language*; conversely, it can be *given* a set of alternatives. With respect to these it can express *subjective* or *systemic* uncertainty.

When the model for the language oriented system has been used prescriptively, as in writing an artificial intelligence programme, the observer can give an operational interpretation of subjective uncertainty and of the corresponding subjective information measure; for example, the degree of uncertainty with reference to problem P is the amount of computation required to solve P or the amount of computation which the system estimates will be needed at the moment it makes an utterance. When the model is used descriptively, an operational interpretation is not generally available and the asserted subjective uncertainties both may be and can be regarded as primitive measures. For instance, if man is a language oriented system, it is legitimate to take confidence estimates, obtained by the veridicial scoring technique of Shuford and his colleagues†, as *primitive* indices of the system state. Objective indices, which may, of course, be closely correlated with them are, according to this point of view, indirect state descriptors.

(5) The model for a taciturn system is identified with reality (for example, in the context of an experiment) by setting up a material analogy‡ between

* To crystallize the idea of the observer's alternatives and the system's alternatives, consider a human subject as the system. If the subject is asked to respond on a five-point scale in Osgood's semantic differential test, the alternatives (words at the ends of the scale) are chosen by the observer (as a matter of fact, as a result of a prior analysis of the statistical response tendencies of a population of subjects). In contrast, the alternatives obtained and used in the Kelly grid technique are system alternatives. They are determined, in the framework of an object language, by the human subject himself. A similar point is made by Bannister and Mair[51].

† Consider an experiment in which the subject is required to respond, at the nth trial in a sequence by choosing one of M alternatives. It may be the case that the subject is uncertain about which alternative to select (in order to satisfy a goal). If so, he is required to state M numbers, $r_i(n)\ i = 1 \cdots M$, such that $\sum_i r_i(n) = 1$. The $r_i(n)$ are interpreted as his degrees of belief in each of the alternatives presented at trial n and it is possible to score the subject over the sequence of trials as a function of the $r_i(n)$ and the alternatives he ought to have chosen. Shuford and his colleagues have introduced scoring schemes with the property that if the subject's real degrees of belief are $p_i(n)$ at the nth trial, then his mathematical expectation of score is maximized if, and only if, $r_i(n) = p_i(n)$. The same technique can be employed when the subject, rather than the experimenter, invents the alternatives. (See Shuford *et al.*[52]).

‡ For example, the sort of relationship which exists between an analogue computer model for a plant and the plant itself.

the model and the thing. Further, the observer or experimenter is solely responsible for determining and maintaining this relationship. Thus, stimulus signs are carefully delineated, responses are carefully observed, and the system is isolated from extraneous parameter variations by efforts to maintain constant and repeatable conditions. In contrast, all language oriented systems are based on models which are identified with reality in the normative framework of the object language; either the natural language of a human subject or an artificial language which he understands.* For example, the human subject is asked to participate in an experiment and he agrees to do so. Normative rules are set up which determine the nature and designation of problems, the class of solutions and so on. Above all, a goal is specified either by the subject or the experimenter. To put it succinctly, an experimental contract is established between the observer or experimenter on the one hand and the human subject on the other. The whole experiment makes sense and the model itself is identified within the framework of this contract. It follows, of course, that *both* the subject *and* the experimenter (or observer) are jointly responsible for determining and maintaining the identification.

General Statement

Theory construction in the large is a generalization of the identification or interpretation process of (5) of the preceding section (in the sense that a class of models are interpreted, not just one particular model in one particular experiment—clearly in the general case the "observer" becomes the "scientist"). Hence, we have two sorts of theory; a theory of taciturn and a theory of language oriented systems.

The theory building process is an open-ended control process in the conduct of which a cybernetic system (by definition a *control* system) is established. Hence theory building is, in one sense, "control of control". But the higher level (open-ended) control process is not formally modelled and possibly any attempt to model it would end up in a (vicious) indefinite regress.

If the cybernetic system to be established is taciturn, then the observer (scientist) is alone responsible for it. If the system is language oriented then

* We emphasize a point mooted earlier. The concept of language is very broad indeed. Pictograms or images are just as good a currency as words or mathematical expressions.

the object language of the system is itself used as the metalanguage involved in the higher level "control of control" and the subject becomes an active participant in theory construction[53].

This is particularly obvious when we notice that the great majority of experimental contracts (preceding section, (5)) are not *really* established by one way instruction giving but are *compromise* solutions arrived at by dint of conversations about the experiment in question.

The Psychological Domain

Let us crystallize our attitude. *All the systems of human psychology are language oriented systems and all the models proper to human psychology are language oriented models.* This follows from the definition of a language oriented system and the discussion on pp. 25–27. To deny the assertion it would be necessary to cite a psychological experiment that does *not* depend upon an experimental contract.

As phylogenetic development proceeds, there is a tendency for the language oriented system to become apposite; it would be absurd to see primitive animals in this way but with adult man we have argued it is the only legitimate point of view. Similarly, there is an ontogenetic development beautifully illustrated by Luria's work[54]. The control function of language unfolds as a child grows up and, with it, the cogency of the language oriented system. In contrast, most functional and physiological systems are taciturn: for example, the autonomic system, simple conditioning, and non-symbolic adaptation.

At first sight, we seem to have come round a full circle and returned to a type of dualism; on the one hand there are language oriented systems ("mind" systems), on the other taciturn systems ("body" systems). But the impression is illusory. "X is a language oriented system" glosses the complete statement, "X is *observed* as a language oriented system" (and must be so viewed if the observer is a psychologist). The price to be paid for the convenience of systemic monisim is that of keeping the *observer* as an integral part of *all* observations[53,55].

With that caveat, we can often observe a human being as a psychological and a physiological system at the same instant. Grey Walter's work (see section on pp. 45–56) provides an admirable instance. On the one hand, he views the human subject as a taciturn (physiological) system. On the other, he views him psychologically, for example, in an experiment where the subject is required to entertain expectations. The psychological system is language oriented, even if the subject only "expects" clicks or light flashes. He *becomes* so, just *because* certain physiological mechanisms are brought into play.

DEPARTURES FORM THE SIMPLE PARADIGM

Let us idealize the cybernetic concept of man as it has so far been presented. An individual human being is a language oriented system (for short, an "L.O. system") occupied with one fully specified goal at once. Any change of goal is guided by a plan, in the sense of Figure 2, which determines the immediate subgoals of a still fully specified overall goal. Such a picture is isomorphic with the operation of a computer programme, the L.O. system, which is embodied in and executed by a computer called the brain.

For many purposes, the picture is a useful approximation to reality but it does not bear close scrutiny. First of all, the brain is not the passive and ductile apparatus which comes to mind at the mention of "computer". It is, indeed, a computer; but, as suggested before, it is a taciturn system in its own right with goals that are not necessarily compatible with those of the L.O. systems embodied in it. Secondly, human beings are not so single minded as the simple picture suggests. Man can often be imaged as aiming for one goal at once, especially when he is making symbolic utterances or is coupled to the observer via the string processing and push down list structures which are characteristic of immediate memory organization[56]. But he is also capable of multigoal operation. This fact opens up the possibility that man is an *evolving* L.O. system and I hypothesize that this possibility is *always* realized.* If so, the consequences are profound and roughly as follows.

An observer who sticks to the rules on p. 24 must see a human being as a system having a purpose *for* and will try to place this in correspondence with a purpose *in* or systemic goal. Now, if the observer elects to see the man as a system with *one* goal, then in certain circumstances, (by hypothesis, in *all* circumstances) the observer will be impelled to say that this goal is *underspecified.* Conversely, if he choses to see a multiplicity of goals (which, in toto, satisfy the purpose *for*) then these may be fully specified *but* the observer suffers an irreducible uncertainty over the systemic boundaries of the

* That, in a non-trivial sense, he is *always* learning. He is built with a propensity to learn[26].

individual said to *have* these goals. For, in reality, this *individual* is an evolutionary process which can be *described* (from the observer's point of view) as a self-organizing system in von Foerster's[57] sense of the phrase.

Equally if, man is defined as an L.O. system, then *he* is able to act as his *own* observer and thus to see *himself*. In that case, he (the individual human being) is in a similar position to the external observer. His evolutionary nature leads him, if questioned, to say either (a) "my goal is underspecified" (even "I have no goal")—here the integrity of the individual is taken for granted by the speaker, *or* (b) "I have a definite goal" (for example, to do running exercises for 15 min at 80 steps/min). But I might choose to aim for a different goal (e.g. writing this paper, solving a problem). Hence the boundaries of the individual are undefined. "I" am, by admission, something that is aiming for a definite goal but also something (undefined) that contemplates other possibilities so that "I" might elect to do differently. The goal, in this case, is *contingent* upon the acceptance of a normative framework, such as the experimental contract of p. 28, or the system of conventions and social mores (accepted, for example, by a devoted problem solver, clerk, or mathematician). *Contingency* arises because the human being *may* and *knows* he may disobey the norms and aim for some other goal, *or* (c) "I have a definite goal at the moment but I realize it is temporary and will give place to another". Here, the human being recognizes the temporal development of the process he *is*. Phrasing it differently, man spends much of his day in goal *setting* (or problem *posing*) rather than goal *seeking* (or problem *solving*).

In practice, the distinctions are less clearcut than (a), (b), and (c) suggest. Even the specific goals of (b) and (c) usually turn out to be underspecified to some extent, i.e. the man who describes the goal state is unable to give it a consistent ostensive definition. The ambiguity of all natural languages allows for the communication of underspecified goals. It is because of this that conversation (in a nontrivial sense) and social development in general are both possible.

We may or may not choose to call evolutionary systems "goal directed"; clearly, if they *are* goal directed at all, then they are directed towards an underspecified and generally open ended goal. Brodey and Johnson[58] have rightly pointed out the dangers of calling an individual or a society "goal directed"; the name suggests a narrowness and specificity which is counterfactual and which may encourage wrongheaded or positively harmful efforts at controlling the system in question. On the other side of the coin, these evolutionary systems are immediately related to simple goal directed systems and it may be a salutory exercise to broaden our notion of goal. One thing is certain; if we *do* use the word in connection with human affairs (and, as cyberneticians, we are prone to do so) then we should be fully aware that goal directedness is rarely, if ever, of the simple-minded sort.

DISCUSSION

The following sections discuss and develop the broader concept of goal directedness, mooted in the last section, i.e. a concept of goal setting as well as goal seeking.

Redundancy of Potential Command

McCulloch coined the phrase "redundancy of potential command" to describe the relationship existing between a set of goal directed systems which compete for dominance. It is clearly assumed that the systems in question (call them the goal directed subsystems) have a value defined on their operation; any one is built to seek an opportunity to operate and command the others and they clearly exist in such a relation to one another (or to an environment) that only one of them can command at once. Generally, the one that wins depends upon evidence from the environment (or from the aggregate of subsystems, or both); there is a tendency for command to shift from time to time in a way that favours the subsystem currently in possession of the most relevant information.*

The multi-goal systems of the last section are parallel computational systems in *this* sense; *not*, for example, in the sense that a perceptron is a parallel system.

As mentioned on p. 21, McCulloch and his colleagues have computer simulated the action of the reticular formation, which is one of the physiological mechanisms involved in directing an organization's attention. This simulation provides a lucid instance of "redundancy of potential command". The goal directed subsystems are, in this case, concerned with the potential *modes* of operation of the organism (i.e. walking, eating, etc.). They interact in the relationship indicated above and the organism as a whole is committed to one mode of activity or the other as command is shifted amongst

* The work of Mesarovic and his associates (for example, Fleming *et al.*[58]) is similar spirited. It is, however, carried out at an abstract level.

them. The selection of the currently dominant system depends upon the weight of evidence in respect to *all* of the modal computations and also upon a feedback from the cortical processes engendered by the immediate commitment. Whilst each of the goal directed subsystems has a fully specified goal (for example, "mediate eating behaviour") the goal of the system as a whole is underspecified ("general stability" or "survivial", or something of the sort).

Here, of course, we are talking about taciturn systems. But a similar picture holds good at the level of L.O. systems, which typically compete for *execution.** For example, the perception of visual illusion figures is frequently accompanied by an oscillation between interpretative programmes†; the Necker cube, seen "infacing" at one moment and "out-facing" at the next is a clear instance of this phenomenon. Here the competing L.O. subsystems constitute a system with redundancy of potential command. But at this level, *co-operative* as well as competitive interaction becomes an obtrusive feature of the process.‡ For example, in viewing a paradoxical figure such as the "tuning fork" or the "impossible staircase", oscillation goes hand in hand with a resolution of the type proposed by von Foerster. The viewer makes an essentially self-referential statement and generates a construct

---

* In computational usage, a programme *is executed*; it does not of itself "compete for execution". Here the analogy with present day computation proves inadequate. The computations carried out in a brain (especially in the "working memory" to be referred to on p. 34) belong to the same class as the computations carried out in a cell. If computers were not so fashionable and cells so unfamiliar, I would have developed the argument in these terms.

To sketch what I mean, enzymes, in particular allosteric enzymes, are the most elementary goal directed systems in the cell. They operate in cyclic transformation processes which are unequivocally programmes (for example, the Krebs cycle). Some of these are protein synthesizing cycles which produce (amongst other things) fresh enzymes: for example, the well-known and unequivocally programmatic organization, "DNA message → messenger RNA; transfer RNA + amino acids → tagged amino acids; messenger RNA at ribosomal site + tagged amino acids → fresh enzymes". Here it is obvious that both simple and complex programmes have an imperative built into them; in the cellular environment they compete for execution and co-operate; in turn, they recreate or reproduce this environment. Mental organization has a similar quality and it is in *this sense* that I use the phrase "compete for execution".

† These programmes match the excitation of a sensory manifold to the expectations entertained by the subject. A similar proposal is made in Gregory[60], which also provides an elegant discussion of the field in question.

‡ "Becomes obtrusive", because, on closer examination of *all* systems with redundancy of potential command, co-operative phenomena are evident in an embryonic form.

involving a further spatial dimension in order to resolve the disparity between the rival programmes[61].* This is *co-operation* in the present sense. Two L.O. subsystems acting in concert can do more than the sum of the two acting alone and a new system is generated as a result of their interplay. If the L.O. systems are cognitive rather than perceptual programmes, then co-operative interaction is identical with Schon's[62] displacement of a concept to produce a new one (see, in particular, the example of the concept "drum", pp. 30–32).

The Individual at a Given Instant

We are now in a position to see the individual, at a given instant, not so much as a *particular* goal directed L.O. system as a collection of L.O. systems bearing (in some sense) the same name† and tied together by the relationship of enjoying redundancy of potential command with respect to an *overall* goal which will be *seen*, either by an observer or the currently dominant system, as a *contingent* or *underspecified* goal in the sense of p. 31 (a), (b), or (c).

Evolutionary Processes

An L.O. system with redundancy of potential command becomes an evolutionary system insofar as its L.O. subsystems must be embodied in a computing mechanism prior to execution, insofar as these *embodiments* are subject to decay or abrasion and insofar as there exists a reproductive or maintenance process that preserves successful subsystems or variants against decay. If so, the basic competition between the subsystems in the population becomes a competition for reproduction and survival.

I hypothesize that the brain, in particular the *functionally* (not physiologically) demarcated "working" memory mechanism, is just such a computing medium and, consequently, hold that the individual is continually evolving. It is exactly in this sense that I sometimes dub the brain (or that part of it) an "organ for reproducing concepts".

---

* von Foerster has studied the matter chiefly in terms of colour vision (working with Dr. Matturana) and constancy phenomena.

† Usually in the sense that the L.O. subsystems are run in the same brain. But note the previous comment, that if an observer tries to identify this class he is liable to an uncertainty about the extension of the individual. Note, also, the comments made later on the subject of conversational interaction.

Any programme being executed in working memory can address information, subroutines, and instructions which are generally lodged in the long term memory of the same brain, but which may also be written as records, in the environment. Equally well, an evolutionary process in brain A can be coupled linguistically to a process in brain B; conversational interaction often mediates exactly the same kind of co-operation as the *internal* process of conceptual displacement. Hence the evolving individual is sometimes distributed rather than localized in a single brain. Recall from p. 31, the observer's uncertainty about the boundaries of an individual.

Goal Setting

The evolutionary process generates a sequence of sets of subsystems having redundancy of potential command. As on p. 31 (a), the goals of the collection, of the whole system is necessarily underspecified. From time to time, the issue of command is temporarily resolved when an individual's goal is definite but contingent either in the sense of p. 31 (b) or (c). Looked at from a slightly different angle, the resolution process is itself part and parcel of the general evolution.

Resolution (and goal setting) occurs in several different ways.

(1) By dint of information received from the environment, which defines a new goal. In deference to Hawkins and Storm, I shall call this "eolithic intervention" (see Hawkins[63]).*

(2) By *external* co-operative interaction or conversation with some other individual.

(3) By *internal* co-operative interaction between L.O. goal directed systems seeking the same goal in different ways.

(4) By competitive interaction.

(5) A special case of (4). The language oriented individual sees his own brain (in particular the programmes run in the limbic structures) as a system

---

* The author recalls and develops the argument in an earlier paper by Storm. The argument is placed in the context of design, which is commonly regarded as a form of problem solving with respect to a fully specified goal. Hawkins points out that a great deal of design is quite different. The designer "has no goal" but encounters some object or method in the environment which suggests a goal; this he calls an eolith. For example, the designer may come across an oddly shaped piece of stone which suggests the goal of making a spade. In our own laboratories bits of apparatus or deeply engrained methods often set the goals for subsequent research proposals.

with goals of its own. These may or may not be consonant with the goals he currently entertains. In any case, this system ("his" computing system) "engages him in discourse".

The Correlates of Conscious Experience

"Man is a language oriented system" glosses "man is *observed* (by the psychologist) as a language oriented system", i.e. he is engaged in discourse. Insofar as the subject states or accepts goals, albeit underspecified goals, he is presumed to be aware and, in potentially communicating his awareness to the observer, to be conscious *with* him. The domain of enquiry defined as psychological on p. 29 is thus a domain of consciousness and it is pertinent to investigate the correlates of conscious experience. Notice, we are not trying to *explain* conscious experience in terms of more primitive events (for example, states of as taciturn system). According to p. 29, that would be an essay in the wrong type of reductionism. Furthermore, I believe it would be doomed to failure because observations of language oriented and taciturn systems are fundamentally different kinds of *observation* (to reiterate the point on p. 29: that does *not* mean there are two sorts of system). However, we *can* usefully set up correspondence between the appearance and even the nature of conscious experience and the operations which go on in (say) an evolutionary process. The following proposals on this score have the form "the execution of such and such a programme in working memory correlates with conscious experience".

Somewhat contrary to general belief, I contend that the human being is *unaware* of the execution of programmes with fully specified goals. He does not *know* when he is acting as an automaton. For example, he is unconscious of the execution of overlearned skills and he is unconscious of the routine and massive searches which must go on in the associative network of long-term memory. On the whole he is unaware of intellectual problem solving when the subgoals are completely specified; he becomes conscious of the process when, though the overall goal is fully specified, some of the subgoals are not, i.e. in general, he is aware of problem *posing* and the process of *constructing* problem solving procedures.

Man can be made aware of some normally unconscious processes if, when asked to describe them, he attempts the dual task of carrying out a procedure and matching an account of it to the observer's understanding (his success in actually producing a description varies widely; he is moderately competent in respect to procedures where there are subgoal points at which he *might* experience uncertainty as there usually are in intellectual tasks; he is utterly incompetent when it comes to describing how he performs an over-learned skill). In general, man becomes conscious when at least two processes are going on at once and these may or may not be internal to his brain.

For example, in skill learning (signal translation, teleprinter operation, etc.) the subject is aware of his errors insofar as (1) he has some rudimentary procedure for making a goal directed response and (2) the experimenter provides an external *co-operative* system which (as it were) does the same computation perfectly and provides the subject with knowledge of results feedback. Later in learning (with no knowledge of results feedback) subjects are conscious of *some* errors but ignorant of others. The *conscious* errors seem to be associated with the following circumstances: (a) there exist some slow but sure response programmes acquired early in learning; (b) these are lodged in long-term memory; (c) a more recent, more efficient but nevertheless more fallible procedure has been learned later for doing the same job; (d) the new procedure is applied (to achieve the goal) in parallel with the old procedure (aiming for the same goal); (e) competitive or co-operative interaction takes place insofar as a comparison is made between the "truth" (old procedure output) and the "actuality" (result of the new but fallible procedure).

Broadly speaking, man is aware of goals which he is asked to or anxious to attain but for which he does not possess the requisite goal seeking apparatus (and has to build it by a concurrent learning process). He is aware of contingent goals and, by the same token, of a mismatch between what he does and what he intends to do, between what he senses and what he expects, or between conflicting interpretations.

My conjecture is thus as follows. *The unique correlate of conscious experience is a state of a process (wholly or partly in working memory) such that (1) there exist two or more goal directed systems (usually in a relation of redundancy of potential command) and (2) these systems interact either competitively or co-operatively; in short, when they engage in discourse. Whilst the discourse in question may be internal to a single brain, it may also involve a system in the environment, in the brain of a conversation partner or in the brain of an observer.** These conditions can be satisfied by the evolution of a language oriented system.

---

* The interaction must be non-trivial. In conversation, for example, the sentient individual must compute what *he* believes the other individual is also computing and there

The conjecture is open to two criticisms. The first, that it says little more than "thought is subvocal speaking" is misplaced. There is no more than a superficial similarity between this dictum and the present conjecture. The second criticism, that the conjecture seems to neglect man's obvious awareness of pleasure, pain and the like can also be refuted. In fact, it would be possible to erect an entire theory of affect on the basis of discourse between programmes (L.O. systems) run in the limbic regions and those run in the neocortex (the sort of interaction mentioned on p. 35). Some recent affective psychologies come close to this stance. But the matter, though interesting, is beyond the bounds of this paper.

Predicting and Controlling Evolutionary Systems

The general *mechanism* of evolution has been computer simulated by various workers, for example, by Fogel and his colleagues[31], by Toda[64], and by myself.* Many of the more dynamic artificial intelligence programmes contain parts that are also "evolutionary". The real difficulty is modelling or representing the quasi-linguistic operation we have referred to as "setting a new goal" and this, of course, is peculiar to the embodiment and execution of an evolutionary L.O. system.

We have a limited understanding of one especially tractable situation involving an L.O. evolutionary system; namely, concept acquisition in a tutorial conversation (recall the definition of p. 20; a concept is a goal directed system). Here, the overall educational goal is fully specified in the sense that the subject (student) agrees to aim for it within the terms of an experimental contract and the whole construct is contingent upon the observance of this contract. Next, the whole of the co-operative interaction which builds up the new concept is assumed to take place via the conversational channel; it is externalized in communication between the subject and the teacher which, either in fact or in effect, is a fully specified teaching mechanism. This machine operates (1) as an external process that co-operates with the student

must be a comparison between the output from *his* model of the other individual and what the other individual says or does. The argument applies, vice versa to the other individual.

* My own work in this field is scattered through the literature, for example, Pask[65–67]. One of the most comprehensive models has been provided by Brieske[68], working at von Foerster's laboratory. Many others (notably Baricelli and Bremarman) have studied evolutionary processes in biological systems.

as he learns and (2) in the role of an observer. Whilst the subject is allowed to propose his own strategies, to set his own subgoals, etc., the acceptance of his decisions is contingent upon and is monitored by this external machinery.

Given all this, the evolutionary process of concept learning can be described by an *heterarchical* model for the subject (student). The original concept is represented by a goal hierarchy or problem solver in the sense of Figure 2; where, for example, G is at a higher level in the hierarchy than either G_1 or G_2. Learning is represented as an operation in which comparable problem solvers act upon the domain of the original problem solver (the original concept) in order to remedy its defects and to write fresh programmes. Clearly, this entails a quite different hierarchy; an hierarchy of control. For example, the original concept is a problem solver at the lowest level of *control* and the problem solvers that operate upon it reside at a higher level of *control*. Since both problem solvers may have the same organization (they need differ only in domain) and since they both have subgoals at various levels in the goal hierarchy, there is an interaction between the hierarchies and the entire model is *heterarchical*, as proposed a moment ago.*

Under these restricted conditions, it *is* possible to predict the course of evolution or learning and to control it by appropriate teaching strategies.

The trick employed is to conceive "goal setting" as higher level goal seeking (higher, that is, in the hierarchy of control). This trick is perfectly legitimate provided that the resulting model is based on the assumption that the goals "set" by the subject are *subgoals* of the *fully* specified educational goal. But the construct becomes completely invalid as soon as the subject departs from the experimental contract (which he *may* do and which he *knows* he may do).

Towards a Theory of Theory Building, i.e. a General Theory of Goal Setting

In general, the generation of new goals involves operations in which the human being becomes his own observer. In the role of observer, he sees himself as a system and defines a purpose *for* this system (in the sense of

* For an outline, see Pask[69]. The most complete statement of the theory is in Pask *et al.*[70]

the section on p. 23) which later acts as a purpose *in* the system (i.e. acts as its goal). In other words, the unconstrained goal setter (for example, "man as a scientist" in Kelly's personal construct theory or "man as an innovator" in all psychoanalytic theory) is his own theory builder (in the sense of the sections on pp. 23 and 24) and the representation of this general case calls for a formal statement of the notions contained in these sections: a theory of theory building.

No such theory exists. But some of its constituents are available, as formal tools, at the moment. The first step towards developing a theory of theories is to muster, integrate and, in some cases, sharpen these tools. The following items are the prerequisites which I, personally, have in mind.

(1) A proper logic of commands and intentions; the germ of it is available in the work of Rescher[18] and Von Wright[11] and in Kottelley's[12] intentional calculus (partly developed).

(2) A formal theory of partially co-operative interaction and conversation. This may be based on Howard's[13] theory of metagames, augmented (so far as the communication problem is concerned) by the ideas of Gorn[74].

(3) A logic of distinction to comprehend the act whereby a goal (or goal like entity) is abstracted from an amorphous flux of development. The problem was clearly stated at a philosophical level by Jung[15] in the 1920s; Spenser Brown[76] has recently solved it and provided an elegant calculus of distinctions which calls for an interpretation in the present field.

(4) A representation for essentially parallel processes. Here, the most promising candidate is Holt's occurrence theory. Within this framework, it is possible to formalize the concurrence of events and the ideas of competition and of information. The phrase "information transfer" has a meaning within occurrence theory that differs markedly from the current technical usage. "Information transfer" between occurrence systems is identical with the co-operative interaction that resolves uncertainty over an underspecified goal[77].*

(5) An axiomatic statement of the notions underlying evolutionary processes. Lars Loefgren has provided the bones of such a thing (the possibility of complete axiomatization is undecided)[78].

---

* Any cybernetic system, in the sense of the section in p. 24 can be represented as an occurrence system. We hypothesize that "information transfer" between goal directed systems of the evolutionary process is the unique correlate of conscious experience. The nub of the problem is, "who interprets or represents the systems in this way?".

Limitations

If a theory of theory building is fabricated, then what sort of theory will it be? As mooted earlier, a *purely* formal theory of the sort that would lead to a casual *explanation* of goal setting and conscious experience, is almost certainly unattainable. But this does not mean that no useful theory can be constructed to *adumbrate* the issues in question in the sense of predicting and controlling the behaviour of evolving, language oriented, conscious systems. The conviction that we can *adumbrate* but not *explain* these systems could be regarded as a doctrine of despair. Personally, however, I see it quite differently; as an indication of the limits and the fascinating potentialities of our discipline.

References

1. G. Bateson, *Naven*, 2nd edn., Stanford University Press, Stanford, Calif. (1958).
2. R. Rappaport, *Pigs for the Ancestors*, Yale University Press, New Haven, Conn. (1969).
3. E. Schwartz, in *Our Own Metaphor, Proc. 1968 Wenner Green Symp. on Conscious Purpose and Human Adaptation* (1968).
4. V. C. Wynne Edwards, *Animal Dispersion*, Oliver and Boyd, Edinburgh (1962).
5. P. Garvin (Ed.), *Cognitive Studies and Artificial Intelligence Research*, Aldine Press (in press).
6. J. Z. Young, *A Model of the Brain*, Oxford University Press, Oxford (1964).
7. C. H. Waddington, *The Strategy of the Genes*, Allen and Unwin (1957).
8. J. T. Bonner, *The Evolution of Development*, Cambridge University Press, Cambridge (1957).
9. N. Wiener, *Cybernetics*, 1st edn., M.I.T. Press, Mass. (1948).
10. W. S. McCulloch, *Embodiments of Mind*, M.I.T. Press, Mass. (1965).
11. W. R. Ashby, *Design for a Brain*, 1st edn., Chapman and Hall, London (1952).
12. W. R. Ashby, *An Introduction to Cybernetics*, 1st edn., Chapman and Hall (1956).
13. K. Craik, in *The Nature of Psychology* (Ed. S. Sherwood), Cambridge University Press, Cambridge (1966).
14. W. E. Hick, "The rate of gain of information", *Quart. J. Exp. Psychol.*, **4**, 11–26 (1952).
15. C. Cherry, *Human Communication*, M.I.T. Press, Mass. (1957).
16. G. A. Miller, G. Galanter, and K. H. Pribram, *Plans and the Structure of Behaviour*, Holt Dryden, New York, (1960).
17. R. Banerji, *Some Results in a Theory of Problem Solving*, Elsevier Press, London (in press).
18. N. Rescher, *The Logic of Commands*, Routledge and Kegan Paul, London, (1967).
19. G. Pask, "Cognitive systems", in *Cognitive Studies and Artificial Intelligence Research* (Ed. P. Garvin), Aldine Press, New York (in press).
20. G. Pask, "The use of analogy and parable in cybernetics", *Dialectica*, **17**, Nos. 2 and 3, 167–202.

21. G.Pask, "A cybernetic model for some kinds of learning and mentation", in *Cybernetic Problems in Bionics* (Eds. H.L.Oestricher and D.R.Moore), Gordon and Breach, London (1968).
22. H. von Foerster, "What is memory that it may have hindsight and foresight as well?", *B.C.L. Publ.*, No.153, Univ. Illinois (1968).
23. H. von Foerster, A.Inselberg, and P.Weston, "Memory and inductive inference", in *Cybernetic Problems in Bionics* (Eds. H.L.Oestricher and D.R.Moore), Gordon and Breach, London (1968).
24. N.M.Amosov, "Simulation of thinking processes", in *Purposive Systems* (Eds. H.von Foerster *et al.*), Spartan and MacMillan, New York (1969).
25. V.M.Glushkov, *Introduction to Cybernetics*, Academic Press, New York (1966).
26. G.Pask, "Man as a system that needs to learn", in *Automaton Theory and Learning Systems* (Ed. D.J.Stewart), Academic Press, New York (1968).
27. G.Pask, *Behavioural Cybernetics*, Elsevier Press, London (in press).
28. M.Minsky, *Computation, Finite and Infinite Machines*, Prentice Hall, London (1967).
29. W.R.Reitman, *Cognition and Thought*, Wiley, New York (1965).
30. E.B.Hunt, J.Marin, and P.L.Stone, *Experiments in Induction*, Academic Press, New York (1967).
31. L.Fogel, A.J.Owens, and M.J.Walsh, *Artificial Intelligence Through Simulated Evolution*, Wiley, New York (1966).
32. F.H.George, *The Brain as a Computer*, Pergamon Press, Oxford (1961).
33. W.K.Taylor, "A pattern recognizing adaptive controller", in *Automatic and Remote Control*, Butterworth, London (1963).
34. A.M.Uttley, "Conditioned probability machines", in *Automatic Studies* (Eds. C.E. Shannon and J.M.McCarthy), Princeton University Press, Princeton, N.J. (1956).
35. L.Festinger, *A Theory of Cognitive Dissonance*, Tavistock, London (1959).
36. G.A.Kelly, *A Theory of Personality*, Norton, New York (1963).
37. R.Laing, H.Phillipson, and R.A.Lee, *Interpersonal Perception*, Tavistock, London (1966).
38. G.Pask, "Learning and teaching systems", in *Survey of Cybernetics* (Ed. J.Rose), Iliffe, London (1969).
39. G.Pask, "Electronic keyboard teaching machines", *J. Natl. Assoc. Educ. Com.*, July (1958).
40. G.Pask, "Electronic keyboard teaching machines" (reprint of above), in *Teaching Machines and Programmed Learning*, Vol.1 (Eds. R.Glaser and A.Lunisdaine), Washington National Education Association, Washington, D.C. (1960), pp.336–49.
41. B.N.Lewis and G.Pask, "The theory and practice of adaptive teaching systems", in *Teaching Machines and Programmed Learning*, Vol.2, *Data and Directions* (Ed. R.Glaser), Washington National Education Association, Washington, D.C. (1965), pp.213–266.
42. G.Pask, "The control of learning in small subsystems of a programmed educational systems", *IEEE (Inst. Elec. Electron. Engrs.), Trans. Human Factors Electron.*, **HFE-8**, No.2, 88–93 (1967).
43. G.Pask and B.N.Lewis, "The use of a null-point method to study the acquisition of simple and complex transformation skills", *Brit. J. Math. Statist. Psychol.*, **21**, Part 1, 61–84 (1968).

44. G. Pask, "Teaching machines", in *Modern Trends in Education*, MacMillan, New York (in press).
45. *Proc. Intern. Assoc. Cybernetics*, Namur (1958 onwards).
46. W. Grey Walter, "Electric signs of expectancy and decision in the human brain", *Cybernetic Problems in Bionics* (Eds. H. L. Oestreicher and D. R. Moore), Gordon and Breach, London (1968).
47. W. Kilmer, W. S. McCulloch, *et al.*, "A cybernetic theory of the reticular formation", *Cybernetic Problems in Bionics* (Eds. H. L. Oestreicher and D. R. Moore), Gordon and Breach, London (1968).
48. R. Lynn, *Attention Arousal and the Orientation Reaction*, Pergamon Press, Oxford (1966).
49. W. R. Ashby, "Set theoretic concepts of mechanism and homeostasis", in *Automaton Theory and Learning Systems* (Ed. D. J. Stewart), Academic Press, New York (1968).
50. R. L. Gregory, private communication.
51. D. Bannister and M. Mair, *The Evaluation of Personal Constructs*, Academic Press, New York (1968).
52. E. H. Shuford, A. Albert, and E. H. Massengill, "Admissible probability measurement procedures", *Psychomerika*, **31**, No. 2, 125–45 (1966).
53. G. Pask, *An Approach to Cybernetics*, Hutchinson, London (1961).
54. A. R. Luria, *The Role of Speech in the Regulation of Normal and Abnormal Behaviour*, Pergamon Press, Oxford (1961).
55. H. Matturana, "The neurophysiology of cognition", in *Cognitive Studies and Artificial Intelligence Research* (Ed. P. Garvin), Aldine Press, New York (in press).
56. R. C. Atkinson and R. M. Shiffrin, "Human memory, a proposed system and its control processes", *Tech. Rept.*, No. 110, Stanford Univ. (1967).
57. H. von Foerster, "Self-organizing systems and their environments", in *Self-organizing systems* (Eds. M. C. Yovitt and S. Cameron), Pergamon Press, Oxford (1960).
58. W. Brodey and A. Johnson, private communication.
59. D. G. Fleming, M. Mesarovic, and L. Goodman, "Multi-level, ... multi-goal approach to living organisms", in *Neuere Ergebnisse der Kybernetik*, Oldenburg, Berlin (1964).
60. R. L. Gregory, *Eye and Brain*, Weidenfield and Nicholson, London (1966).
61. H. von Foerster, private communication.
62. D. A. Schon, *The Displacement of Concepts*, Tavistock, London (1963).
63. D. Hawkins, "The nature of purpose", in *Purposive Systems* (Eds. H. von Foerster *et al.*), Spartan and MacMillan, New York (1969).
64. M. Toda, "Design for a fungus eater", *Behavioural Sci.*, **7**, 164–83 (1962).
65. G. Pask, "The computer simulated development of a population of automata", *Math. Biosci.*, **4**, 101–27 (1969).
66. G. Pask, "A proposed evolutionary model", in *Principles of Self-organization* (Eds. H. von Foerster and G. Zopf), Pergamon Press, Oxford (1962).
67. G. Pask, "Physical analogues to the growth of a concept", in *Mechanization of Thought Processes*, N.P.L. Symp. Proc., H.M.S.O., London (1959).
68. G. F. Brieske, "The analysis and synthesis of a learning teaching system", *Unpublished Ph.D. Thesis*, Univ. of Illinois (1969). Published as Tech. Report No. 16, Biological Computer Lab., E.E.R.L., University of Illinois,

69. G. Pask, "Adaptive machines", in *Programmed Learning Research* (Eds. F. Breson and M. De Montmallin), Dunod, Paris (1969).
70. G. Pask *et al.*, *Final Sci. Rept.*, *Contr. AFOSR*, No. F61, 052, 6,700,010, Systems Research Ltd. (1969).
71. G.H. Von Wright, *Norm and Action*, Routledge and Kegan Paul, London (1963).
72. J.C. Kotetley and H.T. Hermann, "An approach to formal psychiatry", *Perspectives Biol. Med.*, winter (1967).
73. N. Howard, "The theory of metagames", *Gen. Systems Yearbook*, **11** (1966).
74. S. Gorn, "Ambiguity and paradox in mechanical languages".
75. J.C. Jung, *Septem Sermones ad Mortuos* (first published 1925), Stuart and Watkins, London (1967).
76. G. Spenser Brown, *The Laws of Form*, Allen and Unwin, London (1969).
77. G. Holt, *Final Rept. Inform. System Theory Project*, *Contr. A.F.*, No. 30, 602, 4211, Applied Data Research Inc. (1968).
78. L. Loefgren, "An axiomatic explanation of complete self-reproduction", *Bull. Math. Biophys.*, **30**, No. 3 (1968).

CHAPTER M–3

The past and future of cybernetics in human development

W. GREY WALTER

Burden Neurological Institute, Bristol, U.K.

Summary

The contribution of cybernetics to our self-knowledge and self-control has been almost imperceptible. This is not because there has been no contribution; rather, the influence of cybernetic attitudes has been so subtle and pervasive that it has permeated the whole atmosphere of theory and technique. We must acknowledge that this atmosphere is not uniformly salubrious. In 1947 Wiener admitted that he had only a "very slight hope" that the good effects of cybernetics would "anticipate and outweigh the incidental contribution—to the concentration of power—in the hands of the most unscrupulous". Twenty-two years later we cannot honestly feel more optimistic.

In the Jewish cemetery of the ancient city of Prague, where now only the dead are Jewish, there is the grave of the Rabbi Loew. It was he who made the Golem, a magic robot which uttered prophecies; in Hebrew the word means "embryo". The myth of the Golem is in the long tradition of artificial oracles and super-human creations of human origin which were indeed the embryonic ideas which have found mature embodiment only in the last few decades—the electronic omnipotent idiots we call computers.

On the grave of the Rabbi Loew, around the lion motif on his headstone, are little twists of paper stuck in the crevices, scrawled in all languages, with pleas for help in affairs of the heart and business, in examinations and health, in war and in peace. It was not far from this cemetery that Karel Capek coined the word Robot for his drama, R.U.R., and on the walls of the nearby synagogue are engraved the names of the tens of thousands who died for nothing only a generation ago. Here we have the ingredients of our culture: myth, aspiration, creation, superstition, persecution and genocide.

There is an English tale, too, of Friar Bungay, who made a Brass Head which was to utter a definitive revelation; when the moment came it shouted, "Time is!" and exploded into fragments. Deep in our cultural background is a yearning for superhuman guidance and a lurking despair; there have never been lacking those ready to exploit our hopes of salvation and fears of annihilation. In the past the oracles have been at best romantic fantasies, at worst ambiguous frauds; now we must operate in a world of real artificial soothsayers which will answer our questions with ineluctable logic and impersonal gravity.

The story goes that Gertrude Stein, having asked in her last moments, "What is the answer?" and being told that there was no answer, then asked, "What is the question?" What, indeed, is the question that we should ask our computers? If the discipline, or compound of disciplines, that we call cybernetics, is to help us solve the ancient problems of human society, we must think hard before we frame our questions—silly ones will get even sillier answers. Furthermore, cybernetic systems can be mental servo-amplifiers—they can magnify beyond imagination the force and momentum of a passing thought, so that a prejudice or an inspiration can disturb the balance of a million lives.

Just as the ancient world was overawed by the drama and mystery of their oracles so we are in danger of subjugation to the subtle power and elegance of cybernetic technology. Some of us are learning by painful trial and grievous error to distinguish between the Golem or Brass Head and the true white magic of machines. Some of the messages we slip into the tempting orifices of our mechanical slaves are still as pathetic and pointless as those intended for the shade of the Rabbi Loew and some of the answers are even less significant than "Time is!" Nevertheless, if our species is not wholly evil and doomed by insolence to self-extinction, our new-found powers may be turned to good account before our society joins the Rabbi in his crowded graveyard.

We all know that cybernetics means "steersmanship" and is cognate with Government. I need not go over again the history of how Norbert Wiener re-established the term as a portmanteau for the rules and principles that seem to be common to systems found in animals, men, machines and societies. As a physiologist, however, I feel impelled to mention again that the principle which is often considered central to cybernetic thinking—stabilization by negative feedback—was appreciated by physiologists long before the re-birth of cybernetics. In fact the Russian school of physiology headed by Sechenov over a century ago was concerned mainly with the study of reflexive action in the spinal cord and its modification by influences from higher centres. It was from these studies that emerged the Pavlovian school of conditional reflex investigation which has had the most profound and realistic influence on our notions of behaviour and mentality. I allude to this not merely to claim historical priority but as an illustration of how and why the impact of cybernetics has been so imperceptible. The principles of control by reflex action are neither new nor inherently complex, and physiologists have little to gain by referring to the knee-jerk or the eye-blink as examples of negative feedback. At this level cybernetics has little to offer beyond a few rather awkward neologisms and clumsy models.

Even in the field of engineering the hygienic device upon which the health of our urban culture depends, the humble water-closet with its ball-cock, was developed long before the dawn of cybernetics. There is no better material example of the advantages of automatic control by negative feedback, with economy of consumption and even a fail-safe provision in the overflow pipe. Simple though it is, this domestic reflex is not trivial. How many of us could decide instantly, without a moment's thought, whether the water level would be raised or lowered by bending the float-arm up or down? And what a triumph it would have been for cybernetics if the ball-cock *had* been derived from algebraic servo-analysis! Our houses and vehicles are full of mechanical reflexes—thermostats, automatic gain controls, and ignition timers—but none of these was the invention of a cybernetician and all work well enough to preserve the stability of our daily life.

This is what I mean by the imperceptibility of cybernetic influence. Like M. Jourdain who found he had been speaking prose all his life, we realise that we have been depending on cybernetic devices for generations and, what is more important, that the functions of many of these devices are to replicate and even amplify the functions of our own bodies. The thermostat that turns our furnace on and off saves us the trouble of sweating and shivering, of taking off and putting on clothes, of gathering fuel or storing snow. In delegating our physiological homeostasis to an array of mini-computers we have achieved something we call "comfort", and in this process we have revolutionized society. Part of this revolution may be trivial—

our temperature regulation being mechanized we have a market for deodorants and miniskirts—but others are profound. We could not have held a conference on cybernetics without the aid of the cybernetic devices that have protected us on our journeys and maintained our comfort. I repeat this word because it is the obvious gain from cybernetics. Cybernetics, by relieving us of the necessity to battle for comfort, endows us with increasing power of control and imposes on us the duty of selection and restraint. Comfort is the privilege, and kindness is the obligation of the powerful.

If cybernetics were limited to ensuring our comfort and safety we would have few worries, but we know that there are other more sinister applications. As Wiener urged in his book *The Human Use of Human Beings* we may hope and work for the development of machines which can be the willing and harmless slaves essential for the prosperous cultivation of leisure. But, writing at a time when the ignominious slaughter of a hundred million innocent bystanders is a calculated risk, this, as Wiener had to admit, is a very faint hope. As I write, the story of the greatest human enterprise, the Moon landing, is developing. There is no greater material monument to the vision, patience and intelligence of our species than those modest footprints in the lunar dust. But the vehicles that bore three brave men on their impeccable journey are the direct descendants of the vengeance weapons that battered this city only a generation ago, and the twins of those that could transform our verdant planet into a wasteland more deadly than the Moon. Apart from the stupendous power of the vehicles, the elegant precision of every manoeuvre surpasses the imagination of even the most sanguine science-fiction writer. This delicate management of complex information at enormous speed, as well as the control of energy, derives from the military application of the early computers in World War II and we can see why Heraclitus said, "War is the father of everything."

The irony in the contradictions of cybernetic history is more subtle than the mere forging of ploughshares from swordblades. The most challenging feature of cybernetic philosophy is the notion of *purpose*. A goal-seeking device, be it a water-closet, an artificial animal, a nuclear rocket, or a spaceship must by definition include the specification of a goal. It is true that this goal may be trivial—a course rather than a destination—but as cybernetic technology improves we see the specification of higher-order goals. The artificial system replaced first the helmsman who held the course, now the captain who decides the port of call, and then the owner who determines the pattern of profitability of the whole voyage. From the technical standpoint, we now consider the auto-pilot as a simple device. It is more interesting in the air than on the sea because there it has to deal with three dimensions and critical minimum speeds, but it is still not very difficult to design or build. It operates in a well defined space-time continuum and deals with straightforward logical situations. At the other extreme we can imagine a network of autopilots which would not merely take aircraft off, fly them and land them, but attend to worldwide weather, passenger and cargo manifests, bookings, schedules, and servicing. Here there would arise questions of probability, prediction, preference, and profit, as well as simple corrections of attitude and course. Furthermore, long-term considerations would have to be included in each decision. Computation might show that it would be wasteful to fly a single passenger from London to Boston when there was a space on the New York flight but the effect of such a transfer on the reputation and future earnings of the Airline would be an integral part of the program.

I have developed this fancy for several reasons: first, because one of the important current developments of our age is the integration of data processing with business management, the union of the laboratory with the show-room; secondly, because, as H.G.Wells (who was a student of this College) anticipated in *The Shape of Things to Come*, the unique combination of freedom and discipline in the air is a working model of a high-level technocracy in which stability and flexibility must be maximised with minimum risk and waste; lastly, because the higher-level activities of imagination and prediction are still mainly the prerogative of the human brain and it is these activities which cybernetics has tended to ignore.

The essence of these problems is of course man–machine interaction. Cybernetic systems must be designed so that they replace or surpass human functions, yet remain compatible with human physique and mentality. This applies equally to all aspects of cybernetics: the construction of models—"crystallized hypotheses" as I call them—, the elaboration of automatic control systems, or the perfection of human communication. Cybernetic development is useless and may even be pernicious if it is not "person-centred". This is all very well to say, but to ensure man–machine compatibility we must know at least as much about men as we do about machines, and this is still the weak link in the chain of cybernetic evolution.

In the early years of cybernetics the resemblances between some nervous mechanisms and some computer principles were much discussed. The most apt of these is the way in which sensory and motor signals are coded in terms of impulse frequency in nerve fibres. The constancy of impulse size suggested a comparison with digital machines and it was suggested that even the 10,000,000,000 neurons in the brain might act as simple binary relays. We know now that this comparison is superficial and probably misleading; most brain neurons receive signals from a very large number of sources and the fine terminations through which they interact with other elements display various degrees of activity. This grading of activity provides a sort of consensus of input signals so that the probability of a neuron emitting a signal is contingent on some integral function of the signals it receives. It is difficult to be more precise because the transfer function depends both on space—that is the number and provenance of the inputs—and on time—the period over which the effects of the input signals can summate. The situation is complicated by the intervention of a variety of chemical substances, ranging from the elementary potassium ion to complex organic compounds, as humoral transmitters between nerve elements. These are activated and destroyed at various rates by the operation of complex enzyme systems with bewildering reciprocal couplings. It is these enzyme systems which are the targets for the many drugs that alleviate so much mental distress—and also those that are causing so much social confusion.

It may be needlessly discouraging to think of the vast number of individual neurons in the brain—we know that they are expendable and indeed are continually expended and never replaced. But even if the basic unit is a million neurons there are still 10,000 systems with many degrees of interactive freedom. No wonder we know so little about brain mechanisms—in fact it is surprising we know as much as we do. What little we can be certain of—apart from anatomy and the effects of brain disease—is due to the electrical effects associated with the rise and fall of chemical excitation and inhibition. These effects lend themselves to methods of amplification and computation familiar in other fields and a record of electric brain activity obtained in this way is an electroencephalogram or "E.E.G.".

Now, if we return to the aspects of human behaviour which still seem to be beyond mechanical imitation and which I have called simply imagination and prediction, we can break these down into stages: exploration, classification, storage and comparison. The question then is: can we identify in the brain any signs of how and where these four operations are performed? Perhaps the most interesting of these operations is exploration because it is so familiar as a human character, so rare in any machine. Exploration requires some form of spontaneous activity which can be modified by experience; we were at first puzzled, then delighted, to find forty years ago, that the brain does in fact exhibit profuse spontaneous electrical activity. These intrinsic rhythms and discharges are the main features of the E.E.G. and there is still much discussion about their precise origin and function. My own view is that some of the rhythmic components, particularly the alpha rhythms at about 10 Hz, may be a sort of traffic control system, distributing information around the brain on preferred routes to specific destinations. The controls are themselves controlled by the traffic, like pad-operated traffic lights. This process is probably not a passive or purely automatic one, but also involves active interrogation of data sources, particularly those concerned with vision. Here the mechanisms of internal search extend to the outside world in directing gaze and linking this with attention. The complexity of these mechanisms is reflected in the large proportion of brain substance involved in the control of eye movements and refraction. The near perfection of eye control is achieved by a rather surprising feed-back loop, not from the eye muscles, but from the image itself as registered by the brain. Furthermore, the decision to move the eyes is accompanied by a blanking process so that the scene does not jerk with eye movements. If the eyeball is moved passively, this process does not operate and there is a disagreeable sensation of apparent movement. In cinematographic terms, the gaze does not "pan" but "cuts" from one scene to another, unless the movement is very slow. This is why a rapid "pan" seems so disagreeable and unnatural on the screen.

This simple but essential blanking of optical movement is one of the features in the next stage of higher-level brain function—classification. To match it there is the emphasis on external movements and contrast that is typical of the sensory systems. Our view of the world is corrupt at the input, simplified and smoothed by the peripheral sensors and transmission systems. Thus, attention is attracted by differences and transitions, novelty and associations. These effects can be observed both in psychological experiments—as well, of course, as in everyday experience—and also in the electric responses of the brain. This is where one of the applications of cybernetics has been so helpful; the embodiment of hypotheses—or, more modestly, hunches—about brain function in working models, and the projection from the behaviour of the models back to studies of the living brain. This is a long story which has been told elsewhere, but the outcome is a clear picture of the brain as a computer of statistics. The cerebral classifications of events seems to depend on running estimates of their probabilities. These probabilities depend on the unexpectedness or rarity of occurrence—novelty—and on the combined improbabilities of coupled events—association or contingency. Thus, classification requires storage or memory and, again, we can see clear electric signs of the processes of short-term storage in the brain. The items selected for storage are those satisfying personal criteria of rarity or contingency as related to action and survival. The relation of personality and action to these factors of probability is one of the key problems in brain research today. Considering the brain in the simplest terms as essentially a gambling machine, it is obvious that personal habits and strategies vary over a wide range from extreme caution to wild recklessness. Similarly the search for appropriate action varies from stereotyped reflexes to original adventure. The diversity of "human nature" is one of the central difficulties in such research—and one of its most exciting challenges.

In this survey we can see again a complex reflexive or feedback circuit: exploration→classification→storage→comparison→exploration, each step being governed by transition probabilities depending on personal bias and the sum of experiences. The inclusion of exploration in the loop guarantees novelty, and the statistical nature of each junction together with the accumulation of stored information provides what one may call the cerebral uncertainty principle; that any event which affects the brain will have a different effect next time it occurs.

There is one aspect of systems with closed loops that intrigues me, whether they are artificial or natural. I am not sure whether this has been considered before or even whether it is as fundamental as it seems to me, but it is this. Regulation by reflexive feedback implies a functional purpose—some parameter is stabilized within limits to achieve a degree of stability or homeostasis. But it is not always easy—without prior knowledge or experimental evidence—to decide which aspect of the stability is the functionally important one. A Martian, inspecting a cistern with a ball-cock and—presumably—having no notion of the value or use of water, might suppose that its purpose was to maintain the angle of the ball-cock lever. We of course know the purpose of the simple device, *but we cannot assign a simple causal relation to its components*. The introduction of *purpose* blurs the concept of *causality*. The Martian might declare that the water held the ball-cock up; we would say that the ball-cock keeps the water level down. Both of these are descriptions of purpose, not causality; both are true; neither is a complete description. Now, when we come to consider systems as complex as the interaction of diverse human mentalities with the external world, where there are large numbers of reflexive systems with probabilistic couplings and spontaneous activities it becomes quite impossible to define causal relations. Nevertheless, we may quite legitimately assign an arbitrary purpose or function to such a system, or at least to a part of it.

This seems to me to suggest a valuable application of cybernetics; if we make a machine to replace or assist a human function we can and must define its purpose, whether it is one of the primitive self-regulating governors or a computerized factory. In doing this we are assuming cognizance of the purpose of the human functions we are replacing or improving. By comparing and matching the functions of the artificial and natural systems, we can see whether our assumptions about the nature of the human activity were correct. An example of this procedure is familiar to all of us who use computers. The early comparisons between brains and machines were facile and superficial, as I have mentioned, and we realise now that the operations of a logical machine are fundamentally different from those of the living brain. There is not merely the obvious material difference between the "hardware–software" of the machine and the "wetware" of the brain; nor is the difference in scale as significant now as it was 20 years age. The important difference is that the brain is not a logical machine at all. This sounds like a paradox because logic is literally concerned with words and the human brain is certainly a talking machine. What I mean is that, as far as we can see, the algorithms of the brain are not in the least like a Boolean algebra. It would seem that in developing binary computers we have produced not mere high-speed supplements for our brain mechanisms, but the perfect complements to them; they do rapidly and accurately just what our brains can do only after endless practice, with great effort and many errors.

If, then, we consider the future of cybernetics in human development, in terms of both the individual and of society, we can foresee a number of decisions as to whether the main effort should go into supplementary or complementary systems. A range of supplementary devices with obvious but limited value are prosthetic limbs and sense organs. If the current population trend toward protracted senility is maintained these will become of increasing importance to all of us. This is a favourite theme in science fiction (which has proved a better source of prophecy than "serious" essays). At this frontier, I suppose, cybernetics shades into bionics, the analysis, synthesis, and amplification of living systems. Some of the modern prostheses are little more than mechanized crutches, but we may remember how Long John Silver used his crutch as a spear; once an artificial limb can replace the natural one the metal can outdo the flesh. An artificial hand may as well include a fireproof tool-kit in each finger, a signature stamp instead of a finger print, a mini-pistol or a laser beam. The "Possum" system that operates a typewriter by puffs of breath could be magnified into a mouth organ with which a paralysed patient could serenade a computer as quickly and accurately as a Teletype operator. Adequate supplements may turn by degrees into surprising complements; it is this process that promises to transform our lives, however modest the proximate aims of cybernetic designers.

As predicted by Wiener and many of his disciples, devices intended as simple tools are rapidly becoming key factors in a revolution. With the hindsight of history our responses to this challenge, whether participation or counter-revolt, are more self-conscious than during the Industrial Revolution. We should be better informed and more moderate than the innovators and Luddites of a century ago but I doubt whether we appreciate all the possibilities of triumph and disaster. Even in the utilitarian field of prosthetic design the replacement of a hand or a leg with a "new! improved!" version may require direct coupling to the nervous system and this again may suggest "improvements" at a higher level. We have already played with a system to control a computer directly from the electric waves that precede a voluntary decision, and know that our control would be better if our surgical friends would implant electrodes in suitable parts of our brains. We can imagine quite easily the evolution of a hybrid robot, a cybernetic Centaur or even Minotaur—a man half machine, machine half man.

On the social scale we are already dependent on a vast array of systems which, as I have said, are cybernetic in essence although not always in origin. As well as comfort, communication at a distance and extensive travel are now available to everyone; we should recall how recent this development is. We should remember, too, how delicate and vulnerable these systems are, and how close are the margins of efficiency and disaster when signals traverse many junctions. Wiener pointed this out long ago, but it is worth repeating. No transmission system is perfect; at every junction there is a possibility of failure. If we require say 99% overall performance, with only one chance of failure in a hundred trials, then with 10 junctions each junction must have a risk of failure of less than one in a thousand, assuming that the probability of failure is equal and independent at each junction. Even with a system which is less crucial, in which a performance of say 90% is tolerable (Wiener uses the example of a telephone system in which one would be satisfied if only one call in ten went astray) with only five junctions the chance of failure at each junction must be less than 2%. If, as I have suggested, the junctions in the brain are numerous and probabilistic in this sense, then it may seem surprising that the overall failure rate is no higher than the 10% which is said to be the risk of mental breakdown in Western societies. The reason for the success of the living brain as a cybernetic system is presumably the very high redundancy of the pathways in which the unreliable junctions are interposed. But simple redundancy is not enough. If a communication system is to be protected in this way against failure it is essential that the diverse channels be as nearly as possible independent. They should not, for example, have a common power supply. In the brain, some degree of independence is provided by the network of blood vessels known as the Circle of Willis, at the base of the brain, by collateral circulation and anastomosis. The functional domains in the brain, as far as we can see, are not congruent with the metabolic domains. This arrangement is by no means perfect, however, as anyone who has had a migraine headache can testify, and the heart is one of the few organs that is not paired or duplicated. The great redundancy of brain tissue is no protection against circulatory failure or the effect of general poisons which reduce the statistical efficiency of every junction. If we are to replace or improve natural functions, we must learn the long lessons of organic evolution and guard against the abrupt and unpredictable failure of complex systems in which the risk of disaster is a power function of the number of elements. In the telecommunication systems and airports of big cities we can see already how suddenly this process can operate and how difficult it is to restore normal operation once the limit of tolerance has been passed.

In our social organization we try to guard against breakdown by exploiting the redundancy of a large population in the name of democracy. We know only too well, however, that definitions of democracy differ, and in a rapidly changing world the majority are bound to be mistaken. But which of the many minorities is most likely to have the solutions to fresh problems? Few of us expect or hope for the creation of a contemporary Golem—although we may suspect from the growing political chaos that our leaders may be putting their trust in some arcane oracle. Nevertheless, I feel that somewhere in the future of cybernetic technology may lie the solution to the perpetual paradox of government, the precarious balance of freedom and safety, of self-expression and self-control.

Goethe said, "In the end we all depend on creatures of our making." The crucial question is: shall we be able to make creatures that we can trust better than we trust ourselves, immune to the hazard of central confusion and the social poisons of superstition and bigotry?

CHAPTER M-4

Information flows within co-ordinated systems*

W. ROSS ASHBY

University of Illinois, Urbana, Illinois, U.S.A.

Summary

Every coordinated activity, whether in the movements of a tight-rope walker's limbs, or in the traffic flows of a big city, requires an internal flow of information between the parts being co-ordinated. Once the co-ordination is well defined, the minimal quantity of internal information flow is determined numerically. An example is given to illustrate the principle.

This numerical quantity can be partitioned in various ways, corresponding to the various organizational ways for managing the co-ordination. So one can relate a proposed organization, in city or brain, to its resources for internal communication, to see if they are compatible.

Effects with delay ("memory") can be included in the formulation without essential change. Demands for memory, in co-ordinated activity, can be met in a variety of quantitatively different forms; so a designer can select among them for the most appropriate. An example is given as illustration.

We have brains primarily so that our bodily activities may be co-ordinated: so that our left hand shall act properly in conjunction with our right. Co-ordination and integration have long been recognized in physiology as the brain's highest functions, but cybernetics today is equally concerned with co-ordination in systems of other types. Big cities need co-ordination in their traffic flows; the prevention of smog requires that many preventive and

* The work on which this paper is based was partly supported by contracts AFOSR 70-1865 and OEC-1-7-071213-4557

remedial actions be co-ordinated if one remedy is not to be nullified by another; and in social problems, too, the activities of welfare agencies need co-ordination. The purpose of this paper is to show that all co-ordinations require that information be transmitted within the system (a proposition that might be thought obvious), but particularly to show that the transmissions can be measured quantitatively. Every well-defined co-ordination specifies a basic total quantity of transmission, such that less than this quantity makes absolutely impossible the achievement of the co-ordination. It will also show that this total quantity can be analysed (partitioned) in various ways so that we can see how much is required between the components. In regulating traffic flow, for instance, it would show how much transmission is required between point and point. In the brain, it would show how much transmission is required between cell and cell, or between centre and centre.

Co-ordination is essentially a holistic phenomenon, discernible only over the whole. The method of information analysis presented here is also of this type. It does not say that between points X_i and X_j so much transmission must occur: it treats all the transmissions as a complex interrelated set, and allows the transmission between (say) X_i and X_j to take almost any value provided that suitable adjustments are made in the other internal transmissions.

The method could be presented in formal and abstract symbols, leaving the reader to find their application. I prefer to present one example, perhaps oversimplified, to show the method at work. The reader should have little difficulty in adapting the example to his own needs. The example is artificial; I would have preferred to have analysed real data, but it seems that no one has yet collected data on co-ordination with sufficient breadth to make this type of analysis possible. Perhaps when the existence of this method is better known, the experimenters will supply appropriate data.

THE TIGHT-ROPE WALKER

As example, let us consider that classic type of co-ordination shown by the tight-rope walker. The focal condition (Sommerhoff, 1950) is obviously that his four limbs must always have positions such that their centre of gravity lies vertically over the wire. (To keep the example simple I here ignore such complications as their angular momenta.) The unskilled person may well be able to move his limbs through just as wide a range as the expert, but the unskilled person will use combinations of position, all four limbs to the left say, that the expert would avoid. Thus the contrast between the unskilled and the expert may be shown by the fact that the expert confines his actions to a particular subset of those anatomically possible.

The suggestion is not, of course, derived simply from tight-rope walking. As Sommerhoff (1950) showed extensively in biological examples, and as Ashby (1967) showed in the terms of set theory and binary relations, the identification of "co-ordination" with "deviations from statistical independence in an n-dimensional frequency table" is both broad and rigorous. Given any well-defined co-ordination between n variables, there is implied a frequency distribution over events in the n-dimensional space to which Shannon-type measures of information are applicable.

It is simpler now to proceed by example.

To avoid infinitesimals, suppose each of the four limbs can go to one of five places situated at distances -2, -1, -0, $+1$, $+2$ from the central plane. Thus, if the limbs L_1, L_2, L_3, L_4 are at (respectively) -1, $+2$, -1, $+1$, the average is $+0.25$, and their center of gravity is away from the central plane. If we allow two or more limbs to be at the same distance, there are 5^4 possible distributions (postures) of which only a subset would be used by the expert. It is easily verified that of these 625 postures, 85 have the zero average of the co-ordinated posture (61 in the 6 types of symmetrical distribution such as 00400, 01210, etc.; and 12 each in the asymmetrical 10120 and its reflection).

To obtain the necessary frequencies (or probabilities after dividing by the total) we may proceed on either of two assumptions (that lead, in fact, to just the same numerical results). One way is to assume that the 625 postures of the unskilled and the 85 of the expert are actually equiprobable, a very arbitrary assumption that may well be false when we consider actual people. The other way is to think of the case where the facilities for transmission will have to be provided, and to ask: suppose the worst happens—that without transmission all 625 postures occur, and that the expert (for other reasons) may be forced to produce all the 85: what is the least quantity of transmission facility that we must provide to be safe? This second form of the question seems to be free from objection (unlike the first), so I shall treat it as the question to be put.

With the frequencies now assumed to be equal (or the probabilities after dividing by 85) we can now find the basic entropies. These are defined in the usual way, by

$$H(X) = \sum_i p_i \log\left(\frac{1}{p_i}\right)$$

As we shall be using frequencies here, however, an arithmetically more convenient method, if the frequencies are $n_1, n_2, \ldots, n_i, \ldots$, summing to n, is to find $H(X)$ as

$$\frac{1}{n}\left(n \log n - \sum_i n_i \log n_i\right)$$

(With the n_i all integers, much interpolation may be avoided.)

In the general case, these entropies would be found by whatever method was appropriate. In this example we can soon find that L_i has the following frequency distribution, in the co-ordinated postures.

Value:	−2	−1	0	+1	+2	
Frequency:	15	18	19	18	15	Total: 85

So $H(L_i) = 2.315$ bits/posture. By symmetry, this is also the value of $H(L_2)$ etc.

$\langle L_1 L_2 \rangle$ has the following distribution, over its 25 values.

L_2:	+2	5	4	3	2	1	15
	+1	4	5	4	3	2	18
	0	3	4	5	4	3	19
	−1	2	3	4	5	4	18
	−2	1	2	3	4	5	15
		−2	−1	0	1	2	
				L_1			

So $H(L_1, L_2) = 4.544$ bits/posture. All the 85 values of $\langle L_1 L_2 L_3 \rangle$ are different, so $H(L_1 L_2 L_3) = \log_2 85 = 6.409$. Similarly $H(L_1 L_2 L_3 L_4) = 6.409$ bits/posture. If a posture is significant over a time span of (say) 0.5 seconds, then twice these numbers would give the entropies in bits per second.

PARTITIONING THE INFORMATION FLOW

The further analysis uses the methods introduced by McGill (1954) and developed by Garner (1962) and Ashby (1965, 1969). The most important quantity required now is the total transmission, represented and defined by

$$T(L_1 : L_2 : L_3 : L_4) = H(L_1) + H(L_2) + H(L_3) + H(L_4) - H(L_1 L_2 L_3 L_4)$$

It measures the total deviation from statistical independence implied by the co-ordination (with the marginal distributions given). Here its value is

2.850 bits/posture. Its importance is due to the fact that with less than this total quantity of internal transmission the co-ordination cannot be ensured.

It is worth noticing that the total transmission required is not the obvious $\log_2 625 - \log_2 85$ (= 2.878) but a quantity smaller by 0.028. The reason is that the larger quantity would apply were each variable L_i to be distributed evenly over the five values. In fact the distribution (in the co-ordinated case) is not even. Thus, if the variable's distribution is changed from 17, 17, 17, 17, 17 to 15, 18, 19, 18, 15, the change would bring the conjoint 4-variable distribution nearer to the co-ordinated form *without any transmission being used between the variables*. Thus the algebraic and numerical analysis has already revealed a possibility for economy and efficiency that otherwise might have passed unnoticed. (In this example the gain is trivial; in other cases it might be of major importance.)

The total quantity of transmission required may be obtained by adding various components. One possible way is to use the fact that $T(L_1 : L_2 : L_3 : L_4)$ is identically equal to

$$T(L_1 : L_2) + T(L_3 : L_4) + T(L_1L_2 : L_3L_4)$$

Such a partition would be appropriate if the total co-ordination were achieved by mechanisms or channels that (1) achieved suitable co-ordination between L_1 and L_2 (between the arms, say) regardless of the positions of the legs, (2) achieved co-ordination between the legs regardless of the arms, and (3) co-ordinated arms and legs in a way not depending on the details of the relation *between* the arms (e.g. if the arm-pair has centre of gravity at +0.5 then the leg-pair must have centre of gravity at −0.5). The three quantities are found to be

0.086, 0.086, and 2.678

(respectively), summing to 2.850, of course.

Such numbers may be useful in various ways. Thus, suppose that only 2-bit channels were available. Instead of taking two such channels to achieve the 2.678, we could try another way of distributing the transmissions. Another way is represented by the partition (of the total) to

$$T(L_1 : L_2) + T(L_1L_2 : L_3) + T(L_1L_2L_3 : L_4)$$

This partition would be appropriate if the co-ordination were achieved by first a constraint holding between L_1 and L_2; second, by the outcome of this constraint (the vector $\langle L_1 L_2 \rangle$) acting to constrain L_3; and then the consequent $\langle L_1 L_2 L_3 \rangle$ acting to constrain L_4.

The quantities required are (respectively)

$$0.086,\ 0.449 \text{ and } 2.315$$

still excessive in the last quantity. We have also, however, that this last quantity may be partitioned further:

$$T(L_1 L_2 L_3 : L_4) = T(L_1 L_2 : L_4) + T_{L_1 L_2}(L_3 : L_4)$$

$$2.315 = 0.449 + 1.866$$

Thus the requirement could be met by an extra channel of capacity 0.449, together with one whose average capacity is 1.866, linking L_3 and L_4; with the coding between them determined conjointly by L_1 and L_2.

Here I have written as a designer might see the matter, and use the equations to guide the design. The physiologist might use them if, say, he knew that no channel of more than 2.000 bits/posture was neuronically possible. Then the analysis would show decisively that any proposed neuronic arrangement using the first mode must be rejected: such a net could not achieve the observed co-ordination.

The coding question will not be treated in this paper as it is still being studied. By Shannon and Weaver's (1949) theorems, the necessary codings will certainly exist, but the theorems assume that successive acts of co-ordination (successive postures here) can borrow signalling capacity in order to make up an efficient code. If this cannot be done, then the actual capacities required may be somewhat higher than the numbers given here. Further consideration of coding can be given only when more details of the particular case are available.

MEMORY

In the co-ordination just described, it has been assumed that the variables specify the positions of the four limbs taken simultaneously. Exactly the same logic, and the same algebraic method hold good when the co-ordination occurs over time: when later events must be co-ordinated with earlier. $H(X, Y)$ may be the entropy of two distant events taken simultaneously, but it is equally possible that X and Y are separated only in time, so that

$X = Z(t)$ say, and $Y = Z(t + k)$. Now, if the system is to co-ordinate X and Y, it must have "memory", in some form, over the time span k. An example will show the method and something of the possibilities. Again it is artificial, for lack of presently existing real data.

Let us suppose that three unmanned vehicles will be landed on a planet, which has five places of interest. It is required that the three vehicles shall (1) at one time go to some three of the five places (no two vehicles to the same place), and (2) at another time, meet, all three, at a place other than those visited singly. (Events (1) and (2) may occur in either order.) And it is required that the co-ordination's demands on memory shall be minimal.

The computations are straightforward. We prepare for the worst case, where all events and distributions are equiprobable. Let the five places be $\{1, 2, 3, 4, 5\}$, and the three vehicles $\{A, B, C\}$. Let A, B, C represent their places on the first occasion in real time (regardless of whether event 1 or 2 is achieved), and A', B', C' their places on the later occasion. Thus the vector A, B, C, A', B', C' would show the defined co-ordination if its value were $\langle 4, 4, 4. 5, 2, 1\rangle$ or $\langle 2, 5, 3, 1, 1, 1\rangle$, and other similar combinations.

The basic entropies, in the "co-ordinated" case, are easily found.

(1) The 5 values of A occur all with frequency 48, so $H(A) = \log_2 5 = 2.322$. Similarly for $H(B), \ldots, H(C')$.

(2) The 20 permitted values of AA' occur all with frequency 12, so $H(A, A') = 4.322$; similarly for $H(B, B')$ and $H(C, C')$.

(3) Of ABC, 5 values (event 2) occur each with frequency 24, and 60 values (event 1) each occur twice. So $H(A, B, C) = 5.114 = H(A', B', C')$.

(4) The 240 permitted values of $ABCA'B'C'$ each occur once; so $H(A, B, C, A', B', C') = 7.907$.

The units are bits per double event.

One obvious way of organizing the system is to co-ordinate within each event at each of the two times, and also to co-ordinate between the two times. The total transmission required over *both* distributions of vehicles, is 6.025 bits, analysed to

$$\begin{array}{ccccc} T(A:B:C) & + & T(A':B':C') & + & T(ABC:A'B'C') \\ 1.851 & + & 1.851 & + & 2.322 \end{array}$$

Another method of organizing to achieve the co-ordination, would be to consider the "trajectory" (or transition) taken by each vehicle, as A might go $4 \to 5$, B $4 \to 2$, and C $4 \to 1$, and then co-ordinate the trajectories. This would demand the quantities

$$\underbrace{\begin{matrix} T(A:A') & + & T(B:B') & + & T(C:C') \\ 0.322 & + & 0.322 & + & 0.322 \end{matrix}}_{0.966} \begin{matrix} + & T(AA':BB':CC') \\ + & 5.059 \end{matrix}$$

The term $T(A:A')$ represents "memory" affecting only vehicle A, regardless of what the other vehicles do; and the same for vehicles B and C. What is striking is that the *three* "memories" of this type demand only 0.966 bits as compared with 2.322 bits demanded by the single, and more obvious, first type. The method thus enables different *functional* forms of "memory" to be examined for various characteristics.

One would, of course, also have to consider the physical method used to achieve the co-ordination between transitions, $T(AA':BB':CC')$. It is sufficient for us here to notice that these numerical analyses refer only to deviations from statistical independence in their quantities, not to any reasons, or physical causes, for the deviations. Thus any quantity T, here called "transmission", does not necessarily need an engineer's communication channel: suitably paired responses to a common signal may well provide the formal "transmission" demanded by these identities.

The coding problem remains, but I am content if I have shown that the fundamental concepts of co-ordination and integration can be measured, and that the measurements may give information about the system that is much deeper than can be obtained by simple intuition.

References

Ashby, W.R. (1965). "Measuring the internal informational exchange in a system". *Cybernetica*, **8**, 5–22.

Ashby, W.R. (1967). "The set theory of mechanism and homeostasis". In *Automaton Theory and Learning Systems* (Ed. D.J.Stewart. Academic Press, London. pp.23–51.)

Ashby, W.R. (1969). "Two tables of identities". *Bull. Am. Soc. Cybernetics.*

Garner, W.R. (1962). *Uncertainty and Structure as Psychological Concepts*. Wiley, New York.

McGill, W.J. (1954). "Multivariate information transmission". *Psychometrica*, **19**, 97–116.

Shannon, C.E., and W.Weaver (1949). *The Mathematical Theory of Communication.* University of Illinois Press, Urbana.

Sommerhoff, G. (1950). *Analytical Biology*. Oxford University Press, London.

CHAPTER M-5

*The cybernetic cytoblast: management itself**

STAFFORD BEER†

London, U.K.

Summary

The interdisciplinary cybernetic science of control is many-footed. It has a foot in the brain, and a foot in the computer; a foot in ecology, and a foot in economics; it deals alike in social facts and engineering artifacts. This is because cybernetics had demonstrated the generality of so many subtle control mechanisms in viable systems of every kind. If all these things are thereby connected, and we say that they are, where is the central nub? It turns out to be whatever we can identify as managerial principle amid a welter of control manifestations. And certainly any such managerial principle has little to do with the "principles of management" as we hear them discussed by the pundits. Management itself, indeed, and in whatever context is the cytoblast: the nucleus of the cell. It is the organizing principle by which viable systems evolve and grow, learn and adapt. How far have we gone towards metabolizing our knowledge of cybernetics within the cell of the business enterprise or of government? We have not even begun.

* Chairman's address to Congress.

† Stafford Beer is Visiting Professor of Cybernetics at Manchester, President of the Operational Research Society, executive member of the UK Automation Council, and Lanchester Prizewinner for his book *Decision and Control.* He is reverting to consultant status with IPC, where he has been Development Director since 1966, to extend his other activities for government and industry.

SETTING THE SCENE

According to the official trade figures for July 1969, the country's exports were the second highest they have ever been. Moreover, the level of imports fell as compared with June. This was not so gratifying as it might have been, because the trade gap increased by twelve million pounds.

However, the total import–export value (which must be measured to assess this fluctuation) is two orders of magnitude greater than the movement itself. The difficulties in measuring it at all are notorious. Indeed it was recently discovered that sums of this same magnitude had been "lost" from the national accounts every month for the last six years. And in any case, the fluctuation was less than half the average for the year.

What was the outcome? I quote the *London Times*. "The trade figures shocked the Stock Exchange so badly that gilt-edged prices suffered their biggest fall for any normal day, excluding only days of such high drama as devaluation or the outbreak of war."

This story is enough to make a control engineer weep: he would not reckon to design a servomechanism which operated like a steam hammer in response to changes occurring *well within the noise level* of his system. As for the cybernetician, the story is enough to make him beat his breast in public. For cybernetics has more to say about the management of large and complex systems than to attempt their control in terms of a single-loop negative feedback in the first place.

We people have been studying complex systems of every kind for a quarter of a century now. There are neuro-cyberneticians here today, who have kept us profoundly informed of progress in brain research. We have learned from ecology and from genetics; we have dealt alike in social facts and engineering artifacts. All manner of cyberneticians have been closely identified with the development of the computer from the very beginning. What we hope we have done is to demonstrate the generality of many subtle control mechanisms in viable systems of every kind.

If there is a science of cybernetics at all, is has to be thus general. We are surely seeking to identify *managerial principle* amid a welter of control manifestations. For management itself, however this may be defined in the context of a particular complex system, must enshrine the organizing principle by which any viable system may evolve and grow, learn and adapt. It is here in the cytoblast, the nucleus of the cell, that we must needs identify whatever matters about the direction of affairs.

My case is this. If we have spent a quarter of a century in what is essentially research, it is time to turn towards influencing affairs. We may by now take for granted a great deal that we know we know. From Shannon's theorems of communication to Ashby's laws of variety, from McCulloch's neural logics to Pask's principles of self-organization, we have a corpus of theory. From the underexploited experimental mechanisms of Grey Walter to the overexploited digital computer, with analogue machines intervening, we have a considerable mastery of mechanisms. We have models galore, which actually work, from the industrial and urban dynamics of Forrester to the artificial intelligence algorithms and heuristics of Minsky, Michie, and Amosov. Conscious of this armoury, and encouraged by the penetrating insights of many other workers, it is time I repeat for a conference such as this to adopt a stance in relation to the societary problems of our time.

For it seems no exaggeration to say that society is in a state of crisis. There are many deeply disturbing symptoms of this, widely beyond the scope of such anecdotes as that with which I began. There is nothing less than massive revolt against established norms by the young everywhere in the world. There is everywhere a growing manifestation of violence and disorder. An ecological crisis is upon us—in terms of pollution: by pesticides, by noise, by carbon monoxide. There is an exploding population, an exploding technology, and in general a change in the rate of change beyond our ability to grasp. Society is faced with major perturbations arriving with greater mean frequency than the relaxation time of any social homeostat devised by man.

If all this sounds alarmist, i t is because we ought to be alarmed. And in my view the most alarming feature of the whole situation is a kind of paralysis in the alarm machinery itself. Modern communications have procured a rate of reflex in the body politic which amounts to clonus: and I have elsewhere[1] sought to show how clonic reactions make society spastic, causing hunting—and even the *reversal* of what were hitherto regarded as causatory mechanisms. But without turning to pathological models, it should be noted that the perceptual battering sponsored by the mass communications media of today leads assuredly to habitation. We have produced a situation where the continual restatement of profoundly shocking truths leaves them shocking no longer. And therefore there seems no way left in which to draw attention to what are truly lethal threats.

Other sorts of scientist are doing their best to draw attention to various kinds of environmental, political, and economic crisis. The particular form of societary crisis which cyberneticians understand and are qualified to speak about concerns the growing instability of social institutions themselves. For we know that as society becomes unstable the mechanisms of viability themselves will be denatured. We can expect explosive rather than damped responses, uproar in place of homeostatic equilibrium, a loss of adaptability, and so forth. The symptoms are there.

How might we approach a general cybernetic problem of this magnitude? The future of society is a topic to which thinkers have always devoted effort, but not with much sense of urgency. Scientists and administrators alike have been more humble, and have addressed themselves to the particular problems of subsystems. But, if cybernetics has the tools, and if the urgency is there, we ought now to be developing a general systems model which would illuminate the whole scene. One paper such as this cannot take us very far, but I shall do what I can to start the development which appears to me so vital.

THE ESOTERIC BOX

Think of any social institution you like, and imagine it to sit within a box. You might take, for example, "schooling" or "the general practice of medicine" of "penal establishments". What is going on inside this box is an established order of things. There is a complex arrangement of subsystems, a strange set of relationships between people of standing, and a recondite mode of behaviour. These features—their complexity and unintelligibility to the outsider—justify the box's adjective "esoteric". One does not gain admission to the box's activity without the appropriate passport.

This is not to say that the box is a closed system, only that it is self-organizing and self-regulating. It has inputs and outputs, which are usually a stream of members of the general public. These people pass through the box, or are operated on by the box, but they do not change it at all. Whatever else happens, the box goes on; it is very powerfully arranged to maintain its own internal stability—and therefore, indeed, its survival. How can this be?

The social institution, the esoteric box, is an ultrastable system in equilibrium. The stable state at any time is determined by a set of systemic variables. Any change in one of these variables causes a shift in the position of equilibrium to a position that tends to counteract the change that has happened. That this should be so is an obvious consequence of the equilibrium-seeking character of the system. In physical chemistry, this is called *Le Chatelier's principle.*

Le Chatelier suggests to us how an esoteric box in equilibrium is likely to react to any kind of prodding from outside. It does this not by "fighting back", but by moving its point of equilibrium along the scale on which that point may be measured, so as exactly to counterbalance the efiect of the prod. The result of this is, of course, that if you physically assault the esoteric box by changing one of its state variables, it neither collapses nor violently reacts: it simply adjusts its own internal relationships slightly so that the effect of the change is offset. In behavioural terms, this response is recognizably an ultrastability reaction (under Ross Ashby's terminology).

Reformers, critics of institutions, consultants in innovation, people in short who "want to get something done", often fail to see this point. They cannot understand why their strictures, advice, or demands do not result in effective change. They expect either to achieve a measure of success in their own terms or to be flung off the premises. But an ultrastable system, especially one enclosed in an esoteric box, has no need to react in either of these ways. It specializes in equilibrial readjustment, which is to the observer a secret form of change requiring no actual alteration in the macro-systemic characteristics that he is trying to do something about. As to beating off the attack, that would be crude indeed. Only systems in unstable equilibrium need to do that, and our society is now more mature. Reformers are not often sacked, still less burned at the stake, these days. This is a measure of the extent to which institutions have developed self-sustaining techniques of ultrastability.

But it is interesting that, just because reformers do not achieve success in their own terms, they may not as a matter of fact have failed to reform. Let us learn something more from the operation of Le Chatelier's principle. It provides one very extraordinary limiting effect. Suppose, for example, that heat is evolved in a chemical reaction. Then suppose that the prodding involves increasing the temperature. The system's response to this kind of prod (depending on the actual quantities involved) may have to result in *reversing* the reaction. The system maintains its equilibrium, and the esoteric box may look much the same from outside. But the extent of the hidden drama within is not realized. Even so, the system *has* maintained its integrity; and if necessary reform really does require that this particular system should be abolished, or should merge with other systems into new social configurations, then we are in serious difficulty.

In general, cybernetic models of social institutions predict three characteristic outcomes of their ultrastable response to serious perturbation. They are as follows.

First: the institution survives as an integral system; it is still recognizable as the esoteric box.

Second: the institution maintains equilibrial relationships with other institutions, although their form may change. (Indeed, the elaborate forms of the relationships between two esoteric boxes may justify their description as a third esoteric box to which these three remarks would then also apply.)

Third: realignments *within* the esoteric box required to achieve this integral survival of the institution in society may well be extremely dramatic. They may in particular involve actual reversals of role internally, in the specific conditions where that role is defined in terms of the variable most responsive to environmental change.

Testing these conclusions against observed events, I offer four very different examples for consideration, without further detailed discussion. They are (i) the institution of management in any business firm which comes under extreme pressure; (ii) the institution of marriage—and the family in Western culture since 1900; (iii) the institution of the English Public Schools since the first Labour Government in 1929; (iv) the institution of the Roman Catholic Church since the election of Pope John XXIII. Then this is the ultrastable system-for-surviving, and this is what we ought to expect from the social institution as an esoteric box. No system can survive an assault so massive that it is plucked right out of its setting. No system survives environmental perturbations which count as annihilating. Up to this final degree of disturbance, however, ultrastable systems are machines-for-surviving—and they do indeed survive.

Perhaps at this point we should briefly consider the mechanisms within the esoteric box which procure the behaviour under discussion. They must surely be social mechanisms, and we at first seem to be in the realm of social psychology. We are dealing with a professional or a managerial group which is self-organizing, which shares a belief structure and a complicated set of conventions—especially as to appropriate ways of describing situations and appropriate methods of debate. Any management scientist who has worked in an experimental situation with such a group could tell his own stories about the tenacity with which that group coheres—and adheres to its corporately stable states.

For example, I was administering a management game in which, half-way through, it was discovered that one of the teams had been fed a large amount of false information. As a consequence, though by coincidence, they had ranked the profitability of four new products in exactly the reverse order from the solution we knew to be correct at this stage of the game. I therefore told this team what had happened, and apologised. I explained that, in order to keep the game going, the best thing to do appeared to correct their ordering of the products to show correct current estimates of profitability, and to carry on from there—although obviously this would give them an advantage over other teams. They thanked me politely for my intervention, but declared that they were highly satisfied with the ways things were going, and would prefer to ignore my advice. They carried on as if nothing had happened.

The fact is that we might do better to think in terms, not of psychology but of social anthropology. It might be more helpful to talk about tribal customs than about logical analysis, about myths and taboos rather than understanding and intellection, about conversion rather than conviction. This is not rudely to say that there is no rational content in policy making and decision taking within the esoteric box: of course there is, and of higher or lower quality. We are talking here simply about those potent mechanisms which result in the group's behaving in an ultrastable fashion.

There are lessons in this, no doubt, for people who wish to influence the group, particularly when such people do not belong to the tribe. But this sort of discussion is not our present concern. It is our concern to appreciate the strength of the binding forces and the power of the interlocking feedback loops involved. I would like to emphasize that these remarks are not made in a critical spirit: these are not matters for rebuke. It is important to society that its social institutions *do* cohere, *are* ultrastable. For although these characteristics may make life difficult for reformers, they are what gives society its strength as well. We should quickly be in chaos if our institutions toppled under pressure from maniacs or anarchists, or even through the incompetence of individuals concerned with operating the system.

STRINGS OF ESOTERIC BOXES

Now most definable activities which matter in modern society may be described as a collection of esoteric boxes connected together under some rule, and thereby constituting a social metasystem.

For example, the social activity of "education" might be depicted as a chain of three esoteric boxes, respectively called "the school", "the university", and "the employer". Formal education passes a human individual through each of these esoteric boxes, and within each an operation is performed on that individual. We can at once understand what sorts of statement count as rules for the connection of these boxes. In England, "passing the 11-plus examination" constituted the rule opening the door into the box labelled "grammar school". The label "x O-levels and y A-levels" works the filter between the school and the university. "A good-class degree" is the name of the filter operating between the university and the professional career. The stringing together of esoteric boxes by rules of this kinds marks the present state of social organization.

We may note at once that, if the established filters turn out to be unworkable or to promote a total social behaviour which is unsatisfactory, then society characteristically tries to change the filter. "Do away with the 11-plus": that has almost happened, and other criteria are being substituted. "Is the A-level a suitable criterion for university admission?": that has become a question of moment. "What does industry expect from a university graduate?" is an enquiry calling the last kind of filter in question also. But whatever we do about the filters, which is to say, however we manipulate the rules by which the esoteric boxes are connected, the effect of momentous social change on the integrity of the esoteric boxes themselves is almost non-existent. Thank Le Chatelier for that.

We may look at a variety of examples, and begin with education—since that was our first topic. It is evident, I should argue, that the educational system grossly fails either to meet the contemporary social need or to exploit what is now technologically possible. It follows that there is much adjustment of the filters between the three boxes already nominated; and (because educationalists are sensible people) there is also much discussion of the *interfaces* between the boxes. Schools are developing a new concept of the sixth form which reaches out towards the university. And universities, for their part, are developing concepts of a "foundation year" which reach back towards the sixth form. Here is a sensible attempt to bridge the gap between the two esoteric boxes of "school" and "university", without damaging the integrity of either. On the other side of the diagram the university reaches out towards the employer, by instituting "vocational courses" aimed at preparing people for their future life in paid employment. The employer, in turn, has instituted training schemes for graduates, in such fields as research and development, or management induction. These are explicitly attempts to bridge the second gap without compromising either the university's purity of purpose in the pursuit of knowledge itself, or the business intention to make a profit.

Government supervises all these endeavours with benign benevolence, setting up bodies to facilitate the operation of new filters—by, for instance, giving grants to students in the school–university transition and by endowing, for example, business schools to facilitate the university career transition. All this is very fine. But the ghost at this particular banquet is still the ghost of Le Chatelier. Do not expect the schools themselves, the universities themselves, or industry itself, actually to *change*. All these esoteric boxes will do is to modify their equilibrial state to account for the impact of societary manoeuvres which affect their major variables.

Let us take some further examples, in which the situation is perhaps less enlightened by a wish on the part of all concerned to produce the optimal synergistic result.

Three esoteric boxes have, since the advent of the aeroplane, between them constituted the national capability for defence. They are of course the institutions called the army, the navy and the air force. Now it has been clear to all concerned for a very long time that any modern military operation is an amalgam of these three; because operations are conducted on land, by sea and in the air *at once*. This joint operation is called a battle. Moreover, any scenario prepared to discuss the future is bound to treat of situations characterized in these three dimensions simultaneously. In this case a still bolder solution has been adopted than the gap-bridging solution used in education. It is not simply to adjust the interfaces between the three esoteric boxes, but to declare (since all three services come under the command of one government which in turn represents one sovereign) that military might is indivisible. Then let there be simply a "green uniform" force amalgamating the national fighting prowess by land, sea and air.

The British Government, following American precedent, took some kind of step in this direction. Notionally, the armed forces are coalesced into an integral defence capability, under a Minister of Defence (political) and a General Staff (services). Yet a little observation suggests that the Le Chatelier family still has senior officers serving in three different uniforms. And we may note little change in those three extremely ultrastable institutions called the army, the navy and the air force.

From this example, we may escalate to government itself. Now we shall call our three esoteric boxes: first ministry, second ministry, third ministry.

Why are these three ministries quite separate? Do not they all bear upon the social need of the citizen? Of course they do. And therefore we now have (as a result of the Fulton Committee) a new Civil Service Department competent to cross the boundaries between them. There is no doubt at all, in my mind, that this development will produce great and important changes. Even so, I suspect that the three computer installations (for example) required to administer these three ministries were ordered quite separately, on the authority of three different officials (even if they did have the same surname, which we know by now). I mean quite simply that ministries one, two, and three are ultrastable, and will be with us for some time yet.

If we move from government examples into industry, the whole business becomes yet more dramatic. *Obviously* we could not have our energy requirements met by (i) coal, (ii) gas, and (iii) electricity—without regard to their interaction... And so, by nationalizing each of the industries concerned, and by making all three responsible to the Minister of Fuel and Power, we should obtain as a nation synergistic benefits. And so we should have done. But it is not as easy as that. Occasional high-level meetings between three major industries will not produce an integral operation. And in our kind of economy competition must be kept alive. Then there is the classic contention that one ought not to impose a plan of national scope on an industrial management which is to be held *accountable*. This is exactly the same argument as one meets in the large corporation which finds corporate planning impossible because its component firms are called autonomous.

Many of the arguments used in this area of organizational thinking are confused and muddle-headed. Often they discuss bogus problems. We may nonetheless predict with confidence that the debate will be kept very much alive. For so long as the issues are kept nicely confused they are grist to the ultrastability mill of any self-respecting esoteric box.

Then what about the industries which have not been nationalized, which depend upon a free market to adjust the relationships which exist between their esoteric boxes? The extent of their collaboration is notoriously inadequate; and therefore government must intervene to ensure that good sense prevails for the national well-being. The industrial training boards were one fruit of this thinking; the Science Research Council, the National Research and Development Corporation, the Industrial Reorganization Council, and other government-sponsored bodies (having call on national resources precisely in order to lubricate the interfaces) have been set up. But the esoteric boxes go their self-organizing way. Who expects management to change? Who expects the share-holders or the City to adopt new postures? Is it likely that trade unions will adjust to the new challenge? Too many managing directors, too many dividend seekers, too many general secretaries operate under the Le Chatelier alias.

And so the argument goes on. We might finally consider an entire industrial complex, for example publishing, as illustrating the point. Until recently, publishers used to publish entirely through the printed word. It has recently become clear that technology has moved on since Gutenberg and Caxton: we are in an era of electronics. Thus the three esoteric boxes which will determine the future of the publishing industry in the next ten years are: the publishers, the telecommunications and computer industries, and the G.P.O. Yet the chances of changing the interfaces and the dynamic relationships between these three esoteric boxes appear slight indeed. Each will adjust its own equilibrium point to the new challenge, however each may see that challenge; what we can be sure of is the inability of the total system to restructure itself. This is because there *is* no total system, except in the eye of beholders. In the end, it is the beholder who makes take-over bids on the strength of his metasystemic perception who changes this aspect of society.

These are six examples of strings of esoteric boxes: I have compounded the string of three elements in each case, simply for convenience, and have certainly not carried the discussion of any one example to the extent of making it a case history. But perhaps our minds are sufficiently prepared by this quick review of education, defence, the civil service, the national energy problem, industry generally, and one industry in particular, to be ready for the general argument which must now be pursued. It is already obvious that we cannot sack Le Chatelier from any of the esoteric boxes over whose ultrastability he presides. Nor would we wish to do so. But we cannot leave him to operate within esoteric boxes as the last lone survivor of an antedeluvian age. We need him badly on the national payroll.

THE SPURIOUS METASYSTEM

When we first began to talk about stringing esoteric boxes together, I said that they thereby constituted a social metasystem. This means essentially that the total system of which (arbitrarily) three esoteric boxes are the parts, is something more than their sum. "Something more" again essentially, refers to what, following Hegel, could be called a higher-order synthesis. This speaks a language capable of utterances which cannot be expressed in the languages of the esoteric boxes themselves—a language which, following Gödel, could be called a metalanguage. A metalanguage is one in which we may decide issues which are undecidable in the lower order languages; and the metasystem ought to be a social institution capable of surviving when its component institutions (despite their ultrastability) finally fail.

All this, for the cybernetician, says much more than we have so far said in talking about "a string of esoteric boxes". A string of pearls is stable, but not ultrastable. If the cord is cut, nothing is left but the individual pearls—and they are scattered on the floor. What we need is a viable system of esoteric boxes, more cleverly connected, more elaborately structured, than happens simply by talking about them in the same breath. Let us firstly attempt to abstract from the six examples so briefly limned out the rules of the metasystemic game as it is played by society today.

The boxes of our examples, just because they are esoteric, are initially free agents. They can ignore each other; or they can recognize that they belong to some kind of metasystem—and in this case they have an alternative. Either they can act responsibly to their own initiates, which may well mean repelling contiguous boxes with which they ought to have transactions; or they can act responsibly towards society at large, and undertake rapprochements with their neighbours. Thus we approach a first rule of observed behaviour.

First rule: Esoteric boxes may recognize themselves as part of a higher-order system, and this affects their behaviour at the interface with other boxes. They may seek or rebuff sensible collaboration.

Social institutions, however, are each a minor aspect of society at large, and we may safely say that there is always a superior authority which has the capacity to influence the situation, or even to intervene in a definitive way. In the case of public institutions, this superior authority is probably government itself. Where private institutions are concerned, no one can be sure (and this is part of the fun of the private sector) who the superior authority may turn out to be. A trade association may suddenly exert itself; a public petition may suddenly be transformed from a joke to a threat; above all, there is the obvious or perhaps totally unexpected takeover bidder. However it may be constituted and however recognized, the superior authority is defined as an influence making the attempt to subsume a string of esoteric boxes under one head. Thus we obtain the second rule.

Second rule: A superior authority may subsume a string of esoteric boxes under a single controller, with the injunction that he rationalize and then optimize the integral behaviour of this group.

There is yet (according to our six examples) a third mode of conduct. This again involves the superior authority, but it takes account of the fact that in our mature society the too frequent invocation of the second rule usually rebounds. A government which normally invokes it is called fascist; an industrialist who works by the second rule is called authoritarian; and in

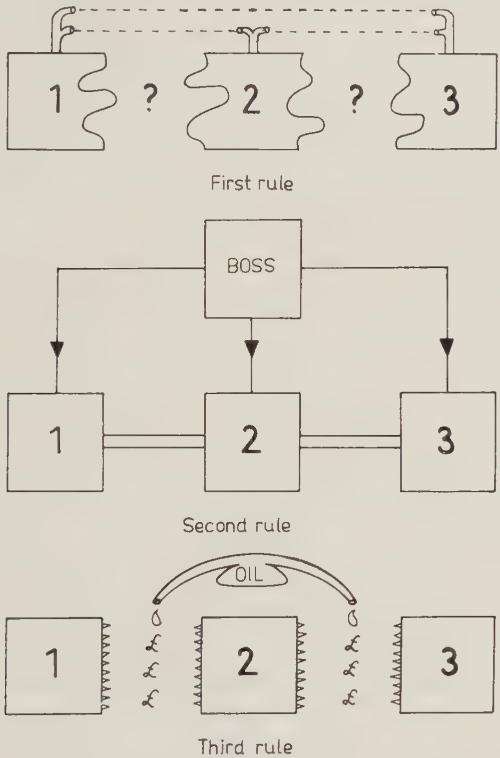


FIGURE 1 Spurious metasystems: pictograms indicating the three observed rules of behaviour

general management eminent consultants stand ready to denounce the second rule organization as over-centralized. Hence, and precisely because our society is mature, there is a third rule which is in truth a sophisticated version of the second—although it operates quite differently.

Third rule: A superior authority, protesting that it wishes not to intervene but merely to facilitate, sets up eminent bodies or specialist committees to lubricate the interfaces between a string of esoteric boxes.

Let us recapitulate our examples, noting the operation of these behavioural rules. In education, the first rule is used, with some help from the third rule. In defence, the second rule is supposed to prevail. In the civil service, the third rule has achieved its apotheosis—especially as a result of the Fulton Committee. In the case of the energy example: the public perhaps expects second-rule operation, sees clearly that it does not happen, and so hopes that it has a third-rule operation—whereas the likelihood is that everyone relies on the first rule. In the case of general industry, we have a third-rule image and—increasingly—a second-rule reality. Finally, in the case of a particular industry, we took an example where none of the rules manages so far to work, and therefore nothing happens. What will eventually happen is therefore predicted to be a drastic operation of the second rule; and when that rule operates it will look more like piracy than democracy in action.

The formulation of these three rules seems to be as much as we can squeeze out of our understanding of what at present goes on in the supervision of strings of esoteric boxes. From the point of view of government, which must accept the major onus for the sensible handling of social institutions, one or two political comments can be made. Firstly, from a game-theoretic standpoint, government ought to attend to the deployment of first-rule mechanisms—which they cannot themselves actually use (by definition), but which they can very well foster in roundabout ways. The next best strategy for government is to use the second rule (because it is effective) masquerading as the third rule (because that looks better). And to ensure that second-rule organizations (which are powerful) do not seriously disbalance affairs, it could well be convenient to think of them in pairs, and to arrange stated objectives and actual personalities in ways which result in opposite outcomes. The Monopolies Commission and the Industrial Reorganization Corporation, for example, do seem in practice to procure antithetic effects.

The key feature of our disclosures to this point seems to be this. There is an established group of methods used by society for relating esoteric boxes, which allows them to maintain their integrity, while society itself feels entitled to hope that its methods will result in adaptation to all kinds of change.

What kind of metasystem have we here? Given the criteria used at the start of this section, I can tell you the answer to that. We have a *spurious* metasystem. Because what is happening is that the manifest need for metasystemic control of the situation is not being met. The rules of the game provide authority with a perfect excuse for avoiding drastic action (on the ground that society might collapse if drastic action were taken), and to justify this by a flurry of specious activity which is in fact designed to paper over the cracks in society, represented by the interfaces between esoteric boxes.

A GENUINE METASYSTEM

A genuine metasystem does not use its power and its insight to *pretend* that a string of esoteric boxes is effectively strung, using the three rules which we have uncovered. On the contrary, it ensures that a new and ultrastable metasystem is actually created which will meet metasystemic criteria of performance. *But this would involve disrupting the integrity of the esoteric boxes.*

Here then is the crunch. We have and we need ultrastable social institutions. These are required for their own sake; but also because they are the components, the building blocks, of higher-order systems. It is in service to the interests of that higher-order system that the integrity of the original institutions becomes threatened. To implement the threat ill-advisedly would mean the collapse of some part of society. To succeed in a redefinition of the metasystem, and therefore of the roles of a string of esoteric boxes, would probably raise the payoff of the system by an order of magnitude.

It is obvious, from our discussion, why those in authority fear this challenge. But it is time to make explicit the reasons why the challenge must be met. I think there are two major components of the problem. Both are contemporary matters, which have not really confronted society before. The first component is the collapse of both organized religion and respect for temporal power. After all, here *was* a genuine metasystem. Society had an external skeleton, as well as an internal skeleton. And if the bones of the internal skeleton were not strong enough to support the system of interrelated social institutions, each esoteric box was individually hooked on to the external framework. Today, because the boxes maintain their integrity, they still incline to acknowledge the metasystem. But increasingly the public, and especially the young public, do not.

The second component of the problem is the rise, the exponential explosion indeed, of technology. Let us never forget or ignore the fact that the characteristic behaviour of every social institution is primarily conditioned by the technology available to its purposes *at the time* its characteristic behaviour gelled. Today is certainly some other time than that. There is no knowing what would be the most suitable configuration within the esoteric box, even for its selfish purposes given today's technology. That is because the box resists the attempt to redefine its role—because it is ultrastable.

So now our established practices are adrift in two dimensions at once. We have largely failed to stuff modern technology into the esoteric box and to reap its benefits there. As to tomorrow's technology, we are not seriously allowed to contemplate it. But much more importantly, there is no acceptable language in which to comment on the way in which the metasystem relating strings of esoteric boxes might best be designed. This is because the genuine metasystem of authority has gone, and its ethical language with it; while the metasystem of power with which we are left turns out to be spurious—and no metasystem at all.

Inexorably we are driven to the need to discuss, in an unknown metalanguage, the wholesale redesign of social institutions. We have ancient, possibly archaic, esoteric boxes, whose modus operandi is governed by anachronistic technology. These boxes are strung together, not in genuine metasystems, but in accordance with the rules which are themselves the product of an esoteric box: the social institution labelled "this is how we conduct affairs around here".

Now one of the lessons which I would say is most apparent from the whole history and practice of cybernetics in every context is the following. What really matters about a system is its structure—and especially its metastructure: the way in which the parts of the system are interrelated at every level of discourse. You may change the parameters of a system; you may vary the inputs to a system over a wide range; you may especially (and most surprisingly) allow the transfer functions which characterize the behaviour of the individual boxes to range across a very wide spectrum. The amazing thing is that the critical outputs remain invariant—so long as the systemic structure is maintained.

This comment, which is no demonstration but just a comment, assumes that the structure itself is hierarchical, exceeding complex, full of interaction, and rich in negative feedback. If so, I repeat, it is primarily the metastructure, the basis on which that whole hierarchy of systems is organized, which guarantees the characteristic behaviour of the whole. The simplest explanation of the reason why this is so begins with perhaps the very first major discovery of cybernetics, made clear by Norbert Wiener himself. I refer to the way in which a high-gain amplifier is rapidly dominated by its error-actuated feedback rather than by its input.

By the time we have moved to a multilooped metasystem, the *interactions* of such feedbacks at every level determine the dominant behavioural characteristics. Anyone who has worked with such cybernetics, in whatever context, knows how difficult it is either to predict outcomes or to intervene in the system to predictable ends. Thus is society in a trapped state of crisis today.

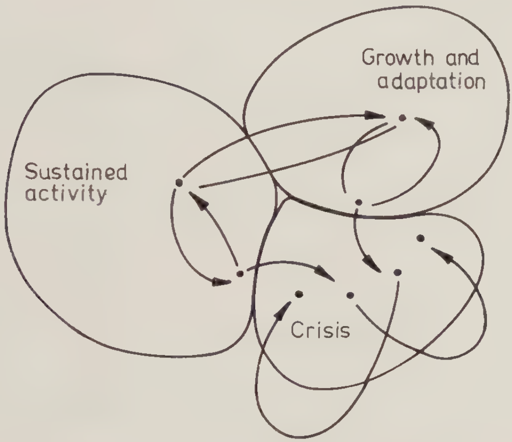


FIGURE 2 Society as a polystable system

Considering how this happens, we may think of any society as a polystable system operating in three main modes. There is growth and adaptation; there is merely sustained activity; there is crisis of one kind or another. As the diagram shows, there is more than one equilibrial state in each mode, and some cycling does no harm. Further, some alternation between the growth mode and sustained activity may be beneficial. But some equilibrial states are near to the boundaries of crisis. If we are not exceedingly careful in responding to pressures at this boundary, we end in the trapped state of crisis—where every attempt to get out leads straight back in. Evidently, some kind of planning process is required to resuscitate the system. It is no longer even possible simply to yield to pressure in the trapped state, if only because there is no known way of doing so: the gradient is in the wrong direction.

If we have not understood the planning process sufficiently to escape, it is due to the absence of a genuine metasystem in which to plan. Instead, the desire for participative management, in itself not only legitimate but necessary, has caused planning to happen at a subsystemic level. Hence "planner" has become a pejorative word; because planning is seen not to be rising to the demands of the total system. Moreover, the planning process itself becomes complicated to the point where everything seizes up. The need to preserve initiative at every managerial level results in an incapacity to show initiative anywhere.

A POSSIBLE ANSWER

We have taken a long and perhaps tedious route to reach any kind of proposal. The fact is, however, that the proposal has no hope of acceptance if the reasons why it is made are imperfectly understood.

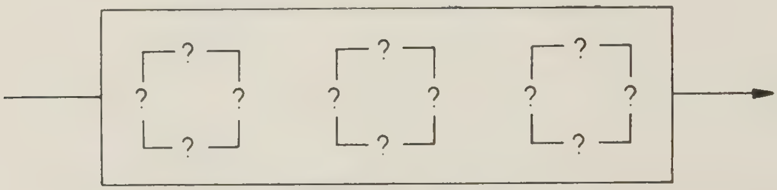


FIGURE 3 The threatening (and possibly megalomaniac) approach of a total rethink

I am asking that societary problems be tackled urgently at the metasystemic level. My fear is that the proposal will *appear* simply to say that they ought to be tackled on a wider front. But this is quite a different proposition. If we say that, for example, "education" ought to be looked at as a total system, and rethought from start to finish, we shall get nowhere. For any such plan would involve many years of work by large teams of experts, the very existence of whom would profoundly threaten the integrity of the relevant esoteric boxes. There would be professional uproar. Moreover, the whole planning, administration, and implementation of the project would inevitably become a bureaucratic nightmare.

On the contrary, we ought to use the self-organizing propensities of the string of esoteric boxes, and to harness their professionalism, knowledge and energy. To do this, we have to create an environment in which these boxes do not turn defensively inward, using up their potency in internal squabbles and the effort to inhibit change. Instead, we create a metastructure and supply a metalanguage so that the organizing power of the boxes themselves is released to give change effect. This means that whatever is the "superior authority" in any given social situation operates like a judo expert—who uses his opponent's energy, rather than his own, to achieve his ends.

Looking at the final diagram, Figure 4, we may think through how all this is supposed to happen. We can check that this systemic architecture does not attempt to repeal the natural law of Le Chatelier: the three prerogatives of esoteric boxes disclosed earlier are all preserved. Next we may trace how the three rules of observed behaviour *could* still operate in the region above

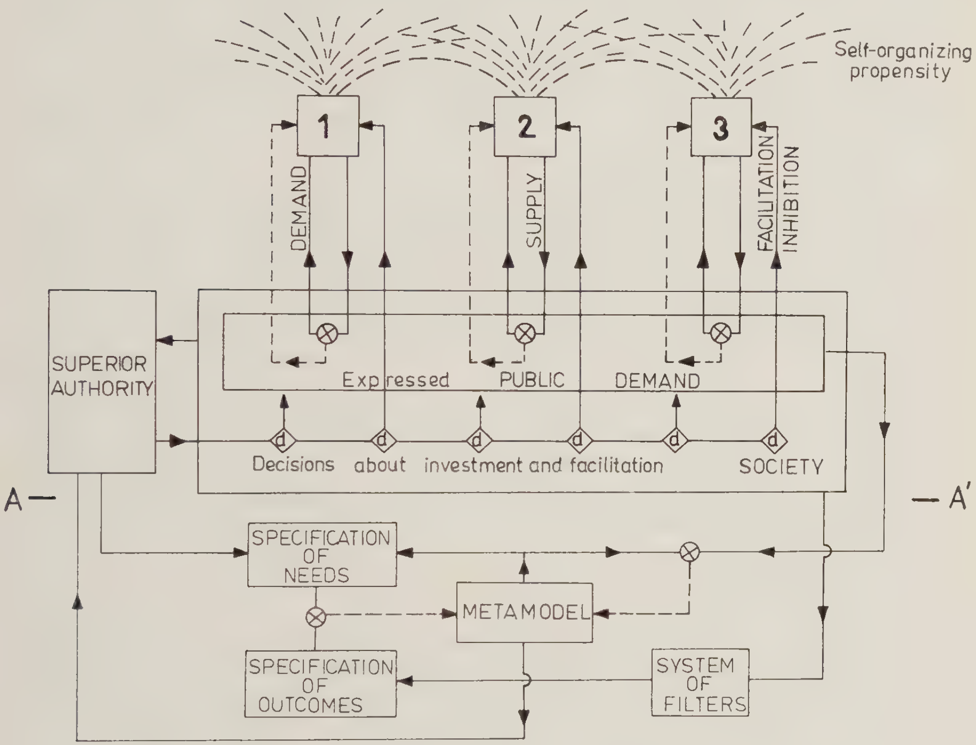


FIGURE 4 Societary control by self-organization under metasystemic governance. Dashed lines are error-actuated negative feedbacks

the line marked *AA′*. For here we have the superior authority taking sets of decisions about investment, in the activities of social institutions, and about the facilitation or inhibition by other means than money of their activities. Authority undertakes this task by direct observation of the operational facts: society expresses a demand which the esoteric box attempts to meet, and the control system uses the systemic language to assess the results.

When the metasystem is introduced, however, according to the diagram below the line *AA'*, authority begins to operate in a new metalanguage competent to discuss the systemic structure itself. The main results of this in practical terms are that the expressed demands of society are replaced in the metalanguage by the specification of needs, and the assessment of results in systemic language is replaced by a specification of outcomes as determined by an appropriate system of filters. Thus, for example, society's *need* for scientists will never again be confused with the mere aggregate of its stated requirements, as it has always been confused—with bizarre results. Nor, to take an output example, will society's raw and emotional reaction to the trade figures ever again be confused with a metasystemic evaluation of their effects.

Then the question arises: What actual innovations must be made to implement the entire model? The answer is that everything structural *below* the line *AA'* needs to be freshly installed. At present it does not exist. What does exist is the capability to do the things required of the below-the-line structure: indeed, some of the relevant work has been done already by professional groups scattered throughout government, the universities, and industry.

The conclusion I draw is that these activities should be institutionalized forthwith. It is not a particularly expensive nor difficult operation. In the firm, and in other large corporate institutions, we should find the realization of our plan in a new kind of group—which might be called a Development Directorate. At the national level itself, we should be calling for a new kind of Cabinet Office.

In both places a continuous planning and evaluation process would begin, with feedback, in metasystemic terms—and working exclusively on the metastructure of the industry or government concerned. Such an organization would replace the teams of advisors and pundits who currently offer their subjective views at the highest level. Now properly organized, continuously presented, real-time displays of situations would be made; together with the metasystemic elucidation of their consequences, the action alternatives which are available, and their likely outcomes.

The comparison which immediately suggests itself contrasts this picture with the "war room" of World War II. I cannot refrain from mentioning that I began speaking to you to the very minute on the thirtieth anniversary of the outbreak of that war; nor from pointing out that we now have a technology competent to exploit the techniques then developed, and since lost, which were at the time fumbling manual operations. (The whole of this metasystemic intention is equivalent to the fourth level of managerial hierarchy laid out by Stafford Beer[1].)

This address is too long, and I apologize for that. In conclusion, then, I content myself with a reiteration of my opening remarks. Society is in crisis. The way of life we have known is vanishing. This is a fact, which we ignore at our peril. The symptoms are all there. It is then open to a congress of this kind to make a cybernetic prediction, and to offer a solution.

My prediction is that a new kind of society will emerge, with new modes of control—which no-one actually chose, and which we probably will not like. The solution is to decide what to do, and how to do it, in the light of cybernetic discoveries as to the nature of systemic control. I have offered the makings of a blueprint. But progress requires collaboration of both government and management, and an informed public opinion. This in turn entails a new forum in the press, on the air, and in parliament; for unaccountably these matters are rarely discussed. Progress also requires a new approach to planning, and the guts to accept that change is needed—I mean change entailing actual alteration.

Alternatively, the gross instability of society, of institutional relationships, and of our economy, will de-nature the civilization to which we have attained. The society we have known will either collapse, or it will be overthrown. It is of course open to anyone to sit complacently in his own esoteric box, and wait for the sky to fall.

Reference

1. S. Beer, "Prerogatives of system in management control, *8th Ann. U.K. Automation Council Lecture*, January (1969).

CHAPTER M-6

Cybernetic medicine (with emphasis on a new method of automatic diagnosis)

A. MASTURZO*

Naples University, Naples, Italy

Summary

A general view of cybernetics applied to medicine and biology is given with description of simulators "AM 1" and "AM 2". Such simulators have permitted a cybernetic representation of diseases by autoaggression and explained the pathogenesis of rheumatism. A new method of automatic diagnosis is presented, studied on the basis of the principles of medimatics. It consists of setting up "medigrams" and processing them with the aid of a computer. Experiments carried out in the field of rheumatology have proved the method very useful and applicable to every field of medicine.

INTRODUCTION

The contribution of cybernetics to the progress of medicine and biology is confirmed by the results obtained in employing simulators for the study of biological phenomena.

We have recently designed and built two electronic simulators, "AM 1" and "AM 2" (Figure 1), which have been of great assistance in explaining immunobiological phenomena, and pathogenesis of rheumatism and of other diseases by autoaggression. A detailed description of the circuit is omitted; it is simply reported that the machine is capable of elaborating the input signals and of emitting output signals that are an electronic "transduction" of the studied phenomenon.

* President of the International Society of Cybernetic Medicine.

This gives us more knowledge of the biological facts and helps us to formulate certain predictions.

When we introduce in the simulator information corresponding to a given biological function, we obtain at the output other information, resulting from a logical elaboration operated by the machine.

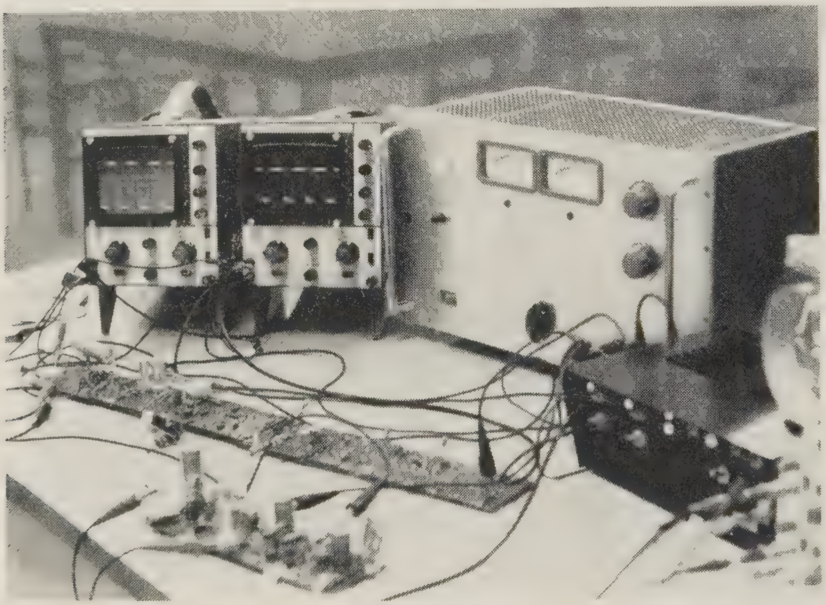


FIGURE 1 The simulator "AM 2"

If we introduce information corresponding to a normal function, we obtain on the screen a sinusoidal wave. Such a wave, indicated by the symbol φ (Figure 2), represents a biological system in equilibrium. On the other hand, if we introduce pathological information, it will on the screen appears a square wave, indicated by the sumbol ϑ.

The defence information will appear as an inverted square wave, that is *minus* ϑ $(-\vartheta)$.

By introducing ϑ and $(-\vartheta)$ we obtain a line coinciding with the isoelectric line, that is to say that the two informational expressions have been cancelled. If we introduce, at the same time, information corresponding to the normal function (φ) plus information corresponding to the pathological one we obtain a new wave, which has the shape of a capital M (Figure 3). Such a wave represents the result of the elaboration of a normal function plus a pathological function.

By introducing simultaneously the three signals, $\varphi + \vartheta + (-\vartheta)$, we obtain the normal function φ, but then affected by symptoms of a cured disease, that is φ'.

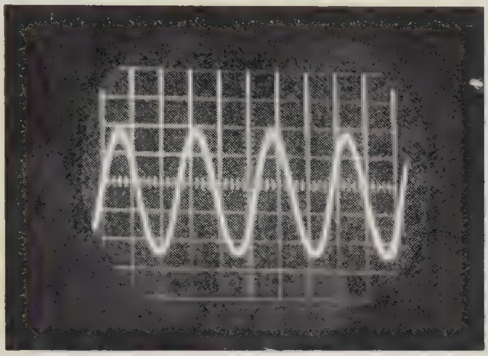


FIGURE 2 The φ wave, corresponding to a biological system in equilibrium

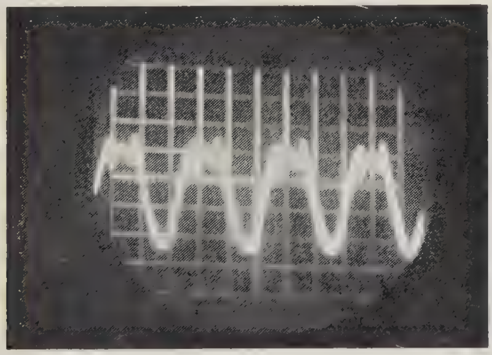


FIGURE 3 *M* wave, signal of a pathological situation

If we introduce now the signals $\varphi + (-\vartheta)$, we obtain a special and caracteristic wave represented by a lance-shaped wave, indicated with the symbol ξ (Figure 4). Such a wave represents the phenomenon of autoaggression.

Information $[\varphi, \varphi', \vartheta, (-\vartheta), M, \xi]$ can be stored as $[\varphi^*, \varphi'^*, \vartheta^*, (-\vartheta^*),$

M^*, ξ^*] and employed by the organism for the pathogenetic mechanism of special diseases, as dysnomopathies. Our simulators have helped to explain many pathological facts and formulate some predictions, which have been confirmed both by clinical observations and by laboratory experiments.

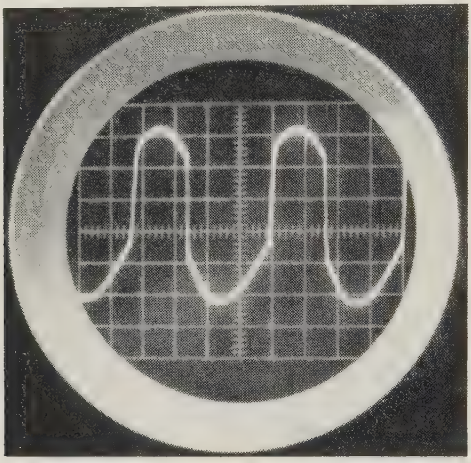


FIGURE 4 ξ wave, corresponding to the phenomenon of autoaggression

AUTOMATIC DIAGNOSIS

As regards the problem of automatic diagnosis, we are presenting the method we have studied and experimented on, especially in the field of rheumatology.

Working from the premise that all diseases constitute a logical and organic set, and that all biomedical data constitute another set, equally logical and organic, and also considering the fact that the two sets can be subjected to logical operations, we reached the conclusion that the following problem can be solved mathematically. Given n symptoms, establish all the diseases that can be described by m symptoms and classify them in order of probability. By the term "symptom" we mean all the informational data from the medical history, none excluded, such as electrocardiograms, X-rays, results of biological and seriological tests, etc., and not only these data, but also their relationship to pathological and biological data. In other words, the symptom is an informational unit with an operational content.

The "analytical" problem can be stated in this way. With n symptoms available, theoretically C diseases can be discerned. But $C - Z$ really exist in nature, while Z remains in the "area of possible diseases".

Let us imagine a matrix with all the known symptoms listed along the first column and diseases, obtained by the combination of such symptoms, in the first row; we shall then have in every cell a number corresponding to the entropy of each symptom related to each disease, that is to say its "diagnostic weight". Such a diagnostic weight has been evaluated by statistical investigations. In this manner we can establish a scale of diagnostic weights from *zero to hundred*, expressed as percentages, and from *zero* to *minus hundred*, also in percentages.

The negative diagnostic weight is calculated on the exclusive value of a given symptom with regard to a given disease. For example, the complete sacralization of the fifth lumbar vertebra excludes the presence of disk herniation between L_5 and S_1. By comparing all the possible diseases with the diseases known by the pathology we can list the latter on the first row and have in the cells the diagnostic weights of each symptom compared with each disease.

In order to make the logic of our method even clearer, let us give an analogous example which, even though posed in general terms, will give an idea of the operational dynamics that are at the basis of our method for automatic diagnosis by means of weighted data, or medigraphic diagnosis, according to the principles of medimatics. Let us imagine three geometrical solids: a cone, a cylinder, and a sphere. Let us suppose that they have the same volume, the same colour, but different weights, since they are made of different materials, and that these respective weights are 100 grams, 200 grams, and 300 grams.

Let us divide each solid into m equal parts, 10, for example, which all have the same weight. By mixing the resulting parts of the three solids, we shall have a heap of thirty pieces. The problem is to restore the three solids by recombining the pieces belonging to each. If we choose ten pieces at random from among the thirty pieces, it is possible that all ten would belong either to the cone, to the cylinder, or to the sphere. But it is certainly unlikely that we shall be lucky enough to choose all the pieces belonging to the same solid, without making a single mistake. So we shall proceed in the following manner. If we want to put the cone back together we shall try to find the ten pieces by attempting to recognize those that most probably came from the cone. If by weighing the ten pieces we obtain a weight of 100 grams, we shall have solved the problem, otherwise we must be prepared to accept an approximate solution, which will be as approximate to the correct one as the weight is near to 100 grams. With successive attempts and relative weighings, we can get closer and closer to the exact solution.

But, if we wish to proceed in a rigorous mathematical way, we should do as follows. Fix all the possible combinations, without repetition, of the 30 pieces, taken ten a time, utilizing the formula

$$C = \frac{n!}{m!\,(n-m)!}$$

We shall weigh each combination and thus find those which weigh exactly and respectively 100 grams, 200 grams, and 300 grams. In this way we can restore the three solids exactly. However, we shall have to carry out an enormous number of operations, which involve a lot of time and a high risk of error. If, however, we have a computer to carry out the operations on the orders of a given program, the time element will be practically nil and the risk of error practically zero, or exactly zero. The preparation of the program is the "hot phase" in the entire procedure and requires a logic that we have recently established. The informational point of departure for automatic diagnosis by means of weighed data is made up of an "operational biomedical form", or medigram, if one wishes to use the term indicated by medimatics.

Medimatics, which we propose as a new scientific discipline and which has been already discussed in other publications, concerns the application of all sectors of mathematics, both the traditional and the "new" ones, to medicine and the use of automatic calculations that are almost always carried out by a computer. Medimatics could be utilized to solve various medical problems concerning research, clinical, and social applications. The medigram collects all the informational data on the patient and expresses them in numerical or alpha-numerical language transcribed on punched cards or magnetic tapes.

The initial medigram is enriched by data given by the diagnostic matrices, which are in turn gathered with the aid of a computer and list all the diseases present in a given population on the first row. These diseases are expressed in alpha-numerical language and are affected by the index of incidence in the population. Along the first column of the same matrix are impressed in code all the symptoms, with none excluded, that can be collected by the methods of modern semeiology. The cells of the diagnostic matrices are filled with numbers that express the diagnostic weight of each symptom for every disease considered. The establishment of these diagnostic matrices, or medimatic matrices, is carried out by research teams, who have the job of evaluating the diagnostic weight of each symptom for each disease, which are established by taking into consideration clinical histories from clinics, hospitals, and diagnostic centres.

At first approximation, the differences between the given values and the "exact" ones are fairly large, but perfection by means of successive "medimatrix" corrections, with the help of the computer, leads to increasingly smaller values in these differences, with a tendency to approach "exact" values. Also, the computer can establish medimatrices alone from exact medigrams gathered from entire populations. In fact, the medigram can give the subject's complete history, from the time of birth and throughout his entire life, condensed into informational data; post-mortem data can also be given, and each individual patient constitutes in himself a complete informational array that can be utilized by the computer.

Thus, by processing automatically the medigram of a given patient and all the medigrams in the archives, one rapidly sets up medimatrices and carries out logical and numerical operations in order to arrive at the "medimatic diagnosis" or "medidiagnosis". The proximity of the latter to the exact diagnosis will depend on the accuracy of the medigrams scanned. It is implied that all pathological and hygienic data must be incorporated, together with preventive and therapeutic measures that have undergone wide-scale experimentation, and that these be expressed in unequivocal and universal language. In this way medical statistics and exact statistics are operational and no longer merely a collection of data with diverse criteria and without any operationally logical relationship.

Our technique for "medimatic diagnosis" is as follows. We operate on the medigram of the patient under examination and on medigrams collected on a large number of patients seen in the past until they contain all the data, not excluding those relative to the biomedical findings, up to the most recent. We then prepare the program, which includes the medigram of the patient under examination and the medigrams in the archives, as well as the instructions directed to the calculator. We also draw up the flow diagram or flow medigram (Figure 5), which tells the computer the sequence of commands to follow.

The first command or statement is that of counting the medigrams and setting up of the "medimatrices" (Figure 6). Successively, we order the computer to register all the data of the medigram of the patient under examination as asterisks along the rows of the matrix, leaving a mark in every cell. We then order the computer to sum up the diagnostic weights in the marked cells of every column and to report the results on the last row.

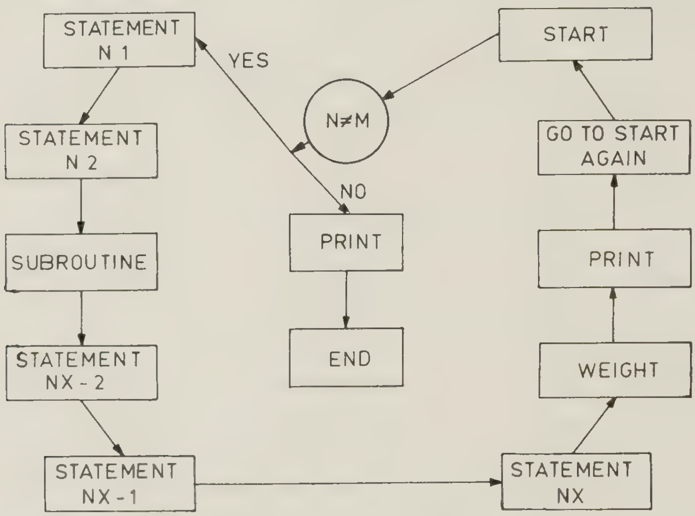


FIGURE 5 Flow medigram

Lastly, we order it to indicate the column that shows the highest diagnostic weight; the latter corresponds to the medimatic diagnosis. In case there is more than one diagnostic weight with the maximum value, the computer is ordered to research at the level of the first row, which gives the index of frequency of the two or more diseases indicated, and the one with the higher index corresponds to the exact medimatic diagnosis.

When data weighing is finished, the computer prints and produces the result, with any eventual suggestions for repeating the informational cycle with the addition of order data, for the purpose of approaching the "diagnostic goal" as closely as possible. The medimatic diagnosis that is produced is an "informational whole" that contains all the data regarding pathology and semeiology, disease therapy and prevention, sociology, and those pertinent to the patient under examination, in an "operative form". The medigram of the latter will be enriched by all the medical knowledge and will contain the decision to adopt all therapeutic means for the case under examination, as well as all preventive measures for analogous cases.

In addition, this will contain operational elements for classifying the diseases on a firm basis and not on theory or on more or less valid conjecture, which is sometimes imposed only by the academic authority of a certain author and certainly does not constitute a scientific basis in the strict sense of the term.

Another advantage of medimatic treatment of medical information is the possibility of uncovering the existence of nosological complexes that have not yet been described.

Medimatic diagnosis constitutes the most accurate instrument for following the development of disease in a given patient and predicts in time

	N	A	C	I	R	L	B	M	Q	V	D	O
	9	8	7	7	7	6	5	4	4	4	3	3
1	10	8	2	1	1	0	3	4	6	2	7	0
2	3	1	7	10	4	9	5	1	6	3	8	2
*3	3	10	1	0	0	2	11	2	3	2	14	12
*4	1	8	2	0	1	2	9	1	4	1	6	7
*5	4	9	1	3	0	1	14	2	7	2	8	10
6	2	4	7	0	3	8	1	9	2	0	4	3
7	9	0	10	7	1	5	3	8	5	3	1	9
*8	3	6	2	2	1	0	15	9	2	6	13	15
*9	3	9	1	2	0	2	15	6	10	1	9	12
10	3	7	2	9	0	10	7	2	4	5	7	4
11	9	3	0	6	2	7	1	9	3	7	1	10
*12	1	0	2	2	0	0	10	7	9	1	15	9
13	4	9	1	5	9	6	3	5	7	3	1	0
14	5	2	6	1	10	5	2	7	3	1	9	4
*15	5	8	1	0	0	0	6	6	5	2	10	15
16	2	5	3	1	10	3	7	4	9	1	6	5
17	1	2.5	9	0	3	5	1	3	2	0	4	8
18	5	1	7	5	3	2	8	7	5	4	8	0
...	...	...	...	...	...	...	...	...	...	...	...	...
	20	50	10	9	2	7	80	33	40	15	75	80

FIGURE 6 A simplified example of a medimatrix

any possible worsening of the patient's disease, with indications for preventive measures. It gives us a better knowledge of the action of therapeutic measures and their collateral or armful effects and suggests new therapies that can be evaluated in a strictly scientific manner by medimatics.

References

1. A. Masturzo, "Quelques aspects pratiques de la médecine cybernétique", *Nucleus*, **9**, No. 5 (1964).
2. A. Masturzo, "Calcolo elettronico in diagnostica", *Proc. Natl. Congr. Social Med., Naples* (1965).
3. A. Masturzo, *Cybernetic Medicine*, (Ed. C. Thomas), Springfield, Illinois (1965).
4. A. Masturzo, "Method of automatic diagnosis by biomedical data weighing", *J. Cybernetic Med.*, No. 3 (1968).

CHAPTER M–7

On ways of automation of natural sciences development

VICTOR M. GLUSHKOV

Institute of Cybernetics, Ukrainian Academy of Sciences, Kiev, U.S.S.R.

Summary

It is complex automation that is said to be the main distinguishing feature of modern science. One of the aspects of automation is the complex automation of experiment data processing. The main problem is to make the data input automatic. There are various conditions of data insertion: off-line, on-line. Examples are given of data processing systems in geology, biology, medicine and library science.

INTRODUCTION

At present, computers are playing an increasing importance in developing the entire scope of natural sciences. While computers have been employed before mainly in the exact sciences (mechanics, physics, astronomy), a considerable place in the general volume of computer applications is now occupied by biology, medicine, geology, and other (descriptive in the past) sciences. We have entered the epoch of the all-round automation of sciences and primarily of natural ones. But what are the principal paths taken to automate the development of science?

First, it is the all-round automation of experiment data processing. "All-round" means in this case the automation of all stages of processing and not only of some of them. It is well known that with the traditional methods of input (by means of cards or tapes punched manually) it is reasonable to use computers for solving fairly complex problems only. The complexity is understood here to be the quantity of operations accounted, on the average, for one number inserted into a computer. This class, however, does not involve most problems which have arisen at the stage of the so-called primary experiment data processing. Related to this are various problems of averaging, finding various statistical characteristics, interpolation, etc.

An essential expansion of the realm of computer applications at the primary experiment data processing calls for automation of data insertion into a computer. Such an automation is brought about in two different ways. The first way (off-line mode of operation) consists of providing the computer with a set of devices enabling it to read graphs obtained by various recording measuring instruments.

Off-standard forms of representing graphic information in measuring instruments is a severe obstacle to this. This circumstance necessitates either that one must have a great number of special input devices for a computer, or one must excessively universalize this device. But the latter not only raises the price of the device but also results, as a rule, in the deterioration of performance in each individual case. An alternative consists in finding a reasonable compromise between the convenience of recording data and the convenience of their subsequent insertion into a computer. To do this, it is necessary to establish a rather small number of standards for a form to record graphic data at outputs of various devices, and to obtain a set of dependable and relatively inexpensive equipment for automatic insertion of these data into a computer. Note that the development of microelectronics makes the digital record on magnetic medium (tapes or cards) one of the most widespread standards for fixing measured data.

The second way of automating the insertion of experimental data into a computer is based on the use of the on-line mode of computer operation with measuring instruments. In this case the best results are obtained for experiments involving expensive equipment (accelerators, radiotelescopes, etc.) and only then when measuring instruments and their link with a computer are designed together with the whole object of dealing with experiments. Thus, a computer accounts for an organic part of this object.

An intermediate version is also feasible when a self-contained computer is used, equipped with special facilities for communication with objects, and capable of operating under time-sharing conditions. Communication facilities serve to transmit to a computer the data drawn directly from measuring instruments, as well as to transmit signals derived from a computer to an object undergoing an experiment, if required.

As far back as 1961 in the Soviet Union at the Institute of Cybernetics A.S. Ukr. S.S.R. in Kiev the computer DNEPR-1 had been developed, which proved capable of creating systems of automating experiment data processing. Many such systems were built on the basis of this computer for various fields of science (mechanics, physics, biology, and others).

AUTOMATION IN OCEAN RESEARCH

Let us refer to one of automation systems installed on the Soviet research ship "Mikhail Lomonosov" in more detail.

The system is intended to automate processes of data gathering, of their handling and of controlling hydrophysical investigations conducted in the ocean.

The system solves the following problems.

(1) Automatic reception of data from measuring instruments and complexes.

The number of information sources in all-round investigations of the ocean amounts from 400 to 500. Part of sources of data requiring speedy processing and analysis aimed at scheduling the experiment is immediately connected to a computer. The second part of data arrives for processing from buffer accumulators linked with information sources by radio, by acoustic channels, or via cable communication lines.

Data obtained by self-contained deep-water measuring complexes are fixed by buffer accumulators which, on raising them on board ship, are removed from containers and connected to the system.

Most of instruments employed for studying ocean processes have digital outputs.

The total amount of information reaching the system for processing during one voyage lasting 4 months ranges from 10^9 to 10^{10} bits.

(2) Processing of observation data including a fairly complete set of programs for the statistic processing of scalar series (calculation of means, dispersions, distribution functions, autocorrelation functions, functions of spectral density, selection of periodic components, etc.), as well as various parameters characterizing interdependence of two and more random processes (mutual correlation, cross-spectra, etc.).

(3) Representation of processing results in the form of graphs enabling the experimenter to estimate, through use of processing results, the correctness of choice of a processing program made by him.

(4) A task control of the experiment course conducted for some investigations in the ocean.

For instance, in hydroacoustic investigations an experiment strategy is chosen depending on obtained results of the speedy analysis of data gathered.

(5) A number of "service" tasks are solved associated with the research ship position finding and linking the observed results to her position.

(6) Tabulation of processing results.

A wide accessibility of the data processing system to personnel lacking knowledge of programming and of solving problems on a computer is attained by an appropriate organization of software.

All processing programs are designed as discrete subprograms. The experimenter is to choose from the available list of subprograms the required set and to specify it in the form of a certain code sequence. The system executes the given sequence of instructions making an automatic transfer from one subprogram to another as instructed. The experimenter is allowed to interfere with the course of processing, to change plans, to introduce new instructions, and so on.

The system is improved technically, the scope of solved problems is widened, the software is advanced.

The creation of experimental data processing system on the base of direct connection of a computer with measuring instruments promotes another step—automation of experimental control. In many cases situations are encountered when the obtaining of some data during experimentation influences the further course thereof. Suppose the automation of radiospectroscopy is dealt with. The task of controlling the experiment (taking the spectrum) in this case consists in varying a space of frequency change at each succeeding measurement. If it is desirable, for instance, to study the fine structure of spectral lines, a space should be decreased when approaching the next line, and it may be increased in between. Obviously this task is readily solved when data processing is conducted at a rate of experiment progress proper. Such instruments are already operating at a number of physical laboratories.

AUTOMATION AND GEOLOGICAL RESEARCH

More complex is the problem of controlling experiments aimed at finding the structure of space fields. A characteristic method of this kind is given by geology. One of important goals of geology development consists in finding a field structure of distribution of chemical elements, compositions, or rocks in the crust. Various experiments are conducted to solve the problem: boring of a hole, sampling, investigation of fine structure of gravitational, electric, and magnetic fields, etc.

The sought-for fields are not known exactly at any given moment. At the same time, lack of direct measurements at a point does not signify, generally speaking, that a value of the field of interest at the point is absolutely unknown. Current theories of crust origin and various empirical regularities enable us to make hypotheses concerning that value.

As a result we usually do not obtain precise values of a quantity in question at each point but some distribution of probabilities. Thus, at each stage of the geological science development we deal not with determinate but with random fields. Each new experiment decreases the uncertainty of certain distributions. By estimating this decrease by known rules of information theory, we can give the strict quantitative estimation of an experiment value. Finally, by evaluating a cost of various experiments, we can give a quantitative price of an information bit obtained as a result of each experiment run. On making such evaluations in advance, prior to running the experiments, we can determine the optimal order of executing experimental jobs, thus bringing about their control.

Certainly, a number of additional difficulties arise in determining the criterion of optimality. The first of them is that the problem implies finding not one field but many fields simultaneously. This necessitates establishing a system of weights, and using estimations of weighted averages.Furthermore, the existing theories have also, as a rule, only a certain probability of being true. Moreover, in some cases, in place of the general problem of studying the field on the whole, partial tasks of determining some features of this field develop, for instance, concentration places of some useful minerals. Answers to these questions are also initially uncertain and this uncertainty is to be reduced by experiments. Therefore the system of weights should cover such tasks too.

A problem of creating electronic archives of primary experiment data is also of great interest for automation. As it is known, the problem of processing experimental data implies compression of information at the expense of discarding useless data. However, the concept of information usefulness or uselessness is relative. What has been discarded today as inessential or even related to experiment errors, tomorrow may result in significant discoveries owing to new standpoints and concepts.

In exploration, data obtained are usually processed for a certain purpose, say, for oil locating. At the same time those data may contain information about other useful minerals in a region being explored. However, these data may not be utilized either owing to the fact that the task of searching for those minerals has not been stated, or owing to absence of required processing methods at the time. When this purpose arises afterwards, it is not easy to extract the necessary data from the archives. Therefore, it is quite often necessary to repeat expensive experimental jobs (e.g. geophysical explorations) while these could have been avoided.

The way out consists in replacement of ordinary paper archives by electronic archives of primary data obtained in the first place by expensive experiments. It is advisable that all data be memorized digitally on magnetic tapes. Each new file of data is to be given a heading for identification of each file element and experimental conditions. Both the form of headings and the method of recording the files proper should be standardized so as to make their succeeding processing on a computer maximum easy. It is also necessary to have an automated reference-information system enabling us to retrieve desired information files by certain tags.

AUTOMATION AND CLASSIFICATION

An important part in developing natural sciences is played by the problems of classifying and identifying an object as belonging to a certain class of objects. Today this field has sufficiently progressed in the direction of formalization and automation of certain classifying processes. Two main directions of automating the classifying problems have come into being. The first is associated with constructing various learning devices and algorithms. The essence of this is as follows.

Suppose there exist several classes of objects, each object being characterized by a certain set of tags $(x_1, \ldots, x_n)$. The device generates a function $f(x_1, \ldots, x_n)$ of these tags, the various values of the function corresponding to various classes. At first the function f is chosen at random and, therefore, the required correspondence does not exist. However, a certain mechanism of changing the function f is laid down in the device (or algorithm). The change takes places owing to insertion in the device of a system of tags $(x_1, \ldots, x_n)$ of some object and the number of the class to which this object belongs. In other words, an unknown function value at a point arbitrarily chosen is stated and the specified function is changed in the direction of approaching to the one required.

It is desirable that the mechanism of this improvement excludes the necessity of an exhaustive search through all feasible values of tags $(x_1, \ldots, x_n)$ for a good approximation of the function f to the unknown function. It cannot be done, of course, in a general case. However, if the basic system of tags reflects the required classification well enough, we can construct the learning algorithms and devices which are effective to some extent.

Learning devices built on the principle of idealized nerve networks have now become well known. Among such devices the perceptrons should first be mentioned. While the first version of the perceptron (so-called one-layer perceptron) had not ensured good pattern recognition, multi-layer perceptrons, introduced afterwards, proved to be more adaptable to solving those tasks.

The second direction of solving classifying problems is also aimed at initializing the classifying function $f(x_1, \ldots. x_n)$ by its values. But this is done in some other way than in the first case. Assume for certain that all tags $x_1, \ldots, x_n$ are binary, and that the number of classes equals two too. Suppose that there exist p objects known to belong to the first class, and q objects known to belong to the second class. The required function should take on 0 on the first p objects, and 1 on the next q objects.

Identify a function specifying the degree of deviation of the arbitrary function $g(x_1, \ldots, x_n)$ from the required function. One of the simplest ways of identifying such a function F is as follows. Let p_1 be the quantity of first-class objects at which the function g does not equal zero, and q_1 the quantity of the second-class objects at which the function g does not equal unity. State the unknown function by the formula $F(g) = p_1/p + q_1/q$. The closer this function to zero, the closer is the given function to the unknown one.

Now, from the totality of functions $x_1, \ldots, x_n$ and their negations we select one of those functions f_1 which gives the minimal value to the functional F. Then we construct various two-place Boolean functions of the selected function f_1 and functions $x_1, \ldots, x_n$, and fix one of those functions f_2 which gives the minimal value to the function F. We repeat the same construction with the function f_2 as with the function f_1 obtaining the function f_3, and so forth. The function F value is not less for each new function f_i of this kind than for all preceding functions f_j $(j < i)$. The process is over when the function value at the next step does not diminish. If the set of basic tags $(x_1, \ldots, x_n)$ has been sufficient to separate the foregoing classes, then the functional value will equal zero at the last function of the kind and, consequently, this very function will be the unknown separating function.

It is known from experience that in this case the basic tags $x_1, \ldots, x_n$ are related in essence to the required classification; the function f, constructed as described above, accomplishes the correct classification of objects not only known before, but of many new objects not participating in its construction.

If the basic tags are not binary, the construction of binary tags of the form $x_i > a$ or $x_i < b$ is effected. When so doing the values of constants a and b are so chosen as to make the constructed binary tags give the least value to the function F. Similarly, if there exist more than two classes, they are divided first into two subjects, each of which is divided into two more, and so on. As a result it appears to be possible to carry out the classification by the system of Boolean functions as by principles analogous to those above.

Availability of the above potentialities of formal classification opens the way for mathematization and automation not only of exact sciences (giving exact quantitative declaration) but also such sciences as biology, medicine, geology, handling numerous qualitative criteria. The described methods of formal classification represent essentially a particular case of mathematical modelling of objects and processes studied by natural sciences. The mathematical modelling is said to be the declaration of an event by means of mathematical notions allowing us to deduce the event features unknown *a priori* at the instant of drawing the declaration. For instance, specifying the law of gravitation, masses of the Sun and planets, their initial relative positions and velocities, we obtain the mathematical model of the Sun system.

ALGORITHMS

During the pre-cybernetics epoch, only such relatively simple mathematical models were recognized which were described by formulas and equations. It is, however, well known that the mathematical formula language is not comprehensive. Being abstracted mainly from mechanics (in the first place—from celestial mechanics) it lends itself well to describing a comparatively narrow range of natural phenomena.

A much broader class of mathematical declarations is represented by algorithms. However, as distinct from the formula language, it was not until fairly recently that the general algorithm language began to develop. That is why it is studied to a much lesser degree. The formula language has many abbreviations specially adapted for describing important processes in appendices (trigonometric functions, derivative, integral, etc.). Object properties expressed by these abbreviations are well studied. We are able to transform equivalent methematical expressions. All of these means are still in embryo in the general algorithmic languages. It is not surprising then that the investigator feels himself not quite contented when encountering a complex algorithmical declaration of some phenomenon (and not simple formula description).

Most of the processes being examined, for instance, in biology, *a fortiori* have no simple formula descriptions. It is relatively not difficult, however, to construct algorithmical models even for such complex biological processes as evolution. It goes without saying that such models will have rather unwieldy declarations despite extensive simplifications. It is impossible therefore, in practice, to obtain corollaries from them by conventional methods (solution of equations, transformation of formulas, etc.). But it is not only the complexity of the declarations, for sooner or later, as the general theory of algorithmic languages progresses, the declarations will be simplified. Essentially evolution and other biological processes are materially much more complex than, for example, processes encountered in celestial mechanics. This complexity shows up, firstly, in the necessity of carrying out a much greater number of elementary operations for transforming the information in order to obtain interesting corollaries from their mathematical models than in the case of simple models.

Owing to this, the value of complex non-traditional mathematical models is dependent in the first place on the possibility of automation of experiments with such models. Such a possibility is given by modern data processing systems based on the use of computers, the model being specified in the form of a computer program. Experimentation proper with the model consists usually in the gradual spreading of a process development (to be exact—of quantities characterizing the process) with time. As differentiated from classical mathematical models, quantities characterizing the process may not only be ordinary numerical ones but logical ones too. They can, consequently, declare availability of qualitative tags of objects under investigation.

Today there already exist many experiments successfully accomplished by non-classical mathematical models. As an illustration we can point out the simplest models of the evolution process or models of the protozoa nervous system (worms, and the like). Note specially one class of non-traditional mathematical models to be referred to hereafter as *automaton models.*

Suppose an object being investigated consists of a quantity of elements, each of which can be in a finite number of states. The human organism can be taken as an example. Elements involved are various organs (e.g. endocrine glands) and regulation systems. A state of the object under study is a vector, the components of which are states of all elements composing the object. The model is composed of a totality of rules. Each rule declares an effect of states of some elements and some external influences on a state of an element. These very rules are established by various particular experts studying the given object. The rules should indicate not only the law of transition from one state to another but also an approximate time during which these transitions occur.

Stating now a sequence of external influences and using the above rules we can sequentially, step by step, determine a change in states of the object in question. Naturally, the system of rules here is to be complete, i.e. to be able to declare transition of all object elements. In principle, the new result obtained during experimentation with such models consists in the complexity and integrity of studying the object: from particular dependences derived by various experts. the general dependences of the model show up, which take into account the influence of the entire diversity of factors.

Reciprocal problems may also be solved in some cases with the aid of such models by utilizing the high speed of up-to-date computers. The sense of reciprocal problems implies the finding of a sequence of external influences to bring the object under test to a specified state (or to some given set of states). A still greater difficulty is encountered—which is most interesting though—in finding that sequence which effects the transition during the shortest time. In the case when the object to be investigated is a human organism, the solution of such reciprocal problems makes it possible to find the best methods of treating diseases (by the use of combinations of current methods).

The great difficulty of organizing experiments with mathematical models is the laborious process of declaring certain programs. The point is that the use of ordinary computational algorithm-oriented languages (ALGOL, FORTRAN, a.o.) leads to materially worse results when modelling algorithms are handled. It is firstly due to the considerable loss of time and memory that usually occurs during translation from such languages when handling variables taking on finite multitude of values. Moreover, the ordinary list of standard procedures in languages of the ALGOL class is first adapted to computational algorithms. The modelling algorithms need quite different classes of standard procedures.

At present a large family of modelling algorithm declaration-oriented languages have already come into existence (SOL, SIMULA, SIMSCRIPT, SLANG, a.o.). However, to our regret, all these languages have, in the first place, a proper computer, as an object of modelling, and data processing systems based on using computers. Meanwhile the description of most objects in these languages, for the natural sciences we are interested in, is as difficult as that in computational problem-oriented languages.

It follows then that one of important problems of the automation of developing the natural sciences is to develop a family of algorithmical languages where the emphasis is placed on the description of mathematical models of objects which are typical of certain sciences. Two ways are open here. The first way is the more traditional. It is aimed at constructing translators from certain languages. The model is described in a problem-oriented language if a translator is available. However, a corresponding computer program is designed by the translator automatically.

This has a number of shortcomings particularly noticeable in translating the modelling algorithms. The trouble is that the mathematical model under investigation often needs various amendments during experimentations. It is sometimes necessary to organize joint operation of the experimenter and computer during which the former cannot only change the model but also mock-up parts of the model or external environment thereof.

However, with the present state of translation art the introduction of an even insignificant alteration to the basic algorithm description necessitates retranslation of the whole program or its large unit. This makes the man–computer intercourse difficult and reduces the possibilities of effective experimentation with mathematical models. The way out is in the use of interpretation instead of translation. In interpreting, the initial declaration of the algorithm is preserved in the computer memory (with the accuracy determined by the method of coding). Therefore the task of introducing changes to the source algorithm and of organizing a man–computer job is solved trivially.

This, however, has drawbacks of its own as well. These mainly result, as a rule, in a marked reduction of computer speed during interpretation. To cure the trouble we must bring about not the program interpretation but the hardware one. In other words, computers proper are to be so constructed as to enable their integral language to coincide with or have maximum approximation to the source problem-oriented languages. This (which is only applicable to computational tasks and those of formulas transformation so far) is successfully realized at the Institute of Cybernetics A.S. Ukr. S.S.R. in Kiev. Note that the problem of enhancing the computer "intellect" is solved at the same time.

AUTOMATION AND MATHEMATICS

The development of methods of mathematical modelling signifies an essential increase in the role of deductive methods of developing natural sciences. Therefore a natural question arises about the ways of automation of the deductive method and the development of deductive sciences. A classical sample of a deductive science is, as it is known, mathematics. This makes the problems faced in attempts to automate the development of mathematics to be characteristic for all other deductive sciences.

The automation of mathematics development incorporates two main points: first—automation of the proof process (deduction from premise data) of a formulated proposition; secondly—automation of the process of concept formation and of statement of new interesting theorems to be either proved or refuted.

There now exist many attempts to solve the first problem—automation of proving theorems that have already been formulated. All these attempts have one of the same features. Each time they try to construct a sufficiently universal proving program capable of proving all (or almost all) theorems in some field of mathematics without a man's assistance. These attempts turned out to be successful only when applied to the first sections of the mathematics foundation—propositional calculus and predicate calculus.

This result is easily understood if one remembers that nobody tries to solve the problem in a much simpler field—computational algorithms—to construct the universal computational algorithm. Such an algorithm (being relatively universal at that) may only be found by combining thousands and thousands of particular algorithms constructed on a long run of development of computational mathematics. All the more this inference is right in a considerably more complex field than algorithms for proving mathematical theorems.

It seems to me that the most correct way of solving the complex problem of automating the deduction process in mathematics and in other deduction sciences consists in constructing operational systems capable of ensuring the shared work of the mathematician and computer. The basis of such systems is constituted by two special languages. One of them serves to write various mathematical theorems and their proofs. This language is much closer to the conventional language of mathematics than the language of a current mathematical logic. The second is the language of search for proofs. In this language the mathematician informs a computer of the direction of search for proof in a form similar to that used for directions given to a clever pupil. These directions are employed by a computer to narrow a range of various versions of proof.

The operational system comprises translators or interpreters from these languages, programs for constructing a tree of logical structures, and the so-called "computer algorithm of eventuality". The latter algorithm enables a computer to establish relatively simple consequences from the facts already known to it. In addition, the system incorporates a large information part. It is composed of all known theorems of a section of mathematics (or of several sections thereof) together with their proofs. The proofs should be written with such a degree of detail that each step of the proof would be "understood" by the computer algorithm of eventuality. It is desirable that the system be so organized as to enable it to operate together with many remote panels under time-shared conditions.

Working at such panels the mathematicians must, in the first place, solve the problem of a continuous stocking of the system "knowledge volume". For this purpose they insert into a computer the new results in the chosen mathematics field (in the course of their appearance) together with their proofs. In so doing the mathematician should "chew over" the proof so as to make it understandable to the computer algorithm of eventuality. If the algorithm of eventuality is not too strong, then such a "chewing over" can serve as a criterion of the fact whether the mathematician, working at the machine, actually understands the proof or not. This solves two problems at once—the problem of objective verification of proof correctness of published results, and the problem of unbiased checks for operation of mathematicians (predominantly students and postgraduates) studying new results.

There also appears the possibility of complete automation of verifying the correctness of a proof of new results being published. It is sufficient for this purpose to establish the requirement that all articles sent to mathematical journals be written in an appropriate language with such a degree of detail that each step of the proof could be verified on a computer.

If the mathematician, working at the computer, states a theorem of his interest, the described system allows, first of all, to evaluate the novelty of the theorem. When this is done, as opposed to conventional reference–information systems, this system does not only show the presence in its memory of *quite the same* theorem but it is capable of giving answers of the type "stated theorem is new but is an obvious corollary of such and such known theorems". In case the stated theorem is not obvious (from the computer standpoint), a search is begun of its proof or refutation in a joint man–computer work. Here the whole "non-creative" portion of work on technical handling of proof is taken by a computer. The mathematician is given the possibility to utilize fully his intuition and ingenuity in stating the direction of a search for proof.

With each new theorem arriving at the information part of the system, all proofs are rearranged automatically because the appearance of a new generalizing result may make obvious several particular theorems at once, each of which has needed for a special proof before. The general reduction of memory caused by such a reduction in proofs can serve as a quantitative measure of a *generalizing force* of the new result.

Now a way is open to solve the second problem—the problem of automating a statement of new results. To do this a random statement of new non-obvious results is effected. Then, prior to finding their validity or falsity, a value is estimated of the decrease in proofs of known theorems on the assumption that a given result is valid. This may give most interesting (from the viewpoint of their generalizing force) theorems for succeeding attempts of proof.

The notion of "interestness" of the theorem is, of course, not reduced only to the value of its "generalizing force". Of most importance is the "intuitive interestness" clearing the way for new practical applications of the result. It is not, however, possible to formalize this notion within the scope of one science. To accomplish this, the automation must jump on to a principally new level when the system of computers fully automates the development of the whole complex of sciences and engineering. Still, even in this case their final estimation and choice of the direction of searches are left to man, since the human society will be the consumer of results of this system.

The advantage of the suggested way of automating the deductive sciences is its gradual character and the possibility of obtaining practical results of importance in all stages of developing the automated system. At the same time this way opens unlimited prospects for raising the automation level. With this view in mind we must only continuously perfect the computer algorithm of eventuality. In the course of this improvement, sooner or later, all theorems now requiring complex and cumbersome proofs will become obvious for computers. As a result, most of the new stated theorems will be proved by a computer with no assistance on the part of a man.

Of great interest is not the simple programming of proof retrieval but the organization of a man–computer dialogue. At the Institute of Cybernetics in Kiev a computer has been built which is specially adapted to such a dialogue. So far the emphasis placed in the computer is not on arbitrary deductive constructions, but on the automation of formula transformations in algebra and analysis. The computer remembers basic relations to the type $\sin 2x = 2 \sin x \cos x$, $\int dx/x = \ln x + C$, etc. The result is that most of integrals usually given to students for calculation are handled by the computer (mathematically) with no programming. If a "prompt" is required on the part of man, a light-pencil screen and a panel for insertion of alphanumeric information (e.g. of new identities) are used.

CHAPTER M-8

Cybernetics and industry

F. H. GEORGE
Brunel University, U.K.

Summary

Cybernetics is the science of organization and planning, and also therefore of management and government. There are two important ways in which cybernetic ideas change the manner in which any sort of organization is likely to be controlled. The first is that it changes the attitude of the people in the controlling situation, and it does this by insisting that the system controlled should be regarded as dynamic and in evolution and the control should, therefore, be capable of being adaptive. This statement, of course, does not apply to a certain few static and non-evolutionary systems, but these are, in general, trivial special cases. The second important aspect of cybernetics in its effect on industry is to highlight specific techniques, mostly derived from the mathematical, statistical, and logical background. In particular, such methods as Bayes rule, Markov chains and the like, and in general heuristic methods, can be used within the structure of each dynamic model. The methods can be used independently of the dynamic model, but take on a somewhat different significance when used in it. Some examples of the cybernetic approach to industry will be cited to illustrate the two major themes, as well as bring out a number of other points, such as the relevance of digital computers to modern management techniques.

CLASSIFICATIONS

I believe that a useful start to a discussion of management cybernetics would be to provide a provisional classification of the divisions of cybernetics. I am not suggesting that such a classification is either unique or necessarily acceptable to all people for all purposes, since no classification is ever that.

Any classification should though at least attempt at this stage to try to break down what is now a vastly complicated subject, into convenient subdivisions.

Cybernetics is now traditionally the science of control and communication in animals, men, and machines, and is especially concerned with systems that are adaptive, capable of feedback, and are also in evolution.

The basic idea is that cybernetics is concerned with artificial intelligence, and is concerned with providing models of cognitive systems. The methods it uses involve mathematics, statistics, probability, logic, and natural language.

The principal subdivisions of cybernetics are, so it is being suggested

(1) behavioural cybernetics, (2) biocybernetics, (3) mathematical cybernetics, (4) management cybernetics, (5) educational cybernetics and (6) social cybernetics.

Behavioural cybernetics is considered first. The difference between this subdivision and cybernetics itself and its fundamental problem, that of artificial intelligence, is that cybernetics is concerned with all and every type of cognitive system, whether similar to or not similar to human cognition, whether synthetic or simulatory. Behavioural cybernetics, on the other hand, is only concerned with models that purport to simulate aspects of human behaviour, perhaps including animal behaviour. Animal behaviour might on the other hand be included under biocybernetics, and this is a reminder of the arbitrary nature of such classifications. We should indeed think of these subdivisions as overlapping and having ill-defined boundaries.

Biocybernetics is mainly concerned with neural nets and other software and hardware models of the nervous system and the special senses.

Mathematical cybernetics, as we have called it, is really made up of the field of automata theory, recursive function and set theory, mathematical logic, stochastic processes, and similar related topics.

All subdivisions are concerned with models, whether in hardware or software and management cybernetics is the main topic of this paper.

One word of warning should be uttered at this stage. A great deal of confusion is caused in science, especially in interdisciplinary sciences like cybernetics, because of a misunderstanding over what is entailed. A mathematician, for example, can well look at cybernetic problems which are couched in mathematical terms and claim that they are *not* mathematically interesting. The answer can be brief and to the point. The cybernetician's use of mathematics is in the interest of cybernetics, and we are not in general concerned with whether or not problems are of mathematical interest.

A FIRST LOOK AT MANAGEMENT CYBERNETICS

There are a number of central issues involved in the development of management cybernetics, and a whole world of applications which have still to be realised; indeed here is a mine of almost unbelievable riches merely waiting to be fully exploited.

There are two special aspects of the problem: (1) the attitude that cybernetics should engender and (2) the particular techniques than can be used in the cybernetic context.

The Basic Concept

The basic concept is that all systems which involve intelligent activity are adaptive and in evolution; they are clearly dynamic and involve what we know as both algorithmic and heuristic methods.

We now say that a business is like an organism and its board of directors is like a human brain; it directs and controls its activity, and should do so intelligently.

We have here the first hint of the need for discussion. Are we asserting that a business should be constructed on the same basis as the human brain or are we saying that there is an analogy between the two which is worth following up? We shall now try to answer this point.

The human brain in its relation to the control of the human body exhibits features which a business would do very well to copy. The reason for this is that there is a great deal in common in what is demanded of both systems. The objection to this view on the other hand is that the brain has evolved over millions of years and moreover there is no suggestion that it is the best possible control device. It has merely been sufficient to ensure the survival of the human being. It may be the most efficient available but this is not the same argument as arguing that it is the ultimate paradigm. In any case the details of the system it has to control are not the same in their structure, nor *necessarily* the same in their function as are involved in the control of the human body. We are therefore only asserting an analogy and perhaps a yardstick; therefore we can afford to think in terms of *synthesis* rather than *simulation.* By simulation we mean literally the copying of organismic methods, where by synthesis we mean achieving the same (or better) ends, irrespective of the methods. This point was part of what was entailed in distinguishing cybernetics itself from its subdivision called behavioural cybernetics.

THE ESSENTIALS OF CYBERNETICS

In this section we go back to our central theme of cybernetics, with management cybernetics firmly in mind. The central feature of cybernetics is that of artificial intelligence, and this is best understood as a simulation of human cognitive processes. Two focal points of artificial intelligence are natural language programming and inference making on computers. We should mention that these are two fields likely to make rapid development in the future. The same is indeed true of all the various decision-making processes and models, as well as concept and hypothesis formation.

What this amounts to is, that we can look at the human being as a rational information processor and therefore a typical cybernetic system. We therefore look at the human being and try and extract from our understanding of how the human works what is likely to prove most useful to our synthetic ends.

First of all we think of *input*, and all the ways in which information can be sensed. We can use human eyes, radar scanners, etc. The underlying problem is always that of *pattern recognition*. The problem is indeed to perceive and to recognize objects, processes, and whole classes of events.

The main point about the input is that it is never independent of what is in the central processor or human brain. The brain's job is to compare the snapshots (in the case of vision) with its stored record of previous snapshots, sometimes to achieve an identification, but more often to recognize that the event comes within a general class of events. Sometimes the role of processing is to say that it does not have a class ready to receive this new item and must therefore set one up.

Central processing receives much more by way of symbolized information than it receives by way of snapshots. Human beings learn most of what they know, or think they know, from other people's descriptions than they do by directly seeing, or otherwise sensing, things. In any case, we can set up a symbolized representation of a picture by describing it.

Given that we have a lot of symbolic models in store, we can then make comparisons, look for patterns in events, and perform abstractions to achieve "higher-level" patterns. These models are mostly in language, but many can be in mathematical form, when it is easier for us to draw the necessary inferences. This is why mathematical methods for assessing probabilities and risks, and making deduction inferences are an essential part of any information processing. This goes hand in hand with the ability to converse with other people and other computers. It can be seen from this that "humanlike thinking" is the target, and this is largely brought about by building upon the adaptive feature of learning. The main superstructure is that of symbolization which necessitates a hierarchical model.

It is artificial intelligence that remains the central problem of cybernetics. As a result of which, we still have the problem of bringing together the various cognitive features such as learning, information processing, perception, thinking, and the like into one coherent whole. Cybernetics deals with a whole set of methods in isolation which can be of great use to individuals who use such methods in order to help in the solution of problems which occur in virtually every walk of life. This is a main starting point for management cybernetics.

Cybernetics, on one hand, promotes a way of thinking. It insists on thinking in dynamic terms and dwells on the adaptive and evolutionary types of system. Plans and decisions of the most complicated kind, are often contingent and dynamic, and must deal with changing circumstances, and then the next stage of cybernetics lies in bringing all these detailed features of cognition together to make a complete set of artificially intelligent systems. The methods themselves are available therefore are available collectively as well as in isolation.

Another way of putting the same point is to say that in the future we shall be passing more and more into man–machine co-operation, and gradually this will be superseded by a "machine only" controlling system.

The reason for the transition from "man–machine" to "machine only" depends simply on the fact that as men interact with machines so in time the men, who stand to gain a little and lose a lot will leave the "decision" increasingly to the machine. This psychological factor is highlighted by the fact that the machine's evolution is inevitably far faster than human evolution.

Our next focal point is more completely on management cybernetics and this is dealt with in the next section.

THE ESSENTIALS OF MANAGEMENT CYBERNETICS

A commercial enterprise is, as we have said, in some ways like a human body; it needs a brain to control it. To achieve effective control requires eyes, ears, and the other senses to pick up details of the environment. This refers especially to *changes* in the environment, and at this point we should be reminded of the importance of "management by exception", that is the need for control where change occurs.

Any company needs to try to anticipate events. Thus sales estimating, market research, motivational research, accountancy records, and advertising anticipate, and attempt to pave the way for a successful business. They need to know what to make and how to make it, and they need to know what to store and how much of it to store. But this is possible only if they can successfully anticipate future demands. This is like the behaviour of an intelligent person's brain, and just as we lay plans as individuals, so any management group, must lay plans for the success of its company. This process of planning forms a major part of cybernetics. The communication part is obvious and the control is common, or uncommon sense.

Management cybernetics characteristically caters for decision making under conditions of uncertainty, an uncertainty which is due to incomplete information, in the same way as human beings are forced to do. The methods are usually mathematical or statistical and complicated and often require a computer to put into practice, but the results can be far better than even the best executive can achieve unaided. Business is something for human beings and something which taxes their intelligence and their ability to assess people and situations. Now whether one likes it or not, the world of the amateur planner is giving way to the world of the professional or scientific planner and the professional demands more and more of the techniques of modern science.

Operational research, computer science, organization and method, work study and many other techniques all play their part, but in the end it is the science of control and communication that provides the framework within which they all operate. In a large or a small company, information is coming from many sources and different groups supply and sieve the information. These different groups have different vested interests and biases and present their information accordingly. Senior management sifts this information and tries to reach agreement as to action which should be in the best interests of the company. All of this can be made more efficient by knowledge of the processes involved, by use of techniques which may be complex but are easy to understand and use, by having information sources made easily available and by following a cybernetic line of approach.

We are following the policy of copying human input–processing–output strategies and we are doing so consciously, leaving some of the work to the machine and some to the human. What, in practical terms, we now envisage is an operational control room, where information is sieved and pre-processed and by the use of natural language and inference making, the information relevant to some question is presented to the human decision taker.

Now to the more particular strategies and tactics. For the sake of exposition, we shall assume we are a military commander in a war, rather than a businessman at peace, although the way one method could be put to use in the other context should be fairly obvious. Consider a man in command of combined forces engaged in a battle. The situation is one where you have advanced overland and captured an important city. You have relative control of the sea but you are uncertain about control of the air; but you are clear that you have to push forward and to take further enemy bases on the way into the centre of the enemy stronghold, in order to win the war.

Decisions now have to be taken; you could attack immediately or you could delay to regroup your forces and make them more efficient as a result. As a cybernetician, you will now analyse the situation in which you find yourself in terms of motive and purpose, ends and means, and do it necessarily in probalistic terms. First of all, it is clear that the enemy wants to beat you and you must know something about the deployment of his forces and their strength. This is nearly in the domain of on-line computing.

The first question that arises is: was his retreat the result of us driving him back or is he going back as an orderly retreat or as a bluff or at least to strengthen his defensive position? Geographical factors and historial factors play a part in your assessment here, and the use of *heuristics* is most relevant. We must now explain what we mean by heuristic methods, since they are central to decision taking and planning, which is itself central to management cybernetics.

Heuristic Methods

Two contrasting ways of carrying out any operation are illustrated by heuristics or algorithms. An algorithm is a precise means for arriving at a definite solution to a well-defined problem. If you are asked how many men there are in that room, the answer is to count them. The method is precise and the answer clear; this is an algorithm at work.

If you have a situation such as a game of chess where you do *not* know the "best" move to make at any point of the game, you invoke general principles such as "control the centre of the board", "exchange pieces if it is to your advantage", etc. ... "to your advantage" is vague because you may get the better of the exchange of the pieces and yet lose in terms of overall position. These methods are heuristics. Good heuristics are valuable short cuts or guides and if we carefully record their use and the results of their use, we can steadily improve them and make them more precise. Through science we can sometimes even turn heuristics into algorithms. All of this is rather like saying that we can remove the uncertainty in a situation by increasing our knowledge of it. In a sense this is a truism; in practice it provides a challenge which we have to try to meet.

Off-line adaptive models or simulations are always possible and can have heuristic programs with language and logical facilities superimposed upon them, as well as algorithms playing the role of subroutines.

This whole field is closely bound up with existing methods of systems analysis and conventional computer programming and is part of an attempt to provide overall computer control of any system whatever, be it commercial, academic, industrial, or governmental. The complete plan will also involve an integrated approach to process and production control, although this will not be discussed here.

The first step in the conventional programming of a computer involves providing an overall block diagram or flow chart of the principal features of the system, e.g. of the whole of a company. The second step provides a point of departure which demands a conventional systems analysis of some subsystem or section of the total system. We should normally start with a section dealing with scheduling, planning, or some such activity which is both centrally placed in the system and convenient as a spreading point for further analysis. Such an approach can and must be made coherent with any existing data processing system, even though, or even especially where, payroll, and other well-established computer methods are already in use.

The next point is one of the most vital of all and it concerns the need for "rethinking". Having all conventional techniques such as linear and dynamic programming available, and thinking of our undertaking as involving an overall modelling, then to rethink one's purpose is quite central and vital. The modelling uses, in the main, logico-mathematical methods, and treats any system as being evolutionary. The emphasis, in other words, is on modelling and controlling a changing system. This implies the need for adaptive and dynamic forms of models.

Heuristic Programming

Granted acceptance of conventional systems analysis and programming, and granted agreement as to the philosophy implied above, we now proceed to the stage of heuristic programming. Heuristics are "general rules of thumb" or *ad hoc* principles, applied in situations where algorithms (detailed precise automatic procedures) cannot be applied for reasons of either economy or inherent difficulties. Playing chess provides an example involving economy while sales estimation, since it implies future unknown states, or, in other words, incomplete information, poses inherent difficulties.

In commercial and industrial systems, heuristics, such as exponential smoothing suggests, usually supply such rough guides. Sales models, often supplemented by market and motivational research, supply heuristics for marketing purposes.

Although one- or two-way interrogation in the computer can be wholly independent of heuristics, the very fact that heuristics are essentially linguistic, especially in the form of inductive generalizations, suggests a close relationship. Language is superimposed upon our conventionally constructed model. The idea is to permit easy communications between programmer (or an executive who need know nothing of computing) and the computer.

At this stage, questions and statements to the computer form the one way linguistic transaction which permits easy data retrieval. At present questions have to be in stylized English, but may soon be in ordinary English. They are now limited as to their domain of operation, but may soon be more nearly universal. The plan is to have a verbal description, supplemented by data, over a domain, e.g. the supply section of a company. This process is selective and can quickly elicit answers to pertinent questions by verbal discourse.

Logical inference is now used to supplement the data retrieval search; for example, if one asks how much the company spent on a raw material, say, during May, the computer may infer the answer through knowing the figure for April and also knowing that April was the same as May. This obviously trivial example illustrates a technique which can be made extremely powerful. This form of programming also suggests setting up axiomatic model systems for marketing, etc., which permit one to draw probable inferences, which, when supplemented by empirical data–from surveys and the like–give a clear and useful model of some aspect of the total system.

This is as far as we shall take this argument, since to set up even a relatively simple evolutionary model, where it is assumed that the models can be made adaptive, and the heuristics themselves are obviously capable of being adaptive, is a time-consuming process, although it is a necessary step to computerization of decision making.

Let us look back once more at our hypothetical military commander. The commander who is assessing his enemy's behaviour should regard such matters as vital. He may have a heuristic which says, "Commanders never withdraw voluntarily, or as a bluff, at certain times of the year, because a change of weather is about to occur which makes it vital to hold every yard." If such a heuristic can be invoked, then a probability of a fairly high value (near certainty) can be ascribed to his reasons for retreat; he was forced to.

The drawing of logical inferences clearly enters into decision taking, so we may now say "then he is in some disarray, since forced retreat is a source of chaos". We therefore look for evidence of chaos, which confirms our heuristic, and helps to determine our next move. Another heuristic may now be invoked. "If you are following a retreating army which is in disarray, harry them to the limits." By this time, our commander is ordering an attack by air, after his first reconnaissance flights, and preparing his troops for immediate advance.

But he now has a further consideration. So far, he has thought only in terms of *probable* enemy activity. What is the state of his own forces? They are tired, but capable of the job provided he is right about the enemy not bluffing. The question he should now ask is, "What will be the result of our attack if the enemy *is* bluffing?" The answer may be "complete defeat for us". In other words, "I am taking a great risk and should consider a second strategy". "What happens if I rest at my new position and consolidate it, and bring up some reserve troops?", and so the analysis continues. It is rather like a chess player weighing the alternatives, using common sense and logic. Cybernetics is precisely the process, in its decision taking role, of insisting on the use of common sense and logic. There are though two factors which make the cybernetician rather different from the usual decision taker. They are that he makes his arguments *explicit* and this allows the detail to be checked and verified, and allows him to use a computer for his purpose.

The insightful commander accumulated his experience, but finds it difficult to hand his experience on to others, because it has not been made explicit, and he has, therefore, called his decision making a process of hunch or "having a nose for things", "smelling a rat", etc. But this *can* all be made explicit and as a result becomes much more powerful. Exactly the same argument applies to the business executive.

Management Cybernetics in Practice

Let us look now at another example. Imagine yourself as a senior executive faced with the need to make a decision as to whether or not to take over another company. You want to know what the new company's profits are, you want to know the profit margins in the trade, and you also want to know all sorts of details about their past results, present state and plans. All you have to do is say, "What are ...", and finish the sentence appropriately and your assistant types out the question on the terminal which is "on line" with a time-sharing computer system. The question does not even need to be in specially selected phrases and the assistant need not translate or code it, indeed you can type out the question yourself in ordinary English and you wait only a few seconds to read the typed answer from the terminal teleprinter. This is all made possible by what is called natural language programming, but can be done almost as well without even this facility.

If you ask a question which is not merely a matter of fact but a matter for logical inference or statistical analysis, then the whole situation is obviously more complicated. Now the computer has to perform the necessary analysis—whether mathematical, logical, or statistical—so it may be a minute or so before the answer is available. But this compares with an hour or two if you use a human mathematician. Such methods, which form an important part of management cybernetics, provide ways of streamlining the whole decision making process; it does not matter whether facts or inferences from facts are needed, the result is the same. The only thing that is different is that one takes longer than the other.

Even if your information is incomplete, and even if you have to include guesses and estimates of such things as other people's trustworthiness or other facets of their character, this can be done by using heuristic methods, and appropriately weighting the relevant model.

The mathematical methods play a vital part. There are various forms of decision processing such as are involved by the rules of probability; you may assess the probability of some event and act accordingly, or combine the probabilities for various possible events with the *desirabilities*, and combine them to find a "best" decision. Here you may be prepared to take a risk on the probabilities for the sake of the desirabilities.

Let us look next at what is sometimes called risk analysis. You are asked to place a bet on some event such as horse race or a dog race. If the amount involved is only £1, you may not bother about losing and place the money on an outsider. The reason is that you are risking very little and if you are wrong, you will have little cause for regret. If the amount involved in the bet were £ 1,000 you may feel quite differently and invoke all sorts of ways of defraying your losses; you may hedge your bet, or distribute it amongst the favourites. Now your risk is great and your regret, if you are wrong, can be very great.

In the case of the commander taking military decisions, or in the case of political decision taking—or indeed in any decision taking—minimizing regret (or minimizing risk under circumstances of great threat) is as important as weighing the probabilities of events in the first place. It is easy to see in these circumstances how strategies can range from complete optimism to complete pessimism. What is required is an assessment of an assessment and taking *all* the *relevant* facts into consideration.

It is, at this stage, the man–machine relationship which is so vital. A man works with a computer terminal so that the computer can use its speed, accuracy, and ability to handle complex material, while the human being uses his flexibility and his gift for quick insights and sense of relevance. But as time goes by, more and more will the man–machine phase of development be replaced by the machine phase, as we have already said. The difference though is only one of degree.

Selecting Relevant Information

We are still imagining the man–machine phase of development, where the executive is using a streamlined, yet easily handled, terminal—perhaps with a human assistant to obviate the need for "programming in" the natural language translation. The key issue is relevance.

Both large and small companies receive information from many different sources: large numbers of people hold, supply, and sort information. These people have different vested interests and biases and present their information accordingly. Management itself appraises this information and tries to reach agreement as to what action would be in the best interests of the company. All of this can be made very much more efficient by knowing the processes involved, by using techniques which may be complex but are easy to use, by having information sources made readily available, and in general by following a cybernetic line of approach.

Computers generate a great bulk of paper in the form of computer output. The judicious computer manager ensures that the output is in readable form and does not, if he can help it, overload the receiver. In practice this approach produces tables and other data that leave the executive confused. The result is not always, or even usually, effective. Cybernetics does it very differently. It assumes the board wants only to know what is relevant to a particular problem.

Relevancy is the essence of the whole thing. It is all too easy, as every director knows, to swamp his board and his management with information that is useless. The uselessness stems from many factors: sheets of tabulations are not easy to follow, and whereas one can read the data it is not usually easy to relate those data to the relevant considerations under revue. This is one reason for the development of natural language programs.

Finally, we believe the next phase in this work is the generation of new heuristics ("new ideas") and this will come directly as a result of applying every kind of logic—both deductive and inductive to the model, and necessarily therefore primarily to symbolized models. The word 'primarily' serves as a reminder that we are also likely to become increasingly involved in the translation of symbols to graphs and pictures, and vice versa, as well as such translations as numerical tables into language. But these exciting and challenging steps still remain largely in the future and go beyond the scope of the present man–machine phase of management cybernetics.

SECTION I

The meaning of cybernetics

(Definition and Philosophy)

CHAPTER I–1

Philosophy and cybernetics

WOLFE MAYS

University of Manchester, U.K.

Summary

In this paper I shall examine some of the arguments used to demonstrate the logical possibility or impossibility of constructing machines to simulate the conscious activities of human beings. Although there are no logical reasons why creative intellectual processes might not be mechanized, there could be important physical and economic reasons which might stand in the way of such a project. Even if the proposition "this machine is conscious" is not self-contradictory, the question whether it would be true in some future state of the world is something about which logic tells us nothing. On the other hand, philosophers have tried to show the impossibility of constructing machines to simulate the conscious behaviour of human beings. Some of these arguments will be critically examined in this paper. The most important of these are the following: (a) claims that owing to the infinite variety of the human mind, it is impossible to give a completely deterministic account of thought or any conscious activity; (b) claims that Gödel's theorem shows that a man can transcend a machine insofar as he can see the truth of a theorem, which can be expressed in a formal system even if it is unprovable in it. It hence concludes that a machine cannot be an adequate model of the mind.

INTRODUCTION

My interest in the field of cybernetics dates back to 1949 when together with Dr. Prinz then of Ferranti's I was involved in the construction of an electrical logic machine. This machine is now in the Science Museum, London, as a dated piece of technology should be. Unfortunately, there is no philosophical museum, except perhaps international congresses, for exhibiting outworn philosophical arguments, especially those dealing with the possibility or impossibility of constructing devices to simulate human behaviour.

From about 1950 there has been a considerable amount of discussion among philosophers and cyberneticians, as to whether machines can be constructed to simulate the conscious activities of human beings—their thinking, feeling, and willing. This discussion was sparked off by Turing's well-known paper on "Computing machines and intelligence." Turing identified the question, "Can a machine think?", with the experimental question, "Can a machine be built which would play a guessing game so well, that it would deceive a man into believing that it was a human being?" (cf. Turing[1]; for another point of view see Mays[2]).

Some philosophers have argued that the question whether machines can be constructed which can think, feel and will, is merely a verbal one of how such words as "thinking", "feeling", and "willing" are defined. If we define them behaviouristically then machines do think, etc., if we define them subjectively they do not. In ordinary discourse, however, we generally use such words as "thinking" to refer not only to behavioural data, but also to conscious states. The meaning we give to these terms is not then merely a matter of convention, but depends on what aspects of the psychological situation we are interested in. If we now, for example, define the word "thinking" so as to exclude conscious states, or to refer exclusively to behavioural activities, we thereby exclude a large number of experiences normally put under this head, and thus emasculate the meaning of the term.

On the other hand, the basic strategy of some cyberneticians has been to assert baldly that in principle a machine could perform this or that activity, and then dwell on the difficulty of proving the opposite. "Who is to say that we cannot, in principle, build a machine that has the same emotions that humans have?[3]" Apart from the difficulty of proving the negative of such an assertion, one can always imagine qualities which a machine might have which would bring it closer to a human being. Clearly the only limit to this sort of thing is the flexibility of one's imagination.

As modern technology develops, so the argument runs, it will be possible to construct machines which will not just simulate our behaviour, but also be conscious. The apparent semantic irreconcilability of the terms "machine" and "consciousness" will vanish. With advances in technology it will make perfectly good sense to speak of machines being conscious. "... there is no *inherent* logical contradiction involved in calling a machine conscious."[4]

Philosophers who understandably enough are reluctant to accept the identification of minds and machines, have been criticized on the ground that they are "... people who think they know it is *a priori* impossible for such models to be made to function in the same way as humans."[5] I would, however, not wish to rule out this possibility myself. Very few things are impossible in the sphere of logic, except blatant self-contradictions. There is, for example, nothing logically impossible in the idea of a sheep being a rational animal, although it is incompatible with the present state of things in nature. Similarly, there are no logical reasons why creative intellectual processes might not be mechanized. "... the only reason I know to suppose that the creative intellectual processes of man cannot be mechanized is the weak one that it has not yet been done."[6] There might, however, be important physical and technological reasons why this was not practicable. For example, man might be the most economic exemplification of himself.

Despite its apparent science fiction character, I would not reject the possibility that machines may yet be constructed out of physico-chemical elements, having inner lives of their own. This could well be another stage in the evolutionary process in which social evolution has been conjoined with nature to produce a new sort of being. It is true that it would not be the sort of machine we know today. An example of what this might be like is to be seen in Capek's "R.U.R.", when two robots, male and female, accidentally touch each others hands, and feel a thrill running through them—they have fallen in love. The production of such artifacts might lead to interesting legal and ethical problems. If, for example, such an artefact was destroyed by a human being, would the latter be guilty of murder or the wanton destruction of someone else's property?

There is, of course, nothing new in postulating beings having some or all of the characteristics of human beings; mythology has had a long start over the cyberneticians here. Further, not only may human beings be simulated by artifacts, they may be improved upon. Theologians have argued that it is possible for angels to exist; beings who are in all moral and intellectual respects better than us mortals (except perhaps the subclass of fallen angels). Just as we are better than machines at seeing the truth of Gödel's theorem, so angelic beings might see at a glance the infinite hierarchy of Gödel propositions.

Most cyberneticians seem to accept a brain-dependency theory of the mind – a nineteenth century scientific materialism in which our conscious states are assumed to be dependent on (or identical with) our brain states. When such a statement as "damage to the occipital lobe of the brain involves disturbance of visual perception" is put forward as evidence for the brain-dependency theory, the ambiguity in the phrase "visual perception" is usually overlooked, namely, that it can be given either a physiological or a phenomenological meaning.

This approach has recently received a new formulation in a weakened form by the so-called identity theory, which asserts that it is not logically impossible for mental states to be brain states. This attempt to identify the two overlooks that in some ways we know considerably more about the former than the latter. It would certainly be difficult, for example, to identify a particular ethical decision, such as deciding to be honest in the face of temptation, with some specific brain process. Only in simple stimulus—sensation situations to which the Weber–Fechner laws applies, can such a direct relationship be shown to exist. In the field of cognition, we can usually specify the psychological state much more precisely than we can the corresponding brain process.

Even if we knew what these brain processes were, it might still be simpler, because of their complexity, to use our introspections or observable behaviour. The attempt to imitate the behaviour of a human being by artificial devices is not one whit more scientific than taking into account his introspections. In discussions dealing with the question, "Can machines think?", cyberneticians often tend to talk about machines in general. A universal machine is conjured up having none of the practical inadequacies of any existing machine. As no particular machine is specified, the whole discussion does not refer to actual concrete machines doing specific jobs of work, but rather to ideal machines or pieces of logic.

PHILOSOPHY AND INTELLIGENT MACHINES

Philosophers, on the other hand, as cyberneticians have noted, have attempted to show that in principle machines could not be constructed to simulate the conscious behaviour of human beings. I will examine three such arguments.

(a) It is argued that in view of the unspecifiability or infinite variety of the factors involved in thought activity, it would be impossible to give a completely deterministic account of thinking or any other conscious activity,

"... scientific enquiry can never exhaust the infinite variety of the human mind".[8] At the most, however, this objection is a practical one, that in view of the inherent complexity of the psychological, social and other data involved, it is highly unlikely that we shall ever be able to give a deterministic account of conscious phenomena.

(b) The second type of argument stresses the spontaneous nature of intelligence, and is in a sense complementary to (a). It starts from the position that the proper function of intelligence is to solve problems, in the sense of finding means which are adequate in relation to a certain end. "If the means are never inevitably determined by the end in view, then it will always be necessary to have recourse to intelligence. ... This consideration excludes the possibility of constructing a machine capable of solving any problem whatever."[9] Such an approach assumes that it is impossible in principle to give deterministic explanations of problem solving. It is intelligence alone through its spontaneity which can find means adequate to the proposed end. This kind of explanation is perhaps applicable to the more intellectual types of problem solving occurring, for example, in mathematics. It does not, however, rule out logically the possibility of problem solving in terms of say, Skinnerian learning theory.

(c) appeals to Gödel's theorem, which demonstrates the impossibility of proving the non-contradiction of arithmetic or of any deductive theory solely by methods borrowed from this theory or from weaker systems such as logic. Gödel's theorem, we are told, seems to prove that mechanism is false, "... given any machine which is consistent and capable of doing simple arithmetic, there is a formula which it is incapable of producing as being true—i.e. the formula is unprovable-in-the-system—but which we can see to be true".[10] Man can thus transcend a machine since he can write down a true theorem expressible in the system, even though it is unprovable in it.

Once again we are faced with a Cartesian dichotomy: minds are essentially different from machines, a mind can always go one better than a machine, which is in some sense definite and finite—no machine can be a complete or adequate model of the mind. However, most cyberneticians would be content to build machines capable of simulating some desired piece of mind-like behaviour including the ability to understand Gödel's theorem.

When philosophers talk about minds in this context, they would seem to refer to a normative mind—a transcendental ego—which is in direct contact with universals, not to some actual mind such as that of our next door neighbour's. Minds as they occur in nature come in all shapes and sizes, are of varying degrees of intelligence and are found in very different cultural contexts. We cannot legitimately argue from the fact that a limited subclass of men, namely, mathematicians and logicians, understand Gödel's theorem, to the nature of the human mind in general. Unless, of course, we assume there are non-empirical *a priori* principles common to all minds which brings us back to the mind in general.

Although it is recognized that human beings are not confined to making deductive inferences[10], nevertheless, it is a contingent empirical question whether we think in a step-by-step logical manner, in terms of *Gestalten*, or by flashes of insight. It is not a formal question. In any case, this whole account of the way the "mind" sees logical truths as self-evident is a completely a-historical one, with no reference to the society or culture in which the particular individual finds himself. As Sartre might say, in such a case we put essence before existence.

An interesting feature of Gödel's theorem is that a stronger system can always be introduced in which a proof of a Gödel sentence can be given, and since this is a process which could be continued indefinitely, a hierarchy of machines could be arranged to do this. Commenting on the attempt to show by the use of Gödel's theorem that men can do certain things which a machine cannot, Good argues, "This point of view is essentially refuted by the observation that Gödel's construction could itself be carried out by another (deterministic) machine."[11] This will give rise to a new Gödel construction and a new formal system and so on. He imagines a human operator playing a game of one-upmanship against a programmed computer, where we proceed to higher and higher systems in the manner outlined above. He concludes, "If the mentalists still wish to make a case they must base it on transfinite counting rather than on Gödel's theorem." However, this takes us beyond the discussion of machines and their limitations, and introduces basic questions of mathematical ontology relating to the formalization of transfinite counting.

The one-upmanship game is, however, a little more complex than Good assumes. To play it we have not only to assume a hierarchy of machines, but also a transcendental subject constantly concerned in establishing the non-contradiction of lower-level systems in terms of constructions at higher levels—in fact we need an intellectual artifact. All that the protagonists in this game could then conceivably do is perpetually to checkmate each other.

They could not remain one-up for very long. On the other hand, it might also be argued that the Godelian concept of non-contradiction by subordination to a higher system is only relative to our human and limited way of apprehending mathematical entities. An angelic (Platonic) being might see at one glance the intrinsic properties of these entities. It would seem then that whether or not minds transcend machines is not entirely divorced from the question of the sort of minds one posits.

It would therefore seem to be doubtful whether philosophical arguments in terms of the logical possibility of constructing devices for simulating conscious activities, in any way strengthens the case for or against their practical possibility. As far as logic is concerned, there is nothing to choose between the propositions "machines are conscious" and "machines are not conscious": neither is self-contradictory. The question whether the former proposition could be true in some future state of the world, is something about which logic tells us nothing.

MAN AND MACHINES

Philosophers have recently criticized the view that machines can be constructed to simulate human beings on the ground that a machine cannot refer to itself in the same way as a person can. For example, in the case of a machine reporting the fact that it is malfunctioning, say by auditory or visual signals, it is another machine (i.e. part) which is commenting on this malfunctioning, since machines, unlike persons, it is argued, are subdivisible into parts. On the other hand, when a person reports he has erred, there is no distinction between an erring element and a reporting element. The "I" he uses, as a reporter, namely, "I have made a mistake" refers to the whole of him.

It does, of course, seem a little odd to talk of one part of a machine commenting on another part malfunctioning. The reason for this lies in the fact, that this is simply an interpretation of such a situation by a human trouble shooter. Machines, at least, as we know them today, could not be said to comment on themselves in the way a person does. On a physical level all we have is a transmission of energy, since when the part X malfunctions, it sends a signal to another part of the machine Y, which actuates, say, a light which acts as a warning to a human operator. Nevertheless, since a change in the malfunctioning part actuates a change in the reporting part, it does in a special technical sense "report" the error.

In dealing with ourselves as persons we are admittedly on a different level—that of immediate self-awareness—than when dealing with machines. We may note here Sartre's distinction between "being for itself" and "being in itself". In dealing with machines not only do we take up a third-person approach instead of a first-person one, but we assume that such devices are insentient things, not having the qualities of consciousness and uniqueness, which we would assign to ourselves and other persons. We would say that it would be odd for a machine to apply the first person pronoun to its activities, because we normally identify its use with self-consciousnes.

The concept of a person can be taken in two senses (a) the legal sense, in which there is immunity to temporal discrimination or (b) in the sense of self-consciousness. Historically, however, a *persona* meant an actor's mask and later the concept of a person was applied to the part a man played in legal transactions. Indeed, the social aspect of a person, the part one plays in life—seems to be highly important in the development of this notion. For all practical and legal purposes, despite all the changes a man may undergo, we still say he is one and the same person. Further, we assign certain rights to him immediately he is conceived, so that a human being may have the legal status of a person without necessarily being self-conscious in a reflective sense.

Philosophers usually take (b) as the distinguishing mark of a person, namely, that he is self-conscious, i.e. aware of himself as having perceptions, feelings, thoughts, desires, aspirations, etc. It has indeed been seriously suggested that legal questions relating to abortion might be settled if we knew, for example, the exact moment at which the human embryo became conscious. However, unless one bases the concept of the simplicity and uniqueness of the person on something like a self-evident intuition, and takes this as being the only guide to concrete reality, one cannot rule out that on, for example, the Freudian analysis of the mind in terms of the id, ego, and super-ego a person would not be regarded as simple and unique.

Although it may seem odd to talk of a machine erring, present-day common usage is not always the touchstone of philosophical truth. One can always imagine situations, where it would not seem odd to talk of a machine using the personal pronoun to refer to itself. As I have already argued, I do not think that one can rule out, at least in principle, that a highly advanced type of machine might develop consciousness and introspection. There is nothing self-contradictory in such a conception. Machines might be developed, which would meaningfully apply the first person pronoun to themselves. One must not forget that Aristotle regarded slaves as human tools and Descartes thought of animals as machines.

If all that philosophers mean when they say that machines are not persons is simply that when machines as we know them today, indicate a state of malfunctioning, then they are not self-conscious in a way a person is self-conscious of himself when he has correctly said that he has made a mistake—I would agree with them. On the other hand, I would disagree with them, if they are also legislating for the future, saying that it would be impossible for such devices ever to develop self-consciousness. The difference between us on this matter is perhaps not so great ar may seem. If such types of machine were ever developed, they would be very different from those we know today. The problem would then resolve itself into a classificatory one: ought these devices still be termed machines, or ought we to say that they are really a new type of human being produced by artifice rather than nature?

CONCLUSIONS

What I have tried to do in the main body of this paper, is to make the simple point which Kant made in his criticism of the ontological argument for the existence of God: namely, that from the fact that you can conceive something as existing it does not necessarily follow it does (or can) exist physically. I have also indicated that one must exercise caution in using arguments which proceed from conceptual impossibility to the non-existence of empirical things. Although I agree that the latter type of argument seems to have a greater plausibility than the former kind.

I think philosophers may get themselves into difficulties if they try to instruct natural scientists as to what sort of facts they may discover in nature and what sort they cannot. One merely needs to recall the errors of the philosopher Bergson when he challenged the results of relativity theory on philosophical grounds. Even the laws of physics as established at a particular period are not always sacrosanct. I need only mention Galileo's story of the Jesuit professor of philosophy who refused to look through his telescope to see the moons of Jupiter on the ground that on Ptolemaic astronomy, they could not exist. On hearing of his death shortly afterwards, Galileo acidly remarked: "When the learned Father was alive he refused to look through my spy glass to see my celestial trifles. Now that he is dead and gone to heaven, he cannot help but see them."

Since, as I have already noted, man is probably the most economic exemplification of himself, it is possible that in what one may broadly call the field of cybernetic studies interesting advances may come in the future more from the side of biological or genetic engineering than from mechanical or electronic engineering. After all we now have our heart transplants, and perhaps sometime in the future we may have our first successful brain transplant. These developments would no doubt give rise to fascinating philosophical and legal problems relating to self-identity. I can imagine some future Merleau-Ponty analysing the experiences of some brain transplant patient, and showing their relevance to the phenomenology of bodily perception.

References

1. A. M. Turing, "Computing machinery and intelligence", *Mind* (1950).
2. W. Mays, "Can machines think?", *Philosophy* (1952).
3. F. H. George, "Could machines be made to think?", *Philosophy*, p. 341 (1956).
4. D. Thompson, "Can a machine be conscious?", *Brit. J. Phil. Sci.*, p. 37 (1965).
5. F. H. George, "Finite automata", *Philosophy*, p. 59 (1958).
6. I. J. Good, "Human and machine logic", *Brit. J. Phil. Sci.*, p. 146 (1967).
7. M. Knight, *Sci. News*, **27**, 123 (1953).
8. J. R. Lucas, "Minds, machines, and Gödel", *Philosophy*, p. 127 (1961).
9. E. W. Beth and J. Piaget, in *Mathematical Epistemology and Psychology*, D. Reidel, (1966), p. 118.
10. J. R. Lucas, ibid, pp. 113–19.
11. I. J. Good, ibid, p. 144.

CHAPTER I–2

Cybernetics as a universal philosophical theory

D. B. FOSTER

Consultant to Industry, and Writer, U.K.

Summary

A comparison as between the techniques of electronic computers and the phenomena of the natural world indicate a parallelism and the most general aspect is that the materiality of nature is digitized into particles and the radiation of nature is digitized into waves. It is suggested that the existing fundamental category of mass–energy applied to the physical category is now replaced by a more fundamental category of data and intelligence in which the following correspondences are suggestive.

Particles	Primary marks or digits
The table of chemical elements	Data memories ("alphabet")
Chemical compounds	Data memories ("words")
Chemistry	Data processing
Active compounds such as DNA, RNA, enzymes, etc.	Data programmes
Radiation	Data transmission

Intelligence is defined as that difference between two sets of data which permits one to be programming for the other and the case is made that the natural universe shows overwhelming evidence of such data differentials. The highest intelligence will be associated with the fastest frequencies which have a greater data content capability and this indicates that the "sea of cosmic data radiations" can be equally regarded as evidence for cosmic intelligence of a very high order. The application of such concepts to man and society suggests a modified confirmation of the views of Père Teilhard de Chardin put forward in his book *The Phenomenon of Man* but with his

"noo-osphere" held together by the repeating life of cybernetically evolving individuals. A new theory of life after death is stated depending upon the ability of the individual to make contact with the cosmic intelligence through the window of discrimination which we normally know as "the conscience".

INTRODUCTION

The most general aspect of cybernetics has the most profound philosophical significance for man.

Until this century the most universal law of physics was the second law of thermodynamics which in general expresses the view that the universe is running down to an ultimate and chilly extinction. By way of practical and homely example it would also tell us that if you take one gallon of hot water and one gallon of cold water to make two gallons of tepid water mixing them together then you can never get your one gallon of hot water back again. The second law of thermodynamics is the law of hopelessness for the future since it tells us that the level of order in the universe can only decrease and thus the signpost to the future is marked "chaos".

It was left to the nineteenth century physicist Clerk Maxwell, who was a professor at Kings College, London, my own Alma Mater, to point out that there was one possible flaw in the universality of the second law of thermodynamics in that, if in such a system you postulated a "conscious sorting demon", then that demon could indeed consciously operate a trap door in a middle partition of that two gallons of tepid water admitting molecules of hot water to one side and shutting cold molecules off to the other side so that finally you could recover your one hot gallon.

Clerk Maxwell's voice was a lone voice in the wilderness of physics indicating that the second law of thermodynamics *could be reversed* and that consciousness could decrease entropy.

It was a *wild* thought until the advent of cybernetics but now it becomes a sober thought.

The father of cybernetics was the late Dr. Norbert Wiener and when I met him at a conference in Moscow in 1960 I asked him just what was meant by the word 'cybernetics' which he had invented.

"Why", he said, "it is quite simple. Cybernetics means 'the art of the steersman'," and indeed we must realise that this is the almost same as the art of Clerk Maxwell's conscious sorting demon. I think that then for the first time I understood a glimmering of the profound significance of cybernetics in a world otherwise dominated by the second law of thermodynamics since if cybernetics is the art of the steersman including the art of conscious sorting, then this is a master law of hope to set against that pessimistic second law of thermodynamics and thus there could be two signposts to the future, one stating "to chaos" and the other and cybernetic signpost stating "to hope and evolution".

Today, in my address to you I would like to take that hopeful cybernetical signpost to its logical conclusion.

THE INTELLIGENT UNIVERSE

The case I would present to you is essentially the same as the one I presented to The British Association for the Advancement of Science last year but with a somewhat difference of emphasis and introducing a new theory at the end relating to the possible immortality of man.

On that occasion I presented "the theory of the intelligent universe" which is the idea that, whilst from the physicists point of view the universe is considered essentially as a structure of matter and energy, from the cyberneticians point of view it is now beginning to appear possible to consider it equally legitimately as a structure of data and data processing so that we may look upon an acorn as a programme for an oak tree.

There are outstanding examples to support the view that at least some natural aspects of the universe are of a cybernetical nature and although I shall make the case later that such cybernetical principles are universal let us start with the sector where the evidence is most clear and that is in the newest science of all, molecular biology.

In molecular biology that remarkable large molecule desoxyribonucleic acid ("DNA") has the power of organising organic life in a cybernetic purposeful process so that even acorns get their ultimate instructions from it. Yet the analysis of DNA only indicates a construction of otherwise dead atoms of hydrogen, carbon, nitrogen, and oxygen. And yet the DNA molecule organizes itself into systems which can only be explained in terms of codes and information theory and the molecular biologist is having to turn himself into an information theorist and cybernetician. If you read an article on molecular biology it is full of expressions such as "dual language dictionaries", "information transfer", and so forth.

Thus we have the fact that nature, in the field of organic life, is using cybernetical processes at an extremely high level and we are only just beginning to unravel the codes and data processing involved.

Now, prior to this development, one might have imagined that organic elements such as carbon and hydrogen were mystically utterly different from inorganic materials such as sulphur or chlorine, but it now appears that there is no real evidence for this and carbon will combine with sulphur to make carbon di-sulphide and hydrogen will combine with chlorine to make hydrochloric acid and indeed this latter compound is also used in our digestive processes. Thus it appears that the essential difference between organic and inorganic matter is not related to the existence of special kinds of atoms but is simply due to the fact that there is some organizational command data process at work which we do not yet understand.

Thus what we are really witnessing in the field of molecular biology is the cybernetic process at work whilst in the inorganic field perhaps it is a non-cybernetic process of chemistry falling more under the laws of thermo-dynamics. Somewhere in molecular biology Maxwell's sorting demon is cybernetically at work creating order from chaos.

But let us next turn to the individual atoms from which any form of matter, organic or inorganic, must be constructed, and remembering we are now bringing information theory concepts into the organic field, then one sees a remarkable thing which throws a cybernetic slant onto the whole matter.

In the first place the whole of matter at the atomic level is digitized and it is digitised three times over. By the number of nucleons in the nucleus, by the number of electrons in orbit round that nucleus, and finally by the quantum numbers which hold those electrons together in cosmic rhythm in those orbits. Thus when we talk about nitrogen or uranium one could equally replace these words with the numbers fourteen and two hundred and thirty-five respectively, their atomic numbers. It is a long time since the Mendeleef table of atomic numbers was stated but now, as cyberneticians, we begin to realize that these numbers have a special significance in themselves as pure data with coding potentialities.

Surely it must be obvious that the essential nature of matter is that the atoms are the alphabet of the universe, that chemical compounds are words and that DNA is rather a long sentence or even a whole book trying to say something important such as "elephant", "giraffe", or even "man".

It is bizarre, but I think it makes sense.

Now, if one scrapes the very bottom of the physicists barrel, one arrives at an indivisible something called Planck's constant or one electrical wave and I find it significant that this is also the cyberneticians ultimate unit, the binary digit. Since it is in just that polarized oscillation of a wave as between a positive half-cycle and a negative half-cycle which establishes the elemental differentiation of information. In any sort of electrical information theory the ultimate building brick is one electrical wave and whether this building brick is the foundation for a radio information system or if you are to square the wave off and call the first half "one" and the second half "zero" and thus have your binary system, it is all the same.

Indeed, in this book I have just finished I refer to Planck's constant as "the cosmic binary digit".

Thus we find that those atoms of the Mendeleef table of chemical elements are themselves made up from ultimate units of Planck's constant which itself is the prime form of natural data. I suggest to you thus that it is at least an information universe in an objective sense.

Thus my analysis of the situation is as follows.

(a) Firstly, there is the information concept of "the ultimate mark" or what the physicists would call Planck's constant.

(b) Secondly, the system is organized into atoms which create an "alphabet" for matter.

(c) Thirdly, there are the simple compounds of both inorganic and organic chemistry which are "words".

(d) Lastly, there are the complex constructions of DNA which are sentences and even books with popular titles such as "giraffe" or "elephant".

So far I have only referred to a hierarchy of information relating to matter. But radiation also comes under the same rules since it is built from the individual waves of Planck's constant, and I do not need to remind you that man-made radiation systems such as radio and television are wave constructions for data transmission.

Thus whether one looks at matter or radiation you can certainly state with some certainty that the whole thing looks primarily like an information complex and at the higher organic levels this information complex is being data-processed with some cybernetical intent. At some level Norbert Wiener's steersman or Maxwell's sorting demon are taking the situation over. After all Shakespeare did not start from scratch, he took over the available English language and at some level perhaps some master artist takes over Mendeleef's table and unmodulated radiation to insert a trifle of cybernetic programming.

It is "curiouser and curiouser" but I think it true and there is worse to come.

One establishes a new picture of the universe as a digitized universe, as an information universe, but I think that because of the strong cybernetical influences at work I prefer to call it "the intelligent universe".

Now what I have said so far is based on the facts of science including cybernetics and it accounts for a great deal of the phenomena of the physical universe. But what I have so described is essentially what I described to the British Association last year and is thus by now mere data. I now propose to process that data philosophically.

PHILOSOPHICAL SPECULATION

My starting point for this philosophical speculation is that man himself is certainly by no means the most intelligent level in the universe in that the structure of organic matter reveals levels of technological excellence which man has far from emulated. Thus the structure of the human eye reveals a delicacy of design and function far higher than that of the most complex man-made optical system. This is the famous "argument from design" but now with a difference that credibility for that argument is re-inforced by the analysis I have given as to the whole of matter having a digitized structure of a form ideally capable of containing information and thus capable of being dataprocessed. Thus we no longer postulate a Creator who somehow magically creates "at a distance" from nothing but we are beginning to unravel just what are the coding methods and data processing tricks which are the means if not the prime causes of creation. Indeed my analysis logically cries out for "a programmer" of some capability as the only missing element in the situation.

But most significant of all is that the structure of the DNA molecule and its functions reveal a cybernetic information system of such formidable capacity and utterly outstripping the human cybernetic art as used in electronic computing systems. Thus we are faced with the certain fact that man is not the most intelligent level in the universe. Furthermore when one realises that the tremendous intelligence associated with DNA somehow embraces all forms of matter right down to that ultimate information element of the single electrical wave which the physicists will still insist on calling Planck's constant, then one has little choice but to take the view that the whole universe is somehow riddled and permeated throughout with intelligence. Not only is the basic structure of matter so formed as to be capable of taking part in digitized data processing but the results of this data processing are manifest before our eyes in the extraordinary structures of organic creation.

But where is this higher intelligence, where does it reside, and what form or mechanism does it have, if any?

I think the key to finding an answer to this must come from information theory in that we are now searching for an intelligence which can control matter which itself appears to have a digitized data construction. Thus what we are now looking for is the cosmic equivalent of paper tape programmes. This means that we must cast around for a form of data handling which stands at a higher level than the data of matter. The answer stares us literally in the face in that this higher data can only be radiation. As an automation consultant myself, whenever I design a control system for a process it is axiomatic that the speed of the control system must be faster than that of the motions of the process concerned and that is because the thinking of the control system must be way ahead of what is happening in the process and this can only be so if it has an inherent speed faster than the process. Thus where we look for higher levels of cosmic intelligence is in systems which have higher inherent speeds than the vibrations of matter and we find just such speeds in cosmic radiations of all sorts.

The potential intelligence of a data system is strictly related to the available frequency band width whether of digits or waves. Thus a television set requires a frequency bandwidth of four megacycles whilst its steam radio brother, which cannot transmit nearly so much information, only needs about ten kilocycles.

Now the natural frequencies associated with tangible matter are revealed to us by the techniques of molecular spectrometry and, if I recall aright, are of the order of 10^{12} per second and thus to intervene programmingwise into such a system you would need to use faster frequencies and this is why the faster gamma radiations can affect the structure of matter. This is also expressed by the Compton effect which tells us that one can make red light out of blue light but you cannot make blue light out of red light. In other words, the faster vibrating blue light is programming for red light but not vice versa.

But the radiations of the universe contains many sorts of waves faster than the 10^{12} of terrestial matter and at the highest level of cosmic rays we find vibrations in the region of 10^{22} per second and thus of almost limitless potential modulated information content and certainly potentially programming for matter from this aspect.

Thus when I seek for higher levels of intelligence than man I look for them in space and outer space and in the fact that man is utterly saturated with such cosmic radiations. Furthermore, if Einstein was right that the universe is "finite" but unbounded, a sort of celestial sphere, then something like the optical hollow white sphere principle operates in that the information of cosmic radiations is simultaneously present at all points of space and time. This is proven by the fact that the cosmic radiations (and I am not referring specifically to cosmic rays) are about the same whether you measure them in Britain or "down under" in Australia. What I say is that we live in a radiation flux of constant density and that it could be that this constant density flux is a constant intelligence density flux.

I think it also significant that we live close to the centre of one such source of radiation, the Sun, and that because of the radiations from the sun there takes place that prime photosynthesis of green matter which is the basis for the maintenance of life. Thus sunlight is programming for the maintenance of life as a cybernetic data process and if the sun were to extinguish, then the second law of thermodynamics would quickly run the terrestial situation down to absolute zero of temperature with a lifeless world, frozen seas and darkness.

But in this matter I do not wish to emphasize unduly the significance of solar radiation but rather to treat the totality of cosmic radiations in which we are immersed as the evidence for the existence of levels of intelligence at a transcendental level.

In my as yet unpublished book on "the intelligent universe", I postulate the existence of a "universal cosmic reason" saturating every point of space and time and programming for lower levels of data which we tend to call "matter". Perhaps one can best describe this as occuring in "eternity" if by eternity one means the total integration of all space and time.

This cosmic reason we detect as sunlight, as light from the stars and in our radio telescopes, but also in those faster cosmic radiations for which we yet have no means of analysis except for the fact we know they exist and have rather high quantum energies.

So far we have not learned to unscramble the information in these cosmic radiations which I call cosmic reason because our man-made instruments are far too coarse to do so except in the lowest frequency ranges which tell us that remote stars may have the same sort of matter as our own solar system. But whilst our man-made instruments may be too coarse we are surrounded by other receiving sets totally in tune with some of the waves of cosmic radiation and these receiving sets are simply "matter" and the higher forms of matter which we call "life". Thus I suggest to you that the programming of material systems as revealed in the complexities of DNA and which finish up as "giraffes" or "elephants" comes from the influence of cosmic radiations, cosmic reason. Thus the process is one of radiation chemistry and a very powerful cybernetic sector of data processing with a high degree of steersmanship.

CYBERNETIC MAN

Now what is the significance for man if, indeed, he lives in an intelligent universe utterly saturated with cosmic reason operating at far higher frequencies and intelligence than his own mind? Thus it is likely that man cannot directly communicate with such cosmic reason. Note that I suggest that this cosmic reason is not changing all the time but is, as it were the totality of universal truth and knowledge in a cosmic sense. But am I right in suggesting that man cannot communicate with it and surely on those rarer moments of human truth and belief when we really understand something correctly then could not that be the moment when the humans reason is resonating in harmony with the cosmic reason?

In my book *The Immortal Robot* I have suggested that this is precisely what happens to us all at "moments of truth" and understanding and that furthermore if this process becomes appreciable in our affairs then man develops a something which I have called his "higher data body" and that this body is simply an eternal resonance with a fraction of the cosmic reason. But, being eternal, it cannot die at the physical death and at the moment of death this higher data body is preserved intact due to its resonance with the cosmic reason and instantaneously joins and merges with the higher data body of another living person having about the same degree of understanding and personal development. Thus it is a transfer by vibrations.

In his book *The Phenomenon of Man*, Père Teilhard de Chardin describes the evolution of society in terms of a noo-osphere, a sort of evolving social mind. This was a message of hope and faith but I think my analysis shows just how it could in fact take place.

If the only thing in man and society which can survive physical death is that degree of truth and understanding which corresponds to the cosmic reason then death is a diode which rejects lies and evil which have no correspondance in cosmic reason and only the truth and the good are permitted to move forward into the future.

Thus the mechanism of death is a rectifier for the cybernetic evolution of man. The theory also suggests that perhaps "there are no dead".

Thus I give you the immortal robot. "A robot", because man must submit to the laws of data and data processing, but immortal in that if he chooses the true and the good so that he resonates with The cosmic reason then his future is assured, and on this same very Earth.

CHAPTER I–3

Entropy and philosophy

D. MALIĆ
University of Belgrade
Yugoslavia

Summary

Over a hundred years ago Clausius discovered a new variable of thermodynamics known as entropy (from the Greek, change). He was conscious of the great importance of this function for the transformation of thermal motions into mechanical and other motions. Later on it was realized that entropy increases in the universe. Clausius then predicted that the universe, being an isolated system, will eventually achieve the state of "thermal extinction". This has since provoked violent controversies amongst scientists. This matter is still being discussed at present and concerns the relations between the evolutions of the Earth and the universe, and the entry of man into outer space, guided by the force of science. The problem is to find out whether the principle of increasing entropy holds on Earth and in the universe.

INTRODUCTION

Clausius, the creator of the idea of entropy, in his attempt to extend the law of heat transfer, valid in our earthly conditions, to the conditions of the universe, predicted some hundred and ten years ago "the thermal death" both of our planet and the whole universe. Starting from the assumption that the universe is an isolated system in which the heat is transferred from a warmer body to a colder one, Clausius came to the conclusion that the heat in future would reach such a low temperature at which life would be impossible. He cried out with pain as if he were falling in the abyss of irreversibility together with whole world. "The entropy of the universe tends to a maximum" was his outcry; this caused a great confusion and lively discussions among the physisists and philosophers of that time.

Because of the general trend of science, however, together with the development of the human intellect, Clausius's entropy has played an opposite role. It serves nowadays to explain different, sometimes very complex phenomena and processes in nature, technology, geology, metereology, biology, etc. It is used in different sciences and represents a link and synthesis at a high level of what is common to all sciences. In this lies the deep philosophic sense of entropy.

CONTRADICTIONS OF THE THERMODYNAMIC PROCESS

The thermodynamic process represents the macroscopic expression of external, visible, and also large changes of a certain system, as compared with the thermodynamic state, which again is the macroscopic expression, but of internal, invisible, minute changes and movements in the system itself. Every thermodynamic process is characterized by these parameters—the parameters of process—which express the mutual relation of the system and its environment (the mechanical work and the quantity of heat).

If the system is isolated from its environment, i.e. if the influence of the environment is limited or reduced to zero, then a state of equilibrium is established in that system. The weaker the influence and the constraint of the environment upon the system, the slower is the development of the the thermodynamic process, and the greater is the tendency towards equilibrium of the system. On the contrary, if, in case of a thermodynamic process, the influence and constraint of the environment are stronger than the tendency for a state of equilibrium within the system, then the whole thermodynamic process is in a state of disequilibrium and irreversibility. The influence and constraint of the environment manifest themselves in the tendency of the system to establish a state of equilibrium with the environment too. This tendency, however, is opposite to the tendency for establishing an equilibrium within the system itself. This means that the environment is constantly forcing the system to be in a certain state of disequilibrium. Then, when such state is related to stationary conditions, i.e. when it is not changed with time, it is considered that an *unbalanced stationary state* is reached, in which all possible fluxes (energy, mass, etc.) are independent of time in the system–environment interaction. Such distribution, corresponding to a stationary state, for which stochastic equations are valid, i.e. when deviations from a state of equilibrium are small, leads to a minimum entropy production.

The concept of a stationary state for a closed and an open system respectively can be considered as one of a state of equilibrium for an isolated system. While in case of isolated systems every contraint of the environment is excluded (the systems are in a adiabatic envelope) so that entropy reaches a maximum and the entropy production disappears, in the case of the removal of the adiabatic envelope the system reaches a partial (closed system) or complete (open system) connection with the environment, which leads to a corresponding unbalanced state and the entropy production in the system. An unbalanced stationary state or, in short, a stationary state analogous to a balanced state, should be one to which all real irreversible processes are tending. It should also be the most probable for any irreversible process under certain conditions, i.e. for the corresponding constraint of the environment. This means that the minimum of entropy production for a stationary state should be considered in the same manner as a maximum of entropy for a balanced state, i.e. the stationary state could be defined as a special kind of the most probable state of system. For an exact definition of such state the corresponding mathematical dependency of the constraint should be known, i.e. the connection between the system and environment.

Analogous to hitherto known and studied thermodynamic system — both isolated, closed and open respectively, the system capable of thinking may be called the *intellectual (conscious) system*[1]. Contrary to the constraint of the environment upon the open system (a physical, chemical, or biological system) and the tendency of such a system to reach a stationary state with a minimum of entropy production, the *intellectual system is exercising a constraint upon environment*. In order to realize the contraint upon the environment, the intellectual system consumes entropy. By consuming entropy, the living intellectual being is acting against the environment, changes relation in the environment, and hightens the potential of the organization.

Consequently, any thermodynamic process is, as a matter of fact, evolving in the presence of two contradictory tendencies in the system–environment interaction.

THE FUTURE—TRIUMPH OF HUMAN MIND

Man has invented a number of structures and courses especially during the last hundred years, and has adapted them to his needs and conditions imposed by the environment in the widest sense of the word. The development of the human mind and thought from the beginning, when he produced fire by friction, to the present day, when he is invading cosmic space, represents in fact a continuous struggle to conquer the new and the unknown.

If, for instance, one says that water is flowing by itself only downwards and heat propagates from a warmer to a colder body, one really wishes to emphasize the fact that "downwards" is one of the fundamental characteristic conditions under which we are living. This property can be utilized, and man has utilized it and is utilizing it constantly, for production of work, animal power being too weak and insufficient for it. The inexhaustible power of the human mind was needed—the power of machinery he has invented not only to work for him but to think for him too.

Human thought has succeeded to "force" not only water but also heat "upwards". It is constantly raising the level of the organization of different structures and courses. This is quite understandable, because human thought arises and develops in a highly organized structure such as the cerebral structure.

Starting from the standpoint that entropy is produced and increases in unbalanced, chaotic, and poorly organized structures, but does not change in ordered and balanced structures, and that is consumed in organized and highly organized structures, such as the cerebral structure, in which human thought is born, it is possible to establish the following simple connection between the information and the organization of the human thought[1]:

$$I = K\,n \ln n = K\,Om \qquad (1)$$

where I is information, Om the highest possible organization, n the number of elements of the given structure, and $K = \log_2 e$ bits (binary, digital units)[2].

If along different changes and processes (physical, chemical, biological ones, etc.) in complex structures and courses human thought is also considered, it will be seen that the intellectual process follows and controls different structures and courses, improves them, and directs them. The part of the intellectual process is that different structures and courses would be better ordered and organized, thus actually leading to less chaos and to less total entropy production in them.

Starting from the above standpoint all processes may be classified in the widest sense of the word into the following: *Spontaneous natural processes*, independent from human will and influence, and *organized processes*, which are the product of human mind and the result of his thought. Any structure, and process in the same, is characterized by a fixed level (potential) of organization, a fixed diversity and some chaotic state.

The efficiency of the process (change, transformation) may be expressed in a general form[3].

$$\varepsilon = \varepsilon_{max}[1 - f(H)] \quad (2)$$

where $f(H)$ is the function of the chaotic state or the insufficient organization of the structure in question.

In the case of highest possible organization (Om) we would have $f(H) = 0$ or $\varepsilon = \varepsilon_{max}$, while in the case of some current organization (Oc) it would be $\varepsilon < \varepsilon_{max}$.

The function of the chaotic state $f(H)$ and the efficiency of the structure in question or the course (energy, mass, information, ...) may be expressed ultimately in the most general form as a function of negentropy of human thought.

The future, consequently, should be envisaged as a triumph of science and art—as a triumph of human mind. In this should, in fact, be seen the deep philosophic sense of the idea of entropy as opposed to Clausius's concept of more than hundred and ten years ago[4].

CONCLUSION

The deep philosophic sense of the idea of entropy is that it serves for analysing, explaining, and discovering different phenomena in nature and human creative activity, which can sometimes be very complex.

All processes in the widest sense of the word may be classified into: spontaneous natural processes, independent of human will and influence, and organized processes, which are the product of human mind and the result of his thought.

Because of the general development of science and creative human activity, entropy nowadays enables man to envisage a better future, in which the free personality will triumph, as opposed to the dim future and the "thermal death" envisaged by Clausius some hundred and ten years ago in the light of the idea of entropy.

The idea of entropy is used now in different sciences and represents a link and synthesis at high level of what is common to all sciences.

References

1. D. Malić, "Aspect thermodynamique de la pensée humaine", *5th Intern. Congr. Cybernetics, Namur* (1967).
2. L. Brillouin, *La Science et la Théorie de l'Information*, Masson & Cie, Paris (1959).
3. D. Malić, "Entropie de la vie et de la pensée", *5th Intern. Congr. Cybernetics Med., Naples* (1968).
4. D. Malić, "Future in the light of the second law of thermodynamics", *Intern. Conf. "Science and Society", Herceg Novi* (1969).

CHAPTER I–4

On the relation between cybernetics and general systems theory

GEORGE J. KLIR
School of Advanced Technology
State University of New York at Binghamton, U.S.A.

Summary

Cybernetics and general systems theory are two mutually related methodological movements in modern science which try to integrate contemporary scientific knowledge. Although neither identical, nor inclusive (one contained in the other), nor exclusive (disjoint), both of these movements are devoted to a systematic study of such concepts, principles, and methods which are meaningful and applicable in all classical scientific disciplines. Both research and education in special disciplines benefit from these movements mainly (i) by using a unified terminology, and (ii) by extending special methodological principles by a general methodology, thus improving significantly understanding between specialists from most diverse disciplines and making them better prepared for the team interdisciplinary work so typical of modern science.

INTRODUCTION

This paper is *an essay* devoted to some aspects of integration in contemporary science. In particular, it is devoted to two significant products of this integrational process, namely, cybernetics and general systems theory. Views on cybernetics and general systems theory are reviewed and their relationship is clarified. Although the subject matter, belonging partially to ethics, does not make it possible to converge toward a set of generally acceptable conclusions, this presentation is still justified. It expresses an opinion about the relationship between cybernetics and general systems theory, opening thus a discussion of the topic. Only on the basis of such a discussion, each of us can enrich his own opinion upon the relationship and his views on research and educational planning.

Writing this essay, I have in mind questions like, "Is cybernetics (general systems theory) a scientific discipline or a point of view applied in other disciplines?", "Is there any difference between cybernetics and general systems theory?", "Does cybernetics share something with general systems theory?", etc. In the following sections, I shall try to find unique and reasonable answers to some of these questions. However, the reader should be aware of my warning. The answers must not be considered as true or false but rather as reasonable or unreasonable.

VIEWS ON CYBERNETICS

My first meeting with cybernetics (although not under this name) happened a long time ago. If my memory is correct, it was in 1947, during a high school lecture. My teacher tried to stress that in some engineering systems, like communication and control systems, only the space–time pattern of energy involved is relevant, as a carrier of information, but not the kind of energy and its magnitude. It seemed strange to me at that time that there was no scientific discipline devoted to these informational aspects of energy. What a surprise it was for me when I learned later about cybernetics and information theory.

It is well known that cybernetics was defined by Norbert Wiener, its founder and promoter, as the science of "control and communication in the animal and machine"[1]. Although somewhat restrictive when compared with contemporary views, this classical definition does indicate the most essential features of cybernetics, namely (1) that cybernetics is concerned with systems defined on quite diverse objects from the living through the inanimate, and (2) that cybernetics is concerned with processes associated with information.

As far as the first feature is concerned, we certainly consider animals and machines as quite diverse objects. In addition, we feel that communication and control play an important role in other types of objects too, e.g. in societies and in objects of the arts. Wiener himself says, in his second book on cybernetics[2], that "communication and control belong to the essence of man's life even as they belong to his life in society". Mueller is one of those who have demonstrated the existence of communication and control in the arts[3].

As far as the second feature is concerned, we are certainly right, when we relate cybernetics to information. Indeed, the lack of information makes both communication and control meaningless. There are, however, other processes which are based on information, e.g. information processing, storing, etc. It does not seem reasonable to include some of these processes and exclude others. Hence we prefer to speak about informational processes rather than about control and communication.

Many views on cybernetics have been expressed within the twenty years of its existence and many definitions of cybernetics have been presented. Nearly all of them generalize more or less Wiener's original definition. Some of them go, in my opinion, either too far or in an improper direction in this generalization. For instance, Margaret Mead suggests that cybernetics be considered "as a way of looking at things and as a language for expressing what one sees"[4]. Another example is the Couffignal thesis that cybernetics, simply speaking, is the art of how to approach the solution of various problems or, in other words, how to act under various circumstances[5]. If we neglect these most extreme views, we find that information and systems are two central concepts of cybernetics.

Information concept has a far wider meaning in cybernetics than is assigned to it in ordinary life. There are various aspects of information. All these aspects are studied by information theory, information being considered as a measurable quantity. In the most general sense, information may be regarded as a measure of organization as opposed to randomness. However, the relation between the amount of information and organization is, in general, very complicated and has not as yet been investigated from all desirable aspects. The difficulty consists mainly in that the amount of information depends not only on the quantity of organization, but also on its quality. If the aspects of quality are completely ignored, the relation between organization and information is most simplified. Then, information has the meaning of negative entropy as shown by Shannon and others[6,7]. Information theory, based on this particular meaning of information, has been elaborated in considerable detail and has found important applications in science, engineering, and the arts. There are other relations of information and organization. Some of them are referred to as semantic information, pragmatic information, etc. Unfortunately, they have not been developed to such an extent.

Let us consider now an object which is under scientific investigation. As a rule, it can be studied from many different viewpoints. Each viewpoint identifies a particular system on the object. These viewpoints can very roughly be divided into two classes.

The first class is characterized by consideration of mass and energy. It contains viewpoints from which we examine problems concerning magnitudes of masses and energies, their mutual effects and evolution in the system itself, as well as between the system and its environment, etc. Such viewpoints find their widest application in the natural sciences, economics (production, distribution, consumption, etc.), and other disciplines. Let us call them mass–energy viewpoints. Depending on the kind of masses or energies and/or aspects studied, special scientific disciplines are distinguished.

The second class comprises viewpoints from which we examine problems concerning organization in the widest sense. As in the mass–energy viewpoints, we may examine the amount of organization, its evolution within the system, between the system and its environment, etc. Since organization is the carrier of information and it is the informational aspect of organization which chiefly interests us, we are here concerned with informational problems, like the problems of communication, control, storing and processing of information, etc. Let these viewpoints be called cybernetic viewpoints.

Let me note that, in the majority of cases, the two above mentioned classes cannot be entirely separated one from the other. When investigating metabolism, for instance, we are concerned, according to our classification, with the examination of a living organism from a viewpoint of the first class, i.e. a mass–energy viewpoint. As a matter of course, problems of organization or information cannot be totally disregarded since the very existence of the living organism depends on a certain organization. It is important, however, that in the given case this organization is only presumed but neither its static nor dynamic nor evolutionary properties are subjects of our study. Another example, where the situation is just opposite, is music. Here, cybernetic viewpoints dominate though both of the classes are meaningful.

Let us return now to the class of cybernetic viewpoints. Even though they find their application in all scientific disciplines, cybernetics emerged as a separate discipline investigating properties of systems whose quantities represent information without any relation to a mass–energy carrier. Thus, two aspects of cybernetics must be distinguished: (1) cybernetics as a class of viewpoints, and (2) cybernetics as a discipline.

Specific cybernetic viewpoints together with specific mass–energy viewpoints creates a special discipline. This is a common practice which has been applied in science, engineering, and the arts long before cybernetics were formally established as a discipline. However, this establishment, stimulated primarily by Norbert Wiener in 1948,[1] has had significant effects on the applications of cybernetic viewpoints in many special disciplines.

First, by adopting various terms formalized by cybernetics, the terminology that expresses applications of cybernetic viewpoints within special disciplines becomer more and more unified. Many precise cybernetic concepts have already replaced the specialized, and in many cases vague, concepts that had been used in special disciplines in the pre-cybernetic era. Examples of these are information, entropy, organization, signal, code, information channel and its capacity, feedback, control, noise, memory, redundancy, distortion, etc. Although applied to different systems as far as the mass–energy viewpoint is concerned, each of these concepts has the same meaning in all disciplines. This situation significantly improves understanding (communication) between specialists from the most diverse discipline.

Second, the special methodological principles concerning cybernetic problems in special disciplines are more and more supplemented by general cybernetic methods applicable in all disciplines, creating thus a profession of cyberneticians—interdisciplinary scientists devoted to the study of cybernetic methodology.

VIEWS ON GENERAL SYSTEMS THEORY

About ten years ago, I attended a lecture devoted to self-organizing systems. The speaker started his lecture by saying: "There is no doubt that everybody in this room understands the meaning of the word 'systems' and, therefore, I will not waste time and proceed directly to 'self-organization'." But the discussion after the lecture showed without any doubt that he was wrong. If anything was clear at all, it was the meaning of "self-organization" but definitely not the meaning of "system". After a long discussion, everybody departed more confused about "system" than before the lecture. Still the lecture had one positive effect on me: it stimulated my curiosity about the meaning of the strange word "system".

I found that there was a scientific movement called general systems theory. After I had read all available publications concerning this theory, I was still not satisfied. Either the approach was too general and thus without any content, or it was not general enough and thus not applicable in all disciplines, or it was based on vague concepts leaving too much uncertainty. I was encouraged by a paper of Kenneth Boulding[8], which said as follows.

"General systems theory is a name which has come into use to describe a level of theoretical model-building which lies somewhere between the highly generalized constructions of pure mathematics and the specific theories of the specialized disciplines ... because in a sense mathematics contains all theories it contains none; it is the language of theory, but it does not give us the content. At the other extreme we have the separate disciplines and sciences, with their separate bodies of theory. Each discipline corresponds to a certain segment of the empirical world, and each develops theories which have particular applicability to its own empirical segment. Physics, chemistry, biology, psychology, sociology, economics, and so on all carve out for themselves certain elements of the experience of man and develop theories and patterns of activity (research) which yield satisfaction in understanding, and which are appropriate to their special segments.

In recent years increasing need has been felt for a body of systematic theoretical constructs which will discuss the general relationships of the empirical world. This is the quest of general system theory. It does not seek, of course, to establish a single, self-contained 'general theory of practically everything' which will replace all the special theories of particular disciplines. Such a theory would be almost without content, for we always pay for generality by sacrificing content, and all we can say about practically everything is almost nothing. Somewhere however, between the specific that has no meaning and the general that has no content there must be, for each purpose and at each level of abstraction, an optimum degree of generality."

Boulding expressed exactly what I felt: a system theory should be elaborated which was based on precise concepts and whose generality and content would be in an adequate equilibrium with regard to its applications in science, engineering, and the arts. During the last couple of years, I have tried to apply these criteria to establish a reasonable approach to general systems theory. My efforts resulted in several papers[9-11] and a book[12].

Simultaneously, several other approaches to general systems theory have been studied. Their differences result, essentially, from assigning different meanings to the word "system".

At the highest level of generalization, systems are considered as relations defined on families of sets without giving any meaning to either the sets or the relations. The role of such a theory consists mainly in solution of problems concerning abstract algebraic systems. This is a wide range of problems including problems of decompositions of the given relation as well as those of the type of Gödel's[13] proof. A lot of work has been done in this direction by Mihajlo D. Mesarovic and his group[14–16] at Case Institute of Technology in Cleveland, Ohio (see also some articles in the journal *Mathematical Systems Theory*).

Other approaches to general systems theory have been elaborated. In many of them systems are considered at a level of generalization which ensures that the respective theory is applicable to all classes of systems in all branches of science, engineering, and the arts. At the same time, basic properties of particular systems are completely retained by the theory. Here, we are concerned with properties that are irrelevant to traditional classification of science, engineering, or the arts. Such properties are, for example, behavior, state-transition characteristics, time–space resolution level, structure, etc. Ashby[17,18], Bertalanffy[19], Klir[12], Wymore[20], and Zadeh[21] are among those who have contributed in this direction.

Thus, there are several modes of general systems theory due to different motivations and the histories of their elaboration. Most of them are mutually consistent approaches which can be combined in applications. An exception is, for instance, the unusual approach suggested by C. West Churchman[14]. Despite differences between general systems theories, the essence common to most of them can be formalized as follows. Every general systems theory reflects some of those traits of systems studied by different disciplines which are irrelevant to special properties of the quantities involved. Since there are many problems arising in special disciplines, which can be either completely or partially solved in terms of these general traits, their study is justified. Rules and methods elaborated for the solution of problems concerning a particular set of traits constitute a general systems theory. Once these rules and methods are prepared, they are applicable in every discipline where the respective problems arise.

CYBERNETICS VERSUS GENERAL SYSTEMS THEORY

It has already been shown that cybernetics is characterized by neglecting mass–energy viewpoints when investigating systems properties. This suggests that cybernetics is somehow related to general systems theory. Employing the concept of abstract models, both cybernetics and general systems theory study certain properties of systems; such properties that are irrelevant to the divisions made by classical science. Indeed, this relation is so intimate that in many cases cybernetics has even been identified with general systems theory. However, such an identification ignores the fact that cybernetics is limited to informational processes or, according to W. Ross Ashby, to "systems that are open to energy but closed to information"[18]. While general systems theory defines and investigates abstract models of special systems without any regard to their nature, cybernetics is solely concerned with informational aspects of special as well as general systems.

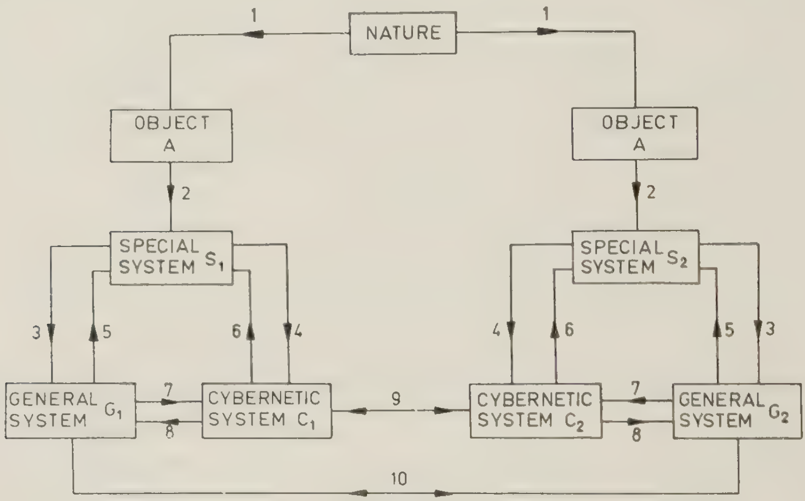


FIGURE 1

As far as general systems is concerned, two classes of problems can be clearly distinguished.

(i) Problems that are irrelevant to the information content of the quantities involved; these problems are treated by general systems theory.

(ii) Problems in which the information content of the quantities involved is relevant; these are problems in which the application of general systems theory is combined with various aspects of information theory. This combination constitutes cybernetics as a discipline.

Our discussion can be summarized in the following four definitions.

(i) General systems theory (science)—A scientific discipline that investigates abstract models of systems without any regard to the interpretation of the quantities involved.

(ii) Cybernetics as a viewpoint in a special discipline—That portion of the special discipline which investigates informational and/or organizational problems of special systems.

(iii) Cybernetics as a science—A scientific discipline that investigates informational and/or organizational aspects of general systems.

(iv) Information theory (science)—A scientific discipline that investigates all of the problems concerning information (organization) regardless whether or not they are "systems type" problems.

The relationship between general systems theory and the two above mentioned aspects of cybernetics is illustrated by Figure 1. The numbers in Figure 1 have the following meanings.

1: Selection of an object for scientific investigation.
2: Definition of a system based on a particular mass–energy viewpoint. This is sometimes meaningless, as in linguistics, mathematics, etc.
3: Abstraction of special systems.
4: Application of a particular cybernetic viewpoint in a special scientific discipline.
5: Interpretation of general systems traits in terms of special systems.
6: Interpretation of cybernetic systems traits in terms of special systems.
7: Interpretation of general systems traits in terms of cybernetic systems.
8: Abstraction of cybernetic systems.
9: Cybernetic modelling[22].
10: General systems modelling[12].

Summarizing, relations 3, 5, 7, 8, 10 belong to general systems theory, relations 4, 6 represent the application of cybernetic viewpoints in special disciplines, and relations 7, 8, 9 belong to cybernetics as a science.

Although cybernetics and general systems theory are two disciplines that may have a great impact on the interdisciplinary cooperation in science, engineering, and the arts, which has become a necessity if the complex urban and technological problems of our days are to be solved, their implementation depends heavily on proper changes in education. So far, education has been quite conservative in making desirable changes. However, some encouraging efforts have been recently made to establish programs supplementing the traditional disciplinary education with interdisciplinary features of general systems theory and cybernetics.

For example, the School of Advanced Technology at the State University of New York at Binghamton has a graduate program leading to the M.S. degree which contains a one-semester course on general systems and cybernetic principles. This course will be offered to graduates in the most diverse disciplines. After this course, they may either continue in their study of a special discipline or switch to a computer and system science oriented curriculum terminating with a sophisticated two-semester course on mathematical aspects of general systems theory.

Thus, there are some activities as far as an inter-disciplinary education on general systems and cybernetic principles is concerned. Although these activities are mainly limited to the graduate level at the present time, we may expect significant changes from this point of view in the undergraduate, the high school, and even the elementary school curricula in the near future. The sooner these changes effect our educational system, the better for the development of science, engineering, humanities, the arts, and consequently, the better for our society.

CONCLUSIONS

In order to express in a condensed form conclusions from our previous discussion, let us introduce the following set of symbols: S_d, subject of a special discipline; G, subject of general systems theory; C_s, subject of cybernetics as a science; I, subject of information theory; C_v, subject of cybernetics as a viewpoint.

A relationship between these subjects can very roughly be expressed by the following formulas of set algebra:

$$S_d \cap G \subseteq G \text{ and } S_d \cap G \subset S_d$$

$$S_d \cap C_s \subseteq C_s \text{ and } S_d \cap C_s \subset S_d$$

$$C_s = G \cap I$$

$$C_v = \bigcup_d (S_d \cap C_s)$$

References

1. N. Wiener, *Cybernetics of Control and Communication in the Animal and in the Machine*, Wiley, New York (1948).
2. N. Wiener, *The Human Use of Human Beings*, Houghton Mifflin Co., Boston, Mass. (1950).
3. R. E. Mueller, *The Science of Art*, The John Day Co., New York (1967).
4. M. Mead, "Cybernetics of cybernetics", in *Purposive Systems*, *Proc.* 1st *Ann. Symp. ASC* (Ed. by H. Von Foerster *et al.*), Spartan Books, New York (1968).
5. L. Couffignal, *Les notions de base*, Paris (1958).

6. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, Illinois (1949).
7. L. Brillouin, *Science and Information Theory*, Academic Press, New York (1956).
8. K. E. Boulding, "General systems theory–the skeleton of science", *Management Sci.*, **3**, 198–9 (1956); reprinted in *Gen. Systems*, **1**, 11–17 (1956).
9. G. J. Klir, "The general systems as a methodological tool", *Gen. Systems*, **10**, 29–42 (1965).
10. G. J. Klir, "Processing of general systems activity", *Gen. Systems*, **12**, 193–8 (1967).
11. G. J. Klir, "An approach to general systems theory", *Gen. Systems*, **13**, 13–20 (1968).
12. G. J. Klir, "An approach to general systems theory", Van Nostrand–Reinhold, New York (1969).
13. E. Nagel and J. R. Newman, *Gödel's Proof*, New York University Press, New York (1958).
14. M. D. Mesarovic (Ed.), *Views on General Systems Theory*, Wiley, New York (1964).
15. M. D. Mesarovic, "General systems theory and its mathematical foundations", *Record of the IEEE Systems Sci. and Cybernetics Conf., Boston, Mass.* (1967).
16. T. G. Windeknecht, "An axiomatic theory of general systems", *Record of the IEEE Systems Sci. Cybernetics Conf., Boston, Mass.* (1967).
17. W. R. Ashby, *Design for a Brain*, Wiley, New York (1952).
18. W. R. Ashby, *An Introduction to Cybernetics*, Wiley, New York (1956).
19. L. von Bertalanffy, *General Systems Theory*, George Braziller, New York (1968).
20. A. W. Wymore, *A Mathematical Theory of Systems Engineering*, Wiley, New York (1967).
21. L. A. Zadeh and C. A. Desoer, *Linear Systems Theory*, McGraw-Hill, New York (1963).
22. G. J. Klir and M. Valach, *Cybernetic Modelling*, Iliffe, London (1967).

CHAPTER I–5

A classification of the sorts of information in cybernetics

N. STANOULOV
Institute of Engineering Cybernetics, Bulgaria

Summary

An attempt is made to analyse a number of existing definitions for information with the aid of the communicative and organizational (syntactic) aspects in cybernetics. Nearly complete confusion exists in this field at present; almost every author tries to give a definition of his own according to his individual tastes and preferences. The author's systematization (the so-called "information tree") is based on introducing five typical branches. Special attention is focused on two last branches; for these, ideas of semiotics are also used. Definitions of asemantic information recently introduced by the author and based on general theoretic–systematic considerations are also included in the classification. The great similarity in the construction of some of the branches is inherent to this "information tree". This allows future investigations to be foreseen, particularly in the field of the pragmatic information from an intellectual point of view. Though the differentiation of various sorts of information is a very difficult task, it proved possible to build up such a classification with the aid of predetermined cybernetic limitations.

INTRODUCTION

Since the formulation of the statistical information theory by Shannon, the concept of information was accepted as a central one in cybernetics. The latter as a science needs a description of its own concepts and generalizations, i.e. a development of the specific apparatus of the science itself. This in part explains the growing interest over the last twenty years in the scientific explanation and essence of information itself. Certainly Shannon's theory is a quantitative theory and, similar to other such theories, a definition of the amount of information only will be enough for its development. However, the information theory differs considerably in character from other quantitative disciplines. This follows from the basic objective of each communication—to restore at the receiver (or destinator) the message generated from the transmitter[1] (communicable aspect in cybernetics). But this theory does not answer the questions (i) what information from the received message a given destinator is able to obtain, and (ii) what are the characteristics of the information content[2]. Simultaneously with the development of cybernetics, and mainly of its applied branches (computing technique, information systems), a systematic study of the information itself was found to be necessary. However, such an approach requires that the "technical" representation of the messages (information), referring to definite objects (and phenomena) in nature, must be related to the states of these objects for which the messages are transmitted. In such an investigation the organizational (syntactic) approach in cybernetics is also inescapably involved.

Many of the difficulties in working out a classification of the sorts of information come from (i) the evolution of the scientific concepts (such as the concept of information), and (ii) the nature of the differentiating features of the sorts of information. With the classification suggested here an attempt is made to put in order the numerous concepts and definitions of the information known to the author. Furthermore, only some of the more typical are discussed in this paper, however, but not in their original form. No doubt many other concepts exist with which the author is not familiar.

The basic branches (types of information) in the classification are as follows (see Figure 1): concepts and definitions of information of an *indefinite* (various) (1), *substitutional* (2), *philosophical* (of a cognitive viewpoint) (3), *subjective* (mental aspect) (4), and *objective* character (5). The first refers to concepts of the type where information is well understood as something requiring no explanation[3]. The second branch refers to cases when the concept of information is used instead of amount of information or the reverse[1,3,5,6], or "information is substituted by 'communication'"[1,7,8], and by "variety"[9]. The third branch is concerned with the concepts of the relation between information and reflection[10–12]. Details may be given for almost all of the above-mentioned branches. For instance, in the third branch the concepts of information as a measure of reflection is today specified only to a degree of comparison—a less more or complete reflection[13]. However, a quantitative measure of reflection does not exist.

The main attention here will be given to the following two branches. The fourth branch comprises the definitions of the information in which the latter depends on man and in addition may have sense and usefulness (value).

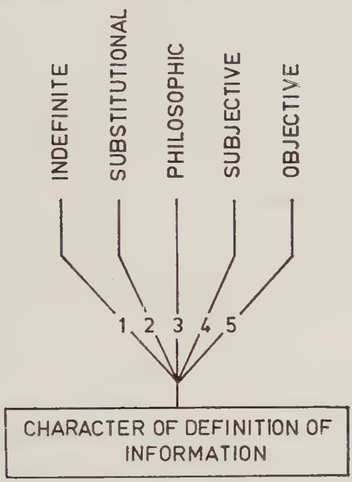


FIGURE 1 Classification of the five branches of the information scheme

The dependence is expressed as follows: since the messages could be represented by symbols (i.e. the technical representation of messages mentioned above) then (i) to these symbols particular meanings are assigned (semantics), i.e. a different interpretation of the messages may be given; and (ii) the response of the destinator to the message (pragmatics) is not single-valued. This dependence expresses the subjective character of the information.

The fifth ("objective") branch refers to concepts stating that the information does not depend on man. Now also the messages have a technical representation in which the organization (i.e. the combination of symbols) of the messages is consistent with the corresponding states of the objects for which these messages (information respectively) are transmitted. It is apparent that this organization for which the destinator receives a message does not depend on him. Such independence expresses the objective character of the information. However, the "objective" information besides being syntactic may also be semantic and pragmatic. It is evident that not only the asemantic information (in the sense of Shannon's theory) but also the semantic and pragmatic information envisages the presence of a certain symbol organization of the message following definite rules. Here the reasonable and always increasing tendency for artificial mechanisms to be used in the capacity of information receivers (and destinators) on a syntactic, semantic, and pragmatic stage is taken into account.

The consistency and interdependence of the two aspects in cybernetics (the communicable and organizational) are easily seen from the foregoing. Thus, for example, the first aspect makes no sense without the second—all the same the destinator does not understand his role (as partner) in any information system; however, the reverse is also true. In the presence of an organization of the messages such a system cannot exist without the communicable aspect. It seems clear that these considerations refer to the three above-mentioned stages of the information whatever its character (subjective or objective).

SORTS OF INFORMATION

The following concepts are typical for the fourth branch: By information we understand the designation of the content received from the outer world in the process of our adaptation and the adaptation of our senses to it[14]; we consider intuitively that we receive information when we learn about an event whose realization has not been predetermined[15]. The concept of information contains in itself the idea of some knowledge[16]. This branch mentions also the following varieties of the information characters.

(a) *Semantic:* A formal model is produced whereby by Shreider[17] the process of the semantic analysis of a text is described. The model contains a description of the "idea of the outer world" (the so-called thesaurus) to which the message refers. The process of the meaning analysis is interpreted as a change of the thesaurus influenced by the given text, and a set of admissible operators over the thesaurus is introduced. The degree of change of the thesaurus θ influenced by the operator corresponding to the text T is called an amount of semantic information $I(T, \theta)$ in T about θ. A semantic interpretation of the statistical concepts of information and entropy is suggested by Voyshvillo[18]; here the developed concepts and relations in the semantic theory are special cases of Shannon's theory. Information is referred to judgements, while entropy is treated as being a characteristic of some question or problem. In such a case the information containing a judgement of some problem is defined, depending on how much the proof or admission of the truth of this judgement reduces the entropy of the problem. This information does not depend on the unexpectedness of the problem solution.

(b) *Syntactic:* The definitions of the semantic information are in some extent syntactic definitions also. Thus a language[19] with a finite number of symbols is introduced (nouns, adjectives and a verb) subjected to a finite number of logical connections (or rules) necessary to form sentences.

(c) *Measurable:* Based on the methods and symbolics of the mathematical logic the amount of information in a given sentence is defined by Carnap and Bar-Hillel[19]; the semantic information is also determined as metrical information referring to the absolute increase (return) of knowledge[20].

(d) *Pragmatic:* The amount of valuable information is defined[21] by the increase of the probability for achieving the goal after information has been received. The amount of valuable information is a special case of the amount of semantic information introduced by Voyshvillo[18]. Measurement methods for the amount and value of the information contained in scientific papers are developed by Miles[22]. A connection will be established between the persons A and B if the message delivered by B changes the state of A (i.e. the probability that A will choose a definite action). If the message influences the probability for choice of action, it is said, the message informs A; if it changes the probability that this action will lead to a definite action, it is said, the message instructs A, and if it changes the significance of the result, it is said, the message motivates the action of A. An information measure for the "difficulty" of a definite task[23] is introduced. In order to solve the problem the destinator carries out an experiment based on the trial-and-error method. The "difficulty" is represented by a known function of the number of trials needed for the problem solution. Through the communication channel of the information system the destinator receives messages on the problem and in this way may change the amount of the trials and consequently the "difficulty" of the problem. Useful information is defined by the difference between the *a priori* uncertainty of the problem with respect to the given method of solving it and its *a posteriori* uncertainty.

The following concepts are typical for the fifth branch: in a cybernetic system information is a message to change the characteristics of the plant (to be controlled) and the external conditions influencing this plant[24].

Evidently the *syntactic* character of information is a property forming the basis of all other varieties of information in this branch.

(e) *Causal:* In the paper by Markov[25] it is said that the event A contains information on the event B with respect to the family of the laws of nature M if on the basis of the family M in case the event A is present a conclusion may be made for the presence of the event B.

(f) *Selective:* The transmitted message is a message chosen from a definite set of possible messages[1]. The process of receiving information is identified[26] by the process of reducing the uncertainty, because from a definite totality of possible phenomena in a given situation a phenomenon is selected which is in fact realized. Therefore in the concept of information the occurring phenomenon itself is not essential, while its relations to the totality of phenomena that could occur is.

(g) *Measurable:* If the set of the possible messages is finite then the number of messages or an arbitrary monotonous function of this number may be considered as a measure of the information created by choosing a message from this set; if a set of possible events is given, whose probabilities of occurrence are known, then by satisfying some mathematical conditions there exists a unique measure (the entropy function) however large the "choice" from such a set of events is or how indefinite its result[1]. With respect to the knowledge of the experimental procedure[27] a measure of differentiation is introduced for the independent categories "in terms of which alone the result can be described"; this measure is called prior or structural information. "The amount of structural information in a result, the logoncontent, is thus the number of independent categories or degrees of freedom precisely definable in its description." On the other hand, posterior (or metrical) information arises when the experimental procedure is carried out. "Each of these measures is quantal in its communicable aspects though, of course, the respective 'quanta differ qualitatively'."

(h) *Systematic:* Let the source of information in an information system be represented by the set of its output states and the rule (or set of rules) for their organization, while the destinator (possibly also the receiver) of information be represented by the set of its input states and the rule (or set of rules) for their organization. Thus a definite type of similarity (homomorphic, isomorphic, etc., with respect to the corresponding states and rules) arising in the destinator[28,29] is called information. This definition of the information is called systematic because it uses essentially the ideas of the recently developed systems theory for the building up of the above-mentioned mathematical models of the information of the destinator and source in an information system.

CONCLUSIONS

The information classification scheme using the deductive approach is given in detail in Figure 2. (For completeness of the scheme the third branch is also given in some details.) The blocks drawn in dashed lines indicate non-existing (but possible) characters of definitions of the information. Evidently the block "causal" has its place in the fourth branch also—all information processes

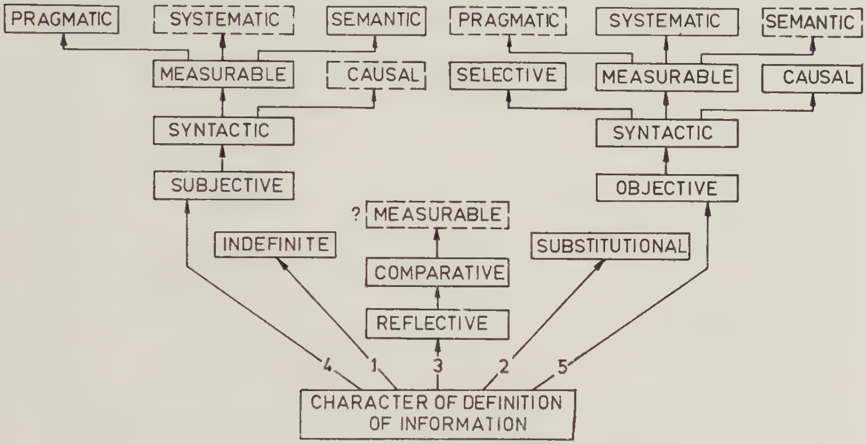


FIGURE 2 Detailed information classification scheme of the various sorts of information in the five branches from Figure 1

are subjected to the causality principle. The scheme presented shows that the fourth and fifth branches have, in general, the same structure. But some of the definitions cannot be strictly differentiated with respect to these branches (and some of the characteristics). However, as it is evident both are not quite completed. Without an analysis of the advantages and disadvantages of one or another concept for information an attempt was made here to generalize and to order a part (as we hope, the greater) of the known characteristics of the information in its various aspects. The aim of this attempt is thus to make for the clarification of the investigated problem.

References

1. C. Shannon, "A mathematical theory of communication", *Bell System Tech. J.*, **27**, 3–4 (1948).
2. U. A. Shreider, "O semanticheskigh aspectach teorii informacii", *Sb. Informacia i Kibernetika Sov. Radio* (1967).

3. U. A. Shreider, "K voprosu ob opredelenii osnovnigh poniatii semiotiki", *Kibernetiku na Slooshbu Kommunizmu, Energhia*, **3**, (1963).
4. A. Gharkevich, *Ocherki Obshey Teorii Sviazi, Gosteghizdaat* (1955).
5. A. M. Yaglom and I. M. Yaglom, *Veroyatnost i Informacia, Gosteghizdat* (1957).
6. H. von Teichmann, *Der Entropiebegriff in der Informationstheorie*, Der Fernmelde-Ingr. (1963) p.10.
7. C.V. Yablonsky, "Osnovnia poniatia kibernetiki," *Problemy Kibernetiki*, **2** (1959).
8. G. Glinsky, "Teoretiko-informacionnie problemy teorii nenadedgnigh avtomatov, *Teoria Koneshnigh i Veroyatnostnigh Avtomatov, Nauka* (1965).
9. W. R. Ashby, *An Introduction to Cybernetics*, Chapman and Hall, London (1959).
10. I. Novik, *Kibernetika*, Gospolitizdat (1963).
11. *Filosofskie Voproci Kibernetiki*, Sotzecgiss (1961).
12. N. Stanoulov, "On the concepts of information and control", *Progr. Biocybernetics*, 3 (1966).
13. F. P. Tarasenko, "Vvedenie v kurs teorii informacii", *Izv. Tomskovo Universiteta*, Tomsk (1963).
14. N. Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*, M.I.T. Press Mass. (1948).
15. A. Feinstein, *Foundations of Information Theory*, McGraw-Hill, New York (1958).
16. A. A. Krasovskiy and G. C. Pospelov, *Osnovy Avtomatiki i Technicheskoy Kibernetiki Gosenergoizdaat* (1962).
17. U. A. Shreider, "Ob odnoy modeli semanticheskoy teorii informacii", *Problemy Kibernetiki*, **13** (1965).
18. E. K. Voyshvillo, "Popytka semanticheskoy interpretacii statisticheskigh poniatiy informacii i entropii", *Kibernetiku na Slooshbu Kommunizmu Energia*, **3** (1966).
19. R. Carnap and Y. Bar-Hillel, "An outline of a theory of semantic information", *Brit. J. Phil. Sci.*, 4 (1953).
20. P. Neidhardt, *Einführung in die Informationstheorie*, Verl. Technik (1967).
21. A. A. Gharkevich, "O cennosti informacii", *Problemy Kibernetiki*, **4** (1960).
22. W. Miles, *The Measurement of Value of Scientific Information*, *Operations Research in Research and Development*, Wiley, New York (1963).
23. M. M. Bongard, "O ponyatii 'poleznaya informacia'", *Problemy Kibernetiki*, *Vyp.* **9** *Fizmatgiz* (1968).
24. E. A. Yakubaytis, *Osnovy Technicheskoy Kibernetiki, AN LATV SSR* (1962).
25. A. A. Markov, Shto takoye kibernetika, konferencia po filosofiskim problemam kibernetiki, *Problemy Kibernetiki*, **9** (1963).
26. V. M. Glushkov, *Cintez Cifrovigh Avtomatov*, Fizmatgiz (1962).
27. D. M. MacKay, "Quantal aspects of scientific information", *Phil. Mag*, **41**, 312 (1950).
28. I. Stanulov, "Opredelenie asemanticheskoy informacii v idealnoy informacionnoy sisteme pri pomoshi iszomorfizma", *Compt. Rend. Acad. Bulgare Sci.*, **28**, 8 (1967).
29. I. Stanulov, "Opredelenie asemanticheskoy informacii v realnoy informacionnoy sisteme pri pomoshi gomomorfizma", *Compt. Rend. Acad. Bulgare Sci.*, **20**, 9 (1967).

The meaning of cybernetics in revolutionary systems (those with rapid or discontinuous shifts in standards)

C. MUSES
Research Center for Mathematics, Morphology, and Morphogenesis
California, U.S.A.
and
Foundation for the Study for Consciousness
Philadelphia, U.S.A.

Summary

The author examines the following topics: Wiener's dilemma and contributions; cybernetics in various systems; hypernumbers; higher operator algebras and cybernetics. It is shown that higher operator algebras could make fundamental contributions to biological and psychological cybernetics. In part 3 epistemological operators are discussed.

INTRODUCTION

Norbert Wiener's philosophy of cybernetics is discussed in the light of his interest in significance and the deepening of research. The place of the cybernetics of systems not merely undergoing shifts in output, but with discontinuously changing standards of control and guidance is discussed, leading to an analysis of discontinuity itself. This analysis in turn leads to the consideration of hypernumbers and higher algebraic operators in the mathematics of self-regulative systems with observed discontinuities. The resolving of such discontinuities is then investigated in the light of the operator algebras generated by successively higher forms of unity of the identity operator. Finally, the epistemological operators contained in hierarchically highest form of unity are described. One theorem deduced from this investigation is that higher operator algebras generate mathematical logic, and hence mathematics, in this ontological operator sense, contains logic, which does not exist in a vaccuum but derives from experiential reality. Another conclusion is that with u_1, u_7 (the seventh form of unity) and thereafter, in u_9 the axiom of restriction* breaks down (i.e. unity breaks through) since for u_7 and u_9, $e(e) = e^2 = e$ where e may here denote either set membership or an elemental form of unity. And L. Löfgren[1], though he is not aware of hypernumber methods, has shown by logical set analysis that the negation of the axiom of restriction is to be associated with what he calls symbiotic and atomic self-reproduction, the latter being a stronger, i.e. more unified form. Higher operator algebras have fundamental contributions to make to biological and psychological cybernetics[18].

PART 1

Since one of the purposes of this Congress is to commemorate the twenty-first anniversary of the founding of cybernetics, it may not be inappropriate to mention a few words about the aims of its founder. In 1962, while we were co-lecturing at the University of Naples, Norbert Wiener asked me one day over refreshments, "Do you think I should have written the book? I sometimes wonder, for it might lead to the misuse of gadgets to stupefy and control men rather than free them." This example from personal experience shows how deeply Wiener was involved with the sociological and psychological aspects of the meaning of cybernetics, even to the extent of a probing concern for the feedback effects of cybernetics itself on the structure of society and human relations.

Norbert Wiener's life was an ever deeper and widening search for more profoundly integrated meaning. Thus the origins of cybernetics stem not only from his mathematical genius but from his fundamental concern with the demanding philosophic problems connected with the unity and continuity of the world on all levels—a unity of which feedback and autocorrelation were for him simply two consequences.

---

* The *axiom of restriction* states that: given a non-empty set or system S, a member s of S exists such that no member of s is a member of S; symbolically, $e(eS) = es \neq eS$, or $e^2S \neq eS$, where e denotes membership in a set.

Though his scientific contributions lay in mathematics, Wiener's greatest human and historical contribution consisted in just this unitary point of view, focused and implemented by his creation of that new interdisciplinary discipline he named cybernetics, which is most simply defined as the science of regulative processes. At this point we do well to remember that Norbert Wiener won his doctoral degree in philosophy, and his interest in *meaning* was always prominent.

Again noting this interrelatedness of the various aspects of the world by the very fact that they all are aspects of the *same* world, there is a quotation from Wiener[2] which I feel is most basic to his viewpoint, (cited also in the obituary essay I was called upon in 1964 to write by the National Research Council of Italy) the gist of which he paraphrased many times in conversation, though the remark appears in published form first in "Mathematics of self-organizing systems", his chapter of a 1962 symposium *Recent Developments in Information and Decision Processes* (p. 20):

"It is extremely important to have a broad basis in very different sciences for one's intellectual work so that one can follow the problem wherever it leads, even though it crosses boundaries... If a problem leads us into a new field in which we have no knowledge, we should acquire such knowledge. It is no excuse, when working on a problem, to say 'but that's not my field'."

The expository foundations of cybernetics lie in mathematics, as do those of all the other sciences, mathematics now being considered not only quantitatively but qualitatively, as the study of all possible forms and relational structures in any context.

The sciences have always progressed by men's attempts to understand more precisely and profoundly the nature of some experiential problem, verifiably and repeatably encountered. Inevitably this process of understanding invoked new mathematical methods, which is to say new concepts and new ways of reasoning, without which the encountered experience could not be resolved or clearly comprehended, but with which it could be unambiguously conceived and expressed—far better than with words alone, although words are a necessary first stage rocket.

Seen in this light, the progress of science is a sequence of increasingly successful attempts to bridge or resolve discontinuities of understanding with respect to human experience. In this way the cybernetics of systems with discontinuously changing standards is seen as fundamental to the discovery process and to the development of science itself.

Imaginary Numbers

One of the greatest integrating leaps across fields of knowledge in order to solve a problem—the kind of interdisciplinary joining leap that Wiener so clearly and insistently advocated for cybernetics in the quotation previously cited—was made in the latter part of the 19th century, when it was found that the so-called "impossible" or "imaginary" numbers, based on $\sqrt{-1}$, could be used to interpret Laplace's equation in two dimensions more deeply and originally, thus resolving many intractable problems of fluid flow and electric circuits. Almost concurrently the application of these same hypernumbers in the theory of wave propagation of all kinds, as well as to alternating current theory, was learned, in the late 19th and early 20th centuries.

By providing a new kind of number, gaps that were thought to be discontinuous were able to be bridged continuously with respect to a more profound context stemming from a deeper understanding of the structure of experiential reality and its mathematical mapping. Without the use of the higher form of unity, $\sqrt{-1}$, many of Norbert Wiener's mathematical findings would not have been expressible, since the entire field of the functions of a complex variable could not then be conceived.

This first new form of unity i, where $(\pm i)^2 = -1, i^0 = +1 = |\pm i|$, made possible a whole new realm of algebra and hence of function theory, in turn leading to new types of space or extension and their topologies, the so-called complexes spaces. Functions that showed unbridgeable gaps when only the usual form of unity, ± 1, was admitted, become continuous when the additional form of unity, $\pm i$, is also admitted. Thus the gap between the isolated point and the continuous part of the curve $y^2 = x^3 - x^2$ is closed by means of i; and the semicubical parabola $y^2 = x^3$ becomes continuously existent in complex 3-space for negative values of x; also the ordinary parabola $y = x^2$ becomes continuously existent for negative values of x, and is seen as one section of a hyperbolic paraboloid or saddle-shaped surface in complex 3-space. In fact there are no real functions with gaps that cannot be closed when the basis of the number system is deepened, and in this sense it may be said that all discontinuities (and hence all discontinuously changing standards) are discontinuous only relatively, being resolved when a profound enough basis for experience is reached. All previous discontinuity will then vanish, now being bridged by a connected path or paths.

It was Hermann Weyl[3] who first pointed out in 1917 that $\pm i$ are the natural units for the first *negative* dimension, which must be the one assigned to time.* And so it is in the Minkowskian 4-space adopted by relatively theory, now the geometrical basis of field physics. Weyl, however, did not stress this conclusion from Minkowski enough, since he was not too interested in emphasising that space–time was not isotropic or homogeneous, and that time played in it a significantly and mathematically very different role from space—as different as $\sqrt{-1}$ is from 1.

With all these rich results for mathematics and cybernetics consequent upon the adjunction of a new form of unity in relation to the mapping of reality, it is natural to ask in what ways, and how far, the concept of number can be extended beyond $\sqrt{-1}$.

Hypernumbers

As early as 1876 the English mathematical genius of William Kingsland Clifford, inspired by the profound findings of William Rowan Hamilton and Hermann Grassmann in 1843 and 1844, conjectured[4] that there must be a number, let us call it ε, such that $\varepsilon^0 = 1 = \varepsilon^2$ but $\varepsilon \neq \pm 1$. It also turns out that both the length of the vector ε and its absolute value are unity,

---

* In Weyl's development of general relativistic space we have n positive dimensions and one negative dimension. One finds the same structure in generalizing Euler's and Descartes' formula

$$N_0 - N_1 + N_2 = 2$$

$N_{0,1,2}$, being the respective numbers of vertices, edges and faces (i.e. zero-, one-, and two-dimensional elements) of any simply connected, or topologically sphere-equivalent three-dimensional form (i.e. genus = 0), including concavo-convex polyhedra and forms with curved surfaces. The vertices, edges and/or faces may be at peaks (convex) or in valleys (concave). Since every such three-dimensional form contains exactly one three-dimensional element, we have $N_0 - N_1 + N_2 - N_3 = 1$ where $N_3 = 1$. But this is not the best generalization, which should show the genus on the right-hand side. Hence we must write, where g is the genus, in this case zero,

$$-N_{-1} + N_0 - N_1 + N_2 - N_3 = 0 = 2g$$

For n dimensions,

$$\sum_{k=-1}^{n} (-1)^k N_k = 0, \text{ where } N_{-1} = N_n = 1.$$

We thus notice that every such n-dimensional form possesses a single unique negadimensional element, i.e. a characteristic time or frequency pattern.

i.e. $L(\pm\varepsilon) = |\pm\varepsilon| = 1$, which was true also for i or $\sqrt{-1}$.* This new kind of number we have found to be a necessity in formulating the structure of algebraic number fields of positive or hyperbolic basis, just as i is essential for number fields of negative or elliptic basis. At the turn of the century, both Felix Klein and Henri Poincaré almost saw this connection, and came just short of formulating it in their writings on the nature of infinite lattices or networks in algebraic number fields and in particular on such meshes being governed by ellipses or hyperbolas.

Very little was done with Clifford's suggestion of ε, however, except in theoretical kinematics, which today is subsumed under geometric transformation theory. One reason for this historical block is that even as late as 1968 it was being insisted that ε did not have a square root.

Obviously such an impasse would prevent any algebraic development or useful application. In February 1968 we were able to prove that ε had all the even roots it should have, including the square root, which is $\pm\frac{1}{2}(1 + \varepsilon - i + \varepsilon i)$. Observe that ε thus requires a 4-space for its powers (i required only a 2-space) and that i is as involved in the mathematical existence of ε as it is in that of (-1). No more strikingly could be demonstrated the *continuity* of all mathematical concepts and hence of the mapping of the nature of reality, which mathematics has through the centuries never failed to do with increasing precision. We now have the important exponential relation

$$e^{\pm\theta\varepsilon} = \cosh\theta \pm \varepsilon\sinh\theta$$

which is readily provable by the ordinary infinite expansion of the exponential function in conjunction with the algebraic rules for ε. Just as i was essential to the mapping of time-as-change so ε maps the counter-aspect of time-as-duration. The first is circular; the second, hyperbolic.

This exponential relation makes much of the work in hyperbolic geometry very simple, and also opens the way toward the demonstration that waves in the form of $e^{i\sqrt{\theta}} = e^{\pm\varepsilon i\sqrt{\theta}}$ may contain an ε as well as an i component. This new mathematical fact augments the entire basis of the Schrödinger wave equation and the Dirac–Heisenberg–Schrödinger development of quan-

* That $L(n)$ and $|n|$ are not in general equal, where n is a number of any kind, is seen from the example that $L(1 \pm \varepsilon) = \sqrt{2}$ but $|1 \pm \varepsilon| = 0$. The *absolute value* of a number is its real "perihelion distance", that is, the extension on the real axis, taken as positive from the origin, to the real perihelion, i.e. to the point of the number's orbital intersection with the real axis; whereas the *length* is the extension from the origin to the place of the number in its orbit, i.e. the radius vector of the number. If the orbit does not share a point with the real axis, an absolute value in this sense does not exist.

tum theory, as well as the related theories of probability and systemic stability, the latter playing of course a great role in cybernetics. In recent months the physicist Eugene Wigner has written to me that he was most interested in the fact that the algebra of ε was now made workable through the explicit expressions for its fractional powers, and in particular its square root.*

Clifford also conjectured that there should be a number,† let us say p, such that $p^2 = 0$ although $p \neq 0$. Eduard Study in the early 20th century tried to develop the algebra of p, but failed because he followed Clifford's one mistaken conjecture that the zeroth power of p should be unity, whereas the truth of the matter is that $L(\pm p) = 1$ but $p^0 = 0 = |p|$, as we have recently shown. Further details are given in Part 2 of this paper which is concerned with the more detailed aspects of the ideas we are discussing.

---

* We recently proved also that the idempotents $(1 \pm \varepsilon)/2$ furnish the mathematical basis for the two projection operators (for positive and negative energy) used in the important Brillouin–Wigner expansion theorem in modern quantum physics.

† It is often misunderstood that Grassmann never progressed to the explicit formulation of p as a number; for him it remained an unanalysable arbitrary symbolic mark such that $p^2 = 0$ and was related to his symbolization of the line AB as the "product" of the two marks A and B denoting its end points. Then as the length of the line shrinks we have in the limit $A = B$ and $A^2 = 0$. But this is a far cry from the algebraically manipulatable number $p \neq 0$, such that $p^2 = 0$, as Clifford first conceived it. The crucial difference, not made clear by either Clifford or Study is that $|\pm p| = 0 = p^0$ but $L(\pm p) = 1$, where L is vector length as usual, denoting the operator that yields it.

Henry J. Smith[4], Savilian Professor at Oxford, one of the great mathematical expositors of the last century, had some very penetrating remarks to make on this question in his introduction to the *Collected Papers* of Clifford, published in 1882 after the latter's untimely death in his thirties. On pages lxiii–lxvii Smith shows that Grassmann's *Ausdehnungslehre* (of either 1844 or 1862) actually does not contain the concept of Hamilton's extension of $i \equiv i_1$ into i_2 and i_3, such that $i_n^2 = -1$, $i_n i_m = i_m i_n$, i.e. $i_1 \neq i_2 \neq i_3$. Clifford tried to extend the index arbitrarily, but overlooked the fact that after i_3, multiplication is in general no longer associative, just as after i_1 multiplication can no longer remain commutative—facts often overlooked by those who discuss "Clifford algebras". Thus if $i_1 \neq i_2 \neq i_3$ is assumed, where $i_a i_b = i_b i_a$, inconsistencies will arise at once, since $i_1 i_2 = i_3 = i_2 i_1$ implies $i_2 i_3 = i_1$ and also $i_3/i_2 = i_1$, i.e. $i_2 i_3 = -i_2 i_3$, whence $i_2 i_3 = 0$ which is a contradiction. Smith on p. lxvii notes this contradiction but does not state as he should that it is fatal to any assumption that a commutative i-algebra of more than one i-element can exist. Similarly, no i-algebra of more than seven i-elements can exist without containing divisors of zero. Bertrand Russell's view of mathematics as giving unrestricted license to assumptions is again shown to be wanting. Assumptions are definitely limited by the structure of mathematical reality which in turn maps that of experientiable reality. Finally, expository and experiential reality merge, and explanation becomes experiencing itself. The ultimate definitions are operational.

Suffice it to say that the false assumption that $p^0 = 1$, and the associated (and also erroneous) assumptions that $p^{-1} = 1/p$ and that neither effectively existed, have held up the development of p algebra just as the false assumption that ε had no square root prevented the development of its algebra.

That the higher form of unity expressed by p is by no means a mere mathematical curiosity becomes at once evident when fractional dimensions are considered; and fractional dimensions are inherently involved with fractional integration (and differentiation) which enter essentially into certain problems concerned with transient or suddenly perturbed states of systems. They are essential to continuity.

Another great mathematical discoverer, Oliver Heaviside, after whom the principal reflecting layer of radio waves from the ionosphere is named, was the first to prove the utility of fractional differentiation in a problem of transients or perturbations in systems comprised of inductance, capacitance, and resistance. And the entire subject of perturbations or transients and their decay is bound up with feedback control in self-regulative systems (we prefer the more precise term *self-regulative* to the less satisfactory and more confusing "self-organizing"), in turn intimately related to the theory of systemic stability and the consequent analysis of differential equations.

These fractional dimensions, which provide the geometric continuity between integer dimensions, are seen so be related to matters of profound cybernetic interest. When one begins to ask precise questions about the structure of fractional dimensions and the geometry and topology of their possible forms, one finds very quickly that there are discontinuities which must be resolved to answer these questions, and that these discontinuities cannot be resolved by simply using 1, i or even ε; rather, p must be employed. And thus the (proper) $\sqrt{0}$, the third form of unity, u_3, which we have designated as p, is as much a mathematical necessity as $\sqrt{-1}$. *The proper nth root of z* is defined to be such that it contains neither $\pm z$, and yet yields z when multiplied by itself n times. Thus ε is the proper square root of $(+1)$, as p is of zero; and i, of (-1).

Actually i, ε, and p are higher forms of the extension operator, and we shall see in the following part of this paper that the most economical and adequate way of regarding these higher forms of unit extension together with the great deepening and mathematical power they afford, is to consider them in the light of the primal elementary operators which generate higher operator algebras, whose constituent operators are in general not bounded (see p. 191, end of Part 2). John von Neumann's work is very important here.

Just as *i* provided the mathematics for electromagnetic fields and waves, so the primal elementary operators or kinds of extension beyond *i* will map the reaches of nature that extend reality beyond and deeper than electromagnetic fields. It is not too much to hope that the self-regulative processes of life and mind may thus be brought within the reach of accurate and penetrating human comprehension,* carrying advanced cybernetics to heights as yet unscaled and even now unguessed.

The key to this exciting and challenging development I venture to predict will consist in our increasing grasp and understanding of the more powerful and profound kinds of number, together with the operations and conceptions they manifest and explain. Armed with such puissant means, man may ask of this Great Computer even his most far-ranging questions concerning life and mind, and sanely dare to expect real and usable answers in the forseeable future. The conclusion† of Morris Kline[11], Professor of Mathematics and Director of the Division of Electromagnetic Research at the Courant Institute in New York that "what is needed is a new evaluation of mathematics" is eminently justified. Higher forms of the number concept appear to point to the natural and workable way toward such a revaluation, which also well may illustrate the meaning of cybernetics in that most revoutionary system called "the 20th century".

---

* The London forces, named after F. London[5] who first clearly explained and noticed them in 1930 and in the 1936 *Transactions of the Faraday Society*, are a case in point.

They arise from the quantum-indeterminate variety of the van der Waals forces; and at close distances they control the apolar attraction between self-complementary molecules[6] like DNA and RNA. Such forces thus play a key role in meiotic synapsis[7] and mutagenesis[9] (*cf.* H. J. Muller[7,8] and the useful recent summary [9]of the work on London forces by H. Jehle). That role, in the light of the London forces, permits the arising of an epigenetic, noetically guided RNA mutagenesis, a possibility Muses[10] first pointed out in 1962 and again in 1969, confirmed by C. C. Lindegren.

Our second proposal is that London forces can mediate all types of noetic guidance of physiology by the principle of a target-seeking process able to modify and modulate the distribution of a random fluctuation[18]. Such random fluctuations, inherently unsuppressable and by definition physically undeterminable, are the basis of the London forces and their phenomena.

The occurrence of a six-dimensional relationship in the governing operations of the London forces and an independent mapping of noetic energy on **W**, the domain of w, suggest homomorphic relations to the behavior of the hypernumber $w \equiv u_6$, a primary operator that will be discussed mathematically in Part 2.

† p. 31, *Science in the Sixties*, a symposium published by the University of New Mexico, Albuquerque, 1965, and sponsored by the U.S. Air Force Office of Scientific Research.

PART 2

The Forms of Unity

Each one of the hypernumbers u_n, which are primal forms of unity i.e. of length $L(u_n) = 1$ running perpendicular to each other and to the real axis, generates a distinct operator algebras in characteristic operational spaces, and a consequent function theory. Considering that fact, and also that we have found nine primal forms of unity, it would require a good volume for their detailed treatment.

There is a wealth of new data and results thus far gained, with a great deal more still to be found. The present occasion does not permit giving all the theorems, except to state the governing fact that they arise most readily from the structure of the orbits of the various kinds of number. We can give here but a few of the salient and most interesting phenomena characteristic of the various hypernumbers. All such phenomena, it may be observed, would be much more difficult to understand without the powerful tool of orbital analysis. As the mathematician B. Segré has well noted in the foreword to his published lectures of 1950:[12] "It is highly desirable that more intimate contacts and new relations should be established between algebraic geometry and arithmetic." H. Minkowski in 1910 was one of the first who established such new relations with his "geometry of numbers."

The algebraic geometry of hypernumber spaces, which can be efficiently developed by the conception of number orbits, carries into new dimensions of meaning the enterprise of Minkowski, Segré and their heirs. For the present it will suffice to define the orbit of u_n, $n > 1$, as possessing a maximum of points on the plane formed by the u_n axis and the real (or u_1) axis; each one of these points being a number generated either by a real power of the hypernumber u_n or by the exponential $e^{\theta u}n$. For example if $n = 3$, 4 or 5, the orbit is generated by the exponential; and in the case $n = 2$ i.e. $u_2 \equiv i$, the orbit may be considered to be generated in either fashion since i^k may be expressed in terms of $e^{\theta i}$ and vice-versa.*

---

* Much was cut from the original text by space limitations. See the supplementary note, p. 199, for a few more details. The 9 primal forms of unity are denoted by $u_1, \ldots, u_9$; and their number domains in connection with ordinary ("real") numbers, by $U_1, \ldots, U_9$, where $u_1 \equiv 1$ and $u_2 \equiv i$.

Conclusions on Algebraic Structures

From the analysis, as shown in references 12–16 and the foregoing, it is evident that the usual definitions of an algebraic, vector, or operator space are not adequate. Even the most general of such definitions assumes a commutativity of addition and of multiplication by scalars; as well as a distribution of multiplication by scalars, whether on the coefficient *(antilog)* or exponent *(log)* level. Thus $x(a+b)y$ is assumed to be equal to $(xay + xby)$ where a, b are real; as also is assumed $(x^a y^b)^c = x^{ac} + y^{bc}$.

Yet even in the domain U_2, $(i_1^{1/2} i_2^{1/2})^2 \neq i_1 i_2 = i_3$, since $(i_1^{1/2} i_2^{1/2}) = \frac{1}{2}(1+i_1)(1+i_2) = \frac{1}{2}(1+i_1+i_2+i_3)$. And as we have seen, in the domain U_8, $(a \cdot x)$ is no longer equal to $(x \cdot a)$; moreover, again returning to U_2, $\log i_1 = \frac{1}{2}\pi i_1$ and $\log i_2 = \frac{1}{2}\pi i_2$; hence $\frac{1}{2}\pi i_1 + \frac{1}{2}\pi i_2 \neq \frac{1}{2}\pi i_2 + \frac{1}{2}\pi i_1$ because the antilog of the left side is i_3, whereas that of the right side is $(-i_3)$. On the *log* level, even in U_2 addition is noncommutative.

Thus the prevailing definitions of algebras are in fact restricted to two primal classes of algebras: that of $U_1 \equiv R$ and $U_2 \equiv I$.

The commonly found "general" definitions of complete vector or operator spaces are misleading because they are needlessly restricted in space without stating they are. Perfectly well definable classes of algebras and arithmetics (including U_1 and U_2 and growing out of them) exist and operate under far less restricted conditions, which we now define, noting that the u_n generate *classes* of algebras, and not merely algebras, and moreover furnish the actual arithmetic for each class, as follows:

(1) Every pair of finite, non-null, single-valued elements x, y in the space or algebra, whether x, y are scalars i.e. $\in U_1$, or not, has a finite single-valued product xy, and $x, y = y, x$ is not assumed, nor is $(x, y), z = x, (y, z)$. Similarly, x and y have a single-valued sum $(x + y)$, and $x + y = y + x$ or $(x + y) + z = x + (y + z)$ is not assumed. Moreover, if $x = yz$, the operation that furnishes z, given x and y, is called *division*, the inverse of multiplication, and is expressed by $1/y \cdot x = z$ or $x \div y = z$, where z is the *quotient* or result of x *divided by* y, the term $1/y$ being called the left reciprocal of y (if we had had $x = zy$ to begin with, the right reciprocal would have developed); and if $x + z = y$, the operation that furnishes z, given x and y, is called *subtraction*, where z is the *difference* between x and y in the sense "y removed from x", written $(x) - (y) = z$; $-(-y) = +y$ is not assumed.

(2) The identity operator or real unit element $u_1 \equiv 1$ is either explicit or

implicit (i.e. isomorphically and isometrically adjoinable) in every algebra, i.e.

$$x \Rightarrow 1 \cdot x \Rightarrow \text{existence of } u_1 \equiv 1 \tag{1}$$

(3) There is one uncountably infinite set, 0, of null elements in every algebra such that

$$(+x) - (+x) = -(-x) + (-x) = 0 = 0^n \tag{2}$$

where n is positive, real and finite, and where the equality sign denotes a lack of finite difference, i.e. only common membership in the set 0 and not member identity.

(4) There is one uncountably infinite set, ∞, of real unbounded elements such that

$$0^{-1} = 1/0 = \infty \tag{3}$$

where the equality sign here denotes merely membership in the set ∞, and not member identity; and any u_n may be adjoined to the set ∞. Observe that not even commutativity or associativity of scalars is assumed.

Thus, only 1 (explicitly or implicity) and the quasi-fields 0 and ∞ are needed, together with determinate operations of multiplication, addition and their respective inverses, division and subtraction. Reciprocation or division into unity is thus included, and exponentiation arises out of self-multiplication. In higher operator algebras, negative exponentiation or inversion is no longer equivalent to reciprocation of positive powers, and already in U_3 we see that $1/(1 + \varepsilon) \neq (1 + \varepsilon)^{-1} = (1 + \varepsilon)/4$. Thus reciprocals and inverses are not in general equivalent. Again, in U_7, $1/\Omega^{-n} \equiv \overline{\Omega}^{-n} \neq \Omega^n$, as we have also seen. Thus the class of algebras generated by even the third form of unity ε, already goes far deeper than that of complete normed vector spaces over the real or complex fields (the so-called Banach algebras), in which inverses and reciprocals are undifferentiated.

It is increasingly clear that the forms of unity are the primal operators of all mathematics, and it is their behavior and operations under various specifiable conditions that determine all possible classes of arithmetics, and their algebras, function theory and functional analysis. Rather than endlessly sailing round and round in proliferating seas of assumption sets whose interrelations are not perceived, and developing details (often trivial) concerning real and complex spaces—when we become aware of the primal operators and their use, we can steer a minimally entropic course to any place in the ocean of universes comprising mathematics.

Even in the comparatively but deceptively simple domain $U_2 \equiv I$ there is current confusion, ever since W. R. Hamilton wrote the complex number $(a + bu_2) \equiv (a + bi)$ as "(a, b)", whence it was erroneously concluded that i-algebra had been "reduced" to expression in the real domain.

But the symbol (a, b) is only that, and is not an explicit mathematical expression, and i is not mathematically reducible to parentheses and commas. Indeed U_2 cannot be expressed in terms of U_1 because of the fact that i-multiplication cannot be written in terms of iterated addition, as can r-multiplication. For example, $3 + 3 + 3 = 3 \times 3 = 3^2 = 9$, and $\underbrace{n + n + n + \ldots + n}_{n} = n \times n = n^2$; whereas in U_2 we have $i \times i = i^2 = -1$, but $i^n \neq \underbrace{i + i + i + i \ldots + i}_{k}$ for $n > 1$ and for any $0 \leqslant k \leqslant \infty$, and k must be real because it represents an enumeration. Thus $(2u_2 \times 2u_2)$ can never be expressed as an addition of like terms, as can $(2u_1 \times 2u_1) = 4 = 2u_1 + 2u_1$. Hence it would be not only misleading but false to say that Hamilton's n-tuple symbolism can express the operations of U_2 in terms of those of U_1, for that can never be done. Only real numbers can denote enumerations or iterations, and that is their main distinction.

On the other hand, by recognising the great operational difference between ordinary nunbers (i.e. U_1) and the higher classes of algebras severally generated by the forms of unity beyond u_1, much crudeness, confusion and lack of creativity and rigor may be eliminated in mathematical thinking and conception.

We shall now conclude the discussion of these primal operators with the last, u_9.

$M \equiv U_9$

We have mentioned reciprocals from time to time, and observed that their behavior is often quite distincive. Using our usual notation $\bar{x}$ for $1/x$, if may be seen (references 12–16)

$$\pm\overline{(u_1 \pm u_3)}, \pm \bar{u}_4, \pm \bar{u}_5 \in N(\infty u_{3,4,5}) \tag{4}$$

where $\overline{u_1 + u_3} = 1/(1 + \varepsilon)$, $\bar{u}_4 \equiv \bar{p} \equiv 1/p$, and $\bar{u}_5 \equiv \bar{q} = 1/q$. Also $\bar{u}_7 = u_7$, i.e. $1/\Omega = \Omega$; and where u_8 is the integration operator, $\bar{u}_8 \in U_8$, e.g.

$$\frac{1}{\int_1^t \log t \, dt} = t = \int_1^t u(t) \, dt \tag{5}$$

Also $(1/\sigma^n) \neq \sigma^{-n}$ and $(1/\sigma^{-n}) \neq \sigma^n$. Indeed

$$\left(\frac{1}{u_n^k}\right) \neq u_n^{-k} \tag{6}$$

for $n \geqslant 3$, and $k > 1$.

It also true that all the reciprocal domains through

$$(\bar{U}_5 \equiv \bar{Q}) \in (iU_7) \tag{7}$$

for example

$$\bar{q} \in i\Omega \tag{8}$$

Also

$$\bar{U}_7 \in U_7 \text{ and } \bar{U}_8 \in U_8 \tag{9}$$

And

$$\bar{u}_6 \in U_6, iU_7 \tag{10}$$

But $m^n = m \equiv u_9$ is not a member of any other domain thus far considered. The equation (11) yields a totally new operator, m, and a correspondingly new domain M, which is thus the exceptional domain U_9, where $m \equiv u_9$. Its orbital form is given by Figure 1, whence we see that each km (k real) has a different orbit, and there are six *classes* of shapes: a circle for $|k| = \infty$; a Cassinian oval for $\sqrt{3} \leqslant |k| < \infty$; the indented or violin-shaped Cassinian oval for $\sqrt{2} < |k| < \sqrt{3}$; a Bernoullian lemniscate for $|k| = \sqrt{2}$; the paired, separated Cassinian ovals for $1 < |k| < \sqrt{2}$; and a point pair for $|k| = 1$, whence $(\pm m)^u = \pm m$ for $0 \leqslant n < \infty$. And

Orbit (u_9):

$$\varrho^2 = \cos(2\theta - \pi) \pm \sqrt{c^4 - \sin^2(2\theta - \pi)} \tag{11}$$

whence the m coordinates (k_1, k_2m) of any orbital point are given by $k_1 = \varrho \sin \theta$, $k_2 = \varrho \cos \theta$, where $e^{\pm \theta m} = k_1 \pm mk_2$, where θ is the angle to R formed by ϱ, and where $\varrho^2_{\theta=0} \equiv \varrho_0^2 = c^2 - 1$. Thus if $\varrho_0^2 = 1$, $c = \sqrt{2}$ and $\varrho_{\pm\pi\,2} = \pm\sqrt{3}$; if $\varrho_0^2 = 0$, $c = 1$ and $\varrho_{\pm\pi\,2} = \pm\sqrt{2}$; and if $c^2 < 1$, ϱ_0 is no longer real, although $\varrho^2_{\pm\pi/2} = c^2 + 1$ is real for all real values of c. Note that $\frac{1}{2}(\varrho_0^2 + \varrho_{\pi\,2}^2) = c^2$. If $c = 0$, then $\varrho_{\pi\,\pm 2} = \pm 1$, i.e. the pupils of the two "eyes" of Figure 1 In this case we have, where $2\theta - \pi \equiv \phi$,

$$\varrho^2 = \cos\phi \pm \sin\phi\sqrt{-1} \tag{11a}$$

or

$$\varrho = \pm e^{\pm\frac{\theta}{2}\sqrt{-1}} \tag{12}$$

which shows that there is an orbit connecting $(\pm m)$ which is not wholly in either (m) or $(m\varepsilon)$ space; and there is a similar orbit in $\sqrt{-1}$ connecting $(\pm m\varepsilon)$.

Also $(m)^2(\sqrt{3})^2 = 3m$; and $(m)^2(\sqrt{2})^2 = 2m$; but (see Figure 1 and Equation 11)

$$(\pm m\sqrt{3})^2 = -1 \tag{13}$$

$$(\pm m\sqrt{3})^0 = +1 \tag{14}$$

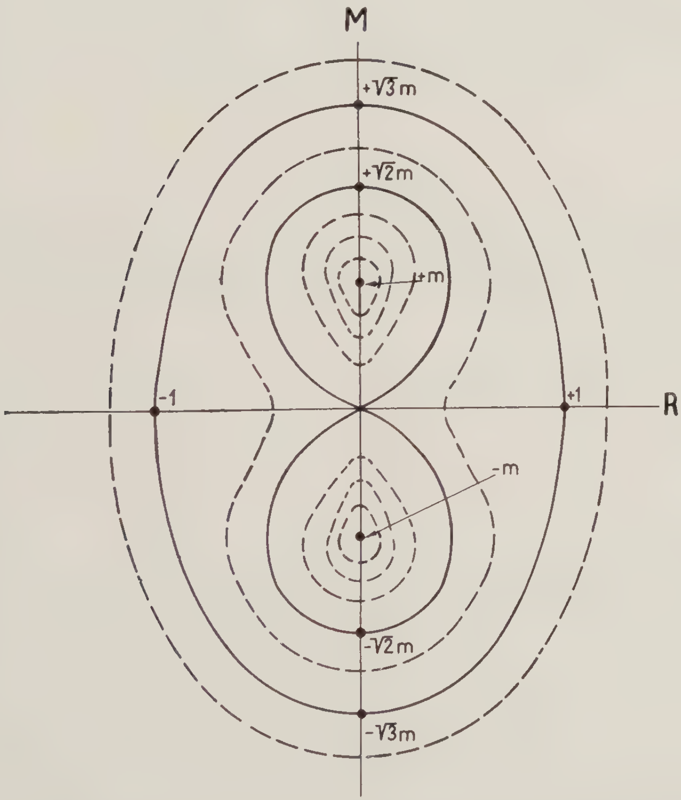


FIGURE 1 The orbits of $m \equiv u_9$ form the complete infinite set of Cassinian ovals of six species. The orbits in real polar form are given by Equation (11).

and

$$(\pm m\sqrt{2})^0 = (\pm m\sqrt{2})^2 = 0 \tag{15}$$

$$(\pm m\sqrt{2})^3 = (\mp m\sqrt{2})^3 \tag{16}$$

Also

$$L(\pm m) = 1 \tag{17}$$

$$L\,(\pm m\sqrt{2}) = \sqrt{2};\; L\,(\pm m\sqrt{3}) = \sqrt{3} \tag{18}$$

But

$$|\pm m\sqrt{2}| = 0;\; |\pm m\sqrt{3}| = 1 \tag{19}$$

and $|m|$ cannot be expressed in terms of any other thus far known domain, since the path of m^k cannot so be derived or elucidated, since its orbit does not intersect the real axis. Actually, it is found that the set of values m^k and $(m\varepsilon)^k$ lead respectively coincidentally on a hypersphere whose radius ψ is such that $|\psi| = \psi$; $-\psi = +\psi$; $\psi^n = \psi$; and (see Part 3)

$$1/\psi = 1,\; \psi \neq 1 \tag{20}$$

The operator $(m\varepsilon)$ forms a set of orbits in the (ME, R) plane perpendicular to the m orbits (Equation (11)) in the (M, R) plane. Moreover $m(\sqrt{3})$ and $(m\sqrt{3})$ can respectively replace (i) and (εi) and U_2 can thus be superseded by U_9 as a meaning of $\sqrt{-1}$.

Now it also can be shown that

$$|\pm m| = 0\psi \tag{21}$$

and

$$|\pm m\varepsilon| = 0\psi \tag{22}$$

which indicate are the minimal distances from the origin on the (m) and $(m\varepsilon)$ orbits. All the relations stated for (m) are true for $(m\varepsilon) = (\varepsilon m)$. The absolute or perihelion values of (m) and $(m\varepsilon)$ are no longer expressible in terms of real numbers, although their lengths are each unity (Equation (17)). Note, however, that $|mk|$, $k \geqslant \sqrt{2}$, is expressible as a real number.

That there are no elementary operators beyond (m) finds further support in the pioneer work of J. von Neumann on unbounded operators, i.e. whose domains are not necessarily confined to a finite number of dimensions, for example the differential operator, $u_8^{-1} \equiv \sigma^{-1}$.

In a very deep paper* von Neumann[17] showed that any such operator could be transformed into an extension of any other in not more than nine steps. This maximal number corresponds to the nine classes of elementary operators u_n, $1 \leqslant n \leqslant 9, L(u_n) = 1$, that we have been considering throughout. They may be called the primal forms of unity.

PART 3

Higher Operators than the u_n

The length operator L is independent of all the nine u_n and therefore its units cannot be in terms of any of them. The unit resulting from the $L(\ )$ operation is ψ, defined in Equation (27), and hence ψ is the radius of a hypersphere of at least nine (meta)dimensions, one for each u_n. Since ψ is the unit of the $L(\ )$ operation, we now see that

$$L(u_n) = 1\psi = \psi = L(\psi) \tag{23}$$

Within the hyper-metadimensional sphere (i.e. more than three metadimensions) of Ψ are contained a hierarchy of higher operators† governing the operators resulting in measure, permutation, combination, systemic programming, proof, and the rest of the operational alphabet of epistemological (including logical) processes. These processes include those of logic,

---

* Recently, Michael Bernkopf, Professor of Mathematics at Pace College, N.Y., has given an excellent summary of von Neumann's key work on operators. Bernkopf rightly underlined the fact that von Neumann showed matrices were not the suitable approach to the investigation of higher types of operators, but that operator algebra was, thus effectively demonstrating Oliver Heavide's vision. Indeed, hypercomplex numbers $\sum_{k=0}^{n} a_k i_k$, where $i_n\,(i_n^2 = -1) \in U_2$, cannot in general be represented by real or even complex matrices, as is so often stated or implied in the literature. Matrices will serve only through $n = 4$. Beyond that, because matrix multiplication cannot handle nonassociative factors, the matrix approach to representation breaks down even for Cayley–Graves numbers, i.e. $n \leq 7$. And for unbounded operators, e.g. $\mathrm{d}/\mathrm{d}t \equiv \mu_8^{-1}$, matrices completely break down, as von Neumann demonstrated[17] with such elegant irony. It is our present proposal—and much more remains to be done—that hypernumber operators can make abstract (and concrete or applied) operator algebra as much more effective as operator algebra can make improvements on matrix algebra—and mostly in the same context: that of workability, actual interpretation and significance, and the entire domain of representation theory.

† See *Part 3* and Figure 2.

for logic is that branch of epistemology concerned with how we know that a conclusion is valid; and epistemology in turn is rooted in ontology, the nature and structure of *experienceable* reality—which is scientifically wider, in an evolutionary and potential sense, than any merely current *experiential* reality.

We have found eighteen basic operators on the circles of Ψ (Figure 2 and the following section on the epistemological operators).

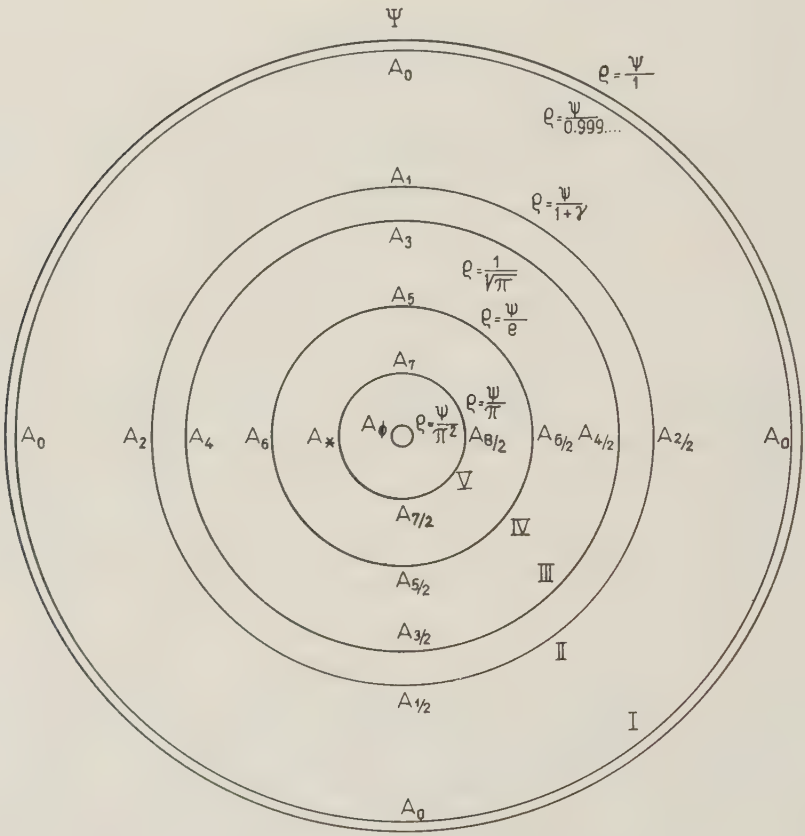


FIGURE 2 The circles of Ψ showing the schematic placement of the epistemological or loom operators (see Part 3). Between A_1 and $A_{7/2}$ inclusive lie the operations of logic; beyond its lower limit lie A_0 and Ψ, and beyond its upper limit, $A_{8/2}$, A_*, and $A_\varnothing$. For a description of the eighteen operators contained in Ψ see the text of Part 3. (Figure not to scale.)

Resuming the discussion on Ψ, an interesting phenomenon with far-reaching consequences is that

$$0\psi = \psi/\pi \tag{24}$$

which is a reiterative rather than a reversible equation. Hence if $0 < k < 1/\pi$, $k\psi < 0\psi$. Moreover, the hypernumbers $1/\psi/(1 + \gamma)$, ψ/e, and $\psi/\sqrt{\pi}$, where the gamma denotes Euler's constant, together with ψ/π, are of importance in defining principal operational circles in Ψ (*cf.* Figure 2).

There is not space here for these developments. Suffice it to say that ψ is the supreme idempotent unity operator, adequate to the bridging of all discontinuities in mathematics, and forms an immanent and transcendent idempotent algebra which generates, contains and maintains all other possible operator algebras, thus exemplifying that continuity is ultimately prior to discontinuity and that all examples of the latter are, with respect to ψ, merely relative and not intrinsic; transient and not permanent. Per contra, ψ demonstrates that a "yes" is always there, even before the first "no", as well as after it, and that ultimate continuity lies at the heart of reality, even though relative discontinuities are a necessity to all more peripheral reality.

Thus revolutionary systems can be seen to be evolutionary—or degenerate (inverse evolution)—if one looks deeply enough. And if one looks widely enough in the universe of systems, evolution supersedes degeneration. In this kind of seeing wisdom, both scientific and human, lies.

The Epistemological* Operators Briefly Characterized (cf. Figure 2)

(1) A_0: The generator, or number-species generation operator, which in explicit form is the collection of all the distinct kinds of unity, u_n, $1 \leqslant n \leqslant 9$, without interactions, but with the adjunction of the quasi-field 0 (zero).

(2) $A_{1/2}$: The measurer. Auto- and hetero-adjunction of the u_n. Addition

---

* The word *epistemology*, connoting the systematic and scientific investigation of how we know is rooted in the Greek *ἐπισήμων*, "knowing, intelligent", and more literally "during (and after) the weaving" (from *ἐπι* + *στήμων*, the warp in the upright loom). The word roots tell us, then, that the pattern is revealed and so known in the weaving. The answer to how the mind works was thus given by ancient thinkers in terms of the metaphor of a loom weaving patterns by one fixed and one movable set of orthogonal parameters (Descartes' most probable source of inspiration was a loom): the warp which

and measure begin with repetitive or iterative adjunction, leading repeatedly applied unit measures and their resulting extension into number axes and spaces of any assignable length. Arithmetically and algebraically $A_{1/2}$ reflects as the operator L which furnishes lengths of vectors.

(3) A_1: the distinguisher or extremal-measure operator. With A_1 we have the first distinguishability in measure, the primordial quantification of awareness or knowing in terms of the presence of some peak measure which thus becomes the *point de repère* for an external value.

All sensation and awareness is composed as envelopes of sequence patterns of pulse heights or external measure. Forming and assessing these envelopes is the process of awareness itself.

Thus A_1 is the primal awareness operator, the first measure or knower of distinctions. Arithmetically and algebraically A_1 reflects as the external or absolute value operator, which furnishes norms or "perihelion" distances to the origin (cf. footnote on p. 180), from the point of intersection of a number orbit with the real axis (unless no such intersection exists).

(4) $A_{2/2}$: this is the orthogonalizing or perpendicularizing operator, which orthogonalizes the primary awareness of A_1, by assuming a new orientation, turning either the view or the viewer through 90°, and in any case attaining a new and independent awareness from that which preceded it.

This operator creates multi-dimensionality of awareness or knowing thus introducing the result of *distinct contexts*. $A_{2/2}$ may be called the semi-rotator.

(5) A_2: view reversal or full 180° rotation. With this operator, which may be called the rotator, reflections and the operation of subtraction starts. Lengths and values can now fluctuate, and hence the notion of a changing relationship or function is an element of A_2. Arithmetically A_2 results in the reciprocation of a number, e.g. $A_2 u_n \to 1/u_n \equiv \bar{u}_n$. Hence the adjunction of the quasi-field infinity ($\pm\infty$) is now possible, for $u_n A_2(0) \to u_n\,(1/0) = \pm u_n \infty$, and $A_2(0)\,u_n \to (1/0)\,u_n = \pm\infty u_n$.

(6) $A_{3/2}$: the matching or fitting operator, the semi-inversor. With $A_{3/2}$ one object or process can be adjusted to match, resonate or fit with another.

---

is fixed, and the woof (or weft) which changes as the flying shuttle makes its numerical patterns of under- and over-crossings, revealing in turn the pattern of the web. Sir Charles Sherrington in a now famous lecture seems by his own insight to have reached the same archetypal image of the mind as a luminous loom, the speeding shuttles being the neuron-firing sequences and paths. The epistemological or A operators may therefore be regarded also as the *loom operators*, the lords of the loom.

$A_{3/2}$ thus governs resonances and anti-resonances (interferences) of all kinds, and the operation that can achieve and govern them. Hence the notions of tallying or pairing arise.

(7) A_3: the full inversor. Just as A_2 could rotate a view of awareness through a full 180°, so A_3 can invert a process fully, i.e. generate the inverse process P^{-1} that will undo the given direct process P, so that $P^{-1}P = P^0$, or the process P brought back to its initial point of no results. Negative exponentiation becomes fully developed in A_3.

(8) $A_{4/2}$: the transporter, mover, or distributor (and de-distributor). Here we have motion within a set of terms. Thus an algebraic form of $A_{4/2}$ is the operation of distribution: thus $A_{4/2}\,[a(b + c)\,d] \rightarrow abd + acd$, and its inverse: factorization, de-distribution or separation, e.g. $A_{4/2}^{-1}\,(abd + acd) \rightarrow a\,(b + c)\,d$, where in general $(abd + acd) \leqslant a\,(b + c)\,d$.

(9) A_4: the deleter or remover, A_4, is an indispensable item in the focussing or concentrating of awareness, as distractions (whether sensations, percepts, or memories) must be removed from the fields of aware consciousness to secure the necessary restriction or one-pointedness of address.

In logic A_4 governs the operation of removing irrelevances, or in general, data, from consideration, and hence the whole field of operations of the rules of evidence. If the data removed are not irrelevant, then error arises.

It is easy to see that an indiscriminating or superficial use of A_4 can degenerate its operation, in that what is designed ideally for the removal of errors and obstructions is perverted into what creates them. A_4 can operate to remove a process or an object of awareness.

(10) $A_{5/2}$: the ordinator, or enumerator or counter of elements. This operator thus introduces also *recognition*, and long-term memory which are necessary conditions for the performance of enumeration.

What began as *tallying* in $A_{3/2}$, now achieves a higher dimension as *enumeration*, which is a temporal line or *sequence*, expressible as a collineation of points in space.

Arithmetically, $A_{5/2}$ can furnish any enumerative result.

(11) A_5: the permuter or changer, A_5 governs all operations or acts of arranging in sequences or linear spatial and/or temporal positions, and hence all *ranking* or *hierarchy* generating operations.

What began as primal, extremal valuation in A_1 now blossoms into full-fledged hierarchical valuation.

(12) $A_{6/2}$: the class distinguisher and enumerator. Distinguishes, recogni-

zes, and counts *classes* of elements, i.e. given combinations of defining characteristics (i.e. class-elements or subclasses, subsets, subcollections, or subaggregates). Classes can also consist of sets or systems of members possessing given characteristic subsets or subsystems.

Since every set has not merely members but interaction rules or conditions for its members, every set is in this sense also a system and set theory includes system theory.

(13) A_6: the combiner. Just as $A_{5/2}$ could *enumerate* the results of given changes or sequences of change, but could not *produce* such changes and sequences (which is done by A_5), so $A_{6/2}$ can enumerate the results of classifying or combining processes, but the operator that performs and produces them is A_6.

This important operator governs the act(s) of creating or forming classes, sets, or systems. Here for the first time *imagination* and *creativity* (new pattern forming, new definitions, or the perceiving and conceiving of new class and subclass memberships) may arise.

(14) $A_{7/2}$: the programmer. This very powerful operator governs all operator interrelations, both in process and form, regarding any results of any of the previous operators, culminating in $A_{5/2}$, A_5, $A_{6/2}$, and A_6.

Conclusions as to A_7 are afforded by these operators in the domain of Ψ; for all the A_n are necessary components of the ultimate context unification and idempotency attained by, and characteristic of Ψ, which thus and literally—as the radii of Figure 2 show—contains them all. In brief we may write symbolically

$$A_{7/2} = (A_{5/2} + A_{6/2} + A_6A_5) + (A_{5/2} + A_{6/2} + A_5A_6) \qquad (25)$$

which expresses the interesting fact that all programming, whether of a national administration or any other system, including a computer, or even the mind itself, in the sense of $\left(\sum_0^6, \prod_0^6\right) A_n$, can be broken down into two distinct but related master operations, indicated by the two parentheses of Equation (31).

These are first, the master operation of formulating various *combinations* of permutations, changes, or sequences (A_6A_5), accompanied by an ever-present enumeration of such changes and combinations ($A_{5/2} + A_{6/2}$) whenever needed. Secondly, there is the more difficult operation of arranging such combinations of changes in *hierarchies* or *sequences* (A_5A_6) both of time and/or scope or power, together with the ever-present-when-needed enumeration operators ($A_{5/2} + A_{6/2}$). Such hierarchies express judgements of program value. Their rules are not easy to find as anyone will soon learn who has ever tried to determine a rule for arranging, say, the circular permutations of seven things in serial order.

$A_{7/2}$ itself is the master operator (like the operon gene in a set of genes for a particular physiological program) which operates so as to "turn on or off" all the other operators when and as required by the program, i.e. by its ultimate end or goal or end result, to which all the network of performance is subordinated. Thus $A_{7/2}$ governs *strategy* or the deployment of all subordinate operators.

(15) A_7: the relative proof operator. This operator governs existential proofs, e.g. that something exists or is true throughout a given universe of applicability expressed in terms of time, space and/or energy.

A_7 thus governs the structure and formation of methodologies and the theory of proofs on its first level: i.e. "that exists or is true for such and such a universe, under such and such conditions".

The weakest form of a result by A_7 is a guarantee of something existing or being true only here and now, and this extreme form of the relative proof operator points up the epistemological and logical need for the next operator.

(16) $A_{8/2}$ or A_∞: the continuing proof operator; operates to guarantee the unlimited *continuance* of appropriate results of A_7. Since such a guarantee requires not only the fact and realisation of endlessly continuing consistency ($A_{8/2}$) but the enduringness of ontological, and hence epistemological and logical, foundations.

It is therefore clear that $A_{8/2}$, in order to achieve its operation, involves the support of a still deeper operator, the next.

(17) A_8 or A_*: the consistency and foundation operator, "*A* star". This operator, the power of which may be referred to conveniently as "star power", is very close to the nature of Ψ itself, for the maintainence of the foundations of consistency and reality is an essential part of the ultimate unity and reality that Ψ connotes.

In vividly metaphorical terms, *A* star is the ontological flower that distils the elixir of epistemological immortality connoted by $A_{8/2}$.

(18) $A_\varnothing$: the origination operator. There is now only one operator left to complete the dynamic and infinitely individuating unity that is Ψ. The only further essential ingredient is that of constant newness, as a careful re-reading of the descriptions of the previous operators will show.

This power is conveniently symbolized by the stylization of an egg that never ceases to break open and hatch an original manifestation. Thus not only enduring and self-confirming foundations of reality are ultimately called for, but also an ever-present and burgeoning newness. This newness is not merely one of example, reiteration, or reproduction (which is a property of A_*)—but of the deepest root conceptions, processes, and possibilities.

Such an eternal newness of possibility itself is the essence of the description of the eighteenth A operator, $A_\varnothing$, at the center of the circles of Ψ (see Figure 2), the operator that stands behind the result $\Psi 0 = \psi/\pi$, and the ultimate negentropic support of the self re-creation of Ψ, which may be considered as the nineteenth of the A operators. The relation between them and the numbers is given by

$$(0A_1)\,(u_9) \rightarrow \frac{\psi}{\pi} \tag{26}$$

where $u_9 \equiv m$ was previously discussed.

The operational categories are now completed, the last ($A_\varnothing$) being ever incomplete because it is ever self-surpassing. Thus ends, and ends not, the alphabet of reality, and of the mind that maps it by virtue of the same alphabet.

References*

1. L. Löfgren, *Bull. Math. Biophys.*, **30**, 419 (1968).
2. N. Wiener, in *Recent Developments in Information and Decision Processes*, Wiley, New York (1962).
3. H. Weyl, *Zeit, Raum und Materie*, Teubner, Leipzig (1917).
4. W. K. Clifford, *Collected Papers*, MacMillan, London (1882).
5. F. London, *Trans. Faraday Soc.*, **33** (1936).
6. M. Weissbluth and B. Pullman (Eds.), *Molecular Biophysics*, Academic Press, New York (1965).
7. H. J. Muller, *Am. Naturalist*, **56**, 1922.
8. H. J. Muller, *Proc. Roy. Soc. (London), Ser. B*, **134** (1947).
9. H. Jehle, *Ann. N.Y. Acad. Sci.*, **158**, 240 (1969).
10. C. Muses, *J. Study Consciousness*, **2**, 22 (1969).
11. M. Kline, in *Science in the Sixties*, University of New Mexico, Albuquerque (1965).

* Note: the present paper is to be regarded as superseding any previous publications by the author on matters herein discussed, in those portions which may overlap.

12. B. Segré, *Arithmetical Question on Algebraic Varieties*, University of London Press, London (1951).
13. H. Minkowski, *Geometrie der Zahlen*, Springer-Verlag, Berlin (1910).
14. C. Muses, in *Functional Analysis and Optimization* (Ed. E. Caianiello), Academic Press, New York (1966).
15. C. Muses, *Ann. N.Y. Acad. Sci.*, **138**, 657, 901 (1967).
16. M. Bernkopf, *Trans. N.Y. Acad. Sci.*, [2], **31**, 516 (1969).
17. J. von Neumann, *Z. Angew. Math.*, **161**, 208 (1929).
18. C. Muses, *Internat. J. Biomed. Computing*, **1**, 75 (1970).

Supplementary Note

Divisors of zero are common in domains beyond that of $i = u_2$, and even in U_2, in complex spaces of $n > 7$ i-elements. Thus in U_3 we have, where $u_3 = \varepsilon$, $\varepsilon^2 = 1$, and $\varepsilon \neq \pm 1 : (1 + \varepsilon)(1 - \varepsilon) = 0$, although neither factor is zero. Other phenomena such as $1/(1 \pm \varepsilon) \neq (1 \pm \varepsilon)^{-1}$ also appear, and these phenomena tend to increase with the hypernumber index. Thus in $U_6 \equiv W$ (see ref. 15 for field form) we have $w^0 = 1 = L(w)$ with $w^2 = -1 + w$ but $(-w)^2 = -1 - w$, $w^{\pm 3} = -1$, and $w^4 = -w$, $(-w^4) = w$; also $(w)(-w) = 1 - w$ and $(-w)(w) = 1 + w$, where the left-hand factors are operators, and the right-hand ones operands. In $U_4 \equiv P$, we have $(\pm p)^2 = 0$, $p \neq 0$, and $p^3 = -p = p^{-1} \neq 1/p$, which is in the neighborhood of ∞. Similarly, in $U_5 \equiv Q$, $q^4 = 0$, but $q, q^2, q^3 \neq 0$; also $q^5 = -q = q^{-1}$, but $1/q$ is in the neighborhood of ∞. In $U_7 \equiv \Omega$, we have $\omega^n = \omega$ and $(\pm\omega)^{1.2} = \pm 2\omega$, as well as other algebraic phenomena, easily shown in the field form or orbital form for ω, which is based on a vertical array of 4 infinite sets of tangent circles along the ω-axis, the points of tangency being at $\pm\omega$ and at the origin. The editor has invited the author to publish the full set of orbital forms, with their related equations and algebraic phenomena in the forthcoming *Journal* of the newly proposed World Organisation of Cybernetics and General Systems. If we have time we will be glad to do so. The reader is also referred to vol. **3** no. 2 (July–December 1970, in press) of reference 10 and to our paper *The Use of Higher Operators in Bio-Medical Computing*, reference 18, volume **1**, number 2, April 1970.

CHAPTER I–7

Cybernetics and education

T. C. HELVEY
University of Tennessee Space Institute
Tullahhoma, Tennessee, U.S.A.

Summary

The interaction between cybernetics and education is on two general interfaces: (1) what to teach and how to teach cybernetics from the lower grades, through the college level into adult education, and (2) how to apply cybernetics to foster, to improve, and to facilitate educational processes or pedagogy in general. *Re* (1) the programme of the first Graduate School of Cybernetics is presented, which will lead to the degrees of M.S. and Ph.D. in Cybernetics. Also various curricula are described, which are recommended for high schools, for undergraduate colleges, and for adult education, especially to upgrade the background of high school teachers of the biological, physical, and social sciences in all aspects of cybernetics. *Re* (2) techniques are surveyed and methods suggested for the utilization of cybernetic principles in the teacher–pupil–machine–environment quadruple system, with feedback loops between all four parameters and some design configuration for self-organization in the learning process.

INTRODUCTION

It is very unfortunate, and I am sure many of us here feel nostaglic about the fact that pure academia are getting out of fashion. Sure, there still are theoreticians throughout the university community but in order to qualify for a true promoter of the state-of-the-art one must be past middle age, have a moustache and a beard, preferably gray, be absentminded, and have a few select disciples who adore the "big ole man" who smokes a pipe and works in a totally cluttered study. The more or less esoteric subject in which a theoretician is indulging may someday become practical enough that the Dean and the Curriculum Committee are forced into permitting a course being added to the curriculum, provided his peers in the department do not object too violently, which they probably will, because that would relieve them of the frustration of being ignorant in the subject.

Of course, there is an attractive argument that has elements of validity and refers to the shaky thesis that the university's prime task is to educate professionals to fulfil available job requirements for the greater glory of the promotion of the aims of the departments and "for the betterment of mankind". The only exceptions to this attitude are the training curricula for the arts and may be physical education.

I have been preaching and practicing for the past 20 years the interdisciplinary approach to science curricula. The practical results were satisfactory, the preaching much less so. You may imagine the reception I get when I am proposing and demanding that all science curricula, and especially the engineering curricula, must have, at the undergraduate level, a mandatory course on the fundamentals of general system theories and cybernetics, and on the graduate level, a mandatory course in biocybernetics. I have never had a rebuttal when I explained to the authorities that every engineer is working directly or indirectly on research and development of artifacts which are used by man, consequently, the logic of pedagogy requires that an engineer know something about the characteristics of his customer and his interface with the artifacts he is designing. However, the wheels of the educational evolution are grinding slowly. As a matter of fact, far too slowly to keep pace with the changes which we encounter in our life. And what is education if not the preparation of the new generation for fitting best into the dynamism of the daily life, thus becoming a true instrument of happiness?

CYBERNETICS AND CURRICULA

Cybernetics is in the transition of becoming a practical commodity without which certain aspects of physical and social technology would be very cumbersome to handle. The advances of automation and the strides towards artificial intelligence are on the way to being integrated into the explanation of alien environments such as deep space, submarine and subterranean technology. Furthermore, the need of dealing with large and ultra-large systems and the comprehension of complex man–machine interactions, self-organizing systems and social problems have opened avenues for the practical application of cybernetics. As a direct consequence of this trend even sub-sciences are already well defined such as engineering cybernetics, biocybernetics, cybernetic medicine, societal cybernetics, etc. Special professional societies are in the process of formulation and special journals provide a forum for the publication of achievements. Thus the time has come when cybernetics should become part of not just adult education but the regular college curricula. A difficulty arises due to the fact that no "proper home" could be established in the university structure for such a *par excellence* interdisciplinary science as is cybernetics. So far, our distinguished cyberneticians have been "integrated" into the departments of electrical engineering, mathematics, industrial engineering and even in departments of social and political sciences and others. Science compartmentalization is a great handicap for progress. When I had the task to present and defend my course offerings in cybernetics I was almost defeated owing to the fact that I am not a member of any of the orthodox departments of science and engineering and therefore no prefix can be applied to the course number in the catalogue. I achieved my goal by being willing to join any department where the majority of my peers are thoroughly familiar with cybernetics. Because there were no takers I suggested using "Cy" as a prefix. This has been unanimously accepted by the Council. Presently we have in the graduate program for cybernetics the following courses: (1) general systems and cybernetics fundamentals; (2) cybernetic biophysics; (3) cybernetics of human behaviour; (4) applied cybernetics and bionics. Besides these we offer a course entitled "special problems in cybernetics", which is a laboratory course for experimental work. Should the research in this laboratory course, which the student takes for credit in a sequence of semesters, become significant enough it can be developed into a thesis leading to a Master's or Ph.D. degree, provided additional courses are taken by the candidate. Regarding the fact that so far no proper undergraduate training is available the practice is for the candidate to get his degree in mathematics or physics or engineering with a minor in cybernetics. As more and more courses are officially authorized by the echelons of the university which are, in the case of the University of Tennessee, a) the Faculty, b) Curriculum Committee, c) Graduate Council, d) the Senate, the program is developing into a major in cybernetics.

A number of courses which I propose as a requisite background for a Master of Science or Doctor of Philosophy in Cybernetics can and will be taught by various departments of the university whereby some of the courses are and will be offered by the staff of the cybernetics program. In the framework of the presently valid requirements of the educational structure in the University of Tennessee Space Institute I have proposed the following courses as prerequisites for the background of the candidates.

(1) General systems and cybernetics fundamentals:

systems philosophy
requisite variety
decision theories
coding and decoding
the Turing machines
the heuristic approach

(2) Information and communication:

semantic information
sensory communication
limits of information
concept of channel
Shannon's theorems
entropy concept

(3) Man–machine interactions:

manual control
human transfer functions
reaction time
information displays

(4) Applied biophysics:

anthropometry
biostatics
biodynamics
biokinematics
neural activities
sensory modalities

(5) Human anatomy and physiology:

man as a system
integument
nervous system

skeletal system
muscular system
circulation
respiration
digestion
excretory systems
endocrine system
reproductive systems
embryology and genetics

(6) Sensory physiology:

vision
information processing in the eye
audition and stereophonics
sensing linear and azimuthal acceleration
kinesthetic sensation
mechano-receptors
tactile sensation
pain sensing mechanism
sensory deprivation

(7) Servo systems and automatic control:

feedback systems
Nyquist and Bode presentations
stability analysis
non-linear systems
design considerations

(8) Probability and statistics:

sample space
random variables
laws of large numbers
multivariate analysis
covariance matrices

(9) Special topics in mathematics:

theory of groups and transformations
elements of Boolean algebra

ergodic theory
Laplace, Fourier, and Z transforms
elements of complex variables
distribution functions and H-theorem

(10) Linear active circuits:

transistor theory
circuit theory
amplifiers
complex newtorks

(11) Optimization techniques:

calculus of variations
linear programming
elements of dynamic programming
Markov chains

(12) Bionics:

electronic and computer analogs
human amplifiers
elements of intelectronics (artificial intelligence)
implantable bionic devices
myoelectric servo control

(13) Applied cybernetics:

engineering cybernetics
biocybernetics
cybernetic medicine
societal cybernetics
man in a fully automated environment

(14) Human behaviour:

psychomotor functions
sensory overload
emotional parameters
genetic aspects of behavior
personality dynamics
group dynamics
small group interactions

(15) Elements of robotics:

pattern recognition
stereophonic orientation
learning networks
vertical sensor and control
mental control of artifacts

(16) Environmental physiology:

acceleration (linear and aximuthal)
vibration (subsonic)
heat, cold, humidity
noise and silence
nuclear radiation
illumination
atmospheric physiology

(17) Analog computer programming:

solution of non-linear differential equations
operational amplifier
servo-division
function generation
time scale change
magnitude scale change
forced vibrations of a linear oscillator

(18) Control theories in physiological systems:

application of frequency analysis
the root-locus plot
first- and second-order systems
steady-state errors in servo and regulator operation
visual control system
reflex function of the nervous system
regulation of the body temperature
mathematical model of heart rate control

(19) Brain function:

steady potentials of the brain
synaptic ultrastructure and organization

cortical topology
electroencephalography
studies on learning

(20) Mathematical sociology:

equilibrium states of social systems
Poisson-type models
contagious models
change and response uncertainty
hierarchization
local implications
diffusion of social structures
probabilistic mathematics versus deterministic mathematics

(21) Perceptrons and neurodynamics:

classification of perceptrons
linear and non-linear transmission functions
code optimization
adaptive pre-terminal network
fixed and conditional response sequences
choice mechanisms
awareness and cognitive systems
isomorphism of structured information
memory localization
mechanism of motivation

(22) Biotechnology:

human decision making
intended and extracted human output information
personal equipment design
comfort engineering
engineering of micro-environment
consoles and cockpits
man–machine task allocation
test procedures

(23) Reliability theory:

wearout failures
Bayes' theorem in reliability

component reliability measurement
confidence limits
failures and system stress
series and parallel systems

(24) Elective in sociology, or philosophy:

The candidate is free to elect any graduate course

(25) Elective in economics of business administration:

The candidate is free to elect any graduate course

(26) Elective in chemistry or biochemistry:

Besides these courses the candidate must have the following prerequisites

(1) Physical sciences:

general physics
general chemistry
electrical circuits

(2) Life sciences:

zoology
botany
psychology

(3) Mathematics:

modern mathematics
elementary calculus
probability theory

(4) Various:

introduction to philosophy
sociology
elements of economics

In our University, for the Master's degree, the candidate should have a minimum of 45 credits which include the required research with 9 credit hours. For the Ph.D. degree the usual required credits are 108 which include the 36 hours for the thesis.

It might be of some interest that every student in every class in cybernetics has to deliver a lecture of an hour's duration and complete a term paper. Appendix 1 lists a few of the topics. Appendix 2 is a selection of research project titles which are selected as possible thesis topics.

CYBERNETIC PRINCIPLES AND EDUCATION

Let us now turn briefly to the utilization of cybernetic principles in education, which is rapidly becoming a separate science. In the last years the coupling of the student's mind with a teaching machine has become a distinct possibility. The machine will read some electrical brain output which would change its signature after the student has memorized some facts or symbols which are displayed by the machine and which the student is able to recall correctly. After this, the machine will display the next item to be learned by the student. No question that such a technology is forthcoming, however, in the meantime, other types of less satisfactory teaching machines will be used as a stopgap measure. Teaching machines can be well utilized even today when applied according to cybernetic principles. If one designs an environment–student–teacher–machine system then the complexity of the individual systems will, no doubt, satisfy the Ashby principle of requisite variety and will become, through numerous feedback and feedforward channels, a self-organizing system. Later, let us say in 50 years from now, when advances in transportation and communication will bring about the trend of deurbanization, the schools and universities as we know them today will fade away and group instruction in the flesh will become the exception. Already today, when the highly efficient teaching aid, the spanking, has been abolished, the interaction between teacher and student is restricted to audio and visual channels. Therefore, by utilizing the already developed color holographs, coupled with two-way audio channels, all the sensory information, liminal and subliminal, can be transmitted between teacher and student, even if they are many miles apart. The three-dimensional audiovisual channel will not be essentially different from today's educational milieu except the teacher and student will not be able to smell and touch each other, which can be considered a distinct attraction for such future systems. Therefore, in the future when commuting to the school from a vast low population density region will become impossible, the instruction will take a different dimension.

It is beyond question that the efficiency of instruction would be highly enhanced if every student would have his own professors for every discipline. Today this is economically and technically impossible and therefore we have to resort to group instruction. In a large class the information which flows from the teacher to the pupils is often greatly reduced in its efficiency because of the unfavorable signal-to-noise ratio. Here the noise is not just meant to be at the approximate 50–60 dB noise level caused by voluntary and involuntary activities of the audience such as coughing, sneezing, murmur, dropping objects on the floor, etc., but also psychological noise such as mini-skirts, attractive smiles by females and not less attractive bushy sideburns and other secondary sex characteristics by males. In the future, through the holograph, the professor will be addressing every student personally and individually. This will significantly enhance attentiveness which becomes decreased today by the fact that the student feels impersonalized in a large class and just a group element, since the young instructor is addressing mostly Mary in the front row.

It is quite obvious that in such an advanced de-urbanized social structure, as we envision it for the admittedly far future, it will be an absolute necessity that primitive subject matter facts be indoctrinated in the students by the teaching machines which they can do at their own learning speed and according to their learning capability. Thus, the teacher whose job in the best Socratian pedagogy is the teaching of concepts, interactions, consequences, problem setting and problem solving will have a student audience with an equal background of the fundamentals.

From the foregoing it is quite obvious that the new information and interaction sciences, such as cybernetics, will play an increasing role in the provision of an optimizing educational system, which will be specially adapted to all levels of comprehension, thus can be utilized to teach human beings from the early infant age to senility.

APPENDIX 1

Student Term Papers

The legal system in 2068 A.D.
Communication and the computer
Automated democracy
The rule of three
Future applications of computers

"2068"
Predictions
A cybernetic look at the V/STOL wind tunnel
Systems approach to supersonic transport
Cybernetic view of future transportation development
Cybernetics in engineering education
Interaction of spacecraft with man
Unique biological systems in lower life forms
Automation and the future
Man and his machine—flying
Theology in the cyberculture
Cybernetics and my homeostatic niche
Toward a unified theory of the states of matter
Man–machine interface in light aircraft
The future of work
The scope of our society in 2068

Student Class Lectures

The Nyquist criterion
Human engineering
Toward the cybernetic factory
Teaching machines
Psychology of language in human interactions
Probabilistic aspects of information theory
Logic elements
Negative rodex number system
The neuristor
Basis of Boolean algebra
Information sequence in language to describe environment
Reliability and maintainability
Physical analogs of biological models
Numerical control
Focus on behavior
Cybernetics of the vascular system
Man–machine symbiosis
Probability of English letters
Homeostasis in man

Second and third generation computers
Adaptive controls
Noise
Information displays

APPENDIX 2

Research Project Titles

Experiments with man on form recognition
Cybernetic study of comprehensible questions
Modelling of automated mill for furnace welding of pipes
Investigation in the area of navigation and orientation
Cybernetic study of foetal distress
Investigation in the area of aerohydrodynamics
Cybernetic investigation of human temperature dynamics
Investigation in the area of man–machine interface
Cybernetic study of cardiac functions
Mathematical approach to endocrinological diagnosis
Feedback in endocrinology
Amplitude–time classification of images
Calculating protein binding of drugs with computers
Mathematical study of the exitation of nerve tissue
Cybernetic regulation of cell numbers in the epidermis
Implantable electronic organs and components
Electro-hysterographic research
Respiratory system as a biological feedback regulator
Swallowable intestinal transmitter
Cybernetic therapy of deafness
Cybernetics of the nervous control of muscles
Monitoring, prognosis, and control of the state of man–machine systems
Universal analyser of a multi-channel diagnostic system
Simulation of a monitoring and decision unit for ship navigation systems
Bioelectrical control of technical objects
Cutano-galvanic stimulation as information generator
Study of the formation of visual images
Experiments in the discrimination of complex set of images
Procedure for conversion of images into sound

Simulating human hearing
Functional simulation of nerve elements
Cyclic processes in the biosphere caused by cosmic factors
Bionic model of colour vision of man
Use of computers to process information from organismic loops

CHAPTER I–8

*Cybernetics—a fundamental science and philosophy of "existence"**

LOUIS CHALLIER
International Advanced University "Cybernetes"
France and Switzerland

"Pity us who always fight at the frontiers
of the ultimate and of the future."
(GUILLAUME APOLLINAIRE.)

Summary

Life is a topological space of atoms. As human beings we are a topological and functional "ensemble", governed and self-governed. Man is a dynamic unit endowed with feedback, his destiny is to survive and realize his potentialities before his death. The cybernetics of life is a complex "know", the bionics studying the various types of different types. Some approximate studies of the cybernetics of man have been carried out by the great cyberneticians, N. Wiener, Louis de Broglie (my teacher), W. Grey Walter, D. Aurel, W. Ross Ashby, L. Laborit, D. F. Gonseth, and others. Medicine and the study of advanced sciences, the diseases of behaviour and those of adaptation have opened new vistas in the field of cybernetics; other disciplines, such as endocrinology, cellular biology, immunology, etc., have also contributed in this field. In other ways, cybernetics with its logical perception, dynamics of evolution, and unique approach is a marvel of methodology of thinking and scientific research. The aim is to understand the human adventure with a faith in man that is "humanism".

* Dedicated to Louis de Broglie, teachers and friends.

INTRODUCTION

Cybernetics, a science twenty-one years old, perceived and discovered seventeen centuries ago, concerns the 20th century. After a long silence, it had a brief history, and like all histories it consists of an interlaced jumble, a delta... Here are mixed together varieties and verities, twists, turns, and meanderings, and the marshes of semantics and philosophies created by men.

For in congress after congress words, carriers of concepts, are always as difficult to separate from their covering as any Assyrian pottery.

Freeing ourselves from all preconceptions regarding cybernetics, and reconsidering the situation, I would say, following an old saying, that of all the sciences it is a *new humanism.*

Within twenty years, starting with a number of hefty jabs of the thumbs, like those of a sculptor premodelling his clay, it has been given its basic orientations, its major lines of force, its main features, its versions, and its present aspect. These were laid down by Plato's writings, Ampère's, and Norbert Wiener's (whose august memory I salute as in my mind I see him, recalling his galaxy of knowledge and of perceptive and intuitive intelligence) and also the Paris Congress of 1951 on various electronic computers of the period, and their use. It was, I believe, during the meeting he then presided over that *Prince Louis de Broglie sketched out one of the first exhaustive concepts of the coming cybernetics.*

Then came another Paris Congress in 1953, still on computers as an excuse, where *Mr. Ashby demonstrated his first homeostat*, while *Grey Walter's tortoises* were trotting around, happy and full of whim! Next came the International Congress on Automation, the first, organized by *F.H. Raymond, that great expert on data processing;* I remember giving a great deal of myself to that Congress, which almost immediately after gave rise to the first International Congress on Cybernetics properly speaking, that at Namur in 1956. I skip the *Teddington Congress in 1958* which *was more specifically philosophical* but which, for some, brought water to the mill of cybernetics. Next there took place *an extremely important but restricted*, very full and very focalized, symposium on cybernetics, organized by my friend *F. Gonseth of the Zurich Polytechnicum.*

This symposium had considerable repercussions on cybernetics from the intellectual and philosophical aspects, and it is a great pity that the proceedings could not be published, for their impact could have been considerable on account of the cross-fertilization of ideas that it brought about, as I recall distinctly. Then came the cybernetics research schools—British, Soviet, American, French (for instance, my Campbell group), though the latter was not grouped organically, which was to be expected, and was very... French!

Nor should we forget the *Karlsruhe group* of which my good friend *Hellmar Frank* was the great animator, nor the *Artoga group of our friend Pask* (I believe), which has laboured greatly. In time came the periodical meetings in Namur, and the Congress of the Association of Medical Cybernetics founded by Mr. *Masturzo.*

And now cybernetics is alive! It is everywhere. It is at last mentioned in dictionaries and in certain universities. In this connection I believe I was the only one foolhardy enough to found, in Switzerland a cybernetics-based advanced international University named Cybernetes, together with its Swiss pioneer, my friend R. Deluz.

NATURE OF CYBERNETICS

Cybernetics, the *Science of sciences*, on account of its all-encompassing, synthesizing, integrating aspects that foreshadow the other sciences yet to come, and *interrelates them*, gently binding them together. It explains, too, the materialistic and spiritual destinies of science—notably through bionics, now considerably developed, strategy, simulation, games theory and technical forecasting, founded by Gaston Berger, so carried off by untimely death from our friendship—that it is, in the full sense of the terms, an integrator, a new humanism.

It has fertilized all the sciences and, through its most varied implications it will attain to the understanding, the awareness of millions of humans who (like Monsieur Jourdain in Moliere's piece spoke prose and poetry without knowing it) will use cybernetics without knowing it.

Take, for example, on the Paris–Lausanne rail line, the self-regulating control loop and traffic on the single, two-way working tracks (because of tunnels on the Dijon–Vallorbe line). Again, take certain sorting yards on the French Railways, cybernetized to the maximum with "*self-governing*" means, or the instant seat reservation facility for Air-France planes, from anywhere on the globe, etc.

It triumphs also through the reality laid bare by its applications, and through the discovery of the uses nature makes of it. Consider the following:

in biology, the work of Jacob Monod and Lwov, of the Pasteur Institute, on cell cybernetics which earned them the Nobel Prize; the work of Professor Laborit, eminent researcher on pre-operational anti-shock artificial hibernation, also work by the same Laborit showing that the former explanation of the function of chlorophyll was false, and formulating a new and correct one, based on cybernetics; the discoveries of Rostand and other researchers on hormones and enzymes, the biological bricks of the *cybernetics of the living creature;* the discoveries on the hormonal processes of differentiation and de-differentiation of living tissue, by a British colleague working on the carrot parenchyme, starting from a tissue with an elementary basis; the recent discovery of synthetic hypophysary hormone, controller of super-controllers, the hypothalamus, that ultra-sensitive centre the determines the whole psychological and intellectual make-up of man.

It is precisely in the area of *medicine, the closest approach to the living being*, and also in the area of psychology, that cybernetics has a rendezvous with man—with the human being, since concurrently research is also being done by looking for possible similarities, analogies, correlations and simulations of complex human actions, on intellectual (i.e. cerebral) behaviour and sense (i.e. emotive) behaviour. I am using these terms in their full sense.

Take, for example, *Grey Walter's experiments on cerebral typology, the telecontrolled attitudes postures and behaviours of semi-complex patterned reactions*, conditioned by the voltages invoked, by means of teletransmitter cerebral micro-electrodes. Or take again *the computer and its interrogation console*, the latest outcome of progress under way, which will enable a number of complex human, professional and/or sociological behaviours to be simulated in parallel. And it does not matter whether these are logical or are based on the paralogical relations of psychology.

All this proceeds alongside the practice of decomposing the concepts taught to us with their processes of intellectual linkages called into being by teaching machines—and what a blessing they are!—and of human actions after biological actions. Soon it will be the turn of *mental intellectual behaviour* to be investigated. We must not of course forget man's contribution which is imagination, fantasy, freedom, emotivity, sentiment, humour! These cannot be put into equation, or else we have to accept that, to live and to live happily, one has to meet the conditions required by the equations! Otherwise the world of insects awaits us (with six thousand million known types of wild insects) and we cannot accept it: it remains the greatest danger for man in the biological and in the humanist sense of the term.

LIFE AND CYBERNETICS

The most recent scientific data seem to show that the material phenomenon *of life*, the living cell, shows a quite different pattern compared with the specific topological pattern of atoms which is inanimate matter. It is a topological set, based on the carbon cycle and its valences, of identical atoms (of C, H, O, N) and the set is found to be particularly functional and subject to entropy, which miraculously enough, however, may be circumvented by rejuvenation of its tissues. (Four or five methods have been discovered so far, others are surely on the way); entropy may even be made negative (*neg-entropy* as formulated by our former teacher, L. Brillouin) by means of reproduction *(exclusive property of living beings)*, and this homeostated functional set self-governed; in certain types of individuals it manifests itself as a cybernetic system subject to "fate".

The equilibrium represented by life (disease being the destruction of life) and health in *the living being is dynamic*. This synthetic dynamic appears to be the master government whether it be merely materialist in an elementary individual, or cerebral in a thinking being, and it is *an integral of the general behaviour* of states made up of differential equations.

And the stability of this governed topological set, in contrast with the free muscular movements caused by antagonisms, is conditioned by the laws of probability. *The latter synthesize the diversity of the unstable*, the fleeting, and the divergent (as for instance the kinetic theory of gases, actuarial insurance tables, the statistical process of ocular neurology in coloured vision, the natural chances of surviving an epidemic, Heisenberg's uncertainty principle, and so on...). Probability governs the world by giving it a stable base, which apparently it takes as its invariants. (Plato's theory of shadows on the back of the cave well symbolizes this illusion.)

The living being's equilibrium, whether man or animal, is dynamic and governed, its *primary object being to "survive"* (the automatic or lucid instinct of conservation) in order to realize itself, to be, and to reproduce itself (i.e. abolish entropy) before dying by retarding this due date as much as possible. (Degeneration may be retarded by biological products of cell rejuvenation, for in principle a cell or tissue do not wear out.) Carrel had shown this before the last war by causing a chicken embryo to survive in Paris at the Palace of Discovery. When it dies it is from accident or lack of nourishment.

The finest power of life and of survival appears in the interlacing of its *elements*, as well as in its "ENTIRETY" *with the great syntheses of mental life*, much like a large capacity memory computer with self-programming. (See in this connection the fine work of the University of Milan involving graphs of the visual examination of an art work by a tracked human eye in combination with ultrarapid photoelectric cell sweeping.)

To illustrate the above, note the innumerable discoveries of bionics, as for instance, the history of the nerve and the sense organ elaborated out of the simple, *already* structured, elementary vibratile eyelash or cilium, and also built up of successive complexizations into sense organs of incredible performance. Then, too, at a higher degree of complexity, *the Pavlovian nerves and organs of perception*, now become, by coupling up with the cortex, selectively or functionally conscious by selective computation, an adapted order/response. At a still higher level, note the "thinking" psyche, thought of various degrees no doubt, but still thought: that is, all the mental and psycho-affective faculties of an individual, sui generis, with his genetic capital of tendencies, his temperament, plus his reasonable education, sometimes critical of knowledge of facts and of their value, of causes, of effects and of processes; plus the cortex and memory (dynamic too) distributed throughout all stages. All this boils down to the cerebral faculty of being able to carry out vast mental syntheses, that is thought! To the computive faculty and creative ability has been added thought under volition.

I have mentioned creative thought (organizational, constructive, inventive, task-performing, conceptual); language and culture are instruments serving the intelligence of the thought, they will help it considerably, and will soon have their effectiveness multiplied by interrogation console computers and appropriate software generated by electronic computation. The latter will undergo a quite staggering advance in time due to intelligent aids.

In combination too with the latter, *bionics* helps to study and integrate the structuring of fragmented cybernetics whose countless corridors throng our instinctive conduct and those of the animal world; this is somatics!

Masterly accounts on "*destiny*" in the cybernetics of the living human being have been given by the great cyberneticians of this century: by my teacher L. de Broglie, and our friends Grey Walter, Ashby, Laborit, Aurel David F. Gonseth, G. Pask, and so on; ... it involves computation, memorization, learning, conduct, decision making, precision, acquired experience, volition, and the intelligent and or affective, spiritual and or artistic elaboration of complex behaviour *with* awareness of what is being done, and the WILL to do it.

PROSPECTS

Everywhere the trace and imprint of cybernetics are found; in advanced medicine, immunology, virology, gynecology, experimental embryology, endocrinology, medical and biological electronics; in new sciences such as cell biology, genetic diseases, adaptation diseases (social stress) and behavioural disease, biological mathematics, tissue culture, geriatrics, genetics finally in absolutely novel concepts introduced in advanced science over the past twenty years, and which have replaced the fomerly so fruitful concepts of viscosity, scatter equations, splitting apart of matter, such as that of field, even at sub-corpuscular scale. All such sciences and advanced concepts are new windows opening out on the cybernetic mystery (a thousand new laboratories ought to be created to carry out absolutely new research).

Knowledge, especially of the life processes that I mentioned timidly in Namur, *all the software of life*, will be primordial in the final edifice of cybernetics; primordial also in coupling up computer and intelligence in such a manner that the machine does not eat up the man! Especially when this involves the highest level of creative thought, and the lowest level or ordinary and infinitely multiple life of *the man–computer dialogue* of the interrogation console, and off direct display by means of an electronic pencil of an exchange of concepts, ideas, shapes, of action to be taken... And further, when man's comfort will be involved, his liberty, his planetary and cosmic activity... Certainly we are at the threshold of a fabulous era in which man's intelligence must not, in the West, lose control, under penalty of wiping out all human civilization at one stroke, in favour of either a civilization of apes, of insects, or of robots, which is the same thing. Cybernetics truly appears *as the science of sciences and as the safeguard of man!*

Starting with a simple observation on communication it ends up with a synthesis of 1201 sciences and technologies and of 100,000 years of miscellaneous philosophies, which during the past 200,000 years of the known past have not been able to explain either man, nor save him from his excesses, his weaknesses, his blind spots, his nightmares, and still less, properly manage his destiny of the celebrated *homo sapiens*.

Cybernetics is that safeguard. It will give it forms of government for its survival and its continuous evolution, without aberration and without monstrosities. Its severity, its flexibility, and its dynamic force with self-governed evolution, provide us with a most marvellous *methodology of thought and of* *research, of coherent and valid intellectual behaviour*, hence a philosophical outcome.

This will at long last enable us to attempt validly to comprehend *the human adventure*, obstinately and unalterably determined to have faith in man and his shadow, HUMANISM. For it is the final and only dignity of man even when he is dying: to attempt to understand why and how he is dying. Pascal had already stated it earlier and better than I have.

CHAPTER I–9

*The impingement of cybernetics on religion**

SISTER STELLA MARY

St. Dominic's Priory, U.K.

Summary

A remarkable phenomenon in the academic world today is the appearance of unquestionably great men best described as interdisciplinary personalities. The late Prof. Wiener was one of these. A mathematician who took an extraordinarily wide view of science, he brought the penetration of his powerful mind to bear upon the problems of many other disciplines. A man of sterling character with a conscience sensitive to the ethical implications of the impact of cybernetics on society, Wiener felt it incumbent upon him to give warning against the possible misuse of the science he had brought to birth and named. He first expressed his misgivings in *Cybernetics or Control and Communication in the Animal and the Machine*. Later, in *The Human Use of Human Beings*, he elaborated the ethical and sociological implications of his previous writing and finally he gave us the quintessence of his thought in the book which bears the evocative title of *God and Golem Inc.* It is on Wiener's evaluation of the impingement of cybernetics on religion that my paper is based. Following his example, I would avoid entanglement in metaphysical subtleties and keep religion within a cybernetic frame.

* Throughout this paper wherever the term "religious sense" is used it is to be taken as meaning the natural human aptitude for religion. This is, as it were, the lowest rung on the ladder of man's religious and moral life. Religious sense may be classified as part of man's "common sense"; that is, part of his store of primordial fundamental certainties which belong to natural reason. The question of grace as perfecting the natural religious sense is obviously outside the scope of this paper.

The title of this paper is unashamedly derived from the subtitle of *God and Golem Inc*[1], the book into which the late Professor Wiener poured the quintessence of his thought on a subject so immeasurably vast in scope that much of it is still left to the exploration of many future generations.

Perhaps the most remarkable phenomenon in the academic world of today is the appearance of unquestionably great men best described as interdisciplinary personalities: Norbert Wiener was certainly one of these. This is not the place for a detailed biographical appreciation but, if we are to understand his contribution to both the material and the spiritual advancement of the human race, we must first realize the perfect continuity of his all-round development from childhood to full age. Of him it could be truly said: 'The child is father of the man.'[2] He was one of those rare scholars who give evidence of exceptional intellectual gifts in their early years and whose subsequent development never brings disappointment to those who have watched hopefully over their first steps in learning. Fortunately for us and for him, his father—a Harvard professor of Slav languages—recognized his son's genius and planned his educational programme to meet his special needs and foster the brilliance of his intellect. The product of such training was a mathematician who was not exclusively interested in mathematics; one who took an extraordinarily wide view of science and brought the penetration of his powerful mind to bear upon the problems of many other disciplines besides the one with which he was mainly concerned.

In this dynamic personality we see a man of sterling character with a conscience sensitive to the ethical implications of the impact of cybernetics on society, who felt it incumbent upon him to give warning against the possible misuse of the science he had brought to birth and named. He first expressed his misgivings in *Cybernetics or Control and Communication in the Animal and the Machine*[3]. Later, in *The Human Use of Human beings*[4], he elaborated the ethical and sociological implications of his previous writing, and finally he gave us the ultimate refinement of his thought in the book which bears the evocative title, *God and Golem Inc*. (Let it be said in passing that the adjective "evocative" is a carefully deliberated choice and not a concession to the word-fashion of the day. The two nouns "God" and "Golem" do in fact call up things stored in the deeps of memory, ready to be connected up with the chain of religious and cybernetic information more lately acquired. This was probably so with Wiener himself.)

SCOPE OF ENQUIRY

In the last-mentioned work, Wiener has defined the field in which the enquiry into the impingement of cybernetics on religion is to be carried out[1]. He makes it clear that he is not discussing religion and science as a whole; he is confining himself to a limited area connected with certain points in the communication and control sciences which verge on the frontier where science impinges on religion. He has no intention of becoming entangled in metaphysical subtleties and he wishes to keep his comments on religion within a cybernetic frame. If we are to follow him, this means steering a different course from that of the theologian and detaching ourselves from the absolute notions with which theology is concerned. Not for us is the treatment of knowledge only in terms of omniscience, of power only in terms of omnipotence, of worship only in terms of the One Godhead. Knowledge is a fact, power is a fact, our religious sense expressed in worship is a fact, and such facts are open to scientific investigation.

There is nothing heartless and inhuman in a coolbrained approach to this subject; on the contrary, the findings of such investigations will dispel the fears of many who shrink from sifting out the facts for themselves. Few are called to be astronauts; ultimately all mankind will benefit by the courage and tenacity of purpose of those who explore outer space. In like manner, the man of questing intellect who ventures out on uncharted ways, plunges into unfathomed depths or hangs in temporary suspension over some crack of doom is risking his more robust faith that the weaker faith of his fellowmen may be rebuilt or underpinned. We must bear in mind, however, that the men who take these risks must first undergo long and arduous training—they are those who are "commended for the gift of discipline"[5].

Investigators such as these will hold with dogged persistence to a tunnelled path until they break through its terminal wall to find themselves face to face with explorers of another type—the mystics. Here we are on the threshold of an unexpected development in cybernetics; but to pursue this aspect of the subject any further would lead beyond the limits of this paper. The way lies open for free investigation according to the way in which the religious sense may have developed with greater or less freedom to be venturesome in the search for truth. It is to be hoped that at least some cyberneticians and some mystics will collaborate to share with others the fruits of their comparative study of the methods of control and communication adjunctive to their respective techniques.

POINTS OF IMPINGEMENT

It is precisely as a man of questing intellect, as a mathematician able to lend the insights of his own discipline to those with problems in other fields, that Wiener sets out to examine with penetrating lucidity, yet with unfailing kindness and humanity, the philosophical and religious implications of the new science. In *God and Golem Inc.* (Chap. 2) he suggests that there are at least three points where cybernetics may be found relevant to religious issues. These arise in connection with (1) machines which learn, (2) machines which reproduce themselves, and (3) co-ordination of machine and man. To these may be added (4) machines which teach.

When Wiener died in 1964 teaching machines were in their infancy and were, as yet, somewhat crude and incomplete. Today they have become part of the ordinary equipment in many spheres of education and it is not difficult to see that their use or misuse is not without ethical significance.

LEARNING TO PLAY

Taking first of all machines which learn, we may also take it for granted that there are today computers which can be programmed to play games, e.g. a game of draughts (checkers in U.S.A., dames on the Continent).

Can such a computer learn by experience to improve its game? The answer seems to be in the affirmative.

The ability to learn is usually attributed to living systems. In its most characteristic form, we see it as a phenomenon observable in man. Learning is a human attribute which has a close connection with man's religious sense; indeed, a being with no capacity for learning, at any rate in an elementary degree, could scarcely be credited with any concern for religion. So, at first sight, machines which only imitate human conduct by "learning" to play games may seem to be remote from the problems of religion. Nevertheless, there is a religious problem which is not entirely irrelevant when we are studying the behaviour of machines which learn. This problem is that of the game which we see played out between the Creator and his own creation.

From age to age the minds of men have been intrigued with the theme of the Book of Job, that amazing poetical work which is included in the Hebrew Bible and admitted as canonical by the Church. And what is this book but the dramatic presentation of a game in which God plays against one of his creatures, Satan, with the soul of Job at stake?

Now, if we accept the view that Satan is one of God's creatures and that there is a perpetual factual disharmony between Satan and his Creator, we see some motive for such a game; yet, as Wiener points out[1], to play games with an omnipotent, omniscient God is the act of a fool and the Devil is no fool, but a master of subtlety. Any contest between God and one of the rebel angels, even Satan their chief, is foredoomed to failure. Why then did Satan enter the game? Did he see some chance of winning?

Here we remind ourselves that we are not steering the same course as the theologian, we are not concerned with dogmas of omnipotence and omniscience, and so we may, as a supposition detached from any dogmatic faith, image the game between God and Satan as a real conflict and God as something less than *absolutely* omnipotent—he may very well lose the game.

And yet God made his creature according to his own free will, any power of action the creature possesses is derived from God himself. So we are confronted with the following questions. Can God play a significant game with his own creation? Can a limited creator, man, play a significant game with an artifact he has made? (The second question will be dealt with under another heading.)

Before we go further into this question of God playing a game with his own creature[6], let us turn to the subject of play in a religious-cybernetic sense. This necessitates a preparatory cutting adrift from the utilitarianism we may have been taught to admire and the assertion of freedom to accept the challenge of practising a new virtue[7]—the virtue to which the Greeks gave the name of Eutrapelia. He who has gained the mastery of it brings everything into life's game, as he now understands it—the first primrose, the song of a bird, a child's spontaneous laughter, the beauty of language and literature, music and art, explorations into some fresh aspect of philosophy or science, the heights and depths of a meaningful religious practice, the great events of history and the most trivial happenings of everyday life, work (yes, work) hopefully planned, faithfully carried out, and brought to faultless fruition or equally faultless failure—and above all, the interplay of divine and human love.

The man who has achieved a truly human playfulness will have found his way to a deeper understanding of his relations with a living, personal God—a God who plays. In this age, when we suffer much from having lost our sense of mystery together with our faculty of play, we need to let our religious sense be refashioned and disciplined to renewed strength and beauty within a reference-frame that may seem unfamiliar and so, perhaps at first, unwelcome. At this point our greatest need is the courage to see our Creator as *Deus ludens*, God at play, and to recognize ourselves as playing a small part in his gigantic, universal game. We must relearn the art of playing with the seriousness of children, which is so different from the "seriosity" of the blasé grown-up man.

If our play is light-hearted without becoming lightminded, we achieve a balanced humanism[8,9] which escapes the pitfall of "seriosity" without falling into that of levity on the other side of the path. If there is a kind of blasphemy in over-estimating our own importance in the game of life, in trying to make ourselves responsible for what is after all God's responsibility; equally there is base ingratitude in belittling our God-given ability to play the part assigned to us in the all-embracing system of the universe.

THE DANCE WITH DEATH

At this point, instead of exploring into the intricacies of game-playing machines, the construction of which really belongs to another branch of cybernetics, let us pick up the thread of an illuminating line of thought which led Wiener to write some of his most delightful and selfrevealing pages. Evidently this great scientist was one of those rare thinkers who see the whole universe and all that is in it as "all of a piece". For him, the passage from consideration of man-made machines to deep reflections on the subject of the Creator playing with his own creation is an easy one. Between his first book and his last he developed this idea of a game that is also a dance—a dance with death. Had he lived longer he might have gone further; but there is nothing to hinder the later cyberneticians from following the same path.

Dancing has always been associated with religion, though it has sometimes met with disapprobation on religious grounds.The dance with death, which is also part of the game of life, is not the concern of the philosopher alone, however, since as a study in the interplay of communication and control it affords information that brings it within the scope of cybernetics. Wiener gave it his attention from both points of view. In his first presentation of cybernetics as a science in its own right[3] we find him drawing upon his amazing stock of general knowledge to introduce the theory of this kind of dance from the point of view of philosophy and also in connection with the study of machines that learn.

We need not here recapitulate what he had previously said on the subject of learning machines used as game-players nor linger on the consideration of those philosophical ideas which Wiener found exciting. What has direct bearing on the idea of God at play with his creation is the use of the theory of game-playing machines to throw light on a type of struggle activity which is more of a dance with death than a game.

The first example given is that of a fight between a mongoose and a snake. Wiener has seen such a fight himself but he also draws upon Kipling's story "Rikki-Tikki-Tavi" for his information. The fight is literally a dance with death and although it almost invariably kills the cobra, the mongoose is not immune to snake poison.

The mongoose begins the fight with a feint which provokes the cobra to strike its poison fangs. The mongoose dodges and makes another feint, the snake repeats the attempt to strike, and so we get a rhythmical pattern in the fight between the two creatures. It is not, however, a static pattern of activity but one that develops progressively as the mongoose makes its feints faster and faster in the time sequence. At last, when the cobra is stretched out in a position that inhibits rapid movement, the mongoose abandons the feint and makes a real attack, delivering the coup de grâce with a deadly bite through the cobra's brain.

In such a fight the snake's activity is confined to making single darts at its opponent without any progressive development in method of attack. On the other hand, the behaviour of the mongoose resembles that of a learning machine, since the pattern of its action involves at least an appreciable segment of the whole past of the fight. In comparison with the snake, the mongoose has a more highly organized nervous system.

Indeed, all physical contests between living beings and the various kinds of games which machines can be programmed to play are alike in that they have the same element of acquiring experimental knowledge and using it to improve play and move progressively towards the winning position.

Machines may be made to simulate emotions[10], but they have no real emotions and they cannot follow us in a line of thought influenced by our emotive reactions to the consequences of war. Computers may be programmed to simulate ethical behaviour, but they have no real ethics, no religious sense. The "mind" of a computer is a literal mind and it will carry out its program without any compromise. If its findings result in the pressing of the Great Push Button, there will be no reversing or modifying the decision.

SELF-REPRODUCING MACHINES

The second point at which cybernetics may be found relevant to religious issues is connected with machines that reproduce themselves. Here is indeed a startling impingement of cybernetics on territory often regarded as a kind of game preserve where poachers from other fields of knowledge should not dare to tread. However, cybernetics would not be cybernetics if the cybernetician did not venture into far-flung areas of investigation to meet co-workers trained in very different disciplines.

The power to reproduce itself is a phenomenon which we are accustomed to think of as characteristic of a living system. This power of self-reproduction is closely related to the power of learning, because an animal with a capacity for learning has also a capacity for undergoing transformation into a different being by learning to adjust to its environment. Although the ontogenetic learning which occurs during an individual's lifespan is a mode which contributes to environmental adjustment, there is another kind of learning of equal importance to be taken into consideration. This is phylogenetic learning or learning which occurs in the history of the race[11, 12].

Wiener read Darwin at the age of seven and at this point anyone who has had a like experience of confrontation with Darwinism in early childhood will appreciate the consistent development of thought from first contacts with science until cybernetics was brought to light. If Darwin's theory of natural selection is admitted as one type of basis for phylogenetic learning, the passage from learning to self-reproduction becomes comparatively easy. First, let us examine the facts relating to natural selection.

Briefly, the theory rests on the three following facts.

(1) The phenomenon of heredity: that an individual plant or animal produces offspring after its own image.

(2) The fact of variation: that the offspring may differ from the parent and that such differences are subject to heredity.

(3) The fact that according to the Darwinian theory of evolution the over-rich pattern of spontaneous variation is modified by the difference in the viability of the different variations. Comparatively few of the variations contribute to the probability of racial survival, the rest of the variations tending in the opposite direction.

In actual fact the process of racial survival and the changes attendant upon it may be very complicated. For example, we have Mendel's description of the functioning of heredity and variation which is not easy to grasp without facilities for long-term experiment, which may prove abortive. (Presumably Mendel was successful in guarding his peas from pigeons and mice.) Then there is the phenomenon of mitosis which accounts for the behaviour of genes in process of duplication and separation, of their formation into an aggregation of chromosomes, of linkage and so forth.

Yet behind this complexity of processes lies a simple fact: that given a suitable nutritive medium of nucleic acids and amino acids, a molecule of a gene, itself made up of a highly specific combination of amino acids and nucleic acids, can cause the medium to lay itself down into other molecules which may be derived from the same or from other genes with a relatively slight variation from it.

The process of reproduction is too vast and complicated to analyse here. It runs through the whole scale of living systems and finds its highest development in the reproductive activity of man—the greatest marvel in creation.

In whatever way our religious sense is developed or undeveloped, we have now reached a parting in the ways of thought. It is not a question of religion and non-religion—some of the ablest explorers into scientific questions have been people devoted to religion and dedicated specialists in its practice; e.g. Mendel was an Austrian monk and Albrecht, Count of Bollstadt (also known as Albert of Cologne), who became a Friar Preacher at the age of seventeen, was far in advance of the 13th century in his philosophical researches and his experiments in natural science. Had he lived seven hundred years later, he might well have been a cybernetician of the first rank.

The dividing line here lies not between science and religion but between two different ways of approach to all that may be included under the heading of "reproduction". Some people are aware that man possesses reproductive powers but their scientific outlook is restricted by the "stork and gooseberry bush" type of information; yet their practical sagacity enables them to accept the idea that a machine could produce another machine in its own likeness. Others, although seemingly more sophisticated, have factual knowledge of generative production in man and the lower animals; yet they find the idea of ascribing reproduction to machines emotionally disturbing. The religious sense is ineradicably rooted in the human heart; but it is subject to the influences of both heredity and environment. Where there is a tenacious abhorrence of any impingement of science on religion, or on what may be termed a "pro-religion", there is no acceptance of alien lines of thought even under pressure of forcible mental conditioning: "A man convinced against his will is of the same opinion still."

Whatever the fontal spring of their objections to self-reproducing machines, such objectors will not be cyberneticians and what follows will not be of concern to them. Let those who will withdraw from what, to them, would be fruitless discussion on a painful topic.

People who do not believe in God cannot accept the statement that he is the author of creation; they do, however, admit the fact that man makes man in his own likeness and they can proceed from that point to discuss the possibility of self-reproduction in machines. The orthodox believer sees in man's self-reproduction a prototype of the act of creation by which God is supposed to have made man in his image and, unless that believer's mind is closed by prejudice or some other cause of thought-block, he will be willing to consider the likelihood of self-reproduction as a function to be found in a machine. Two searchers after truth who part company at one stage in their exploration may meet again at a later point in progression.

Holding aloof from the devotional overtones under which the realities of religion may be veiled, we can now probe fearlessly and honestly into the question of something like the act of creation occuring in the case of a non-living system. In other words, we ask if a machine can make a machine in its own image? Wiener consistently gave the answer as "yes", even before cybernetics had reached its present stage. His more technical explanation of selfpropagation is to be found in his first book introducing the new science (see Wiener[1], Chap. 9, pp. 177–80). There he points out that a machine is not only matter, but an agency for accomplishing certain definite purposes—it is a thing that *does* something. And self-reproduction is not just the production of a tangible self-image; it is the creation of a replica with a capacity for doing the same things.

Here two questions arise. Is it possible for a machine to have enough parts and a structure adequate to the requirements of self-reproduction? How could such a machine be built? The first question has long since been answered in the affirmative: there is no doubt that, in principle, an artifact could be made to reproduce itself. The other question concerns the actual working out of some operative procedure for building self-propagating machines.

Living systems make other living systems in their own image. If machines are self-reproducing they should be capable of producing other machines in their own image, though not necessarily exactly similar to themselves. What is the image of a machine?

There are two kinds of image: a pictorial likeness and an operative likeness. For example, on the walls of a farmhouse kitchen in the West of England there were framed drawings and photographs of various kinds of agricultural machinery. On a wet day, these pictorial images might delight a child with an engineering turn of mind; but the instructive entertainment they gave was on a small scale in comparison with a visit to the engine-yard outside where machinery at rest could be explored. Still more exciting it would be to see the machines at work on the farm--to see them in operation. The pictured machines on the kitchen walls were pictorial images of what the inventor of each machine had in mind when he designed it; the machines at work on the farm were operative images of the same thing.The operative images were capable of some performance, the pictorial images were not; yet the best of the working machines could not reproduce themselves as the various living systems on the farm were able to do. Men, beasts, birds, insects all had progeny but the machines had not. In all probability the farmer was well content with that situation. If a modern farmer invested capital in a huge combine to help him to solve his labour problem, he would not want it to busy itself with making more combines in its own image: he would want it to get on with agricultural work.

However, what we are concerned with now is not the economic value of self-reproducing machines, but the method of making them if we wanted them to be made. At this point an important fact to be grasped is that, to be effective, a self-reproducing machine must produce other machines with operative similarity and not just pictorial likeness. An operative image may or may not bear pictorial likeness to its original; that is not important. What matters is that the image should replace its original in action —there lies the true similarity. In other words, it is operative similarity that we must look for in a self-reproducing machine.

Next we must decide from what point of view we are looking upon machines—what is a machine? Some will readily give the answer that a machine is a prime mover, a source of energy. That answer is good enough in its right context, but it will not do for the cybernetician. He is concerned with communication and control—his machine is a device for converting incoming messages into outgoing messages.

In his earlier and more technical work (see Wiener[3], Chap. 9, p. 178), he chose the non-linear transducer for working out the theory of self-reproducing machines. In passing, it is interesting to note that for reference he turns to his own "Non linear problems in random theory" and also to

Gabor's lecture[13]. The statistical basis of Wiener's work is slightly different, but there is an essential similarity between the two methods. (Here is an example of human brains working independently in different locations and contributing towards the development of the same science, a subject to which we shall return.)

We need not stop here to enquire minutely into the construction and working of the apparatus needed to carry out the processes involved. Briefly, what the sumtotal of the various operations amounts to is this: an imitation of any unknown non-linear transducer can be obtained by a sum of linear terms, each of fixed characteristics and with an adjustable coefficient. This coefficient can be determined as the average product of the outputs of the unknown transducer and a particular known transducer, when the same shoteffect generator is connected to both of the inputs. Furthermore, it is possible to transfer the coefficients automatically to the pieces of feedback apparatus, instead of relying on human intervention. Potentially, we have the means of making an artifact that could reproduce itself[14]. It must be remembered, however, that there is a wide difference between a logical possibility and its practical development. In practice we might not find it helpful to have machines as prolific as mice; but it is not the business of scientists to restrict their explorations to what comes within the limits of practical politics.

MACHINES AND EVOLUTION

By this time there are few people who are shocked at the idea that creation did not take place in six days of twenty-four hour duration; on the contrary most of us accept it as a fact that evolution is a process continually going on. But what about applying the theory of evolution to man-made machines? The Dutch bulb growers tell prospective customers that "bulbs must be handled as eggs"—the same may be said of this delicate subject. We are beginning to tread on poachy ground. Metaphysical observations on the universe and the assignment to man of a unique place in it are outside the scope of this paper. Nevertheless, we are concerned with commenting on certain points where cybernetics impinges on religion and there must be no evasion of the fact with which we are now confronted—reproduction is not exclusively a function of living systems: non-living systems can also be made to reproduce themselves.

If we admit that living systems that reproduce themselves are of necessity involved in the process of evolution, what are we to say of non-living systems with the function of self-reproduction? Surely if man can produce artifacts which reproduce themselves, he must expect some at least of his artifacts to be capable of evolution. Self-reproducing machines are not obtainable at the nearest chain-store; but they can be designed and even built by way of experiment. Today we may build a computer and see the finished product as the operative image of the computer the designer had in mind; tomorrow we may become aware that our computer is exhibiting behaviour of a far more sophisticated kind than anything its designer foresaw—our artifact has been caught up in the process of evolution and new properties have emerged.

Such an impingement on the territory in which our religious sense operates may be disturbing and lead to differences among those who hold divergent views as to the positioning of the frontier between science and religion and the friendly or hostile guard to be kept there, but the obstacles are not always as insuperable as they appear to be. Two parties to a dispute about the respective territorial rights of religion and science are both concerned about the same thing—the safeguarding of faith. Let them come to an agreement on some fundamental concept which they both wish to safeguard and they are well on the way to understanding of each other's faith. Mutual understanding will lead to mutual aid and a cross-fertilization of ideas that may result in undreamed of discoveries in science and the ascertaining of deeper truths about the very nature of things with which religion is concerned. A self-propagating non-living system caught up in the process of evolution should not be a threat to any man's faith, if he views it from the right standpoint.

TRAVEL BY TELEGRAPH

Now let us enjoy a little cybernetic recreation and play with an idea with which Wiener himself toyed[4]—the idea that a man might be sent by telegraph. The difficulties are so great that there is little likelihood of any such achievement in the existence of man; yet the idea is not intrinsically absurd. The theoretical possibility lies in the fact that there is no absolute distinction between the types of transmission used for sending a telegram from one place to another and the types of transmission conceptually possible for transmitting a living organism, e.g. a human being. The practical difficulties are technical ones, the chief one being the difficulty of keeping the living organism in being during the process of transmission and reconstruction of the "message" at the other end.

The question which presents itself to an enquiring mind is: would the man received at the terminus be the same man who was despatched from the office of transmission[15]? The answer must be in the affirmative, or we have not succeeded in telegraphing our man.

This idea is not only of cybernetic interest; it is also of interest to religion. As a committed Christian one can accept "the resurrection of the body" as an article of one's creed and leave it at that. There is nothing wrong in entertaining a degree of intellectual curiosity about it. Why should there not be some connection between the possibility of telegraphing a man and the resurrection of the human body, even if it had been dissolved into its elements? Though previously mummified like the Pharaohs, or reduced to ashes and scattered to the winds, there should be no difference in the identity of the body arriving at the other end—wherever that place might be. There are times when the faith of science and the faith of religion seem to meet and science and religion may at least occasionally intone the same credo.

This is poachy ground indeed; but the over-cautious are under no obligation to set their feet upon it.

CO-ORDINATION OF MACHINE AND MAN

"Render unto man the things which are man's and unto the computer the things which are the computer's" (Wiener[1], Chap. 6)—that is Wiener's condensed advice as to the intelligent policy to be adopted when men and machines are in partnership. It steers a middle course between that of the gadgeteer and that of the man whose objection to mechanical aids has some sort of religious basis. At both extremes there is an element of fear that can easily lead to cruelty—the gadget worshipper is either dominated by fear of his idol—the work of man's hands—or, what is worse, he uses it to establish a fearful domination over others; the religious objector fears the transference to an artifact of any function of the human brain, as men of an earlier period feared the machinations of wizards and witches, and such fear may well lead to witch hunting. The sane middle path leads to the unbiased examination of known facts about the making and use of manlike machines.

There are two types of scientist who may be attracted to experiment with anthropological machines. One type is eager to accept the intellectual challenge offered to those who aim at making machines with the highest standard of efficiency in the performance of difficult tasks; he will enjoy the excitement of producing something new that may or may not be of benefit to his fellow men. The second type of scientist looks at things from the standpoint of experimental psychology; he will want to build machines in order to gain a better understanding of man. It is this latter type of scientist whose application of cybernetics in the field of psychology is most likely to offend against other people's religious sense—he may find himself permanently labelled as unsound, not only by religious heresy hunters but also by his fellow scientists. It was wounding enough to human pride to be told that man has evolved from lower types of living organism; it is outrageous to be told that machines can be made like men and prove themselves more efficient in some respects. Man, as a limited creator, does not want to play a losing game with an artifact of his own creation.

Those who have the patience to make a fair comparison between the human brain and a computer will discover some reassuring facts. One of the most important of these facts is that whereas a faulty component part will cause a computer to produce catastrophic errors and print out nonsense, the brain does not break down completely, except under certain pathological conditions. Although there may be some inaccuracy in the processing of information, the brain is rarely responsible for complete nonsense. Again, the computer must have its input carefully prepared for it; the human brain processes information very efficiently and provides its own filter for the input.

Of course, there are some men who seem inferior to most computers but the best of computers is subject to the master-minds among men. There lies a moral danger, for men with master minds are all too often unscrupulous and in the new science of cybernetics they can find much to augment their power and secure them dominative control over their fellow men. The partnership between man and machine can become an unholy alliance. Wiener[1] (Chap. 5) saw this type of mastermind as the mind of the sorcerer in the fullest sense of the word.

However, there are systems involving a partnership between human and mechanical elements that do not degrade the men concerned with them. Like all good things, they can be abused but the intention of their designers is to amplify human strength where it needs amplification, while allowing a man to exercise his capacity for decision making and to retain his natural dignity.

The obvious use of such machines is to enable men to amplify their physical strength and to perform actions involving carrying and lifting heavy loads where ordinary carriers and lifters would fail on account of heavy weight or peculiarly difficult conditions.

Another use for manipulators is not as a direct attachment to the human body but as slave "hands" to deal with dangerous work, such as may occur in nuclear research and in some classes of industry. Such manipulators usually terminate in pincers. This is a man's elaboration of a woman's way of dealing with dishes in an excessively hot oven—she puts on asbestos gloves and takes a pair of oven-tongs or pincers so that she may operate from a safe distance.

Possibly the best use made of man-and-machine combinations is in the field of prostheses for lost or paralysed limbs. The old-fashioned peg-leg was a forerunner of the more sophisticated artificial limb made to resemble the natural form of arm or leg. As far back as World War I a man who had a double amputation could be fitted with a pair of artificial legs so well constructed that he could control their movements with a push-button and even ride a horse, with some help in mounting and dismounting. That was already a great achievement but cybernetics has cleared the way for much better things.

It is too early yet to decide how much is to be expected from the new perspectives which have been opened up in connection with the therapy of people partially paralysed by disease, injury or nervous disorder; but some at least of the new techniques will bring new hope into the lives of those who have been disabled.

Because the art of healing has always been so intimately associated with man's religion, we have been touching upon a point where cybernetics definitely impinges upon religion. Since that is so, there may be unfavourable reactions to the best planned efforts of those who are trying to put the latest scientific aids at the service of those whose physique has in some way been impaired. Here the scientist who uses his knowledge of cybernetics to gain a better understanding of human behaviour may be able to do some supplementary work by developing the resources of psycho-cybernetics. This is one of the most interesting of the later applications of cybernetics and it makes such a strong appeal to people who would normally shy away from psychology that it offers great opportunities of helping those who are hampered by a poor self-image. Of its very nature it cannot fail to make a strong impact on personal religion[16].

MACHINES WHICH TEACH

"Delightful task to rear the tender thought, to teach the young idea..."[17]; so all true teachers may think when they are not bogged down in mere drudgery. Teaching is a great vocation and we rightly set a high value on the reality of the relationship between the teacher and the taught. Today, more than ever, we see the need to train our youth to take an integral view of man and his relations to society, to act as responsible, intelligent co-workers, knowing how to avoid boredom and to use leisure well. All this is desirable and many dedicated men and women are making heroic efforts in the field of education, but teaching requires teachers and they are in short supply. This is where machines which teach can be of help.

The idea of programming subjects for learning was first developed in America, but the first English machine[18] soon made its appearance. Today there are several types of teaching machine available and some remarkable results have been achieved with them. Departments in the Services have been pioneers in the United Kingdom in the use of programmed instruction (P.I.) with or without teaching machines. The Ministry of Defence Report[19] on research and development in this field is encouraging and will be found helpful and reassuring to those interested in this special application of cybernetic ideas.

In spite of so much success, teaching machines still represent a point where cybernetics impinges heavily on religion and there can be strong opposition to their use on religious or quasi-religious grounds. Many people feel an understandable fear that teachers will become mere monitors in a mechanized system of education—there is something in such an idea that offends the religious sense. But there is no real cause for fear; if the machines are treated as ancillary, teachers will not be demoted but only assisted in their work. They will be left with time to devote to those who need personal contact with love and sympathy if they are to succeed.

Perhaps the teaching of languages is where the teaching machine can be of greatest help and this is also a department of education inextricably bound up with religion. All the best known religions have sacred writings and attach great importance to accuracy in preserving the language in which they are expressed.

Speaking of languages leads to a further digression (after all, digressions like postscripts and change of mind are a woman's privilege). Wiener did not disdain the use of *Alice in Wonderland* to illustrate some of his points.

Why should not Tolkien's[20] *The Lord of the Rings* be pressed into the service of cybernetics? It may be pure fantasy but fantasy is not to be despised, as Wiener (Chap. 5) himself assures us. If he had used this book it would not have been surprising: it is worthy of a place in every cybernetician's library and not just to amuse the children. One of the most impressive features of Tolkien's work is the language he invented as a means of communication among the elves. That should be of scientific interest. If the cybernetician may walk in Alice's Wonderland, he may also make a profitable journey into Middle-earth with Tolkien's map to guide him[20]. Perhaps some future writer on cybernetics will tell the tale of such journeyings.

IN CONCLUSION

Since cybernetics as a science has barely come of age, while religion is one of our earliest human aptitudes, it is difficult, though not impossible, to bring them together within the same range of thought. Religion touches us so closely and is so shot through with emotion and swathed in reverence that the subject can rarely be approached with anything like scientific detachment. Cybernetics, on the other hand, is still remote from many people's minds, something they prefer to leave to the scientists and the technologists whose competence fits them to deal with such matters. Yet however much we try to avoid it, sooner or later we must meet the impact of cybernetics on at least some limited segment of the sphere in which our religious sense operates. In our modern technological world human purposes cannot be achieved without recognizing various systems of communication and control and without regard to ethics. Since the science of communication and control is cybernetics and the ethical element in it belongs to the normative side of religion, the very real impingement of cybernetics on religion is evident.

The study of cybernetics is the pathway to a common ground where science and religion in their search for truth may meet and engage in a meaningful enquiry into the real nature of things—an enquiry that is to be a real probing into real questions (see Wiener[1], Chap. 1). If those who meet to discuss science and religion within the cybernetic frame are to contribute anything of real value to the advancement of human progress, they must be disencumbered of the prejudices and religious or scientific taboos which forbid the bringing together of seemingly incompatible views. There should be nothing to hinder the honest search for a full-bodied truth that has not been watered down and weakened by compromise where ethical principles are at stake. Many crimes have been and are committed in the name of religion and many have been and are committed in the name of science. To avoid participation in such crimes we must be steadfast in rejecting the temptation to manipulate the divine for human ends; neither must we consent to sell the magic of modern cybernetic achievements to those who would use them for the furtherance of unworthy ambitions or to let loose the apocalyptic terrors of nuclear warfare. In such a misuse of science and the men of science Wiener saw a form of simony or sorcery (see Wiener[1], Chap. 5) which is all too easily written off as extinct in the world of today.

Maybe there are other things besides simony which call up some accusing voice from a past century, putting the conscience to such a question as: "Have you thrust your unwanted daughters into religion?" That item from a form of "examination of conscience" framed by a 16th century spiritual director[21] for some Spanish grandee may be dismissed as just an interesting light on the social practices of another age. But what is the basic element that may still survive? What was the guilty parent doing?

Imagine an ambitious man with a large family to settle in life. He wishes his sons to cut a grand figure at court and one or two daughters to be equipped for splendid marriages. The surplus girls (not necessarily illegitimate but unwanted as pawns in father's political game) may be disposed of with the comparatively small dowry that will admit them to the cloister. There they must go, whether they have a genuine religious vocation or not; in other words, human beings are sacrificed in the interests of worldly success. Such things are not done now? Let us reframe the question. "Have you ridded yourself of those who are obstacles to your ambitions by consigning them to oblivion and giving your action a respectable ethical gloss?" Today the unwanted are not a man's daughters and the place of consignment is no longer a convent; otherwise, the cap still fits and may be worn with modern dress in our own times. Men of promise can be shunted into blind-alley occupations where, bogged down in the petty details of repetitive and boring tasks, they can never hope for more than the partial development of their natural gifts. The Spanish maidens might have found fulfilment in the cloister, their sisters married to exacting husbands might have envied them; but the side-tracked scientists live in miserable frustration deprived of the fulness of life for which they yearn.

Deeply implanted in our human nature is a hunger for that fulness of life, an urge to develop every power that man possesses to its utmost expansion, its greatest intention and extent. The satisfaction of this hunger presupposes the acceptance of life at every level with all that it involves in the faithful carrying out of each and every duty. Caught up in an increasing acceleration of change, we may lose our sense of proportion and ruthlessly reject all that is most deeply rooted in our lives just because it is the growth of many seasons. Certainly what cannot be reconciled with truth, as we now see it, must go, but the going must be quiet and well-ordered. What was good in the old should fall into its new place in a cybernetic synthesis which gives meaning to the fuller and fresher aspect of truth that centres the technological civilization in which so much of the modern world is involved.

At this stage in man's evolutionary progress the builder is of greater use than the demolition squad; but there must be no avoidance of the fact that cybernetics does impinge on religion and that the impact can be a disconcerting and even damaging one. Those whose faith is strong and clear and widely searching in its view will see that collision of science with religion must ultimately shake away the dust and cobwebs on both sides. What is left intact is the conserving element that is always there to preserve the organic unity of the truth. It is in that organic unity of the truth that humanity finds a common heritage.

In this paper I have tried to cover the three points Wiener raised in *God and Golem Inc.*, with additions drawn from and influenced by a somewhat unusual educational background. If the purely personal values of religion seem to have been set aside, it is not because such values lack significance. They have their significance; nevertheless, they are less important than the basic truth which is as much the pith and marrow of science as it is of religion. In that truth is found the freedom of communication without which there is only inadequate living for man as a social being and no fulfilment in his inner life of the higher aspirations of his religious sense.

Let us now end this inconclusive conclusion with a flashback to the "dance with death" that Wiener found so intriguing. It calls for courage to take part in it authentically, but those who dance dauntlessly through its whole cybernetic pattern will find it like a river changing its name as it flows into another country. What begins as the "dance with death" becomes the "dance of life". There science and religion find fulfilment of a promise[22]: "Vincenti dabo edere de ligno vitae, quod est in paradiso Dei mei."*

* Jerusalem Bible (English translation) gives: "Those who prove victorious I will feed from the tree of life set in God's paradise."

References

1. N. Wiener, *God and Golem Inc.*, A Comment on Certain Points Where Cybernetics impinges on Religion, M.I.T. Press, Mass. (1964); Chapman and Hall, London (1965).
2. Wordsworth (1770–1850), "My heart leaps up...".
3. N. Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*, 2nd edn., M.I.T. Press, Mass. (1961).
4. N. Wiener, *The Human Use of Human Beings*, Eyre and Spottiswoode, London (1954).
5. *Bible*, *O.T.*, Wisdom, Chapter 7, Verse 14.
6. H. Rahner, *Man at Play* (English translation), Burns and Oates, London (1965).
7. Thomas Aquinas, *Summa Th.*, II–II ae q. 168, a. 2 (Latin text and English translation), Eyre and Spottiswoode, London (1963); McGraw-Hill, New York (1963).
8. Aristotle, *Nicomachean Ethics*, Book 4, 14.
9. L. Bagrit, *The Age of Automation (Reith Lecture, 1964)*, Weidenfeld and Nicholson, London (1965).
10. J. C. Loehlin, "Machines with personality", *Sci. J.*, October (1968).
11. T. de Chardin, *The Phenomenon of Man*, revised edn., Collins, London (1965), Chap. 2.
12. S. Beer, *Decision and Control*, Wiley, New York (1966), Chap. 14.
13. D. Gabor, "Electronic inventions and their impact on civilization", *Inaugural Lecture*, Imperial College, University of London, March 3 (1959).
14. S. Beer, *Cybernetics and Management*, 2nd edn., English Universities Press, London (1967).
15. Thomas Aquinas, *Contra Gent.*, Book 4, Chap. 84.
16. M. Maltz, *Psycho-cybernetics*, Wilshire Publishing Co., Calif. (1960) (out of print).
17. J. Thomson (1700–1748), *Seasons*, London (about 1730).
18. R. Goodman, *Programmed Learning and Teaching Machines*, English Universities Press, London (1962).
19. *Programmed Instruction in the British Armed Forces*, H.M.S.O., London (1965, 1967).
20. J. R. R. Tolkien, The Lord of the Rings, Vols. 1, 2, 3, Allen and Unwin, London (1955).
21.* Louis of Granada (1504–1588), *Memorial de la Vida Cristiana*, Lisbon (1565).
22. *Apocalypse*, Chap. 2, Verse 7.

---

* Modern Spanish editions and foreign translations omit the original examination of conscience.

CHAPTER I–10

On the fundamental notion of cybernetics

TSUKASA HATORI

Defence Academy Yoosuka, Japan

Summary

The author first considers information and thinking processes, as well as the construction of relevant models. He then analyses creative processes and proceeds to discuss the relationships in systems involving man and machine, and man and man.

TREATMENT OF INFORMATION

The problem consists of studying the methods of dealing with the concept of information. The first task is to express the concept by means of symbols, e.g. letters of the alphabet, binary numbers, etc., and then feed these into a computer, which will deliver the information required on the basis with a certain program. This procedure, which may be demonstrated on a flow chart, enables one to clarify data and to remember and recognize patterns. In addition, other processes may also be included, e.g. learning and even creative processes, termed artificial intelligence.

In the human brain information can be processed and a model constructed. Though the speed and accuracy of machines are superior to those of the human brain, their pattern recognition and allied abilities are inferior. In order to study this problem it is necessary to analyse man's ability to recognize, learn, think, and create.

THINKING PROCESSES

It is possible to envisage three "thinking" modes involved in learning. In the primary mode knowledge of a given subject may be obtained by reading a book or from learning from other people. Experimentation forms the basis of the second mode, whereby one constructs hypotheses on the basis of experiments. Should these be wrong, the information could be used for performing other experiments and postulating different hypotheses; in effect, this involves procedures of trial and error. The third mode leads to correct information via rigorous experiments and feedback of resulting information.

For example, if a stranger wishes to know London, he would—in the primary mode—read books on the geography and history of the city and look at maps and guidebooks. In the experimental mode, he would travel or walk around London, visit the museums, theatres, and other places of interests, converse with the inhabitants, etc. Finally, in the third mode the stranger would obtain a historical and accurate picture of the city by thinking about it in depth and studying the views and comments of experts.

To construct a model of these processes, one could proceed as follows. The information S enters the brain and is transformed by the operator d on S, i.e. $d: S \rightarrow S_1$ (S is transformed into S_1). For other operators, $d_1 \ldots d_n$,

$$d_1 : S_0/S_1, S_2, \ldots, S_{n-1} \rightarrow S_n$$

$$d_n : S_0 \rightarrow S_n$$

The operators $d_1 \ldots d_n$ are converted into a knowledge space as an invariant set in S. Furthermore, one has to note that to find the pattern from a set M of some objects, one has to search for an equivalent relation E by writing $a \equiv b(E)$, where a and b belong to M. Thus the pattern of dividing by 3 a set of natural numbers may be expressed by the relation

$$a \equiv b \pmod 3$$

where a and b are natural numbers. If M is the set of numbers, then the pattern will consist of equivalent units. The theory of algebraic structures, as well as other theories, may be brought in to deal with this problem.

LOGIC

The logic used by computer is symbolic or formal logic. However, there exist problems that cannot be tackled by this type of logic, as illustrated by the Gödel theorem or Russell's paradox, hence the need for a more accurate logic and a re-investigation of the definitions and methodology in mathematics.

CREATIVE PROCESSES

There are two kinds of creative processes. In the first kind one selects and builds an optimum pattern on the basis of past experiences, including random elements, coupled with perception and insight, e.g. composition of a piece of music. It may be possible, in principle, to achieve this by means of a computer.

The second kind of creative process cannot be formed by means of a machine, since it involves great heterogeneity and it is also much different from previous patterns, e.g. Eastern versus Western culture. In the author's view, Western culture is exact and "humanizing", since its mode of thought is descriptive and analytical, while the Eastern variety is spiritually enlightened and "naturalizing", because its underlying base is instructive and synthetical. It is possible that a synthesis of the two cultures may arise by prolonged contacts.

MAN AND MACHINE

The purpose of cybernetics is, as Wiener put it, to control controllable objects by observing the behaviour of uncontrollable objects. Two objects will form a stable system provided they act against one another or have opposite properties. As regards a man–machine system, three cases arise. In the first case, the machine controls man; this is an unstable system. In the second case, man controls objects, while in the third case, one is dealing with the system man–man. This involves personal relations on an individual scale, or national and racial aspects; in the former case a stable relationship could be established by proper communication, while in the international sphere some stable co-existence could also be achieved.

CHAPTER I–11

The role of positive feedback

D. STANLEY-JONES

Full Circle Foundation for Education and Research
Hayle, Cornwall, U.K.

Summary

A car engine operates on two levels of energy, namely chemical and electrical, the fuel that explodes and the flash that starts the explosion. To describe this, we need two terms, namely energy and dynamic. The living body likewise has two levels: the chemical energy of the muscles, and the neural dynamic in control. Kybernetic studies of neural dynamic reveal that release of energy is often associated with positive feedback. A jet engine has no moving parts; it derives its immense power from a pattern of positive feedback.

Other examples are explosions and flames (two-dimensional explosions) of all kinds, chain reactions, atomic reactors (controlled) and bombs (uncontrolled by negative feedback).

There are many examples in biology of positive feedback as a source of neural dynamic: the flight of locusts, lateral line of fishes, gamma bias in spinal cord, decerebrate rigidity in animals and sham rage, and real rage in humans; fight-or-flight reaction and emotions of rage and fear, and many other instances.

FIRST PRINCIPLES

Positive feedback has not attracted as much attention as negative feedback, because negative feedback makes for stability, whereas positive feedback may destroy stability.

Positive feedback, however, has its biological uses. It is, indeed, widely distributed in Nature. An examination of various instances of naturally occurring positive feedback may throw new light on certain problems of biology and medicine, and lead to the formulation of some general principles.

The simplest example of positive feedback, of course, is the thermostat of an electric oven with its connections reversed. The input of current increases with increase of output or temperature, and so the oven gets hotter and hotter, in an accelerated runaway which ends in its own destruction. In the other direction, the current is switched off as the oven cools, and the result is a decelerated runaway toward zero (or room temperature) as an asymptote.

Another example is the fly-ball governor of a steam-engine, with its links reversed. As the output or engine-speed increases, the input of steam is also increased, and drives the engine still faster. Again, the result is a runaway to destruction; or, in the opposite direction, a runaway to zero.

A third example is the turbo-jet engine. Positive feedback is inbuilt into the structure. The greater the output or thrust at the rear, the greater the speed, the greater the intake at the front, and again the greater the output. Input and output chase each other in the never-ending spiral of positive feedback, and the only method of preventing a runaway to destruction is by limitation of the fuel supply. There are no external moving parts, and the immense power of the jet engine derives solely from the pattern of positive feedback used intentionally for this purpose.

From these and other simple illustrations, the following rules have been suggested (Stanley-Jones, 1967):

(1) Positive feedback is potentially a danger, and must always be monitored by some form of control.

(2) If positive feedback gets out of control, it leads inevitably to runaway, either to zero or to maximum. Runaway to zero is tantamount to death. Runaway to maximum leads usually to destruction.

(3) Conversely, runaway to either extreme, and vicious circles in general, invariably betoken positive feedback. This is the kybernetic equivalent of the saying "extremes meet": they share a common factor, namely positive feedback which is running out of control.

(4) A system that is "running away" under the influences of positive feedback has broken free from the monitoring or stabilizing influence of negative feedback, and cannot be restored to normal except by active intervention from outside; for negative feedback cannot regain control unaided.

(5) Positive feedback when out of control can spread without limit to dominate the behaviour of the whole organism.

(6) A system that is running away has lost touch with its environment and, being out of control, is heading for disaster.

(7) A system that is running away comes to rest only when the input or source of supply has been cut off or exhausted.

This last rule is illustrated by the turbo-jet. The system can be controlled only by restricting the input of fuel, that is at the chemical level. No attempt is made to vary the inbuilt pattern of positive feedback, for this is unalterable.

CHEMICAL ENERGY AND NEURAL DYNAMIC

It is helpful at this point to distinguish two levels of organization, which could be named energy and dynamic. There is the lower, chemical level of energy, that is of fuel (in the case of a machine) or of food (in a living body). Above this is another level, which does not actively supply the energy or motive power, but is necessary for its control or release. In the case of the human body, this is known as neural dynamic (Stanley-Jones, 1960, 1966a), a term which replaces the rather vague concept of "nervous energy" which is confusing, because nerves do not in fact supply any motive power to the muscles they control.

For want of a better name, "dynamic" could be used for this upper level of organization, the level of control, leaving the name "energy" for the lower or chemical level which actually does the work.

This two-tier organization of energy and dynamic is perfectly illustrated in a car. On the chemical level, there is the fuel which supplies the energy of motion. This energy is not available as motive power until the mixture has been exploded by the spark. The higher, electrical level, or dynamic, releases and controls the energy supplied at the lower, chemical level.

Positive feedback appears in countless examples in non-living systems. Every explosion is a chain reaction in which the output of one unit of action is the input for the next. In an atom bomb, the output is neutrons; in a chemical explosion, the output is heat. A flame is a two-dimensional explosion, a sheet of burning that takes its dynamic from positive feedback, and without which it could not continue. In the stationary flame of a gas burner, the flame is travelling downwards at the same rate as the explosive mixture is travelling upwards.

In all these instances involving positive feedback, there is no way of controlling the process except by limitation of the intake or fuel supply; that is, on the chemical level, as distinct from the dynamic level with its pattern of input and output linked by feedback.

This concept of a two-tier organization is essential in studying the kybernetic control of the living animal. Chemical energy of the muscles is wholly dependent upon the neural dynamic of the nerves. When the nerves are cut, muscles are paralysed, although the actual quantum of energy contributed by the nerve fibre is virtually nil.

POSITIVE FEEDBACK AS A SOURCE OF DYNAMIC

Neural dynamic depends, for its effect, on the presence of positive feedback. An example of this is to hand in the pattern of control of a flying locust (as reported by Weis-Fogh in 1949). The locust carries on its head several patches of sensory hairs. Stimulation of these hairs by a breath of wind supplies the neural dynamic for flight. If a locust is held fast on a table with its wings pinioned, a jet of air blown onto the front of the head causes vigorous movements of the wings. The response is abolished if the sensory hairs are covered by paint.

This supply of neural dynamic is patterned on positive feedback; for the greater the output or speed of flight, the greater the input or air pressure on the front of the head—exactly as in a jet engine.

The system is prevented from running away to maximum by the limitation of supplies on the chemical level, that is by the availability of glucose as fuel for the muscles. The pattern of positive feedback allows the locust to mobilise the maximal output of which it is capable, and to sustain this for a flight of over 1,000 miles.

Positive feedback here acts in a fore-and-aft direction. The same system is patterned also on negative feedback in the side-to-side direction. If a jet of air is blown onto the side of the head, that is at an angle to the direction of flight, neural dynamic is directed into the flight muscles of the opposite side, thus enabling the locust to correct the angular deviation, and to reduce to zero the error as distance-from-goal. Positive and negative feedback can thus co-exist within the same circuit.

THE ORIGIN OF NEURAL DYNAMIC

Positive feedback accounts for the all-or-none quality of a neural impulse, and in fact supplies it with the necessary dynamic to carry its message. The all-or-none response is a typical runaway to maximum, and is essential to convey the nerve impulse over long distances where a graded or decremental response would fail to arrive.

Passage of a neural impulse involves changes in the permeability of the nerve membrane, especially to ions of sodium. The resting cell membrane is relatively impermeable to sodium ions, and is the site of a sodium pump of unknown mechanism. The continuous expulsion of sodium ions is a relic of our marine ancestry, an evolutionary adaptation to life in the hypertonic saline of sea-water; this demands the pumping out of excess of sodium, and conservation within the cell of the much rarer potassium.

Sodium ions, bearing a positive charge, remain outside the resting nerve cell; glutamic and other acidic radicles, negatively charged, remain within. A difference of electric potential is thereby maintained across the membrane, aided by the sodium pump, with the inside of the cell negatively charged. The difference in potential across the membrane of the resting cell is about 80 millivolts.

If the membrane potential is reduced, the permeability to sodium ions is increased (Hodgkin, 1958; Katz, 1966). The reason for this is unknown. Sodium ions carry positive charges, and by entering the cell they lessen still further the negative charge within the cell, and hence the potential-difference across the membrane.

Here then is the pattern of positive feedback. The first step is a reduction of negative potential, say from 80 to 70 millivolts, in response to some outside stimulus. This increases the permeability to positive sodium ions, whose influx reduces still further the negativity and membrane potential, and allows a further flow of sodium ions to continue the process.

The reverse influence of effect upon cause (which is the essence of feedback) is such that as the effect is increased, so too is the cause of that effect. Sodium ions pass across the cell membrane and increase the membrane's permeability which allows further entry of sodium.

So long as the membrane potential lies between 80 and 40 millivolts, the cycle of positive feedback is kept under control by negative feedbacks, and the system remains stable. At a critical point (Katz, 1966, p. 76) in the region of 40 millivolts, however, a trigger mechanism comes into play: positive feedback gains the upper hand, and the resulting explosion or runaway is propagated along the membrane of the nerve fibre as a neural impulse. A neural impulse, in fact, takes its dynamic from the pattern of positive feedback.

NEURAL DYNAMIC FOR VOLUNTARY MOVEMENT

A further example of positive feedback as the source of supply of neural dynamic is seen in the maintenance of postural tone and in voluntary movement. This is achieved by the interplay of the motor neurons in the ventral horn of the spinal cord. Motor neurons here are of two sizes: the large alpha cells that control the muscles, and the small gamma cells that control the muscle spindles.

A muscle spindle contains a spirally coiled structure which acts as a stretch receptor. Each stretch receptor has its own minute muscle, activated by the small gamma cells. These spindle muscles pull directly on the stretch receptor, and cause a train of afferent neural impulses to enter the spinal cord. This input is directed, in part, to the gamma motor cells, and increases their output; that is, it increases the contraction of the intrafusal muscles that pull on the stretch receptors, and so increases the input. This is a typical circuit of positive feedback in which input and output chase each other in a continually self-exciting circle.

This returning input from the muscle spindles is the source of neural dynamic for contraction of the muscles. There is a bifurcation of the incoming afferent impulses: part goes to the gamma cells and re-enters the circuit of positive feedback, part goes to the alpha cells that control the muscles of voluntary movement. Once again, the source of the nerve impulse or neural dynamic is seen to be directly related to a pattern of positive feedback.

Positive feedback is potentially dangerous because of the ever-present risk of runaway. It must therefore be kept under control by the monitoring action of negative feedback. When this monitoring control from the higher centres is cut off, the result is decerebrate rigidity which is nothing if not a runaway to maximum.

The lower or spinal centres operate also a negative or stabilizing feedback. This is achieved by the interplay of alpha and gamma motor cells; that is, by the same neural circuits that are used for positive and negative feedback. This is similar to the pattern in the flying locust, in which the two kinds of feedback, positive and negative, share a common circuit.

Negative feedback works through the alpha cells, which cause the voluntary muscles to contract. This takes the strain from off the muscle spindles (which register mainly the passive stretching of the postural or antigravity muscles). The returning input from the spindles to the gamma cells is reduced, and this cuts down also the supply of neural dynamic to the alpha cells. Activity of the alpha cells automatically reduces the neural dynamic which maintains that activity.

Further details of this work, which is based on the writings of Granit, Eldred, Merton, and others, may be found in my books *Kybernetics of Natural Systems* (1960) and *Kybernetics of Mind and Brain* (1969b).

Positive feedback appears in the higher neural centres, in the interplay between cortex and reticular system of the brain stem. Stimulation of the cortex yields evoked potentials in the reticulum, one of whose functions is to maintain or to increase the activity of the cortex. It has been suggested that the reticular system can be alerted by thoughts and ideas descending from the cortex, and it then arouses the cortex to further alertness and attention (French, 1955).

The negative feedbacks necessary to prevent this mutual reinforcement from exploding into a runaway have been discovered in the cortico-fugal afferent controls, which permit the level of activity aroused in the cortex to regulate the sensory input. This has been proved for all the senses, as a "built-in negative feedback" (Lindsley, 1958). It extends its effect as far outwards as the first synaptic relay, and in a few cases to the peripheral receptor.

This regulating function may be supplemented also by the sensory suppressor zones (Stanley-Jones, 1966b) which are believed to be aroused by the same ascending reticular activating system: yet another instance of negative and positive refectory circuits operating over the same neural structures.

FEEDBACK IN THE AUTONOMIC SYSTEM

Positive feedback appears also in the autonomic system. Here it seems to be limited to the orthosympathetic. The parasympathetic is based wholly upon negative feedback, as befits its homoeostatic function of maintaining the status quo and preserving the stability of the internal environment.

The orthosympathetic has its headquarters in the posterior hypothalamus. Stimulation here causes secretion of glandular adrenalin into the bloodstream, where it circulates for a while in low concentration. Nerve cells in the posterior hypothalamus are specifically sensitive to low concentrations of adrenalin (Porter, 1952, 1953). The electrical response to injected adrenalin, in cats and monkeys, is limited to the posterior and lateral areas (the orthosympathetic centres) and does not arise in the anterior hypothalamus (the parasympathetic centre).

Secreted adrenalin is thus capable of promoting its own further secretion by virtue of its action on the posterior hypothalamus. Elsewhere the production of adrenalin is not itself caused by adrenalin nor by adrenergic nerves; for the adrenal medulla is controlled by preganglionic fibres which are cholinergic. This serves as a built-in check to prevent a runaway in which adrenalin could stimulate its own secretion.

The runaway, when it does occur, is at the level of the posterior hypothalamus. This is fully consonant with all that is known of the orthosympathetic emotions of fear, rage and anxiety. All of these can precipitate into a runaway, in which the organism loses contact with the environment and, being out of control, is heading for disaster. It is the function of the parasympathetic to prevent this.

When the monitoring controls break down, fear (the emotion of controlled flight) becomes uncontrolled panic, and rage (the dynamic for a fight) breaks down into a loss of temper or a brainstorm. Anxiety marks the point of transition from "under control" to "out of control" (Stanley-Jones, 1957b, 1957c, 1968, 1969b). Anxiety arises from the failure of negative feedback to monitor the potentially dangerous circuits of positive feedback, which are at risk to break out into a headlong runaway to destruction.

Such are the contrasting phases of schizophrenia. Runaway to maximum appears as manic and frenetic displays and running amok, while stupor, silence, autism, katatonic rigidity and the like are the results of an uncontrolled runaway to zero. One aspect at least of schizophrenia may be understood in terms of positive feedback, in which the inter-relations of anxiety, adrenalin and the posterior hypothalamus could profitably be studied by kybernetic analysis.

SPASM AND COLIC

A further instance of positive feedback in a human setting is the uncontrolled spasm of a sphincter muscle. An example of this is ureteric colic, of "stone in the kidney", in which a small stone is impacted in a narrow tube, giving rise to reflex contraction of the circular muscle. This grips the stone still more tightly, and thus increases its own spasm and the associated pain.

All types of colic are examples of runaway under the dominance of positive feedback. The circular fibres surrounding the hollow organ are reflexly stimulated by the presence and pressure of an impacted solid, and they go into spasm. The spasm increases the pressure, and the pressure increases the spasm, and both rise to a maximum.

The vicious circle once started spreads throughout the organism as the totally disabling pain of renal colic. The runaway can be stopped only by mechanical dislodgement of the impacted stone, or by anti-spasmodic drugs which affect the kybernetic circuit from without. Negative feedback and normal peristalsis cannot regain control until the vicious cycle of cause and effect has been broken by interference from outside.

One of the most convincing examples of total spread of a vicious cycle patterned on positive refection is seen in the condition known as vaginismus. In certain types of female patients, local spasm of the vagina not only causes intolerable pain, but the runaway may spread to involve also the whole personality in an attack of "hysterical" shrieking: the patient loses all pretence of self-control and is totally out of touch with her environment—a transient psychosis brought on solely by the unrestricted spread of positive refection.

CARDIAC ASTHMA AS A RUNAWAY

Runaway to zero in one aspect of a system is not incompatible with runaway to maximum. The two conditions frequently co-exist. A further human example is cardiac asthma (paroxysmal nocturnal dyspnoea) in patients who live on the borderland of cardiac failure. Here the sequence of events leading to the sudden onset of runaway refection is fairly clear. During the night the lungs slowly become congested with blood, after the patient has changed position from upright to flat. This congestion is not due solely to gravity, for this would make its influence felt within a few minutes; it takes several hours before it reaches the critical point of breakdown into runaway. There is a slow rise in the volume of the blood, as fluid passes out of the oedematous tissues. The excess of blood affects mainly the thin-walled veins, rather than the thick-walled arteries, and a rise of venous pressure causes over-distension of the heart. "Once this occurs a vicious cycle is produced; decreased output leads to further elevation of venous pressure, and elevation of venous pressure leads to further over-distension of the heart and decreased output." (Youmans, 1961.)

Engorgement of the lungs brings on respiratory distress (cardiac asthma) from back-pressure into the pulmonary capillaries, followed by transudation of fluid from the capillaries into the alveoli. The exchange of respiratory gases is much impaired, leading to reduced oxygenation of the blood and further embarrassment of the overloaded heart.

The vicious cycle is in truth an expanding spiral that exhibits all the features of a kybernetic runaway. Runaway to maximum co-exists with runaway to zero within the same system. Respiratory distress, distension of the heart and oedema of the lungs rise steadily to a maximum, while the movements of the patient as a living organism are reduced to zero if help from outside is not forthcoming.

Once positive refection has gained control, there is unrestricted spread until the process involves the whole organism or is brought to an end by failure at the chemical or economic level. It cannot check its own spread to death or destruction except by interference from outside, and the patient becomes progressively more helpless to bring the attack to an end. Change of posture from flat to upright, and an injection of morphia to depress the respiratory centre, are examples of outside interference or therapeutic measures that can give relief to the patient and bring the runaway to a halt.

The vicious spiral was started when the parasympathetic reflexes and their negative feedbacks became unable to exert their monitoring control because they were taken outside their normal working range, and the kybernetic system of homoeostatic controls broke down into a runaway. The orthosympathetic and its positive feedbacks gain the ascendancy, and with them the emotions of fear and anxiety which dominate the patient's acute mental distress. Negative refection regains control when the venous pressure and the dilatation of the heart are reduced to within their wonted physiological limits. A large part of abnormal cardiac physiology could be written in terms of positive feedback and orthosympathetic dominance, in contrast with the classical account of the normal action of the heart with its many negative feedbacks mainly under the control of the parasympathetic.

RUNAWAY TO ZERO

Runaway to zero presents fresh problems. A number of instances could be cited; but each works on its own pattern, and it is not easy to discern any general laws applicable to all the cases. I have a hunch, however, that elliptic integrals enter into the picture (Stanley-Jones, 1969a); and if anyone here is working on these lines I should be glad if we could compare notes. In the belief, however, that some general principle will eventually emerge, a few examples of runaway to zero will be considered.

The thermostat with reversed wiring again serves as the paradigm. Cooling of the oven (output) brings about a lessening of the current (input) until both have fallen to the zero of room temperature. There is a decelerated runaway approaching room temperature as to an asymptote, but (as with all cooling bodies) never actually reaching it. Is the deceleration as essential an accompaniment of runaway to zero, as acceleration appears to be of runaway to maximum? And what is the mathematical formula that shapes the curve?

The traffic congestion of cities exhibits positive feedback: it is certainly slowing traffic to a standstill. Here is a good example of positive refection leading to zero, culled from a London newspaper: "The traffic conditions with which the buses have to contend militate increasingly against their efficient operation. Traffic congestion, by preventing regular and speedy running, means that equipment is not used to capacity, costs increase, and the manpower resources get overstretched. As the bus services falter more people take to private cars, further increasing the congestion." (*Daily Telegraph*, October 29, 1963).

It is common knowledge that traffic flows fairly smoothly until it reaches saturation. Then there is a *sudden* build-up of congestion and long delays. The point of change-over from freely moving traffic to severe stagnation is the point of change from negative feedback, that is traffic under proper control, to positive feedback with breakdown of control.

The dinosaurs of Jurassic and Cretaceous times are distant by a hundred million years from our present age of stress. Their many species extinguished themselves seriatim by evolutionary and physiological processes in which the latent powers of positive refection had probably got out of control. The giant carnivores *(Therapoda)* enlarged their bodily size and muscular powers of attack to beyond the respiratory capacities of their heart and lungs: the amount of oxygen needed to move twenty or thirty tons of muscle and bone for a brief burst of movement created an oxygen debt which would leave a cold-blooded reptile panting and breathless perhaps for an hour or more, at the mercy of the first predator who ambled into view. Many species of these giant reptiles, such as *Tyrannosaurus* and *Gorgosaurus*, became extinct from over-development of body weight following an irreversible trend of natural selection; the struggle for existence, instead of resulting in the survival of the fittest, issued only as the failure of the heaviest.

These flesh-eating *Therapoda* carried little in the way of defensive armour. Their strength lay in their muscles and teeth, on the principle that attack is the best form of defence. The vegetarian dinosaurs, *Stegosaurus*, *Triceratops* and their kin, developed their defensive armour to such an extent that they too became defenceless under the sheer weight of gravity. "With a heavy skull, a tank-like body, and a crusader's mace arrangement of spines and spikes on the tail, it is no wonder that these animals were vegetarians. They were far too static for any other food." (Swinton, 1962.)

Attack and defence in these opposing groups, the carnivores and the herbivores, appear to have followed one another in the all too familiar pattern of the modern world, each stimulating the other as output and input in a kybernetic circuit or vicious spiral, until the greatest of these giant reptiles became defenceless by the sheer weight of their defensive armour, or impotent in attack on account of muscular and respiratory exhaustion. No doubt endocrinic disturbances played their part in the age-long fashioning of these bizarre monsters; the horns and gigantic frills of the latest of the hooded dinosaurs are clear evidence of over-activity of the pituitary gland, whose role as conductor of the endocrine orchestra is mediated largely by an intricate balance of negative feedbacks on the chemical level. Disturbance of this balance risks the emergence of unmonitored positive feedback, with long-term disturbance of metabolic activities leading to premature ageing of the individual and ultimately to extinction of the species. Runaway to maximum as regards size and weight led to runaway to zero as regards the species and indeed the group as a whole.

Another biological example of runaway to zero is the stoated rabbit. A stoat when seeking food may come across a group of rabbits in the open. One of these rabbits, and one only, falls victim, usually the rabbit nearest to the stoat. The other animals continue unconcernedly in their play, as though they had been frequently the witnesses of similar scenes, and knew both the outcome of the tragedy and their own immunity from attack. The victim screams shrilly (the only occasion when a rabbit vocalizes, other than by low grunts), and remains virtually rooted to the spot, unable to make any movements to escape the oncoming stoat (Vesey-Fitzgerald, 1946).

No physiological explanation has yet been worked out for this instance of runaway to zero. It is often said that the rabbit is "hypnotized" by the stoat—not really a misuse of the word, although there is little in common with hypnosis in man. Comparison seems more profitable with the man who is terrorstruck in the face of imminent danger and finds himself powerless to move. Of this also, there is as yet no explanation either physiological or psychological.

Of immediate interest to human physiology is absence of pain in a total burn. Recent instances have come into the news, of human beings soaking their clothes in petrol and setting alight to themselves, immolating themselves in the flames of their own igniting, and enduring the seemingly intolerable pains of being burnt alive. The writer has come across two such cases of accidental total burns, with a most surprising absence of pain.

A child aged seven caught her clothes alight while alone, for a few minutes only, in an isolated cottage in a village in Devonshire. When her mother returned from the local shop, she met the child running down the lane, her shredded rags still smouldering, her body almost totally scorched. The child gave no signs of pain or anguish, nor even of fright. Her sole reaction during the journey to the hospital (where she died shortly after admission) was one of guilt; she spoke quite coherently on one theme only, that she was so sorry that the accident had happened, and so sorry to have caused her mother all this trouble. She seemed to be suffering no pain whatsoever. The absence of pain cannot be explained by bodily collapse or mental shock, as neither was present. Her dominant emotion was guilt, which is mediated by the parasympathetic (Stanley-Jones, 1957a, 1957c, 1969b). Only the orthosympathetic seemed to be out of action, and with it the orthosympathetic emotions of fear and anxiety, and the physical sensation of pain.

There is no explanation, yet, in terms of physiology or biochemistry, of this unaccountable absence of pain after an extremely painful injury. Kybernetics would suggest that because pain is above all others the sensation which can precipitate a runaway to maximum, it has in these rare cases caused a runaway to zero, possibly as a protective action, leaving the sufferer devoid of any sensations of pain.

I have also seen an elderly woman who tripped up while holding a paraffin lamp. Her clothes caught alight and were almost wholly burnt. About twenty minutes after the accident, she was sitting in her cottage talking fairly coherently with two neighbours. She made no complaint of pain over the parts of her body that had been scorched; she complained only of severe pain with hypersensitivity to light touch, around the right shoulder, which was the only region of her body that was not burnt. There was a runaway to zero, in the absence of pain, in the areas of skin that had been burnt, together with a runaway to maximum with severe pain and tenderness in a small area of skin that had not been burnt.

CONCLUSION

There are many other examples of positive feedback in natural and living systems, both biological and social. All of them obey, or illustrate, the empirical rules that were suggested at the beginning. The main conclusion to be drawn from this study of positive feedback is that it is potentially a source of danger unless monitored by negative feedback; yet in many cases it serves as the primary source of neural dynamic on which the organism relies in both normal and emergency situations.

Conversely, any system which appears to be moving in a vicious circle, which exhibits a runaway to maximum or is freezing to a standstill, may be suspected of harbouring a positive feedback. To restore such a system to normal, it is imperative to identify the negative feedback which has failed in its monitoring duties, and to locate the exact point of change-over from negative to positive feedback, the critical point at which positive feedback has taken over the controls.

The understanding of the role of positive feedback is essential if world affairs are to be salvaged from their present state of chaos. Positive feedback has gained control in many realms of international relations, particularly in the accelerating pace of nuclear armaments and the ever increasing cost of the balance of terror.

The emotions of fear and hatred are patterned on positive feedback. They are the natural source of dynamic for any system that is fighting for its survival. The present and urgent danger is that these emotions, necessary and proper for the economy of a single organism, may wreak irreversible destruction if they are allowed to dominate human society as a whole.

References

Eldred, E., R. Granit, and P. A. Merton (1953). "Supraspinal control of muscle spindles." *J. Physiol.*, **122**, 498.

French, J.D. (1955). *J. Neurophysiol.*, **18**, 74.

Granit, R. (1957). "Systems for control of movement." *Proc. 1st Intern. Congr. Neurol. Sci.*, **1**, 67.

Granit, R. (1968). "The functional role of the muscle spindle's primary end organs." *Proc. Roy. Soc. Med.*, **61**, 69–78.

Hodgkin, A.L. (1958). "Ionic movements and electrical activity in giant nerve fibres." *Proc. Roy. Soc. (London), Ser B*, **148**, 1–37.

Katz, B. (1966). *Nerve, Muscle, and Synapse*, McGraw-Hill, New York.

Lindsley, D.B. (1958). "Common factors in sensory deprivation, sensory distortion, and sensory overload." In *Sensory Deprivation* (Ed. P. Soloman). Harvard University Press, Cambridge, Mass.

Merton, P.A. (1952). "Servo-control of movement." In *The Spinal Cord (Ciba Found. Symp.)*.

Porter, R.W. (1952). *Am. J. Physiol.*, **169**, 629–37.

Porter, R.W. (1953). *3rd Annual Report on Stress* (Ed. H. Selye and A. Horava) pp. 212–19.

Stanley-Jones, D. (1957a). "The structure of emotion: lust and rage." *Psychoanal. Rev.*, **44**, 289–97.

Stanley-Jones, D. (1957b). "The physical basis of anxiety." *J. Nerv. Ment. Disease,* **125**, 247–258.

Stanley-Jones, D. (1957c). *Structural Psychology*, Wright, Bristol; Pergamon Press, New York.

Stanley-Jones, D. (1960). *Kybernetics of Natural Systems.* Pergamon Press, Oxford; *La Cybernétique des Êtres Vivants.* Gauthier-Villars, Paris.

Stanley-Jones, D. (1966a). "Kybernetics of Cyclothymia." *Progr. Brain Res.*, **17**, 151–68.

Stanley-Jones, D. (1966b). "Protopathy, Paraesthesia and Sensory Supressor Zones." *Progr. Brain Res.*, **23**, 200–18.

Stanley-Jones, D. (1966c). "The thermostatic theory of emotion." *Progr. Biocybernetics*, **3**, 1–20.

Stanley-Jones, D. (1967). "Kybernetics of the nervous system." *Scientia (Milan)*, **102**, 150–65.

Stanley-Jones, D. (1968): "The Cartesian diver, the swim bladder, and the nature of anxiety." *Cybernetic Med. (Naples, S.I.M.C.)*, **3**, 6–17.

Stanley-Jones, D. (1969a). "The full circle pendulum." *Math. Gaz.* **53,** 153–156.

Stanley-Jones, D. (1969b). *Kybernetics of Mind and Brain*, C.C. Thomas, Springfield, Ill.

Stanley-Jones, D. (1969c). "Biological origin of love and hate." *3rd Intern. Symp. on Feelings and Emotions* (Ed. M.B. Arnold), *Loyola University, Chicago*; Academic Press, New York.

Swinton, W.E. (1962). *Dinosaurs.* British Museum Natural History, London, p. 31.

Vesey-Fitzgerald, B. (1946). *British Game.* Collins, London, p. 171.

Weis-Fogh, T. (1949). "An aerodynamic sense organ stimulating and regulating flight in locusts." *Nature*, **164**, 873.

Youmans, W.B. (1961). In *Physiological Basis of Medical Practice* (Eds. C.H. Best and N.B. Taylor), Bailliere Tindall and Cox, London, p. 504.

SECTION II

Neuro- and biocybernetics

CHAPTER II–1

Cross-connection neuron network learning and self-learning ability

A. KVITASHVILI
Georgian Academy of Sciences
U.S.S.R.

Summary

The paper deals with some results of simulation of figure recognition learning, using cross-connected perceptron-type neuron networks. Concave learning process curves have been obtained. The problem of "teacher-free" learning after receiving a certain amount of initial information is examined. This learning process may be called "self-learning". Working at a self-learning regime, simple perceptrons can damage the acquired abilities, whereas cross-connected perceptrons improve their abilities during the self-learning process.

INTRODUCTION

The present paper deals with some problems of simulation of a perceptron-type neuron network with cross-connections designed for geometrical figure recognition learning. A perceptron with cross-connections implies a self-learning neuron network permitting all connections between the neurons. Hitherto very few results have been obtained on the simulation of completely connected neuron networks or perceptron-type systems with cross-connections, the topicality of which is indisputable. This involves great difficulties in simulating analogous structures. The first circumstance causing these difficulties involves large masses of information entering the digital computer. The second source of the difficulties stems from the necessity of simulating a fairly complex dynamic system with a very complex transient. The behaviour of such systems is not susceptible to analysis by the methods of the existing mathematical apparatus. However, an attempt was made by F. Rosenblatt to develop a method of approximate estimation of the behaviour of such systems. To do this, he employed the averaged probability characteristics of the neuron network[1].

As a result of interaction with its environment, the connected neuron dynamic system can acquire certain transient-free states, determined by the entire past history of that interaction with the environment. To quote W. R. Ashby, it can change into "a polystable system"[2].

Thus, such systems can have homeostasis determined by the past history of the changes of the interneuron connections during the learning process, which has so far hardly been investigated. This means that cross-connected perceptrons essentially differ from simple perceptrons without feedback connections within the neuron network. The latter have been investigated fairly well, and the mathematical apparatus describes their simple functioning, whereas dynamic neuron networks with any connections create great difficulties in their study.

SIMULATION OF LEARNING PROCESSES

One of the possible investigation methods of cross-connected networks at present, is their ineffective simulation on a digital computer. The present writer has succeeded in simulating a neuron learning process network of 100 completely interconnected neurons and of 50 neurons connected to these[3] 100 neurons.

A binary retina consisting of 400 elements was used as input device. The detailed description of the model will here be omitted; it will suffice to note that for learning, the system was exposed to visual stimuli in such a way that the intersections on the retina between the representatives of the same class equalled 25%, and between those of different classes 75%. Notwithstanding such "poor" pictures, the system learned to distinguish the images as shown in Figure 1.

In the former case the test was conducted according to the pictures that had been presented to the system during the learning process. The time of trouble-free operation of the digital computer amounted to 14.5 hours at the rate of 20,000 operations per second. The probability of correct recognition reached 0.975.

The second experiment during the test envisaged the presentation of half of the pictures previously unfamiliar to the system. These new pictures were obtained by shifting and turning previously familiar pictures, so that the

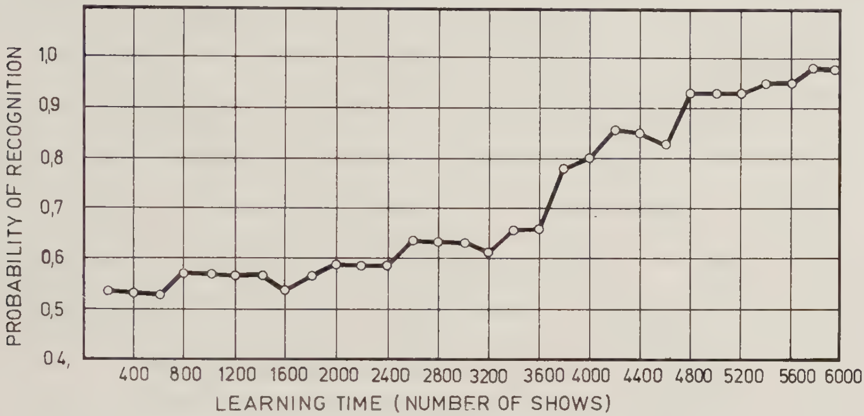


FIGURE 1 Learning diagram of the cross-connection set

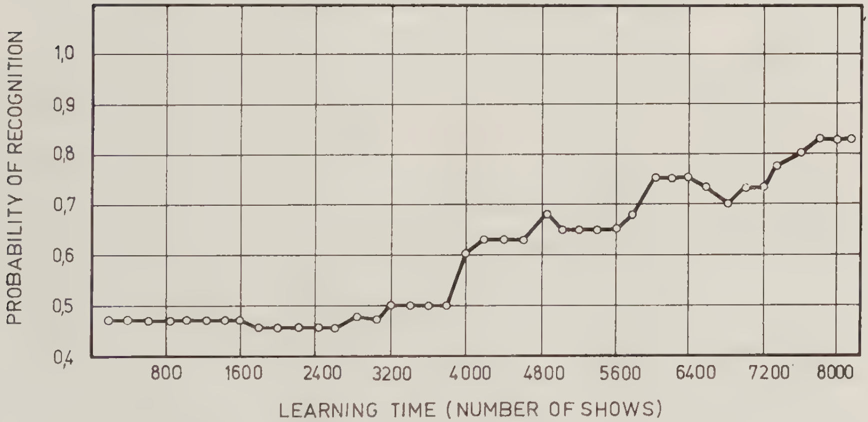


FIGURE 2 Diagram of generalization of the cross-connection set

intersections between the representatives of one class were reduced to the minimum, whereas intersections between the different classes were considerable. As is seen from Figure 2, the system learned a two-class recognition with probability 0.825 after 21 hours of trouble-free operation of the computer.

The following control experiment was carried out: in the computer program the elimination of all the cross-connections was envisaged, and learning and testing were conducted according to experiment 1. This perceptron, devoid of cross-connections, actually did not learn correct recognition of the given pictures. This confirms the expediency of experimenting with the employed patterns in order to illustrate the advantages of the cross-connection system over simple ones.

It is interesting to note that both correct recognition probability curves turned out to be concave. This circumstance, as suggested by F. Rosenblatt[4], points to the analogy of the cross-connected system with living organisms in which the ability for quicker generalization appears as a result of prolonged learning. This phenomenon is not observed in perceptrons without cross-connections in which the probability of recognition curves show a steep rise at the beginning, and then asymptotically approach their maximal value.

The result of the second experiment also appears to be of interest. In simulating a relatively small system (150 neurons in all), the hypothesis on the ability of the cross-connection neuron network to generalize the pictures presented during learning, together with their one-to-one manner transformations[4], proved correct. It is therefore advisable to simulate large cross-connected systems in which the ability for complex generalizations should be more pronounced.

It should be noted that the cross-connection neuron network lays down certain conditions of temporal correlation of external stimuli in both the learning process and testing. Therefore, such a network learns temporal sequences, which would be absurd in the case of systems devoid of cross-connections. A trained cross-connected system forms quite definite dynamic attraction patterns for the representatives of the same class. In other words, during the learning process, if certain weighting coefficients have already been stored, "attraction" of a new picture to some class takes place automatically, forming a complex transient in the system. In these systems there occurs, in a sense, an identification of "pattern-" and dynamic stability. The larger the sequence of pictures of the same class presented, the better will the system learn. And the fewer the intersections on the retina between the pictures, the oftener they will have to be presented during learning. The cross-connected systems learning procedure has not been sufficiently investigated, hence every experiment rests upon the ingenuity and insight of the researcher.

TRAINING SYSTEM OF A NEURON NETWORK

Let us examine the training system of a learning device or neuron network (see Figure 3).

The training procedure consists in the establishing of correspondence—desired by the teacher—between the set of impingements of the environment and the responses of the learning device. The influence exerted by the teacher on the neuron network, until complete learning is achieved, consists in the

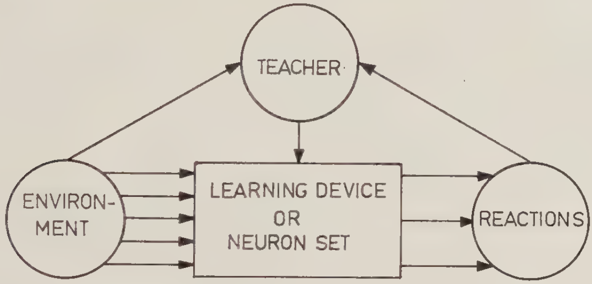


FIGURE 3 General diagram of learning of a neuron set with a teacher

transmission of the necessary information for the correction of the network structure, after which the network effects the classification of the external situations according to the teacher's desire.

"Teacher" usually implies a human being or a neuron network which has its own "desire" or "goal" in its previously established connections. It will be noted here that a learning network may, as a result of learning, surpass its teacher in its acquired characteristics.

Let us assume that the input situation sequences are finite. Then the state of complete learning will be one, in which the probability of correct response reaches its maximum. It is clear that if the volume of memory contained in the interneuron connections exceeds a definite amount of information for the input finite sequences, the probability of recognition will always be equal to unity after the completion of learning.

Now let us imagine that the neuron network has not reached the state of complete learning, but has been trained up to a certain intermediate point, after which the teacher was eliminated (Figure 3), and the training functions began to be automatically fulfilled by the outputs of the undertrained system. Thus, the system with but little experience had recourse to a regime of self-correction or self-learning. It appears that the quality of self-learning depends on both the early training stages and the further sequence of external situations, and may lead to the deterioration of the characteristics acquired by the system, as well as to their amelioration. Depending on this, the procedure of ultimate learning or self-learning by the system may respectively be classed as self-destruction and self-perfection of the system.

Thus, the self-learning of the neuron network or of the learning device is nothing but the procedure of changing in the structure of the system on the basis of the previously received information.

Let us examine the learning procedure of the simpliest perceptron without cross-connections. Let us assume that an imput situation ensemble is finite and consists of representatives belonging to the class M (indicated by small circles in Figure 4), and of representatives belonging to the class N (indicated by triangles in Figure 4).

The connections established through the assistance of the teacher correspond to the dividing plane P (see Figure 4(a)). This plane is perpendicular to the line connecting the centres of gravity of the trained subsets of the representatives of different classes (points m and n, Figure 4), and it passes through its middle point. Now let us assume that the teacher was eliminated from the learning process, and the system was left to its own devices to complete its learning. According to the subsequent random sample, let us suppose that the system was presented with a picture indicated by the point 1 (see Figure 4(b)). Since the point 1 is recognized by the system as the class M, there will be some effect upon the coefficients of weight, so that a new dividing plane P′ will be formed. This will mean a deterioration of previously acquired abilities, for the points 2 and 5 which may further occur in the learning process shifted to areas non appropriate to them. This may be described as a step taken by the system towards self-destruction.

Let us assume now that the system was presented with a picture indicated by the point 2, and not by the point 1. In this case the dividing plane P″ is formed (see Figure 4(c)), and the points 3, 4, 5, and 6, that earlier were located in irrelevant areas, will now be in their appropriate places. This step corresponds to the improvement of the recognizing abilities of the system or, in other words, to the self-perfectionning.

It will be easy to show that the optimal strategy in the self-learning process is the strategy of a gradual transition from known generalizations to more remote ones. However, the self-learning process is characterized by a random nature of environmental situation sampling. Simple perceptrons

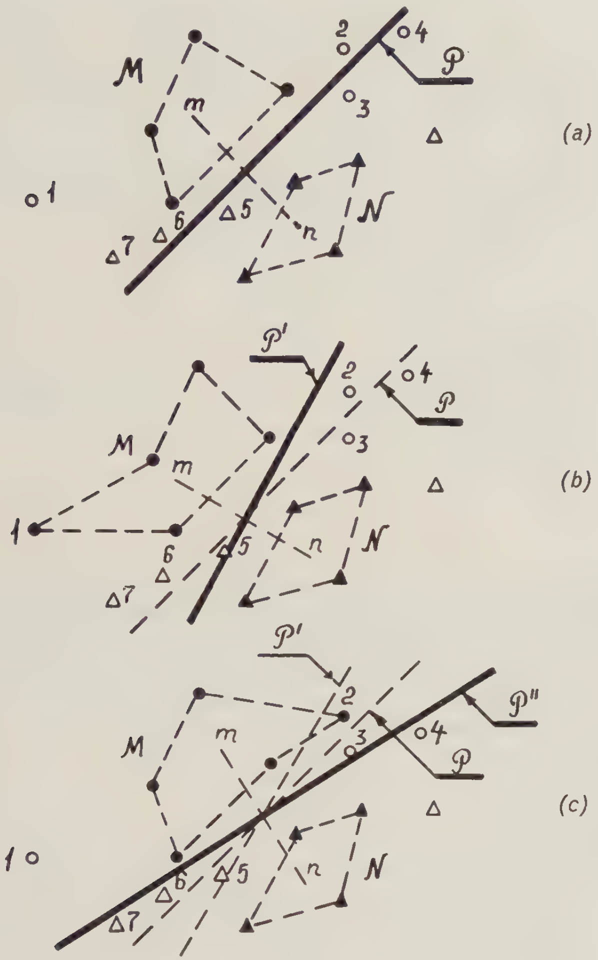


FIGURE 4 The pattern space

change their connections in the same degree under the influence of both "remote" and "close" pictures. Therefore, the quality of the self-learning process may equally deteriorate or improve, depending on the felicitousness of the random sampling.

Apparently, all proceeds differently in perceptrons with cross-connections. In systems of that type, during learning, when stable weighting coefficient values have not yet been achieved, an essential role is played by frequently recurring excitations of certain neuron elements whose connections form dynamic subsets of picture attraction. Therefore, a suddenly appearing "remote" picture is repelled by this attraction center, i.e. changes of its corresponding weighting coefficients do not occur. This shows that a cross-connected network is itself in possession of a strategy of cautious movement or gradual adaptation. The latter circumstance is a very important and interesting fact pointing to the self-perfectioning ability, peculiar to cross-connected systems.

Finally, the present writer would like to observe that cross-connected dynamic perceptrons are homeostatic learning systems which are capable of absorbing from the environment an enormous amount of information—even an amount that may exceed the information originally supplied to it—after which, to borrow N. Wiener's phrase, the machine acquires some of the significant characteristics of a living being.

References

1. F. Rosenblatt, *Principles of Neurodynamics*, Spartan Books, Washington, D.C. (1962).
2. W. Ross Ashby, *Design for a Brain*, Chapman and Hall, London (1960).
3. A. Kvitashvili, Nekotorie vozmoshnoste chetireghsloynovo pierseptrona s pierekrestnimi svyazami, Izvestiya AN SSSR, Teghnicheskay kibernetika, No. I, Moskva (1969).
4. F. Rosenblatt, *Perceptual Generalization over Transformation Groups*, *Self-Organizing Systems*, *Proc. Inter-disciplinary Conf.*, **5–6** *May* (1959).

CHAPTER II–2

Principle of the probabilistic ensembles in neuronal brain organization

A. B. KOGAN
Rostov State University, U.S.S.R.

Summary

The brain stands out against background of all the other biological and technical self-organizing systems by its ability to combine the plasticity with the stability of the function, which secures the wonderful reliability of their performance. Such ability is connected with the peculiarity of the neuronal organization of the highest brain parts. Investigations carried out in our laboratory have led us to the conclusion that the main peculiarity of the highest brain parts organization in vertebrate animals is the presence of neurons of the so-called screen structures which in the course of their activity get tied together in ensembles that interact as functional units. The participation of a given neuron in the ensemble reactions has a probabilistic character and the ensemble itself is a statistical system. The neuronal ensembles of the highest visual analyser parts in animals of different levels of development are described. The decrease of ensemble sizes and the increase of their numbers is in accordance with the analysis becoming more subtle.

INTRODUCTION

Everyone knows the wonderful properties of brain as an extremely reliable, plastic, effective, and economical information-handling system. As the result of the electrophysiological, micromorphological and cytochemical studies of the brain mechanisms we have supposed that such properties are set up by the peculiar principle of the neurons organization in the probabilistic ensembles.

The concept of the neurons ensemble as hypothetical self-organizing structure was used in many abstract models of brain activity by physiologists, psychologists and cyberneticians. In our laboratory it was shown that the ensembles are a specific elements of the neuronal organization of brain higher parts and some of their characteristics were determined.

NEURON ENSEMBLES

The morphological basis of the ensembles, apparently, are so called screen nervous structures. The functional properties of the ensembles have the following manifestations: (1) the synergic responses of compact neuron groups; (2) the balance of excitated and inhibited neuron populations; (3) certain dimensions and configuration of the ensembles; (4) the probabilistic character of individual neuron participation in the ensemble reactions and the "collective" responses of neurons; (5) the dependence of ensemble parameters upon the kind of stimuli and upon the functional state of the brain.

The Synergic Responses of Compact Neuron Groups

This is the most important sign of the ensemble organization. As it is known, in the peripherical parts of analysers among the neighbouring cells there are not synergic but antagonistic interrelations. This rule was clearly shown in the experiments with the retina cells of *Limulus* eye by Ratcliffe in 1961, and was called "lateral inhibition". In the evolution of brain structures the grouping of synergically responded neurons appears only on its high levels, having the screen design.

The typical interrelations of neighbouring neuron reactions in the nervous structures of different level of the central nervous system development are as follows. In the structure of a nodal segmental type, for example, in the subpharingeal ganglion of grape snail, the impulse responses of the neurons have a global synergic character. In the structure of a diffuse type, for example, in reticular formation of the frog medulla oblongata, the impulse responses even of the neighbouring neurons are not synergic but antagonistic. At last unlike to these types of reaction in the screen structure of the brain analysers, for example, in frog lobi optici, the impulse responses of neurons show the local synergism of the limited groups. Similar results were obtained from different analyser zones of the brain of rats, guinea pigs, rabbits, and cats.

Thus, local synergism of the neuron groups distinguishes the ensemble organization from global neurons synergism of nodal segmental organization and from individual neuron antagonism of the diffuse organization.

The Balance of Excitated and Inhibited Neuron Populations

To record simultaneously the responses not only of neighbouring but also of remote neurons in the highest parts of the brain, we require additional characteristics of their ensemble organization. These characteristics consist of the fact that beside each group of the excited neurons a group of the inhibited ones was found. The complex mosaic of the excitatory and inhibitory ensembles arises in this case. But such ensemble mosaic picture occurs only in the highest parts of the brain, having a screen structure.

The typical interrelations of the pair of the remote neuron reactions in the different nervous structure are the following: in the subpharingeal ganglion of grape snail the impulse responses of the remote neurons were synergic as among the neighbouring ones in this nucleus. When the second microelectrode goes out of the nucleus border, the neuron recorded by it more often does not participate in response. In the reticular formation the impulse responses of the remote neurons were more often antagonistic than among the neighbouring ones. Unlike to these types of the reactions, in the analyser centres of the frog, rat, guinea pig, rabbit, and cat brain the impulse responses of the remote neurons were most often antagonistic simultaneously with synergic responses of the neighbouring neurons.

With the help of a computer the coefficients of the correlation of the impulse streams of the neighbouring and remote neuron pairs were calculated. This calculation shows that only screen structure of the analysatory higher part of brain responded by means of the mosaic combination of the synergically excited and synergically inhibited groups of the neurons.

The relation of the general number of the excited to the inhibited neurons in the screen structure responses reveals a very interesting pecularity. It shows a wonderful stable proportion 2 : 1 for different animals (rats, guinea pigs, rabbits, cats), for different analyzers (visual, somatosensory), and even for the reactions which were known as the manifestation of the general excitation (epileptiformic activity) and the manifestation of the general inhibition (slow "sleep potentials"). These relations, perhaps, indicate one of the important regularities of the ensemble balance which consist of holding back each two excited neurons by one inhibited neuron.

Thus, neuron ensembles in screen structures enter among them in complex interrelation, forming the dynamic mosaics of the excited and inhibited neuron groups. However, in this mosaic there appears a general law of the balancing of excitatory and inhibitory neuron reactions in relation 2 : 1, being in very different form of nervous activity.

Certain Dimensions and Configuration of the Ensembles

The dimensions of the neuron ensembles were measured by moving apart the leading off electrodes until the synergism of the neuron responses turns into its antagonism. For example, such procedure in the lobus opticus of the frog brain shows the ensemble dimension to be nearly 200 microns. The same dimension has the zone of the stable antagonism, which coincide to the side by side located ensemble. But the other explanation of this data may be possible, if we assume that each ensemble contains the excited neurons in its centre and inhibited ones on its periphery, which save the central excitation from unrestricted irradiation. Then the diameter of the ensemble must be three times more than the above-mentioned one.

It appears that in the different screen structures of the brain the dimension and configuration of the ensembles may be greatly varied. For instance, in the frog cerebellum cortex the ensemble diameter reached to 500 microns. On the other hand, in the multi-layer cerebrum (big brain) cortex of mammals the ensembles sharply decrease up to 50–80 microns in diameter but they enter the lower layers in depth of several hundred microns. This structure increases the number of the ensembles on the square unit of the brain surface and it is perhaps, the evolutionary means for enhancing the analyzing effectiveness (permission ability).

The ensemble grouping of the cortical neurons revealed the tendency for its morphological consolidation. For example, the pyramidal cells in guinea pig visual cortex were united in groups each of these have 4–8 cells, placed mainly in vertical rows and joined by their apical dendrites in the general ascending bundle. These groups also revealed some cytochemical differences, probably connected with their various functional states.

Thus, the synergic neuron ensembles, acting as functional units, have very different dimensions and configuration. They varied from a flat disposition of the several tens of cells on the territory 200–500 microns in diameter in amphibian brain to the deepening disposition of a few cells on the territory 50–80 microns in diameter in mammalian brain.

The Probabilistic Character of the Individual Neuron Participation in the Ensemble Reactions and the "Collective" Responses of Neurons

The evolution of the nervous system leads to the progressive multiplication of the interneuronal connections. As a result of this process the participation of the each given neuron in the nervous centre reactions becomes less strongly determined and more probabilistic. This is the reason for the increasing of arrhythmia in the background impulse activity.

The distinction between the probabilistic responses and the strongly determined ones is shown, for example, in the reactions of the primitive snail ganglion and the high organized rat cortex. The repetition of the stimulations evoked in snail ganglion stereotypically reiterated responses. But the responses of the rat cortical neurons to the repeated stimuli were different every time, though the reactions of the whole neuronal population, manifested by electrocorticogram, were every time the same. Evidently, the redundancy of the parallel connections in the screen structures gives a wide choice of ways for the signal propagation. It is not excluded that here the principle of the Neumann "random switch" (1956) is realized, which provides the high reliability of functioning.

The comparision of the responses variability in the line of the nodal, diffuse, and ensemble organization shows regular enhancement of their stochastic character.

The randomness of the neuron responses in the probabilistic ensembles hinders the appearance of the individual neuron responses, because such variable responses become undistinctive in the background arrhythmic activity. In these cases the presence of the neuron responses may appear only in its "collective" reactions by the changing of their impulse stream correlation. For instance, two neurons in visual cortex of the guinea pig did not reveal the statistically reliable difference in the mean values of the impulse frequency, interimpulse intervals and in the form of their distribution histogram before and after light stimulation. But the statistically reliable increase of the correlation coefficient clearly revealed the presence of the response to the light.

Thus, the probabilistic participation of interchangeable neurons and the statistical character of the ensemble reactions create principally a new nervous mechanism, allowing to join the polyfunctionality to reliability in its properties. One of the arguments in favour of such properties is the "collective" mode of the responses.

The Dependence of the Ensembles Parameters upon the Kind of the Stimuli and upon the Functional State of the Brain

Dynamic neuron ensembles were generated in the different manner depending on the character of stimulation. For example, the average dimension of the ensemble in the frog lobus opticus evoked by adequate light stimulation of the eye was nearly 200 microns, but the electrical stimulation of the visual nerve generated ensembles up to 500 microns in size. However, the correlation of the impulse streams of the synergically responded neurons in the later case, i.e. with widening of the ensemble territory, was lowered. The disturbances in the normal brain state, for instance, as the result of the narcotics and convulsive drug actions widened, as a rule, the dimension of ensembles but weakened interneuronal correlative connections.

CONCLUSIONS

In conclusion, it must be admitted that in this field of investigation, there are more questions than answers. The first fundamental question is how the signals reaching in the nervous centre immediately organize the randomly discharged neurons in synergetically discharged ensembles. Some data about this problem give the very interesting observation that the synergetic or antagonistic type of neuron reactions may be revealed before the stimulation even in background activity, though statistically unreliable. Consequently, there is some primary organization existing even in the background, on which the working ensembles arise. The second fundamental question is what is a statistical law which governs the random switching on and off of the interchangeable neurons in the probabilistic ensembles. To answer these questions one has to learn the mystery of the wonderful properties of the brain!

CHAPTER II–3

A neurocybernetical contribution to the theory of agnosias

C. BĂLĂCEANU and M. BELIS
Bucarest University, Rumania

Summary

(1) The nervous system works as a data processing system; it disposes of patterns recognizers. Agnosias represents the clinical expression of their disfunction.

(2) Excitations from the surrounding medium are transmitted in a spatial–time code into the receptive areas of the cortex under the form of distribution of neurons in different degrees of excitation and inhibition. These distributions are the patterns which must be recognized.

(3) Analysing different forms of agnosias the authors conclude that to each cortical receptive area are linked special recognition devices. These must dispose of stored patterns classified after certain criterions.

(4) The recognition mechanism implies a succession of operations. The evidentiation of relevant characteristics of the cortical patterns: averaging, accentuation of contrasts and outlines (edging) their isolation, etc. The pattern normalization from the view point of dimension and position. Only after this preliminary stage (preprocessing) the fundamental (processing) stage takes place. The comparison of the relevant elements of the normalized patterns with those from the stocks of complex algorithms.

(5) The authors show how different clinical aspects of agnosias correspond to the alteration of one or more of the operations of the above-mentioned mechanism.

INTRODUCTION

The problem of agnosia is one of the main topics in clinical neurology (where it is implied in the topographic diagnoses of the cerebral lesions) and in neuropsychology (where it appears as a deterioration of the pattern recognition function). It follows that agnosia is a central problem in neurocybernetic and psychocybernetic researches.

It is known that the nervous system ensures the adaptation of living organisms to the changing conditions of their environment. This adaptation implies on the one hand detection of the external stimuli (which represent in fact input signals) and on the other hand selection of the optimal response (the output signal). This represents the classical behavioristic scheme stimulus-response ($S_i - R_j$) which expresses the functioning principle of the neurax, if we realize that the responses R_j are selected not at random but according to a certain strategy based on the finality imposed by motivation and affect.

The pairs $S_i - R_j$ must ensure the optimal activity of the adaptive system. This activity is much more efficient and economic regarding the number of neurons and the speed of operation if the whole stimuli set $\{S\}$ is separated in classes $\{C\}$ so that only the connections between classes and responses are to be recorded. These classes are determined partially by the genetical information but for the major part they are formed progressively by experience, through a learning process.

The establishment of the correct correspondence between the stimuli S_i and the classes C_j is the basic problem of pattern recognition. Deterioration of the mechanism through which the human brain performs this task leads to agnosia. This work is devoted to a cybernetic analysis of these mechanisms.

A GENERALIZED REPRESENTATION OF THE MESSAGE SOURCES

The biologic (and social) environment consists of objects and of their manifestations. Both are characterized by physical properties which represent the input signals of the biological system. It follows that the environment acts on the nervous system through an organized collection of signals. Such a collection is a structure or a pattern and its elements are the recognition indices (m_i). A physical source is then a collection of elementary messages or symbols called indices.

A physical source can be represented by a graph $G(M, \sigma_m)$ whose vertices represent the elements of the finite set of indices $M = \{m_1, m_2, \ldots, m_n\}$ and whose arcs represent the conditional probabilities $p(m_i/m_j)$ of the element m_i in the presence of the element m_j. As $p(m_i/m_j) \neq p(m_j/m_i)$ two

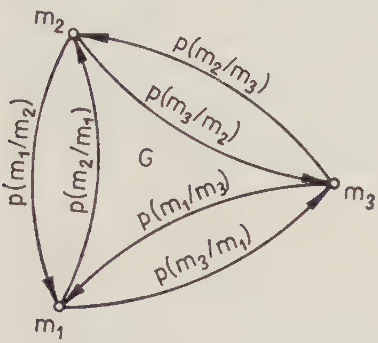


FIGURE 1 Physical source

vertices are connected by two arcs of opposite senses; the graph of a source is symmetric and strongly connected (Figure 1). Each graph has a corresponding matrix (Figure 2). The sum of the terms on a line (for ex. line *i*) represents the first order iterative power of the term m_i:

$$\pi(m_i) = p(m_i/m_1) + \cdots + p(m_i/m_n) \tag{1}$$

A hierarchy can be established between the indices of a pattern, according to their iterative power. If $\pi(m_i) = n$ the index m_i is a fundamental one and if $\pi(m_i) \geqslant \theta$ ($\theta < n$), m_i is a relevant one.

$$\left\| \begin{matrix} p(m_1/m_1) & p(m_1/m_2) \cdots & \cdots & p(m_1/m_n) \\ p(m_2/m_1) & & \vdots & \\ \vdots & & & \\ \vdots & \cdots\cdots & p(m_i/m_j) & \\ \vdots & & & \\ p(m_n/m_1) & \cdots & \cdots & p(m_n/m_n) \end{matrix} \right\|$$

FIGURE 2 Matrix of graph in Figure 1

TRANSPOSITION OF THE ENVIRONMENT INTO THE SYSTEM

The physical reality, which constitutes the environment is projected into the system via the sensory organs; these ones can be viewed as input transducers or encoders for the physical features of the pattern, or as non-linear filters having a sensitivity domain, a transfer function and a noise level. Because of the filtering effect of the sensory organs, only a part $1 < n$ from all the features is transferred into the neuronal network of the system. Detection of a source represents the mapping of the set $\{m_i\}$ of physical features into the set $\{m_i^*\}$ of neuronal changes.

$$\xi_1 (m_i) = m_i^* \tag{2}$$

The neuronal changes corresponding to the physical features detected are partially transposed in the memory of the system through another encoding–filtering process which reduces furthermore the number of features. Only a number $(k < 1 < m)$ of features are to be recorded by the memory elements of the neuron. These is a new mapping of the set $\{m_i^*\}$ of circulatory neuronal signals into the element set $\{c_i\}$ of the fixed memory.

$$\xi_2 (m_i) = c_i \tag{3}$$

The initial graph G is transposed first into the graph G' and finally this last one is transposed into the graph G'' which defines the engram pattern (Figure 3). G'' is also a symmetric, strongly connected graph, whose vertices correspond to neuron cells C (or groups of neurons) and whose arcs cor-

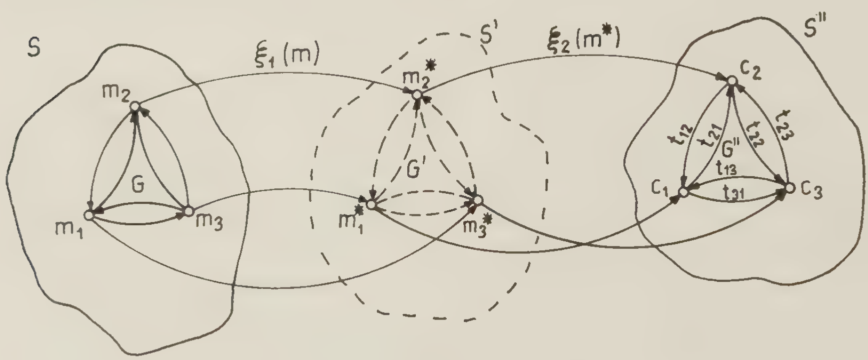


FIGURE 3 The engram pattern

respond to the thresholds "t_{ij}" of the neuronal connections $(c_i c_j)$. The set T of these arcs characterizes the memory of the system. The value of the threshold t_{ij} inversely varies with the probability of establishing a connection between the respective neurons c_i and c_j and $p(c_i/c_j) \neq p(c_j/c_i)$ implies $t_{ij} \neq t_{ji}$.

$$\begin{Vmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & & \vdots & \\ \vdots & \cdots & t_{ij} & \\ \vdots & & & \\ t_{n1} & \cdots & \cdots & t_{nn} \end{Vmatrix}$$

FIGURE 4 Matrix (for graph in Figure 3)

This graph too has a matrix associated (Figure 4) with it, where the different symbols are characterized by their iterative power. We notice that the perceptron is a model of this kind.

CLASSIFICATION THEORY

We have seen that to each exciting pattern S_i corresponds a neuronal pattern S_i' and an engram pattern S_i''. Each of these patterns is a collection of indices. These indices define a pluridimensional, abstract space in which the pattern represents a point. The classification process amounts to the division of the space into two or more decision regions with the aid of hyperplanes or boundaries. Each decision region corresponds to a class C_j. In the nervous system this division in classes is performed at the engram level. The graphs G'' are located in the decision regions of the neuraxial store. In the human brain the organisation of the classes is made easier by the existence of the language which provides code-words. This facilitates not only the formation of the classes during the learning process but also the use of the store (words-memory addresses). If $\mathscr{C}$ is the set of classes $\{c_1, c_2, \ldots, c_k\}$ we obviously have

$$c_i \in \mathscr{C}; \quad c_j \in \mathscr{C}; \quad c_i \cap c_j = \phi \tag{4}$$

where ϕ is the empty set.

PATTERN RECOGNITION THEORY

The recognition process consists in establishing to which of the classes C_j belongs a given pattern S_i. From a neurophysiological point of view the comparison is carried out at the neuronal message level. The pattern S_i transposed in the pattern S_i' is compared with the neuronal pattern S_i'' provided by the store (Figure 5). This implies a mechanism of extraction or activation of the patterns recorded in the long term memory in order to process

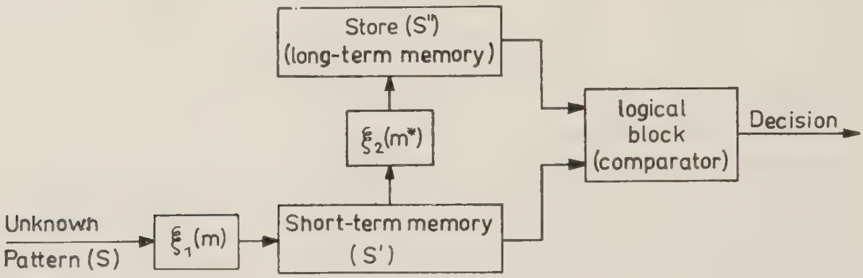


FIGURE 5 Pattern recognition

them in some way. This makes possible not only the recognition but also the evocation of the messages recorded (the representation) and generally speaking the activity of the imagination.

The relations (2) and (3), express the fact that the internal representation of a source differs from its real structure. The comparison is performed between two equally distorted (inner) representations so that the recognition is possible.

Pattern recognition is performed at different levels of the nervous system. Thus the retina is able of such performances (Young; Letvin and colleagues). However, the cerebral cortex is the most likely to carry out this process and agnosia is the clinical expression of lesions in this component of the nervous system. At the cortex level, recognition is performed by using stored information which is gathered epigenetically by learning, unlike the peripheric recognition (as the retinian), which is based on stores given genetically.

The recognition process consists in comparing the unknown pattern with the standard pattern of different classes. Each pattern can be represented by an n-dimensional vector and then the comparison is to be performed between the different features $\{s_1 s_2 \ldots s_n\}$ of the unknown pattern and the features of the standard one. This leads to the concept of intervectorial distance (Hamming, Sheppard, Gradley, Luce) well known in the psychology of pattern recognition.

Two criteria are taken into consideration.

(i) The pattern S_i belongs to the class C_j if the distance d_{ij} between the respective vectors is zero.

(ii) The pattern S_i belongs to the class C_j if the distance d_{ij} is less than a given threshold η ($d_{ij} \leqslant \eta$). This last criterion is used by the biologic systems.

When using this last criterion, the biological systems do not exhaustively compare all the vector components. Only the relevant clues are taken into consideration. If they lead to an ambiguous result the recognition is false (wrong). It is not a matter of agnosia but of a special kind of disorder met in frontal syndroms. Usually in case of ambiguity, the system performs a new verification of the messages received and takes into account additional features. This process can be repeated many times.

In cases of ambiguity the decision is often taken on a related basis, resorting to a likelihood function (Tanner). This function $L(S)$ is defined by the expression:

$$L(S) = \log \frac{P\,(C_j/S_i)}{P\,(C_k/S_i)} \qquad C_j \neq C_k \tag{5}$$

where $P\,(C_j/S_i)$ and $P\,(C_k/S_i)$ are conditional probabilities. They depend upon the genetic information, upon the own experience of each individual, the information available about the environment and the internal state of the nervous system (motivation, affective state).

The decision rule is

$$\begin{Bmatrix} C_j \\ C_k \end{Bmatrix} \quad \text{if} \quad L(S) \begin{Bmatrix} > \\ < \end{Bmatrix} \varDelta \tag{6}$$

where $\varDelta$ is a threshold or a cut point.

This rule leads to an approximation; it explains certain illusions known in psychology or in psychiatry. They are not manifestations of agnosia but erroneous recognitions due to incomplete information, great perturbations and particular internal states (as for example terror).

PREPROCESSING

The process of recognition is usually preceded by different auxiliary operations, neurophysiology has brought forward:

(a) *Contour intensification.* This process was put into evidence in the retina and the skin (Rashewski, Bălăceanu and Nicolau).

(b) *Pattern simplification.* Only the relevant features are used (recognition keys) (Bulloch). The algorithm for selecting these features can be damaged, which leads to pseudoagnosia (M. Botez).

(c) *Standardization of the detected patterns* implies certain transformations as augmentations, diminutions, displacements, rotations. The process is a complex one, based on a related syntax which explains certain phenomena as the dimensional invariance in the perspective (Helmholz, Herring, Nelson). Deterioration of this mechanism leads to distorsion in perception (usually paroxystic) as macropsia–micropsia and others, noticed in the pathology of the occipital lobe and in certain intoxications (ex mescaline).

(d) *Isolation* of the different patterns in the field of perception implies certain algorithms as for example the following of the closed contours. This process perhaps uses some interneuraxial mechanisms of complex sweeping as suggested by certain physiological illusions (moving postimages, Plateau's spiral, etc.).

(e) *Message extraction from the noise.* A message can be detected and processed if it exceeds the noise N:

$$S - N > \delta$$

where δ is a threshold (threshold theory; Bekesy, Stevens).

To improve the detection of messages in the noise the neurax uses two ways.

It increases the difference S–N by a selective amplification.

It performs a statistical average by an autocorrelation function (C. Collin, Mantel, Bălăceanu).

THE MECHANISMS OF THE UNISENSORIAL RECOGNITION

Theoretically the pattern recognition process can be carried out according to two principles: pattern recognition straight away; pattern recognition in successive stages.

Neuropsychological data prove that the neurax works according to the second principle. The distinctive features which define the pattern are recognized in successive stages. Finally these successive recognitions are combined according to an algorithm which leads to a propositional organization (Minsky, Ledley, Grimsdale and colleagues).

The mechanisms the nervous system uses at each stage are not always algorithmic but heuristic. These heuristics are responsible for the wide variety of different individuals. When put into evidence by different tests the heuristics of the cognitive function defines the personality.

Models of successive feature detection and recognition have been studied by Bledsoe and Browning, Uhr, Selfridge, and Rosenblatt. The neurophysiological test of this principle was carried out by Southerland, Young, Hubel and Wisel, Maturana and colleagues.

Neurophysiological and neuropsychological data show that the successive stages occur beginning with the periphery (ex. the retina for sight) and ending in the cortical areas. Likewise they show that the respective operations occur from the periphery toward the centre for each receiving system (hearing, sight etc.). The most complex solutions in the unisensorial recognition are to be found in the parasensorial areas (18, 19, 37, 39, 41, 42, Brodman). This accounts for unisensorial agnosia: visual, auditory, tactile, etc., (Henschen, Foerster, Kleist, Nielsen).

The successive feature recognition accounts for the existence of partial agnosias as for colours, for geometrical shapes, for music, for letters, numbers, theorems or words etc. (Nielsen, Hecaen and Ajuriaguerra). Besides their anatomo-clinical importance, these partial agnosias are relevant in the study of the neural mechanism of pattern recognition.

POLYSENSORIAL RECOGNITION

They recognition process does not stop at the unisensorial level but takes place also at the polysensorial level. Patterns provided by several types of receivers are taken into consideration. The intersensory relations are established for mammals at the cortical level. They have been proved anatomophysiologically (Sugar, French and Chusid). They account for the right $\rightleftarrows$ left transfer (Sperry) and the cross-modal associations (Hilgard and Marquis). These interrelations make also possible the recognition of patterns whose elements belong to several sensory domains.

The polysensorial recognition sometimes concerns the spatial features of the patterns. These features are estimated in the frame of a unique system of reference (the perceptual space). Actually the study of agnosia shows that this localization too is carried out in successive stages.

A localization is established in a spatial system characteristic of each sensory type. Afterwards a localization is carried out in each half of the perceptual space, which explains the agnosia or the ignorance of the left or right half space (Brain, Patterson and Zangwill, Deny Brown, Geschwind, Bălăceanu and Solomonovici). Finally, in the last stage, the localization is achieved in a unique perceptual euclidian space. This operation is carried out in the parieto-temporo-occipital region of the minor hemisphere. It is correlated with the motor behaviour (whence the clinical correlation between spatial agnosia and the different forms of apraxia).

The polysensorial recognition also performs the synthesis of the features coming from the same source but reaching the neurax by different channels (sight, touch, smell, etc.). This synthesis is carried out in the major hemisphere in the parieto-temporo-occipital region and is correlated with the language function (whence its clinical correlation with certain forms of aphasia).

MECHANISM USED BY NEURONAL NETWORK FOR RECOGNITION

We do not know anything about the mechanism used by the neuronal network for recognition. The simplest principle would be to compare successively any detected pattern with the standards of the classes recorded in the memory, until the most similar one is found. Regarding the great many classes and particularly the great many features to be handled, this would be an extremely cumbersome technique, which would take much more time for recognition than actually used by the nervous system.

The operation is made considerably easier owing to verbalization. Language is an instrument for finding the classes (memory addresses, codewords) and exploiting them. But in this case too we have to deal with a set of linguistic patterns with a great many syntagmathic and paradygmatic interrelations.

An useful algorithm could be a pattern identification from the general to the particular. This implies the organization of the neuraxial store in a multilayer system of classes. Such an algorithm would make possible recognition at a more and more general (abstract) level. The more general the level of a class, the less the number of features to be compared. A model of this kind could be the Steinbuch matrices, having three dimensions in the neurax.

CHAPTER II–4

A mathematical model of associative memory

VLADIMIR DROZEN
Faculty of Pedagogy, Hrodec Králové, Czechoslovakia

Summary

The site of associative memory is the layer of associative neurons in a diffuse neuronal net, described by the author in one of his previous papers. When several signals enter the associative neurons simultaneously, their respective synaptic weights become correlated and the diffuse images of elementary signals, which were mutually orthogonal in the beginning, group themselves into topological structures mapping the statistical structure of the outer world, from the point of view of the given system. In order to form associations between signals occuring not simultaneously but one after another, e.g. a typical conditional reflex, the neuronal net must be provided with a "short-time memory" which can be conceived as a system of delay lines enabling the system to transform the signals, originally consecutive, into simultaneous signals, resulting in changes in synaptic weights according to the above mechanism. The neuronal net is then able of mapping time patterns of signals from the environment, into the space of diffuse images, in a degree depending on the parameters of the short-time memory which can consist of closed feed-back loops (reverberation memory). Among other features, this model is able to extrapolate time series of signals on the basis of statistical prediction as well as single experiences.

INTRODUCTION

In a previous paper[1] I described a model of a synaptic memory of the static type, able to record associations between stimuli occuring simultaneously and thus to recognize static patterns. The present paper is devoted to a more general model, endowed with an integrated "short-time" memory (or "dynamic" memory) enabling the model to record and recognize time patterns also. The applicability of the model increases considerably by this method. Starting from the conditioned reflex in its classical form we can now model a wide variety of psychodynamic phenomena, where coding of time sequences and spatio-temporal patterns is involved.

STATIC MODEL OF A SYNAPTIC MEMORY

The static model mentioned above is based on two main assumptions.

(1) The memory has a diffuse (or "distributed") character: any individual stimulus influences all associative elements, or at least a whole group of them.

(2) The elements of the net work and change autonomously—no central controlling device is involved.

The simplest structure of this type has three layers of neurons.

(1) Input neurons X_i which transform various input stimuli into real numbers; to each "word" ${}^{\lambda}\bar{s}$ of the input alphabet corresponds a vector $\vec{x} = \{x_1, x_2, \dots x_i, \dots x_m\}$ as its image.

(2) The second layer consists of associative elements Y_i; each of these is connected to all input neurons by means of synaptic weights u_{ji} and its output is simply the sum of its weighted input values:

$$Y_i(t) = \sum_{i=1}^{m} u_{ji}(t) \cdot X_j(t)$$

Thus each input word ${}^{\lambda}\bar{s}$ has its "diffuse image" $\vec{y}$ which can be expressed as a column-matrix

$$[y_j] = [u_{ji}] \cdot [x_i]$$

(3) Analogously, the third layer consists of output elements Z_k; their respective outputs are

$$[Z_k] = [v_{kj}] \cdot [y_j]$$

In the beginning, the columns of the "centripetal" matrix $[u_{ji}]$ are orthogonal, or, more generally, uncorrelated. The model works on a discrete time scale.

The algorithm of synaptic memory is established thus. At the end of each elementary time interval each synaptic weight is enlarged by a value which depends directly on the intensity of the respective input signal, but also on the intensity of the output signal of the given neuron during the time interval just elapsed. The simplest way of meeting this requirement is to make the synaptic change proportional to both signals mentioned. Taking synaptic weights of associative neurons as an example, we may write

$$u_{ji}(t + 1) = u_{ji}(t) + q \cdot x_i(t) \cdot y_j(t)$$

where q is a proportionality factor.

If several input neurons of such a net are stimulated simultaneously, the result is that the sets of synaptic weights u_{ji} concerned, which had been previously statistically independent, become correlated. Then if we stimulate one of them, there appears in the space of "diffuse images" not only the image of the signal in question, but also, in a lesser intensity, the images of those input signals that had occurred in the past simultaneously with the given signal. With growing experience the memory works as a source of distortion: the system implies signals that are not actually present and applies the principle of incomplete induction.

INTEGRATED DYNAMIC MEMORY

As has been noticed in the introduction, this is a matter of a statistical mapping of the relations between exterior signals on the basis of their simultaneous occurrence, but any regularities in time series, time patterns or associations between successive signals are left unnoticed and unrecorded—the system cannot establish a conditioned reflex of the usual type or make predictions based on time extrapolation: it knows no past and future. These limitations can be surmounted by incorporating into the system a short-time (or dynamic) memory, not of course as a separate part of the system but as an integrated substructure of the whole.

The idea of the integrated dynamic memory is very simple. Working in terms of a discrete time scale, each input neuron X_i has its own delay line consisting of a series of "delayed input neurons" $X_{i\tau}$, the delay between any two neighbouring delayed neurons being the same as the magnitude of the elementary interval of the time scale. Each $x_{i\tau}$ is connected to all associative neurons y_j through its respective synaptic weight $u_{ji\tau}$ as illustrated by Figure 1.

Various types of delay lines are feasible: e.g. such a line can consist of a finite number of delay neurons without any decrement in the intensity of the delayed signals, or an infinite delay line with an exponential decrement,

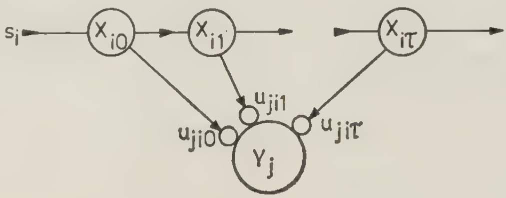


FIGURE 1 Delayed input neurons

or else a finite line with a constant arithmetic decrement. More generally, delay lines of different input signals may have different properties. Also there are some extremely interesting possibilities in applying delay lines with feedback connections. Unfortunately the extent of this paper makes it not possible to follow more closely how the filtering effects I described in an earlier paper[2] may bear upon the behaviour of neuronal nets.

In the simplest case mentioned (finite number of delay neurons with zero decrement) we have the output of an associative neuron

$$y_j(t) = \sum_{\tau=0}^{r} \sum_{i=1}^{m} u_{jir}(t) \cdot x_i\,(t - r)$$

Let us remember that each input delay line is a time analogy of the spatial case of a linear retina (see Drozen[3]). The occurrence of signals with duration of several time intervals is expressed by analogous formulae as the case of an abscissa moving along the linear retina. We may thus speak of the "time topology" of the environment, mapped on the space of diffuse images and depending on the properties of the delay lines in question as well as on the properties of the prevailing input signals.

For illustration of the basic idea let us study the origin of a simplest conditioned reflex based on the association resulting from a successive occurrence of the conditioned and unconditioned stimulus, without any central reinforcement. Let us have an "inexperienced" net where the columns of the three-dimensional matrix $[u_{jir}]$ are mutually uncorrelated, with the short-time memory of the length T and zero decrement. The unconditioned unit signal ${}^a\bar{s}$ consisting of only one unit component ${}^\delta a$ evokes an undelayed output signal ${}^a\bar{z}$ which has only one non-zero component $z_k = 1$. For simplicity's sake let us leave the matrix $[v_{kj}]$ of centrifugal synaptic weights unchanged.

Then

$$\sum_{j=1}^{n} v_{kj} \cdot u_{jio}(0) = 1 \quad \text{if} \quad i = a$$

$$\sum_{j=1}^{n} v_{kj} \cdot u_{jio}(0) = 0 \quad \text{otherwise}$$

This simply means that all input stimuli with zero δ_a component evoke no z_k response.

Let us issue an ineffective stimulus ${}^b\bar{\delta}$ consisting of only one unit component ${}^\delta b$ at the moment $t = 0$ and let it be followed by the unconditioned stimulus ${}^a\bar{\delta}$ at the moment R. For $t < R$ we have

$$y_j(t) = u_{jbt}(0)$$

but if $t \geqq R$ then

$$y_j(t) = u_{jbt}(0) + u_{ja,t-R}(0)$$

According to the algorithm of synaptic memory mentioned above, at the end of a complete cycle the new synaptic weights are

$$u_{jbr}\,(t > T) = (1 + q) \cdot u_{jbr}(0) \quad \text{if} \quad r < A$$

$$u_{jbr}\,(t > T) = (1 + q) \cdot u_{jbr}(0) + q \cdot u_{ja,r-R}(0) \quad \text{if} \quad r \geqq R.$$

Further

$$u_{jar}\,(t > T) = (1 + q) \cdot u_{jar}(0) + q \cdot u_{jb,r+R}(0)$$

We see that associative relations are now established resulting in the origin of a conditioned reflex. Now let us consider the response to the conditioned unit stimulus $b_{\bar{\delta}}$ occurring at the moment $t = 0$.

For $t < R$ we have

$$y_j(t) = (1 + q) \cdot u_{jbr}(0)$$

and the output signal

$${}^b z_k(t) = \sum_{j=1}^{n} v_{kj} \cdot (1 + q) \cdot u_{jbr}(0) = 0$$

If $t = R$, then

$$y_j(t) = (1 + q) \cdot u_{jbr}(0) = q \cdot u_{ja0}(0)$$

and

$$^{b}z_k(t) = q \cdot \sum_{j}^{n} v_{kj} \cdot u_{ja0}(0) = q$$

If $t > R$, then of course the output is again zero.

Now the system reacts upon conditioned signal $^{b}\bar{\delta}$ and the intensity of the reaction grows with the number of repetitions of the given time patterns. In our simple case, the time lag between the conditioned signal and the reaction is exactly R. The situation is different with an "experienced" neuronal net that had been exposed previously to signals of non-elementary duration and thus established a "time topology" in the synaptic weights of the input delay lines in question. In such a case which we shall not consider here in detail the response would be prolonged, starting earlier and ending later, with a maximum at the moment R.

One unrealistic feature of the whole scheme is the unlimited exponential growth of synaptic weights. This can be amended easily but the mathematical formulation becomes then more involved. Another point worth considering would be introducing a threshold effect. And of course the basic structure discussed her might serve as a building material for some higher whole, consisting of more layers or several specialized diffuse nets more or less independent.

CONCLUSIONS

Modelling of neuronal nets has for its main aim to fill the gap between the neurophysiological and the psychological approach in studying living beings and the information processing involved in their behaviour. Yet I believe that some of the results may be useful as a way of approach to the problems of artificial intelligence. Surely the approach by heuristic programming is much more effective as far as nontrivial special problem groups are concerned; but perhaps the study of neuronal structures ought not to be underestimated as it covers the interaction between the system and its environment in all its universality, and even may show some new aspects of the environment itself.

References

1. V. Drozen, *A Mathematical Model of Synaptic Memory* (in Czech.), Hradec Králové, Kulturní dum ROH (1969).
2. V. Drozen, "The integration of a periodical signal as a filtration process", *Slaboproudý Obzor*, **17**, 444 (1956).
3. V. Drozen, "Mapping of topological relations in diffuse neuronal nets", in *Information Processing* 1968, *Proc. IFYP 1968 Congr.* North-Holland, Amsterdam (1969).

A model of inter-nervous system information transmission*

J. ZABARA
Temple University, Philadelphia, U.S.A.

Summary

The model is restricted to information transmission in a universe of two nervous systems. The primary unit of information transmission is called a communication (C). The primitive undefined term relating to a communication is called a pulse (p). A pulse is related to an action potential as it is observed in neurophysiological experiments. The properties of the action potential called all-or-none and conduction are incorporated as pulse characteristics. The all-or-none property indicates a two-state logic for the pulse conveniently described as 0 and 1. A communication is considered to be a class of pulses, which is subject to an ordered relation. Three axioms can be derived based upon the definition of communication, but it must be emphasized that the derivations are not rigorous.

INTRODUCTION

Although information transmission has been a serious concern of both biological-medical and physical–engineering sciences, it has not been readily appreciated that the emphasis is placed very differently in these two areas[1–5]. Contemporary concepts of information transmission were developed primarily for the physical sciences and secondarily applied to the organism[1–2]. The purpose of this report is to present a model of information transmission which is based directly on the nervous system.

* This work was supported in part by a Grant-In-Aid from Temple University.

THE REFLEX

It is about three centuries since Descartes first presented the reflex, which continues to be the primary model of the nervous system. Though the reflex has been amplified with new corollaries and primitive undefined terms, the basic axiom remains unchanged. This axiom can be stated in logical terms as "if a stimulus, then a response". The axiom is based on Descartes' comparison of the organism with an automaton. The corollaries, which are directly based upon experimental observation, include (1) the Bell–Magendie Law of afferent and efferent nerves, (2) the neuron theory of the nerve cell, (3) excitation and inhibition at the synapse—Sherrington's integrative action of the nervous system, and (4) the nerve potential with the Hodgkin–Huxley theory of sodium and potassium equilibrium potentials.

Stimulus and response are treated generally as primitive undefined terms. In addition, a response occurs only as a consequence of a stimulus. The first step toward forming a new model of the nervous system was to consider a response which occurs without a previous stimulus. A response which occurs without a stimulus can be said to be initiated by an autorhythmic event[6]. Autorhythmicity is considered to be related to activity recorded from synaptic regions of the nervous system. Examples of such activity are the electroencephalogram and synaptic potentials of invertebrate ganglia of *Aplysia* and *Dytiscus marginalis* and in mammalian ganglia infected with a suitable virus.

THE PULSATION MODEL

The "corollaries" of the reflex model are a consequence of observations spanning several hundred years. However, these statements are almost impossible to incorporate into a self-consistent formal model, except as a series of primitive undefined terms. But to test them as a series of primitive undefined terms leads to a nonsophisticated model which apparently is all too common in biology. The procedure here was to develop a single undefined term based upon one of these corollaries: the action potential[7]. This procedure might be called an "idealization" of the actual experimental observations such as occurred for instance in the development of the Newtonian axioms of motion. In the present study, the action potential is considered the starting point since environmental stimuli do not act directly upon the central nervous system but activate receptors. Thus, the central nervous system receives information in terms of action potentials and determines responses also in terms of action potentials. However, the term potential may be misleading in this context. Potential is a defined term in physics, and there has been an intensive effort by investigators of the nervous system to erect a physical picture of nerve potentials in terms of charge, thereby establishing a definition of potential identical to that in physics. But, such a definition does not appear to be useful at the present time for a model of nervous system dynamics. In order to differentiate the action potential from the physics definition, the term "pulse" was introduced to represent the two properties of the action potential of conduction and all-or-none[7].

A set of pulses is called a pulsation. The terms stimulus pulsation and response pulsation were considered to relate to specific stimuli and responses. It was possible to develop this conception further by considering the set of all pulses, whether they be related to stimulus or response. The set of all pulses was called the universal set of pulses which are considered as having relations or operations[7].

The three operations on the universal set of pulses are focusing, geometry and order. Focusing (F) is a mapping between pulsations. The reflex model represents a one-to-one mapping from a stimulus pulsation to a response pulsation. Focusing was assumed to be mediated at a point termed the focal point. A set of lines intersecting at the focal point represents the pulsation (P). In the reflex model, the focal point can be considered to be the synapse. Order is represented by $<$ or $>$. Thus, if $P = [P_1, P_2, \ldots, P_n]$, then $P_1 < P_2 < P_3 < P_4 < \ldots < P_n$ is read P_1 precedes P_2, etc.

Based upon these definitions and operations, the following axiom and corollaries were developed.

Axiom: a pulsation conducts in a circle unless it is intersected by another pulsation.

Corollary 1: a pulsation can be intersected only at its focal point.

Corollary 2: a pulsation conducted in a circle is a periodic function.

THE UNIVERSE OF TWO NERVOUS SYSTEMS

A model of information transmission based upon the above axiom is to be developed for a universe of two nervous systems. The two organisms related to each nervous system are called communicators. It is possible to consider three states of the universe called monologue, dialogue, and no communication (the zero state). These three states are considered for the purposes of clarity and are not to be interpreted as final components of the model.

The dialogue is perhaps the simplest state to describe. This is the state which is viewed generally as mutual communication between two communicators. A communication of the part of one communicator is followed in some time interval by a communication by the other communicator, and this process is repeated. Therefore, the dialogue has a definite rhythmicity consisting of periods for the communications of each communicator which are generally unequal. Thus, there is usually no periodicity. However, it is possible to consider the "perfect" communication which precedes according to "perfect" periodicity.

The monologue is the state in which one communicator is the "listener". The "listener" commonly is thought of as being inactive and simply recording communications. The reason for this view is that the observer of such a universe is fixed on the "dominant" communicator. But this is deceptive, because the "listener" frequently exhibits non-obvious communications. In other words, the communications of the "listener" are not generally noticed by an observer or "consciously" perceived by the "dominant" communicator.

The non-communication state occurs only when one or both of the nervous systems are either nonviable or unstable.

Thus, with these general conceptions of a communication universe of two nervous systems to point the way, we shall be able to investigate a communication universe in terms of pulsation and autorhythmicity.

Definition of Communication

P_1 corresponds to communicator 1 (Cl) and P_2 corresponds to communicator 2 (C2), such that a communication (C) is defined as the set consisting of P_1 and P_2:

$$C = [P_1, P_2]$$

$$P_1 = (P_1^1, P_1^2, P_1^3, \ldots, P_1^n)$$

$$P_2 = (P_2^1, P_2^2, P_2^3, \ldots, P_2^n)$$

Three states of communication have been considered: monologue, dialogue, and no communication. In the monologue, P_1 "drives" P_2 where Cl is the "dominant" communicator and C2 is the "listener" or "passive" communicator. In the dialogue, there is a reciprocal relationship between C1 and C2, such that P_1 is alternately "driving" and being "driven" by P_2. The state of no communication denotes the absence of either P_1 or P_2.

No communication $= (P_1 \wedge \sim P_2) \vee (P_2 \wedge \sim P_1)$

Monologue and Dialogue

For purposes of conceptional convenience, we shall continue to approach an analysis of communication in terms of monologue and dialogue. In the monologue state, P_1 is driving P_2 so that P_1 and P_2 are synchronized. P_1 consists of a packet of pulsations to which C2 is responding with a packet of pulsations P_2. If a verbal communication only is considered, then P_1 can be related directly to words or sounds by C1. But with C2 being silent, what can be the response denoted by P_2? Obviously, the response by C2 cannot be considered as verbal, but must be considered a non-verbal response. It is claimed that C2 has a packet of P_2 consisting of a non-verbal response to P_1. Thus, P_2 gives "cues" which C1 may or may not notice. Nevertheless,

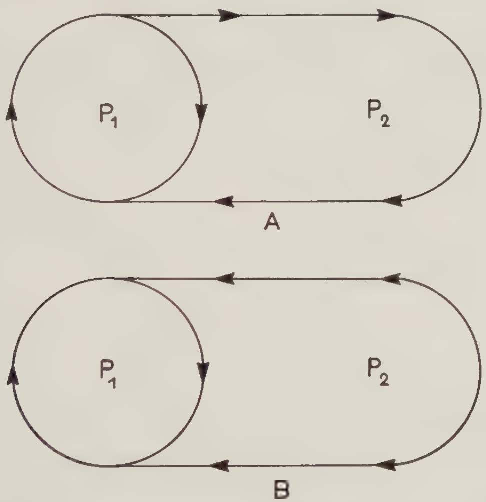


FIGURE 1 The pulsations of the communicators are represented by P_1 and P_2 such that P_1 is driving P_2.

A. P_2 is synchronous with P_1 as shown by the clockwise orientation of the arrows.

B. P_2 is bucking P_1 as shown by a combination of clockwise and anticlockwise arrows it is an inherent part of the communication that C2 respond with a P_2. If there is no P_2, there is no communication.

Again, for the purpose of illustration, let us consider the dialogue in the special case of verbal communication. It is considered usually that C1 makes statements, after which C2 responds with statements, and so on. However, it is claimed in this report that while C1 is making statements, there is a response in C2. This non-verbal response can take the form of facial muscle contractions, limb movements, changes in posture, autonomic changes, etc. During this time, P_1 is "driving" P_2; but then, as C2 begins to speak, P_2 is "driving" P_1. In the usual conception of verbal communication, while C1 is speaking, auditory impulses are arriving at the auditory cortex. Based upon this information resident now in the auditory cortex, C2 begins to speak through efferent nerves innervating the muscles of vocalization. For the purpose of developing a consistent model of communication, this picture is not sufficient.

In this report a communication is developed as a series of simultaneous pulsations, with alternate pulsation-driving (Figure 1). While C1 is speaking, a set of pulsations occurs in C2. While C2 is silent, the C2 pulsations are non-verbal. These non-verbal pulsations in C2 can be considered to correspond to contractions of facial muscles of expression, change in heart contraction rate, sweating, salivary secretion, change in tonus of striated muscle, etc. Thus, while C2 is silent, there occurs, nevertheless, effector activity corresponding to the statements of C1.

The next step in the dialogue is that C2 begins to speak. The pulsations in C2 now are a function not only of the previous statements of C1 but also of the preceding pulsations in C2.

$$P_2 = f_1(P_1) + f_2(P_2)$$

With these general concepts as a background, we are able to proceed to a discussion of operations.

Operations of the Communication Universe

The first operation based upon the definition of communication is a mapping G between P_2 and P_1.

$$P_2 = GP_1$$

where

$$P_1 = (P_1^1, P_1^2, P_1^3, \ldots, P_1^n)$$

$$P_2 = (P_2^1, P_2^2, P_2^3, \ldots, P_2^n)$$

$$G = (g_1, g_2, g_3, \ldots, g_n)$$

For example

$$P_2^1 = g_1 P_1^1$$

The second operation based upon the definition is a mapping F of P_2 or P_1 into itself:

$$F = (f_1, f_2, f_3, \ldots, f_n)$$

For example

$$P_2^2 = f_1 P_2^1; \quad P_2^1 = f_1 P_1^1$$

The mapping F can be viewed as a memory. It is assumed that the combination of G and F, $G + F$, is linear, for example

$$P_2^2 = g_1 P_1^2 + f_1 P_2^1$$

This means that a specific pulsation element of P_2 corresponds to a "simultaneously" occurring pulsation element of P_1 and a previous pulsation element of P_2.

Order in Communication

It is now possible to consider the significance of the definition of a communication in relation to the order operation. Again, we proceed on the basis of metalanguage statements. Elements of P_1 and P_2 can be considered to occur simultaneously with "driving". For instance, when P_1 is driving, elements of P_2 occur simultaneously with those of P_1. Suppose, when P_1 is driving, that there is a lag or lead of P_2 in respect to P_1. What does such a lag or lead mean in terms of the communication? It is assumed that as the lag or lead increases there is a change in the nature of the communication. For instance, if P_2 leads P_1 by 90° where P_2 is driving, then it is assumed that C2 is anticipating C1. If P_1 and P_2 are 180° out of phase, then the communication may become "defective" or "unstable". Thus, an unstable communication can occur when P_1 and P_2 are bucking each other, i.e. the rotation of P_1 is in a negative direction to that of P_2.

Therefore, there are two limiting conditions for a communication, that of potential instability and that of synchrony. Synchrony is where P_1 is simultaneous with P_2 or there is a constant phase difference other than 180°. This discussion of lag and lead in communication has a rigorous basis in the concept of order of a pulsation which has been presented in the description of the pulsation model.

Frequency and Driving in a Communication

The next concept to be investigated in relation to communication is that of frequency. Let us consider the special case where G is a one-to-one mapping from P_1 to P_2. For each $P_1^a \in P_1$ there exists a $P_2^a \in P_2$ where $a = 1, 2, 3, \ldots, n$. If $P_2^a \notin P_2$ for $P_1 \in P_1$, then there is no communication. Therefore, it may be considered that for all states, except no communication, the period of P_1^a must correspond to the period of P_2^a. If $\forall P_2^a \in P_2$ occurs with the same period, then the frequency of P_1 is the same as the frequency of P_2.

THE FIRST AXIOM OF COMMUNICATION

In the foregoing discussion of communication, the special condition of the communication has been considered where the period of every P_1^a is identical and also the same as the period of every P_2^a. Thus, the period of every pulsation in P_1 and P_2 has the same value. The first axiom is based upon this conception:

A communication proceeds with a periodic frequency unless

there occur states of no communication.

The concepts of period and frequency are based directly upon the order relation as presented in the pulsation model. For instance, $P_1^a > P_2^a$ indicates that P_1^a is leading P_2^a, or $P_1^a < P_2^a$ indicates that P_1^a is lagging P_2^a. The emphasis in this model is upon the form, rather than the content, of the communication. In other words, this report represents an attempt to state conditions for information transmission which is independent of the nature of the information.

A COMMUNICATION MEASURE SPACE

It is possible to consider a communication space upon which a measure can be placed. This measure should be considered as a single suggestion among many possibilities. The primary intent of the author is to develop an information or communication space which is directly based upon the nervous system. The measure on the communication space, $\|C\|$, is defined as

$$\|C_a\| = 1 - \|P_1^a - P_2^a\|$$

where $a = 1, 2, 3, \ldots, n$ and $P_1^a \epsilon P_2^a < P_2^a \varepsilon P_2$.

For example, when the lag or lead between P_1^a and P_2^a is zero,

$$\|C_a\| = 1 - \|0\|$$
$$= 1$$

Thus, $\|C_a\| = 1$ demonstrates that the pulsations of the two nervous systems are simultaneous. Further, $\|C_a\| = 1$ is a maximum communication in the measure space.

Let us now investigate the measure space for various values. We will continue to utilize the Fourier type terminology of phase angle and lag or lead for illustrative purposes. When P_1^a and P_2^a have a 90° phase angle difference, let us assume the following value:

$$\|P_1^a - P_2^a\| = \tfrac{1}{2}$$
$$\therefore \|C_a\| = 1 - \tfrac{1}{2}$$
$$= \tfrac{1}{2}$$

Thus, this measure demonstrates that the communication is one-half that of maximum communication. Finally, let us assume that when P_1^a and P_2^a have a 180° phase angle difference, that this corresponds to the following value:

$$\|P_1^a - P_2^a\| = 1$$
$$\therefore \|C_a\| = 1 - 1$$
$$= 0$$

The value of zero for $\|C_a\|$ corresponds to a minimum communication for the two nervous systems.

Thus, there is defined a measure space for a communication which has a minimum value of zero and a maximum value of one:

$$0 \leq \|C_a\| \leq 1$$

This measure space is interpreted in relation to Fourier wave analysis as depending on the lag or lead of P_1^a with respect to P_2^a.

References

1. H.B.Barlow, "The information capacity of nervous transmission", *Kybernetik*, **2**, 1 (1963).
2. J.D.Cowan, "The engineering approach to the problem of biological integration", in *Nerve, Brain and Memory Models* (Eds. N.Wiener and J.P.Schade), Elsevier, New York (1963).
3. D.M.MacKay, "Cerebral organization and the conscious control of action", in *Brain and Conscious Experience* (Ed. J.C.Eccles), Springer Verlag, New York (1966).
4. H.Marko, "Die theorie der bidirektionalen kommunikation und ihre anwendung auf die nachrichtenübermittlung zwischen Menschen", *Kybernetik*, **3**, 128–36 (1966).
5. C.E.Shannon and W.Weaver, *The Mathematical Theory of Communication*. University of Illinois Press, Springfield, Ill. (1949).
6. J.Zabara, "Reflex and autorhythmicity: a formal model", *Math. Biosci.*, **4**, 33–8 (1969).
7. J.Zabara, "Axiomatics for a nervous system", *Math. Biosci.* **5**, 419–426 (1969).

CHAPTER II–6

The brain and its information trapping device

C. R. LEGÉNDY

Westinghouse Advanced Technology Laboratory
Baltimore, Maryland, U.S.A.

Summary

A possible scheme of the capture and organization of information within the living brain is outlined, and arguments are made in support of the various details of the scheme, some weaker, some stronger. An effort is made to understand how the neurons can organize themselves, instead of being organized, and how they can transmit syntactic information among themselves.

INTRODUCTION

It is often assumed that neurons are basically simple and well approximated by a switch of some sort. I claim, and shall line up some arguments behind the claim, that the opposite approximation is closer to the truth. The neurons are not "ideally stupid" but "ideally smart".

That a neuron can do much (rather than only little) data processing is certainly not beyond possibility. After all, there exist whole animals that are no bigger and no more complex than the neurons, and they must even hunt for their food. It is not unbelievable that if a whole cell is specialized to the task of information processing, it could be ideally good at it.

The more worrysome feature of the assumption is that the gross brain circuitry, already too complicated for treatment when the neuron is a switch, will probably be a logically inpenetrable dark jungle when the neuron is more than a switch.

But, as it turns out, this is not so. The complexity of the neurons (as assumed here) is just such that treatment of large numbers of neurons is simple; in fact much simpler than in the other models.

The language of neurons is no more expressive than that of animals, for instance; at least so I shall assume. Animals can make a few characteristic sounds; but it is pointless to look for any code or system beyond that. For all we know, their feelings and imagery are as rich as our own, but their speech can do them no justice; it can only give some vague expression of agitation, or the like. Similarly muted are the neurons, by what could be called "the axonal bottleneck" of information flow.

The input arriving to neurons is not as information-poor as the output. On the input end neurons are more open than on the output end; they can receive information at a rate about 10^3 times higher than they can give it off. But information and knowledge are not the same thing. The input to neurons may be directly interpretable into events or facts of the outside world, but the neuron has no way of finding out what input signifies what event or fact. It has no way of finding that out; nor is it conceivable that it carries, from birth, a "codebook" of sorts, containing the interpretations. The reason is that the brain has no way of foretelling what information would arrive to it during lifetime. (How could any innate codebook foretell, for instance, the firing pattern that would signify the spoken name of Julius Caesar?) All a neuron can do is "sense" it that an input pattern signifies "something".

It will be shown below that this can be done, and that the neurons can use the input patterns in question to coordinate their activity with the activity of other neurons, including ones with which they do not have synaptic contact. They can do so, by use of probability theory.

Changing the subject for a moment, let me predict from the very start that the reader will not believe the arguments of the sequel unless he thoroughly believes that probability theory works, and his belief is not swerved in the least by the fact that people have no idea why it works. The validity of probability theory is of course impressively borne out by experience; so much so that few people even concede that there is cause for surprise. But there is, as is dramatically demonstrated by the success of computer-generated random numbers in simulating the unpredictability of Nature.

These random numbers are usually generated by very simple schemes. For instance, a number is taken as a starting point, then it is multiplied by some number A, then truncated from front and back so as to be reduced to its original number of digits; then it is multiplied again by the quantity A, then truncated again, then multiplied again, and so on. The sequence of truncated numbers is the "random" sequence.

The sequence is certainly not random in the usual sense of being unpredictable, because, once the secret is known, each number uniquely determines the next. Yet it yields Gaussians as a random sequence should, and has in the past served as basis for countless simulation experiments, yielding industrial designs with highly satisfactory results.

I say all these because in the sequel a number of very non-random-looking things will be assumed to be random. (By the way, mathematicians regard the problem of randomness a very real problem, very much susceptible of rigorous solution, and very much unsolved. The only progress made sofar in attempts to prove the ergodic theorem is a recent 400-page proof, for the case of a gas of N rigid elastic spheres enclosed in a rigid elastic box.)

That neurons are capable of "learning the language" of any realm into which fate chooses to throw them is inferred from various lines of evidence, each suggesting that neurons do not always know beforehand the neurons with which they are destined to be in synaptic contact.

Altman's (1967) autoradiographic data shows that most cerebellar granule cells arise after birth. They begin their life no differently from vegetable seeds. As was already known to Cajal (1909), when they start from the external granular layer, they are only a cell body with no branches. Then they migrate down, send out their (arrow-straight) parallel fibers, migrate further, send out more fibers, make contacts and go to work. It is quite reasonable to assume that such neurons do not know beforehand what job would be assigned to them.

One receives the same impression, and even more strongly, from the relative straightness of most axons and dendrites and the fact that they have synapses at every micron or two over their entire length, rather than only on their endpoints. Some axons, like the thalamic afferents to striate cortex, have a zigzagging complicated arborage. It is easy to believe of these that when they first invaded the cortex during embryonic development, they carried with them, as an errand man in a briefcase, a list of hundreds of neurons they were supposed to contact. They then branched and zigzagged in search of these, and knew no peace until they contacted at least a sizeable fraction of them. But about the rest of the neurons it is hard to believe that more than 10% of their synapses are planned contacts. Most of their synapses seem more like "contacts of convenience", made only because the other fiber happened to be nearby and it would have been a shame not to make contact with it.

The enormous number of such "contacts of convenience" again allows us to infer that neurons must possess an ability to figure out what is going on around them and make themselves useful in an environment originally unfamiliar to them.

INFORMATION SENSITITY OF LIVING NEURONS

The Principal Assumption

Neurons of the living brain are not very reliable. Anybody who has any doubts in the matter need only to look at Figure 1, which shows 12 traces of firing by the same neuron, in response to the same stimulus, and on which, of course, no two traces can very well be called identical.

It is reasonable to assume, however, that the few hundred or so neurons

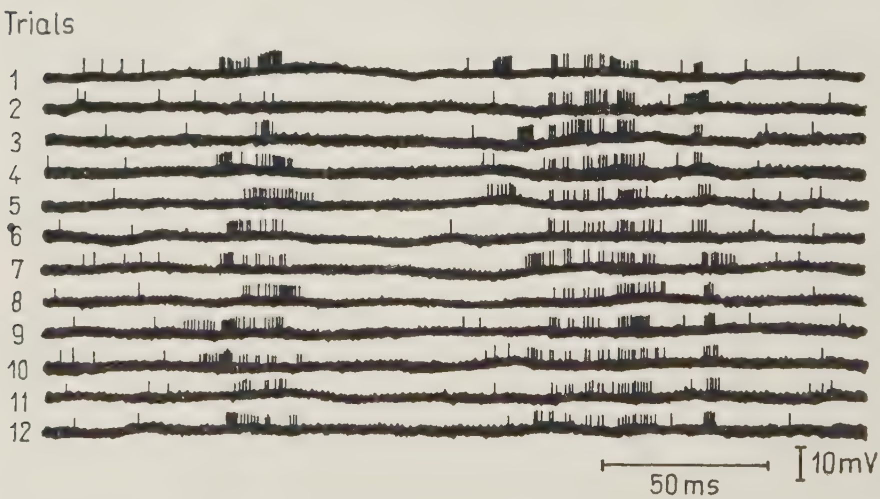


FIGURE 1 Responses of a neuron of the visual cortex to repeated application of a visual stimulus. The stimulus was a dark corner in the right upper quadrant of the visual field. The neuron was in Brodmann's Area 19; the experimental animal was a cat: (Reproduced from *The Neurosciences*, Eds.: Quarton, Melnechuk and Schmitt, The Rockefeller University Press, New York, 1967.)

that listen to the firing of the one of Figure 1 can still recognize the twelve firing patterns as being "more or less" the same thing. How do they do it? What do they look for?

I shall assume that neurons always assume that anything coming to them is random, and are waiting to be convinced of the opposite. In the above traces they can be convinced of that quite easily.

Suppose their initial hypothesis is that the neuron fires at an average rate of 100 firings/second, for such is the rate they find by long-range averaging. Under this hypothesis, we ask, how often will a firing pattern like Trace 5 occur? For easy reference, let us call it "Pattern A" and define it as follows: 20 firings must occur in less than 25 milliseconds, and then after less than 0.3 seconds another burst of 26 firings must enter in less than 50 milliseconds. We assume exponential arrival statistics. (It turns out to make very little difference what sort of statistics we assume; they all lead to the same conclusions.) This means that after any given firing impulse the probability that no other firing will arrive for t seconds is $e^{-\mu t}$, where $\mu = 100\ \mathrm{s}^{-1}$. It can be shown that under this assumption the probability that exactly n firings will arrive in T seconds is

$$P = \frac{(\mu T)^n}{n!}\,\mathrm{e}^{-\mu T}$$

Using this formula, we find that Pattern A is expected to occur about once in every 10^{12} years. (10^{10} years is the age of the Universe!)

Now let us retrace our thoughts for a moment. We assumed that firing arrived at random and at a certain average rate. This assumption allows bunching of the impulses, and assigns various probabilities to various modes and degrees of bunching. These can be said to be the probabilities that the bunching occurs "accidentally". In particular, we found that the probability of accidental occurrence of Pattern A is "essentially zero". Another way of saying the same is that it is "impossible" that it would occur by accident.

But the pattern did in point of fact occur. (Three of the twelve traces meet the specifications of our calculation.) The only possible conclusion is that the occurrence was not accidental. There was a "reason" for it. It is established beyond (reasonable) doubt that the sequence of impulses down the neuron's axon is not a random sequence. Instead, pattern A as a whole must be regarded a separate event, occurring with probability much greater than could be deduced from the random-sequence assumption. It is a "meaningful" pattern, as we shall call it.

If neurons can "do probability theory", they can figure all this out. Thus, for instance the neurons receiving contacts from the neuron of Figure 1 are able to "notice" (as we shall say) that the Pattern A is meaningful. They then readjust their scale of probabilities, i.e. expected occurrence frequencies, assigning a separate probability to Pattern A, and with this little correction return to assuming that everything is random—until they find that, for instance, Pattern A occurs immediately after another Pattern B (possibly arriving at another synapse) more frequently than it should. They then conclude that the combination of A and B is a meaningful one, deserving to be treated as one event and assigned a probability of its own. And so on.

We shall assume that neurons can indeed "do probability theory"; at least as much of it as is necessary for noticing when an input occurs more frequently than it should. It is the principal assumption of this paper.

Probability Sensitiveness of People

It may be remarked that people as a whole are sensitive to low-probability events. For instance, it may happen now and then that we talk about a long-gone friend and run into him in the street five minutes later. Even the most callous of us will find such an event "amusing". The rest of us will conclude that "it cannot be a coincidence: there must be telepathy". But the only evidence for a "reason" behind the event is its low probability! The notion of "telepathy" merely arises in our mind to furnish the "reason" and, indeed, its definition and descriptions give no more detail than that!

It will be noted that systematic laboratory research also often involves similar modes of thought. An event occurs and we ask: "Is the event significant or is it merely a coincidence?" And the question is invariably answered by making a quick estimate of the probability that it is a coincidence.

If the event recurs, the latter probability is much reduced; it is now equal to the probability that both the event and its repetition are accidents. Thus the experimenter's first impulse is to try and reproduce the event. (We shall assume that the neuron tries the same thing.)

Patterns that may repeat themselves need not carry as overwhelming odds against being "accidental" as did "Pattern A" in the above example. Probabilities like 1/1000, even 1/100, are enough to make the neuron "suspect" that an input is meaningful. If the input repeats itself, the probability becomes (essentially) squared; if it repeats itself again it becomes cubed; and so on. Thus due to repetition, the odds become overwhelming, and the suspicion grows into certainty.

Connection with Information Theory

It is seen that it is reasonable for neurons to pronounce an event meaningful if they notice that it has low probability. But there is one ticklish question here: how does one recognize *a posteriori* an event of low probability? Any event, if defined closely enough, appears to fit the description. That a spinning needle stops at a given point is an event of zero probability; yet every time the needle is spun it stops at some point or other.

Clearly, it is necessary to decide *a priori* on a criterion to be met by the input signals, preferably by defining a random variable and demanding that the event be far enough in the tail of its distribution function. The natural choice of random variable is *information content.*

The information density of an infinite sequence of random binary digits is uniform if the *a priori* probabilities of 0 and 1 are equal. It is then always 1 bit/digit. But the same is not the case if the *a priori* probabilities are unequal. For instance, if on the average there are 10 times as many 0's as there are 1's, the information content of a 1 becomes $-\log_2(1/11) = 3.46$ bits, and the information of a 0 becomes $-\log_2(10/11) = 0.13$ bits. Shannon's formula gives us an average information density of 0.43 bits/digit; but there will be fluctuations about the mean. Segments containing many 1's will have high information content. They can also have quite low probabilities of occurrence. For instance the probability that a 100-bit-long sequence contains 50 or more 1's turns out to be 10^{-24}.

Now, firing occupies less of the time than no firing; bursts occupy less of the time than intervals between them, and meaningful firing (at least in the naive brain) less of the time than meaningless firing. Thus if the neuron keeps a record of all the *a priori* probabilities (updating the record every time a new event is noticed as being meaningful) and if it makes itself sensitive to inputs of extremely high information content, it will automatically be sensitive to a well-defined class of low-probability events.

There is much advantage to choosing high information content as criterion for meaningfulness. (i) If meaningful signals have high information content, they have a chance of still being recognizable after noise wears away some of the information (as in Trace 2 of Figure 1), or after some of the contributing neurons are lost. (ii) We will not expect all meaningful events to mean the same thing. In our scheme, everything worth storing will correspond to a meaningful stimulus, or a sequence of meaningful stimuli, or a set of equivalent meaningful stimuli. Certainly, a neuron must be able to tell apart meaningful stimuli if they have different meanings. Patterns handpicked for their high information content have the advantage that many of them can be distinguished. (iii) One event will be "captured" by one set of neurons, the other by another. If neurons always capture information-rich signals, the neurons that capture an event are at a good vantage point to perceive it. For a neuron is at a good vantage point to perceive an event if it is at the crossroads of the information flow pertaining to the event; thus if it receives event-pertinent signals from many neurons, and the signals are well discernible.

The advantages of the extremely low probability of meaningful events may be thought of as being additional to the advantages of high information content. (i) One can make the aforementioned metaphysical inference that for such events there must be a "reason". (ii) When neurons notice an unusual event, they can make the (equally metaphysical) inference that other neurons have also experienced an unusual firing event (stemming from the same reason), have noticed it, and will, when at some later time the event recurs, recognize it, and respond to it. A neuron can use such knowledge to coordinate its activity with the activity of other neurons not even in synaptic contact with it! It is my contention that information trapping by living brains would be impossible if the neurons could not make this latter inference. (Which is why I spoke above on the notion of randomness at some length.)

Capacity and Threshold

It is characteristic of this present scheme that the number of possible signals is overwhelmingly greater than the number of meaningful ones. Only this way can we absorb *anything* the world of realities chooses to offer us. If it were differently, we would be unable to cope with anything that did not fit into the few "pigeonholes" initially provided to us by Nature.

To see how few signals are meaningful, we consider the motor sphere. The number of motor neurons (in cat) is of the order of 10^5. Any motor skill consists of a sequence of activity (or inactivity) for each neuron. If ten new motor skills were learned in every second of a 300-year-long lifetime, the result would be 10^{11} motor skills. But the 10^5 motoneurons are capable of some $2^{(10^5)} \cong 10^{30,000}$ firing patterns, and that does not include the added variety due to time sequencing!

Since in the present scheme all decisions are made by the single neuron, the latter must be able to distinguish potentially a very large number of signals. Let me show that it is able to do so, if it is able to keep track of each of its dendro-somatic synapses, and tell apart different combinations and different sequences of inputs arriving to them. (I shall assume that it is.)

If (for purposes of an estimate) time is divided into 100-millisecond intervals, and in each of them each synapse can tell whether it does or does not receive a burst of firing, then a synapse can receive up to 10 bits/second of meaningful information. If the neuron has 10^3 synapses, and can retain inputs for 3 seconds (the length of the "short-term memory"), then it can distinguish

$$2^{(30 \times 10^3)} \cong 10^{10,000}$$

different inputs (any of them potentially meaningful), which is again a number enormous compared with 10^{11}, the upper limit set by our limited lifetime.

In a way, the neuron's code is like an error-correcting code in which many signals are "beyond correcting", i.e. meaningless; in fact, most signals are.

It is desirable to make at least a crude estimate of the minimum amount of information a meaningful signal must contain if it is to be distinguishable from all other meaningful signals. Since $2^{40} \simeq 10^{12}$, which is larger than our upper limit 10^{11}, a 40-digit binary number should be sufficient for describing a meaningful input. (Actually, a number half that long is probably almost sufficient, as *Webster's Collegiate Dictionary* contains about 10^5 words, and that may not be a bad order-of-magnitude estimate of the number of elementary "things" stored in our brain. Be that as it may, we shall use the more generous estimate.) However, a number of that length is only enough if we put the numbers together methodically, making certain that no number corresponds to more things than one. In our case the numbers are obtained by an independent selection process; they must therefore be considered random.

Thus we require a number n, large enough that if we pull n binary digits randomly out of a hat, put them together into a number, and repeat the process 10^{12} times, it is still improbable that any two of the resulting numbers are identical.

Estimating n is simple. The probability that two given sequences are identical is 2^{-n}. If there are N sequences altogether, the probability that some two are equal is

$$P = \frac{1}{2} N(N-1) 2^{-n} \cong 2^{-n-1} N^2 = 2^{-n-1+2\log_2 N}$$

Thus, for instance $P < \frac{1}{2}$ means that

$$n > 2 \log_2 N$$

with each added digit doubling the safety factor.

In other words, if the digits are random then (more than) twice as many of them are required for constructing N distinct sequences than if they are methodically put together. We conclude that meaningful stimulations must contain about 80 bits of information lest confusion arise. The information threshold of neurons must therefore be of that order of magnitude or higher.

We shall go on the assumption that the information threshold of the neurons which will be denoted by τ, is not fixed but is "flexible"; i.e. can be moved up and down. A state of greater attentiveness corresponds to lower threshold. (If the thresholds are lowered far enough, some meaningless combinations will pass as meaningful.)

Some inputs may be marked by Nature as if they were of extremely high information content incapable of depreciation; they will then make any stimulation accompanying them seem meaningful and worthy of recording. In this way Nature can put rewards, punishments, and various other absolutes as well as innate patterns into the brain. (We will not deal with anything innate in this paper.)

General Remarks

In summary, it will be assumed that the neuron is not a simple threshold device as is often said. It is not merely a door that opens when sufficiently banged upon. It breaks down not to large assaults of force but to large assaults of information.

However, it will be noticed that an extraordinarily large assault of "force" (i.e. excitation) also has a large information content. The conclusion is that the neuron always responds to sufficient "force", but can also respond without it. It will be seen in a moment that in the everyday chores of brain work it usually responds without it. It responds not because it receives strong stimulation but because it receives the right stimulation.

The encouraging feature of the neuron model outlined above is that it can be made to result in a tendency of the neurons to organize themselves in a certain way, which, as far as I can tell at the time of this papper, has all the earmarks of being the right way. One aspect of the organization is a certain grouping of the neurons which shares the attractive features of another similar kind of neuron grouping (Legéndy, 1967, 1968) and overcomes its disadvantages. The attractive features are characteristics of the data structure that are very typically brainlike. (i) Things consist of parts, and the parts of other parts; learning consists, in a way, of piecing together old parts into new wholes. (ii) Things are often difficult to recognize when taken out of their usual context. (iii) Recognition is not very sensitive to the absence of some details as long as enough of the details are present. (For instance we are unaware of the blind spot on our retina.) (iv) The brain is relatively insensitive to damage, loss of neurons, and unreliable firing.

The disadvantages overcome are a limitation on the hierarchy of details within details, inability to handle sequences, relations, syntax, and wholes that consist of few parts, and (partial) inability to account for the fact that neurons organize themselves instead of being organized.

Speculations Concerning the Physical Arrangement

It is reasonable to suppose that inside each neuron labor is divided, that the analysis of single-channel firing patterns (like Figure 1) is done by some neuroperipheral data processing units (maybe in the subsynaptic spaces); and that what is found by those to be meaningful is relayed to a more centrally located data processing unit in the cell body (maybe using the microtubules or the neurofilaments as telephone wires). The peripheral units should be, among other things, in charge of understanding the characteristic firing modes of the various neuron types (each unit, of course, need only understand one neuron); the central unit should probably be the one distinguishing the different neuron types (for instance, neurons processing auditory information should be very sensitive to the time pattern of their inputs). The information storage facilities of the neuron should make use of molecular information storing principles, as in no other way can the neuron have room for the enormous quantity of information accumulated in a lifetime.

THE GROUPING OF NEURONS

The Compactum Concept

Next we outline a scheme in which the neurons "hit hardest" by an event will come to be tied together and form a group; though the ties will lie dormant almost all the time, and show their existence only at times when the proper event recurs. For every meaningful event noticed by the neurons, a new group will form.

The groups will be referred to as "compacta" (singular: "compactum"). I have shown elsewhere (Legéndy, 1967; see also Levy, 1969, 1970) that they could ignite to firing (in a chain-reaction-like manner), and contain their firing, and proved a few more of their basic properties. The assumptions of the earlier paper differ somewhat from the assumptions in this one; but the main conclusions carry over and the proofs are too similar to be worth repeating here. The differences are that the earlier compacta contained fewer neurons (and were made up of certain smaller compacta assumed to be innate), the neurons were not sensitive to information, and their chain-reaction-like "ignition" was followed by steady reverberatory firing. In this paper there are no reverberations, and ignition is a one-shot affair in which every neuron responds only once, even though it may for a while continue to receive relevant stimulation after having responded. The ignition of a compactum will be viewed as the smallest unit of brain activity, and referred to as an "excitation". Compacta in the present paper are simpler than those in the other one in that they are not made up of other compacta. (There are also some other minor differences.)

A compactum of the present paper is very much like a neuron, except of course that it is bigger, and new compacta are constantly forming as new meaningful signals are noticed. Each compactum is assumed to contain so many neurons, and the neurons are assumed to be so well scattered throughout a whole cortical area (or other cerebral structure), that (essentially) every neuron in the area is connected to every compactum with plenty of synapses. (We ignore the fact that in real brains there may be systems of neurons intermingled with certain other systems yet making no synapses with them.) The synaptic connections between compacta are assumed to be intimate enough, or equivalently, the compacta are assumed to be large enough, that if for some reason compactum A needs to respond every time compactum B ignites, it can do so and not make mistakes (where A and B are any two compacta). The result is that any compactum can tell any other "all it knows".

It must be assumed that the formation of a compactum consists of two parts: the "impression", which is a one-time event, and the "consolidation", which is a never-ending process. The descriptions of the two parts below are in many ways unsatisfactory, and must only be regarded as efforts to convince the reader that these process are information theoretically possible, meaning that the neurons are able to find out what they must do at any time, and how they must change their behaviour if they must.

Announcement of Meaningful Events

Impression of an event by necessity starts from those neurons that notice the extremely low probability of accidental occurrence of the event. These neurons are in all likelihood few, and if a compactum consisted only of them, it would be too small to be in synaptic contact with all other compacta.

Since the probability of accidental occurrence is extremely low, there is no need for majority vote in deciding whether the event is meaningful. If a neuron learns that one neuron has seen a meaningful event, it is justified in believing that a meaningful event has indeed occurred.

Thus, the whole brain can be made aware of the occurrence of a meaningful event if all neurons are capable of emitting a special sort of firing, amounting essentially to a cry like "Meaningful event! Meaningful event!" Such a mode of firing will be assumed to exist, and will be called "M-firing". The neurons that noticed the event emit M-firing and all neurons that hear that firing, even if only from one neuron, emit M-firing; thereby, in a fraction of a second, all neurons can learn that a meaningful event has occurred.

They know ar much as if they had been told, "You just saw a low-probability event." Of course, finding out about it in this way is not as good as actually noticing the event, because it still leaves the problem of figuring out what the low-probability event was.

Impression

How the neurons do that is a question not satisfactorily answered in this paper. In principle, there is no difficulty. What may essentially happen is that the neurons scoop up all information that arrived to them in the immediate past, and is therefore likely to be in causal connection with the event. They do so and assume that they captured enough information to recognize the event when it recurs. (Let us note that the neurons could not assume this if the event were not of low probability, and if from this fact they could not make the earlier mentioned "metaphysical" inferences; but as it is, they can.)

Members of the compactum must be chosen from among the neurons most likely to have captured enough information, or, as could also be said, to have captured the event. These are the neurons that received the strongest assault of information.

When such a neuron recognizes a recurrence of the "proper event" of the compactum (as I shall call the event that gave rise to the compactum), it fires. Such firing occurs as soon as the neuron receives enough relevant information to ascertain that the proper event actually recurred. This amount is what has been called τ, and shown to be about 80 bits. The reason is that the recurrence is a genuine low-probability event. The events of forming the impression are (to most neurons) only promises of a low-probability event; the recurrence is fulfilment of the promises.

Consolidation

(The following description, like that of "impression" above, is only a sketch and must not be regarded as anything more.) If neurons of a new compactum captured the proper event well, they are able to ascertain all recurrences of it beyond reasonable doubt. They then undertake to sharpen their skill of recognizing it by collecting every possible set of inputs relevant to the event, noticing early signs, and of disregarding irrelevant inputs. To do that is again possible, as the neurons do have an absolute reference at their disposal: they do possess one (presumably) foolproof way of recognizing the proper event. Consolidation turns neurons from responding to the strongest stimulation to responding to the right stimulation. Many neurons will probably be unable to recapture the event, and many compacta unable to consolidate; the result being a form of "natural selection" among compacta.

Neuronal Knowledge

The notion of "meaningful" stimulus is helpful in telling us how much a neuron can "know". It gives us a sort of upper limit. It is conceivable, in fact certain, that a neuron does not notice all meaningful stimuli arriving to it, and that it does not store all the ones it notices. But it is not possible for the neuron to know anything more than has arrived to it in the form of meaningful stimuli (excluding innate knowledge for now).

We do not imply that the neuron can "know" (in the usual sense) what the meaningful firing means. All a neuron can ever know is that a certain firing pattern is, for instance, the same as some earlier one. But meaningful patterns arise for a "reason", which is why we could assume that a whole compactumful of neurons, each of them recognizing them in its own way, can achieve agreement as to the times at which they recur.

It may well happen that the reason can be described in words. (Although it cannot *always* be described in words!) When it can, we may describe it, and say, purely in a manner of speaking, that the thing we described in words has been "heard" by and is "known" to the neurons that noticed the pattern. Thus for instance, we might say that from a given set of excitations every neuron of a compactum A "knows" that there is a cup on the table, but not that the cup has coffee in it; by igniting, the compactum A "tells" a compactum B all it knows; and the fact that there is coffee in the cup is conveyed to B by a third compactum C. Such a convention of speech is only a sort of "verbal notation" which, however, clarifies matters considerably, as it ignores the fact that every compactum, and within it every neuron, receives and recognizes signals in its own private language.

A neuron will be said to "know" of a certain circumstance if it possesses enough information to respond differently in presence and absence of the circumstance. The notion is indispensable in discussing motor activity. Clearly, motor activity depending on a certain circumstance must be directed by neurons that, in this sense, "know" of it.

Overlap between Compacta

Compacta may overlap. They must, because a compactum, as envisioned here, can contain as many as 10% of all neurons in a cortical area; and at that rate 10 compacta would take up the whole cortical area, if they were not allowed to overlap. As it is, the number could, if there were no other limitations, be as high as about

$$\binom{10^8}{10^7} \sim 10^{10,000,000},$$

because that is the number of ways in which 10% can be chosen from among 10^8 neurons. Actually there can be no more than 10^{11} compacta, because, as pointed out above, there is simply no time in a lifetime to form any more than that. (It is not impossible that for instance in the visual cortex the number 10^8 should be replaced by 10^4, the number of neurons in a cortical column; but even if this is so there is room for more than $10^{1,000}$ compacta.)

It does not disturb matters if a neuron participates in one compactum as well as another which is supposed to inform the first compactum of something, as no neuron draws its information from one single neuron anyway. Therefore, for all intents and purposes compacta can be treated as if they were as many distinct, non-overlapping sets of neurons.

The Unlimited Hierarchy of Compacta

It is a property of compacta, due to the thorough interconnectedness of their members, that whenever a compactum ignites, the knowledge imparted by the igniting stimulus is acquired by every neuron of the compactum (and is by the same ignition imparted to all compacta listening). Thus "knowledge" can be thought of as passing from compactum to compactum. Furthermore, every compactum can impart enough information to any other to "tell it all it knows". (For no knowledge need be more than about 80 bits!)

Because of this, a large collection of randomly interconnected neurons is capable of a surprising feat of information theoretical acrobatics, which may be called "self-augmentation" of information. It is essentially what appears in the example of the cup and the coffee: one set of neurons notices the cup, another notices the coffee, another again puts the two together. The surpising part is that the set of neurons that notices the combination need neither be a larger set than the other two, nor one set apart from them and removed to an anatomically distinct area. It can be like the other two and share many of their neurons.

Thus, as the brain gradually learns about the world outside, some neurons will first notice the correlations most easily noticeable, build compacta responding to them, and thereby make them noticeable to every neuron. Some neurons will then notice meaningful events composed of ignitions of several compacta (events present before, but not noticed until exposed by the punctuating effect of the compacta), and form further compacta responding to those; and so on, *ad infinitum*. All of it will be done by use and re-use of the same neurons.

The brain can in this way accommodate an unlimited hierarchy of details within details, things that consist of components, and components that consist of further components. That this can be done follows from the fact that in spite of their extensive overlap, compacta are as independent as if they were as many distinct, non-overlapping sets of neurons, and from the fact that, hierarchy or no hierarchy, the brain will always contain a negligibly small number of compacta compared with the immense number it could in principle contain.

No Reinforcement Required

The self-augmentation of information is, quite definitely, a property of those neural networks which McCulloch and Pitts (1943) call "networks with circles".

Owing to the phenomenon of self-augmentation, each successive cerebral structure is able to refine the information available to it to as high a level of sophistication as that information can possibly be refined. For instance, Brodmann's area 19 can (probably) recognize all visual objects, simple and complicated, because they can all be noticed without the help of auditory information.

It should be noted that to do this the neurons do not need a separate influx of reward or punishment, nor any other kind of influx of "absolutes". Their only reward is the thrill of having noticed an event of extremely low probability, occurring more frequently than can be explained by the accidents of chance.

Nor do we assume that the neuron's own firing tends to beat the path for future firing (as Hebb (1949) assumed). We shall assume that (at least in passive neurons) the acquisition of a firing habit has only to do with the discovery of certain things about the input, and has nothing to do with the neuron's own output. (In regulators of motor activity the same will not be assumed.)

SYNTAX

The Syntax Problem

When a minute ago, in illustration of neuronal knowledge, it was said that one compactum could know of a cup, another of the coffee in it, and a third could "put the two together", the reader may well have felt that something important was missing; essentially, that there was something unbrainlike about the definition of a thing as merely a bagful of details. A neuron that knows "cup; coffee" does not know all that is implied in "a cup of coffee".

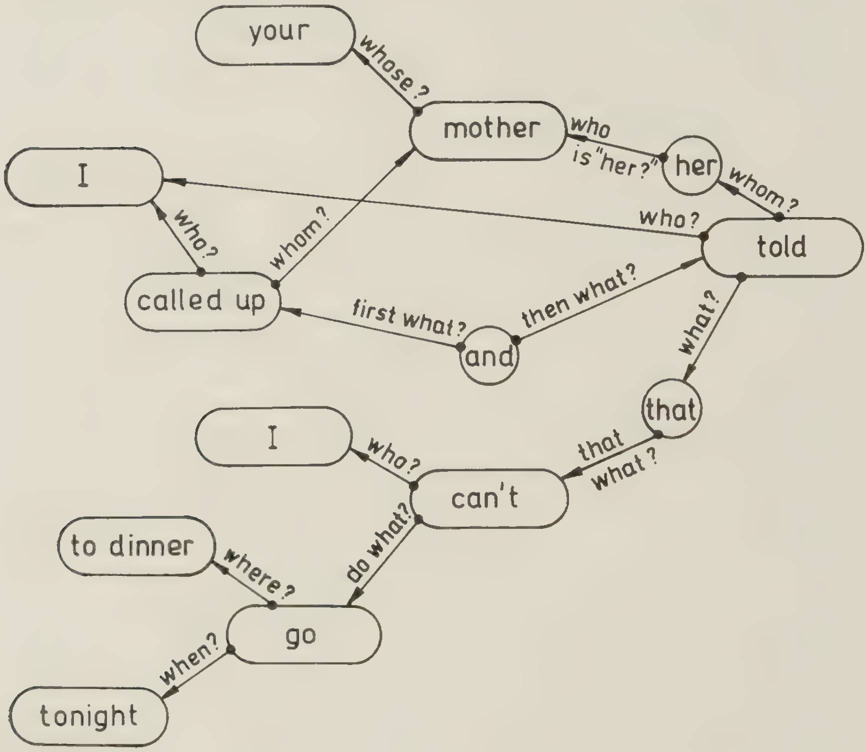


FIGURE 2 Illustration of syntax. The syntactic information is contained in the arrows

There was in fact something missing; the missing ingredient was what will be called "syntax". By syntax in this context I mean the same as is meant by it in grammar. The syntactic information in the sentence of Figure 2 is the information that tells where the arrows start and end.

Implicit in this definition, and in much of what follows, is the assumption that spoken language is ideally suitable for summing up the goings-on within the brain, that its self-evolving, easy-to-learn, disorderly, untidy ways, with few rules and many exceptions, are the brain's own, and are as characteristic of the communication between neurons as of the communication between people. (It may be noted that the information efficiency of the English language is said to be about 0.5, which checks nicely with the "2-to-1" result found in our derivation of the "80 bits". Of course, it is hard to evaluate the significance of this agreement.) Implicit in such an assumtpion, in turn, is the belief (which is at the roots of everything I am saying here) that although different neurons differ widely in their form, behaviour, and role, "basically" they all do the same kind of thing.

Returning to Figure 2, we note that the arrows between the words there do not stand for ignition-transmitting synaptic connections between the compacta representing the words. Such connections do exist inside the brain, as will be seen in a moment; but the figure only talks about connections that exist outside the brain; its arrows stand for "connections" in the figurative sense. Their existence must be regarded as being new information, arriving from the outside world and somehow relayed to the compacta of the sentence, which are then sole possessors of it and must, in turn, relay it to some further compactum. (In the case of the sentence of Figure 2, I suppose, they must relay it to whatever compactum causes a person to explode.)

There is one difficulty here; it is a hidden one, known probably only to those who thought about the problem before. For instance, it may well be known to those persons who ever tried to build a pattern-recognizing machine and in doing so experienced the difficulties that plague all efforts to integrate "parts" into "wholes".

A possible first step in building a pattern-recognizing machine is to build detectors of simple figure elements, for instance several kinds of edges and corners, (build them so as to be independent of position on the retina) and use them to identify objects by allowing them to inform a higher-level processing unit of the number of edges and corners of each kind contained in the objects. It is quite possible to be dissatisfied with this scheme because the higher-level unit is not told which figure element joins hands with which other. In an effort to correct the deficiency the builder may supplement the information arriving to the higher-level unit by adding the outputs of further detectors, built to notice groups of two or more figure elements that join hands. Finally, he may feel a sense of frustration after having gone through all this, because the higher-level processing unit still receives no more than a few additional pieces of is-it-there-is-it-not-there information, and still does not "know" which element joins hands with which other element.

Solution of the Syntax Problem

In our compactum scheme the manifestation of the same difficulty is the fact that it is not possible for a compactum to talk to another compactum *about a third* compactum, if the third compactum happens to be a different one at different times. The reason is that, although the ignition of a compactum contains much information, it always contains the same information. Translated into words, its message is always "this is compactum A speaking" (and the whole information goes into defining which compactum the "Compactum A" is). The only "twist" it can put on the message is the shading of its voice, as it were. (Neurons are known to be capable, in principle, of emitting several different forms of firing bursts; maybe up to about a dozen.)

What is the trick? How do they do it? Certainly, every pair of (figuratively) connected compacta must somehow convey the fact of their connection to every compactum to which their connection is of concern.

The answer is that the trick is simple: one compactum cannot do it, but two can. A compactum C can be informed of connection between two events A and B if the ignitions of the compacta of A and B are *timed* with respect to one another, in a reproducible way, so C can detect the temporal correlation.

To achieve such prearranged timing, a conspiracy must exist between A and B. At least one of the two must have the ability (i) to recognize when the relationship is present and (ii) to recognize the ignition of the other compactum as being the ignition of the other party in the relationship. By igniting in response to the recognition (ii), it achieves the required temporal ties, and makes them noticeable to all compacta listening.

A glance at Figure 2 shows that each arrow starting out from a word (together with the question going with each such arrow) is an inherent property of the word; it is part of its "dictionary definition", as could be said. For instance, whenever the word "told" is used in speech, the sentence containing it is incomplete, unless it somehow answers three questions (the same three questions every time the word is used): who told whom what?

This is convenient, as it allows us to infer that of the compacta at the two ends of an arrow the one that must make the recognitions (i) and (ii) is always the one at the arrowbase. The reason is that it is in a better position to make both.

To make (i), this compactum need not even perform any separate act of recognition. Its own ignition amounts to that recognition, since every time a word (whose corresponding firing event may be called the "primary event") occurs, its proper questions can be asked, and it is certain that some answer exists to them all (meaning that the relationship *is* present). Only the answers remain uncertain; but they belong in the domain of recognition (ii).

The compactum at the arrowbase is also in the better position to make recognition (ii), because the alternative would be to build a recognizing apparatus into every compactum that could potentially represent an answer. The fact that the questions going with a word are always the same has the advantage that this apparatus can be built into one compactum (the compactum at the arrowbase), once and for all. Of course, the "apparatus" is not literally made up of additional neuronal hardware; it is only a familiarity of the neurons at the arrowbase with a new set of events. These neurons need only to develop a sensitivity to a separate "secondary proper event" corresponding to each arrowbase (i.e. to each proper question), amounting to a habit of responding every time an ignition is recognized as representing the answer to the proper question. Every such response is a "recognition (ii)". (The result is that for instance the compactum of the word "told" has four proper events: the primary event and three secondary events.)

It is reasonable to assume that when the compactum first forms, it only has one proper event, a mixture of all the future ones, in which the future secondary events are represented by specific answers to the future proper questions. (For instance, the compactum for "pencil" may at first be a compactum for "green pencil"; only later will it be discovered that a pencil may be of other colours than green.)

The conclusion that the brain cannot be informed of a relation between two compacta unless something amounting to a recognition takes place between the two compacta checks against the well-known fact, so obvious that it is hardly even noticed, that *relations are specific*. For instance, there is such a thing as a cup of coffee, or a cup of tea; but there is no such thing as "a cup of book". In other words, even though the statement "there is a cup of A on the table" leaves "A" unspecified, "A" cannot be any compactum in the brain. The statement narrows down the choice of things acceptable as A considerably (namely to liquids, powders, etc.).

Finally, let us note that it would not be possible to transmit syntactic information by merely arranging coincidences in timing, were it not for the fact that the syntactic information is always relatively little. If, for instance, a compactum ignites once every hour on the average, timing its ignition with an accuracy of 0.1 second could impart no more than $\log_2 36{,}000 = 15.1$ bits of information, which is much less than for instance 80 bits. I propose that speech is broken up into sentences (in every language!) for just this reason. The brain must ingest information bite by bite because otherwise it could not chew it, meaning that it could not keep the syntax straight.

This does not imply that for example, the visual cortex is constrained to such a slow information rate. It is quite possible that the 10^4 cortical columns in the visual cortex are like 10^4 little brains functioning in parallel, each forming its own sentences and sending them on separately.

GOAL PURSUIT

The following section contains an attempt to understand the faculty within our brain that acts as headquarters for goal-directed motor activity.

The Durability of the Brain

The brain is bothered remarkably little by physical damage inflicted upon it. It is well known that many neurons perish during our lifetime of, essentially, natural causes. The usually quoted statement is that from the brain of old people some 10% of the neurons are missing; which means that on an average day more than 10^4 of our neurons die. It is also well known that whole chunks of brain matter can sometimes be removed without the slightest noticeable effect. At other times a person may die of no complaint other than headaches, and then be found to have a brain of which parts are rotted away.

Facts such as these are so puzzling that they have caused many people to give up entirely on trying to achieve a precise understanding of the living brain, and drove others toward the belief that the data processing in the brain does not occur in the neurons at all, but either in the macroscopic electric fields surrounding them, or else in the chemical bath in which they are all immersed. (But the latter ideas are both information theoretically absurd. The electric field does not have enough mathematical freedom to carry the required amount of information in the narrow frequency range of the E.E.G.; and the chemical bath is, among other things, too slow by many orders of magnitude.)

The message conveyed by the preceding sections of this present writing is that the remarkable facts of brain durability do not necessarily compel us to abandon the "neuron theory". Every "meaningful event" is captured by a whole compactumful of neurons, thus by as many as 1%–10% of all neurons in the area (or areas) over which the compactum extends, and whenever the event recurs all these neurons will "know" the fact. The event leaves a trace, as could be said, in all these neurons. Thus the trace is not easy to erase. Removal of one neuron will certainly not erase it; neither will surgical removal of, for instance, half of the area; because the neurons of the compactum are in all probability scattered evenly over the whole area they occupy, and with half the area removed there will still be plenty of them left in the other half.

Formulation of the Goal Pursuit Problem

But we are faced with the question: is the validity of such an explanation of brain durability not automatically restricted to the passive portion of brain function? At the headquarters of goal-directed motor activity, is it not necessary for some neuron to "give the orders" and to be, by virtue of this very fact, indispensable?

It turns out, the answer is no. It is quite possible to distribute the order-giving function between the neurons, just as it was possible to distribute the fact-recording function; and the principle involved is the same, too.

We reason as follows: the brain can acquire new skills by the device of (essentially) piecing together ones already at its disposal. The skill of lifting a glass is made up of the skill of prehension, the skill of shifting the eyes from one object to another, and some others; then the "skill" of making a toast is made up of the skill of lifting a glass and, again, a number of others. (Once again some of the lameness of the examples comes from the neglect of syntax. The syntax problem in goal pursuit is more complicated than its counterpart in passive brain function, and will be omitted altogether from this paper.)

The hierarchy of skills within skills is unlimited (as is the hierarchy of compacta in the earlier sections); but every skill, simple or complex, must by necessity consist of three basic steps. First we want to achieve something; then we do something; then (with luck) we learn from our sense organs that the thing we had wanted became achieved (and that therefore we can go on to the next thing).

First, let us note that the same physical object within our brain in which a "wish" arises, whatever that physical object may be, must, once the wish is satisfied, receive the signal conveying that fact, and "know" from it that the wish has been satisfied; otherwise the wish would persist *ad infinitum*, even after it has acutally been satisfied.

Second, let us note that the same physical object must also be the one to issue whatever order must be issued for bringing about the satisfaction of the wish (or, equivalently, issue some signal from which that order can be uniquely deduced), because, by definition, the object in which the wish arises is (at first) the only object that knows what the wish is.

It follows that there is only one way in which the activity of goal pursuit can be divided and distributed within the brain: wishes can be divided into component wishes (and those in turn into further components, and so on); orders can be divided into component orders, (and those in turn into further components, and so on); and the signals bringing the information of satisfaction can be divided into component signals (and those in turn into further components, and so on). But each given component wish must correspond with a certain component order and a certain component signal; and the basis for each correspondence must be the fact that the same physical object gives rise to the wish as initiates the order, and as receives and registers the signal.

In the spirit of the previous sections, the natural choice of the object to house the apparatus of the wish–order–registration triplets is the individual neuron. Once again, groups will be seen to form, and all members of each group will be seen to carry the same wish–order–registration triplet (although each in a different form, as in the last sections). Thus once again the whole will be reasonably insensitive to neuron loss. We shall refer to a wish–order–registration triplet as an elementary "skill".

Active Neurons

The neurons capable of acquiring skills cannot be like the "passive" neurons we discussed in the other sections; they must be a different breed of neurons. They will be referred to as "active" neurons. Active neurons can fire because they "want" something; in other words, it is possible for the cause of their firing to lie within the neuron, rather than being necessarily outside it, as is the case with passive neurons.

However, as will be seen in a moment, a person observing neurons as they fire, will nevertheless have difficulty telling which one is active and which one passive (even if he is able to monitor the firing arriving to the neurons). The reason is that, as it turns out, even the "endogenous" firing of an active neuron cannot be truly spontaneous, but must await certain incoming clues, in the form of firing; otherwise the order contained in it can never be understood by the other neurons (and, even if it could be understood, it would never be carried out).

Before turning to these matters, let me extend the fundamental principle of this paper, the "too improbable to be accidental" principle, to active neurons. Suppose a neuron observes, through its inputs, three things in sequence: first the world is in a improbable state A, then the active neurons of the brain issue an infrequent order p, then the world changes into another improbable state B.

The neuron can reason that the whole sequence is so improbable that it cannot be accidental. It, taken as a whole, must be a meaningful event and therefore able to occur reasonably frequently. However, this does not mean that fragments of the sequence are also meaningful and occur frequently. In particular, there is no reason to believe that (i) the event where the world changes from A to B without the intervention of the order p, and (ii) the event where the world is in state A, and the order p is issued, but the world fails to change into state B, are meaningful. As far as is known, (i) and (ii) can still only occur by accident, and both accidents are so rare as to occur only once in aeons.

The neuron concludes that whenever the world is in state A, the order *p* is (beyond reasonable doubt) necessary and sufficient for bringing the world into the state B!

Impression and Consolidation

A neuron making such a discovery is justified in being pleased, as the discovery can, if processed appropriately, add a new skill to the brain's arsenal of skills.

The required processing can be reasonably divided up into the same two steps as were sketched in the earlier sections (and will again only be given a sketchy description): impression and consolidation. The impression must once again begin with an "announcement", probably in the form of a campaign of what has been termed M-firing, started by those (probably few) neurons which made the unusual discovery.

Impression of a skill is expected to be one degree more complicated than the impression of an event aimed at achieving no more than a passive familiarity. For, this time, part of the recording must consist of telling those active neurons which had issued the "order p" to issue it again whenever ordered to do so, while another part of it must consist of forming an ignitible group of neurons that must, whenever necessary, order them to do so.

The added complexity does not change the principle. Once again, the M-firing is an order for neurons to scoop up as much recently arrived information as they can; hopefully, enough to recognize the pertinent events when they recur.

The ignitible group, which corresponds to the "compactum" in the case of passive neurons, and which will be called the "chorus" of the wish, must be made up of those neurons that believe themselves best able to recognize the reason for the M-firing. Like their passive counterparts, they are the disciples of the neurons that originally noticed the improbable sequence, they must therefore be able to recognize the aforementioned "state A" of the world, must "know" that, when the world is in that state, it can be brought into "state B"; they must in addition be able to recognize the "state B". They will be the neurons that develop the ability to "want" to change the world into state B, and give off firing (of a certain kind, see below) when they do want to do so, and think they can.

Since the neurons that had issued the "order p", and must learn to give repeat performances of that order—I shall call them the "extension" of the skill (the chorus being the "intension")—must learn to respond to the chorus, the neurons of the chorus must, immediately upon initiation, emit a volley of firing.

To the neurons of the extension the impression means that they must capture as much as they can from the firing of the nascent chorus. Since the neurons of the nascent chorus know that they are just being initiated to a chorus, they are able to modulate their initial firing so that it is recognizable as being the initial firing of a nascent chorus; thereby they can enable the extension to capture it.

The neurons of the extension must furthermore capture as much of the "state A" of the world as is necessary for knowing at future times when an order they receive (for bringing about the "state B") can be carried out. They are not hand-picked for ability to recognize "state A"; but chances are that, already to start with, they are in possession of a fair amount of information relevant to "state A".

The reason is that the extension consists of already established choruses. They represent the earlier-mentioned skills already at the brain's disposal, the skills that the brain pieces together to make up a new skill. The sheer fact that these choruses ignited originally (when they issued the "order p") indicates that they knew the outside world to be in such a state as would allow them to achieve their individual goals. Since the world was in "state A", that state must be similar to the one familiar to them.

Finally, the neurons of the extension must capture the time relations between their own firing and the ignitions of other choruses of the extension. They must attempt to capture these time relations well enough to reproduce the whole time sequence when next receiving the order to do so.

It must be emphasized that the extension is much bigger than the intension, as it consists of many choruses, none of which is necessarily smaller than the chorus constituting the intension. The intension may for instance be the brain circuit issuing the order to sing a certain song, and the extension the brain circuit issuing all the orders (on a certain level) that go into the performance of singing the song. Thus the extension of a skill may well involve every neuron in the area, some of them many times.

The objective of consolidation is the same as in the passive case. The consolidation of motor skills is a somewhat more readily observable process than its purely internal counterpart; consolidation is, essentially, achieved by "practicing" the skill.

Neuronal Will

Because of the axonal bottleneck, a chorus is the smallest unit that can issue an order and expect it to be deciphered. But before a chorus ignites, every neuron of it individually "wants" the goal achieved. (In parallelism with the fact that when a compactum ignites every neuron of it "knows" that the proper event has taken place.)

Now, the notion that the active neuron has a will of its own (which must of course once again be regarded as a "verbal notation") is seemingly in contradiction with the large number of rules and orders with which the neuron must cope.

For instance, a chorus of the extension of a skill ignites "because" it is ordered to do so by the intension. It is not free to carry out the order immediately but must time its ignition with respect to the rest of the extension. In addition, it must previously have received reassurance that outside world is in a favourable state. When finally the chorus ignites, the individual neuron must fire with it; it must not delay, otherwise other neurons will not associate its firing with the rest of the chorus and its efforts will be wasted. (This is why I said that a person observing the firing of an active neuron has difficulty telling it apart from the firing of a passive one. The active neuron does not act like one that has "a will of its own" at all, as it always seems to be responding to some input firing.)

However, the rule-ridden behavior of neurons does not mean that they cannot have a "will". Since neurons are very small, and members of a community of very many, we must assume that they are ideally well disciplined, meaning that if they find that under certain circumstances they cannot get something, they come not to want it under those circumstances; and, conversely, if they find that under certain circumstances and in response to certain orders (which they merely regards as cues) they can get something, they come to want it whenever they recognize them again.

Let us note that under such assumptions we are able to formulate a "pure" sort of goal pursuit, the kind that is without any external reinforcement such as pleasurable or noxious stimulus. Just as the only reward of passive neurons was the thrill of having noticed an event of extremely low probability, the only reward of active neurons is the thrill of having wanted something (again, of extremely low probability), and then, with the help of their own efforts, having gotten it.

The flexible, centrally regulated threshold of the neurons (see also Legéndy, 1967) plays a role in the ignition of choruses, analogous to its role in the ignition of compacta, where it causes greater or lesser attentiveness. When no chorus is stimulated enough to ignite, the brain may lower the threshold let us say from 80 bits to 40, maybe further, until one chorus does ignite. The result is a coarser attempt by the organism to do something. However, it must be noted that even in such a case the chorus that does ignite is in possession of much encouraging information about the state of affairs in the outside world. This is the explanation of the fact that even when we do things by trial and error, our trials are fairly good guesses, and not at all out of touch with reality.

The Firing of Active Neurons

When the proper wish of a chorus has been satisfied, the neurons of the chorus can recognize the fact. Since they do make a recognition, and one that even gives them satisfaction, it would be unreasonable to assume that they made no response; it would not be in keeping with the "personality" of neurons as we sketched it out so far. It would also be wrong to assume that their firing were indistinguishable from the first ones issued. If it were, the neurons receiving it would interpret it as a call for a repeat performance.

The proposition that the same neurons that issue an order must emit a separate acknowledging signal later, when the order has been carried out, may also be inferred from the following consideration: The neurons of the chorus are the ones in which the "wish" was born in the first place; therefore they alone are qualified to pronounce it satisfied (see above). As soon as they do so (if they do), they have an obligation to call off the goal pursuit. It is not enough if they quietly stop wishing their wish; the rest of the brain must be notified. Otherwise it will continue *ad infinitum* to make renewed and varied efforts to secure satisfaction of a wish already satisfied.

The ability of each chorus to emit two kinds of firing is essential for instance in the consolidation of the extensions of skills, because it lets the intension tell them when they did the right thing and when they did not. It plays a role in the impression of skills, too, by telling neurons when their firing (possibly) has been a help.

Translated into words, the two kinds of firing convey the messages "we want A" and "we acknowledge with thanks that the order A has been satisfactorily carried out". Every recipient neuron feels on its own synapses the identity of the "we" in each of these messages (for each neuron is assumed to receive enough of synapses from each chorus to allow a unique correspondence), and the identity of the "we" is the information used by neurons to define "A". Since the same neurons emit the two messages, the recipient neurons can tell when the letter "A" means the same in them both.

There is a further discrimination that is information theoretically possible for the neurons to make: discrimination between firing that comes from active neurons and that coming from passive neurons. Since every neuron is either one or the other, neurons can, if they want to, mark their synapses as ones bringing firing from active or passive neurons.

Making the discrimination can be of use to the neuron, because firing of passive neurons always brings direct information about the state of the outside world. Therefore this discrimination, too, can be expected to find equal use in impression and consolidation.

References

Altman, J. (1967). "Postnatal growth and differentiation of the mammalian brain, with implications for a morphological theory of memory." In *The Neurosciences* (Eds. Quarton, Melnechuk, and Schmitt). The Rockefeller University Press, New York, pp. 723–43.

Hebb, D.O. (1949). *The Organization of Behavior*. Wiley, New York.

Legéndy, C.R. (1967). "On the scheme by which the human brain stores information." *Math. Biosci.*, **1**, 555–97.

Legéndy, C.R. (1968). "A possible scheme of data storage in the brain." *Proc. I.E.E.E. 7th Symp. on Adaptive Processes, Los Angeles, California, 1968.*

J.-C. Levy (1969). *Le Temps Psychologique*. Dunod, Paris.

J.-C. Levy (1970). This volume. Chapter II–7.

McCulloch, W.S. and W.H. Pitts. (1943). "A logical calculus of the ideas immanent in nervous activity." *Bull. Math. Biophys.*, **5**, 115–33 .

Ramón y Cajal, S. (1909). *Histologie du Système Nerveux de l'Homme et des Vertébrés,* Vol. II. Maloine, Paris. (Translated by L. Azoulay (1955). Inst. Ramón y Cajal, Madrid.)

CHAPTER II–7

A neuron model acceptable by computers

J. C. LÉVY

Centre d'Études et de Recherches en Automatisme
Velizy-Villacoublay, France

Summary

While a real neurological system is in continual evolution, a digital computer can operate only step by step. The problem is to define an optimal sampling time allowing the simulation of a continuous process. We define an average density of spikes, received by the neuron during a period Θ. Θ is for the instant, arbitrarly defined; it corresponds to a more or less sharp analysis of the process. Let t be the time of the middle of the period Θ. Then, a function $f(\Theta, t)$ is defined. Now, we use the curve $\varphi(\tau)$ of synaptic potential; this is related to the function. It may be seen that it is not worth it to take for Θ a too little value; this means that the system works as a filter which eliminates upper frequencies of $f(\Theta, t)$. Also, the individual position of each spike, can be neglected in a first approximation. It is then possible to introduce such a model into a computing program by using a corrective term that is easily calculable.

INTRODUCTION

The model which shall be described is to be introduced in assemblies, allowing the simulation by computers of a nervous system's comportment. As the calculus is necessarily sequential, we must define a sampling period $\bar{\theta}$. Let $E(t)$ be a vector defining the state of the system at the time t.

$$E(t) = F\,[E\,(t - \bar{\theta}), \quad E\,(t - 2\bar{\theta}) \ldots E\,(t - k\bar{\theta})]$$

The purpose of the present paper is to define a time base $\bar{\theta}$ and a number k as small as possible, allowing with a good approximation to calculate the evolution of a neuronic system.

Pulse response

We take as an experimental result the curve of variations of the post synaptic potential produced by an isolated spike which is assumed to arrive at time 0.

$$\varphi(t) = \varphi_c(t) + \varphi_r(t)$$

where $\varphi_c(t)$ is a certain function. $\varphi_r(t)$ is a random term which shall not directly be introduced in calculus.

We must be careful not to confuse a spike, with a Dirac impulse function which is a result of the product of the convolution.

Input Signal

In a first step of calculus, we consider an input signal corresponding to a given repartition of spikes. Using a method other than the one used by Jomannesma, we consider that signal as a certain function.

Let $f_0(t)$ be that function and $g_0(v)$ be its Fourier transform. The pulse analysis of such a function is possible even if each spike has a certain width or is considered as a Dirac pulse. Thus, we are able to define a post-synaptic potential (P.S.P.)

$$\Phi(t) = \int_0^\infty f_0\,(t - \theta)\,\varphi(\theta)\,d\theta$$

where

$$\varphi(\theta) = \varphi_c(\theta) + \varphi_r(\theta)$$

The term $\varphi_c(\theta)$ and the random function $\varphi_r(\theta)$ have about the same function because we consider only the width of their spectrum of frequencies, that is to said of their Fourier transform $\gamma_c(v)$ and $\gamma_r(v)$ such as

$$\gamma(v) = \gamma_c(v) + \gamma_r(0)$$

is an Fourier transform of $\varphi\theta$.

We can add to this P.S.P. a random term (Gerstein) $\phi_{rc}(t)$ which is independent of the input signal. It can produce neural noise emited by a neuron without any input signal. Let $\Gamma_r(\nu)$ be the Fourier transform of $\phi_r(\nu)$. We assume that, for a given ν_0

$$\nu > \nu_0 \Rightarrow \gamma(\nu) = 0$$

The theorem of Plancherel gives the Fourier transform of $\phi(t)$:

$$\Gamma(\nu) = g_0(\nu)\,\gamma(\nu) + \Gamma_r(\nu) = \Gamma_0(\nu) + \Gamma_r(\nu)$$

Components of $g_c(\nu)$ exceeding ν_0 are thus eliminated.

First Smoothing

$$\nu_0 = 1/\varepsilon$$

where ε is the width of each spike.

$$\nu \leq \nu_0 \Rightarrow g_1(\nu) = g_0(\nu)$$

$$\nu > \nu_0 \Rightarrow g_1(\nu) = 0$$

$f_1(t)$ which is the Fourier transform of $g(\nu)$ has the shape represented by Figure 1. This curve can be determined in following manner. We take the

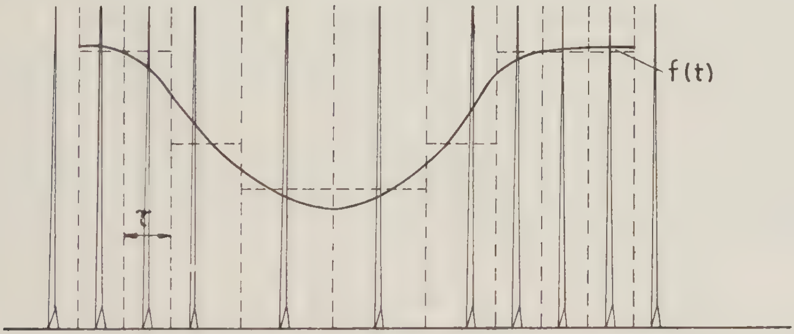


FIGURE 1 Fourier transform of g, (ν)

middle of each interval between two successive spikes .Let τ be the interval of time between two points so defined $f_1(t) = {}^1/\tau$ or $-{}^1/\tau$ according the fact that the spike is an excitatory spike or an inhibitory one. We obtain a step function which can be replaced by a continuously varying curve. The upper value of components ν of $f_1(t)$ are of the same order as τ^* where τ^* is the time between two successive spikes: that means about 10^4 pulses/second.

Such components can be found in the function $\varphi(\theta)$ which should hav the shape illustrated in Figure 2.

In that condition, variations of gap between two successive spikes could influence the P.S.P.: $\phi(t)$. But we have also to consider the comparison between $\phi(t)$ and the threshold S and to take into account the variations of

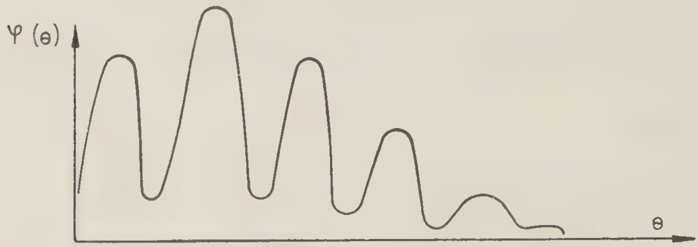


FIGURE 2 Shape of function $\varphi(0)$

this threshold during the refractory period. We do not use here the calculus similar to that one of Jomannesma. But we can assume a very weak correlation between the output signal and the position of each pulse.

But it should be possible to imagine that some components ν_i found in $\gamma(y)$ function, could allow the decoding of the same components in function $g_1(\nu)$.

This should be a very sharp analysis of the phenomena and for the time being, we are only mentioning such a possibility and are proceeding with further simplification.

Second Smoothing

Assume now that $\varphi(t)$ has a shape close to that Rayleigh's law. It is thus evident that ν shall have the same order of value as $^1/_\sigma$ or $^1/_2T$, where T is the time where $\varphi(t)$ has a non-negligible value and σ the parameter of the law of Rayleigh. Thus we obtain Figure 3. *The individual position of each spike has no more action in any way.* Function $f_2(t)$ can not vary more quickly than $\varphi(t)$ and the *neuron can be considered as a pulse density modulated system.*

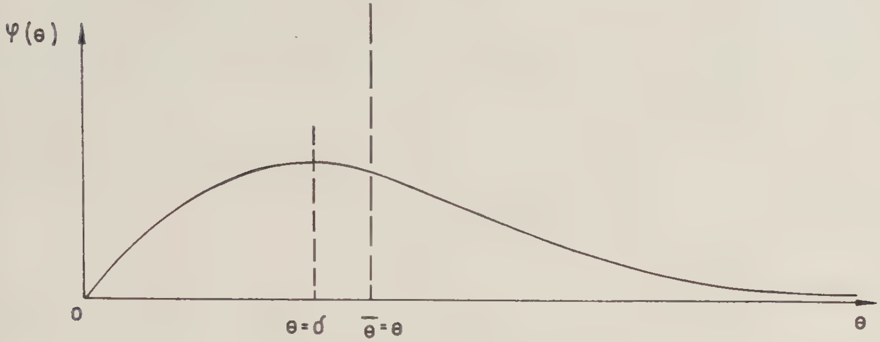


FIGURE 3 Second smoothing

Calculation of the Optimal Sampling Period

We now consider only the function $f_2(t)$ whose minimum period value has the same order as 2.

We take a limited development of the second order around the point t_0 of that function. Let $f_2'(t)$ and $f_2''(t)$ be the first and second derivatives of $f_2(t)$:

$$f_2(t) = f_2(t_0) + (t - t_0) f_2'(t_0) + \tfrac{1}{2}(t - t_0)^2 f_2''(t_0)$$

Let

$$M_0 = \int_0^\infty \varphi(\theta)\, d\theta$$

$$M_1 = \int_0^\infty \theta\, \varphi(\theta)\, d\theta$$

$$M_2 = \int_0^\infty \theta^2\, \varphi(\theta)\, d\theta$$

be the moments of order 0, 1, 2. The product of convolution

$$\Phi(t) = \int_0^\infty f_2(t - \theta)\, \varphi(\theta)\, d\theta$$

becomes

$$\Phi(t) = M_0 f_2(t_0) + f_2'(t_0)\,[M_0 (t - t_0) - M_1]$$
$$+ \tfrac{1}{2} f_2''(t_0)\,[M_2 + (t - t_0)^2 M_0 - 2(t - t_0) M_1]$$

We write

$$\bar{\theta} = t - t_0 = M_1/M_0$$

$$\Phi\left(t_0 + \bar{\theta}\right) = M_0 f_2\left(t_0\right) + \frac{1}{2} f_2''(t_0)\left[M_2 + \frac{M_1^2}{M_0} - \frac{2M_1^2}{M_0}\right]$$

$$= M_0 f_2\left(t_0\right) + \frac{1}{2} f_2''(t_0)\left[M_2 - \frac{M_1^2}{M_0}\right]$$

by noting that

$$\frac{M_1^2}{M_0} = \overline{\theta^2} M_0$$

$$\Phi\left(t_0 + \bar{\theta}\right) = M_0 f_2\left(t_0\right) + \tfrac{1}{2} f_2''(t_0)\left[M_2 - \overline{\theta^2} M_0\right]$$

The first term is the most important. The second term depends on $M_2 = M_2 - \theta^2 M_0$ moment of inerty of $\varphi(\theta)$ around its average value. It can only be considered as a corrective term, which is easy to calculate starting from

$$f\left(t_0 + \bar{\theta}\right), \quad f(t_0), \quad f\left(t_0 - \bar{\theta}\right)$$

It is also easy to calculate a third-order corrective term

$$\frac{1}{3!} f''(t)\left[M_3 + 3\,\frac{M_1 M_2}{M_0} - 2\,\frac{M_1^3}{M_0^2}\right]$$

where, of course, M_3 is the moment of the third order. According to the sharpness of calculus, we can have $k = 1$, or 2, or 3. For $k = 3$

$$\Phi\left(t_0 + \bar{\theta}\right) = M_0 f_2\left(t_0\right) + \frac{1}{2} f_2''(t_0)\left[M_0 - \theta^2 M_0\right]$$

$$+ \frac{1}{\sigma} f_2'''(t_0)\left[M_3 + 3\,\frac{M_1 M_2}{M_0} - \frac{2M_1^3}{M_0^2}\right]$$

A good representation of the function $\varphi(\theta)$ can, with a constant factor, be given by the law of Rayleigh, which, given the distribution of amplitudes of a Gaussian noise, is

$$\varphi(\theta) = \frac{1}{\sigma^2}\,\theta\, e^{-\frac{\theta^2}{2\sigma^2}}$$

where σ is the standard variation of the gaussian noise and allow the calculus of differents moments M_0, M_1, M_2, M_3 and period. The calculus not given here shows that

$$M_0 = 1$$

$$M_1 = \sigma \sqrt{\pi/2}$$

$$M_2 = 2\sigma^2$$

$$M_3 = 3 \sqrt{\pi/2}\, \sigma^3$$

After reductions, we have

$$\Phi\left(t_0 + \bar{\theta}\right) = f_2(t_0) - 0{,}872\, \overline{\theta^2}\, f_2''(t_0) + 0{,}038\, \overline{\theta^3}\, f_3'''(t_0)$$

CONCLUSION

The model which has been briefly described here, can be introduced in computing programs. This possibility results from a simplification which consist taking the neuron as a pulse density modulation system and neglecting the position of each spike.

The consideration of P.S.P. curve shows that this approximation is available.

CHAPTER II–8

The form recognition problem in neurological clinic: a cybernetic approach

M. I. BOTEZ and R. MAGUREANU
Polytechnic Institute, Bucharest, Rumania

Summary

In some visual recognition disturbances of central origin, i.e. visual agnosia for letters, the patient is unable to recognize a letter by static examination (by mere sighting) but can recognize it if the letter is drawn in the air or on a sheet of paper by the examiner's finger with a certain speed, i.e. in an optimum time interval. The same phenomenon has been noted for some geometrical figures or pictures. In forms of dissolution, visual and tactual, (in the Jacksonian sense) of sensory perception modalities, three phenomena appear evident, viz. the optimal time (above-mentioned), the role of motion in the process of facilitation of sensory recognition (either visually or tactually) and the role of the significant cue of the relevant. In the above mentioned agnosia forms recognition of shapes is possible only if the patient "grasps" the detail with the maximum informational value, i.e. the significant cue. A close analysis is made between adaptive and learning systems, on the one hand, and those of recognition of shapes based on the two sensory modalities already mentioned in agnosic patients, on the other hand. The authors consider the dissolutions in agnosias as various degrees of disturbance of the learning system.

INTRODUCTION

The object of the present study is to attempt a cybernetic approach of some shape recognition problems of the clinical field on the one hand, and of some adaptive and learning systems, on the other.

The neurological diseases sometimes provide experimental models of disintegration of the shape discrimination mechanisms. We shall deal here with the role played by motion in the visual perception of shapes and with the importance of the significant cue in compensating for the loss of the ability to recognize shapes, known in clinical neurology as agnosia.

In our previous investigations (Botez *et al.*, 1964) we found that in some visual recognition disturbances of central origin, i.e. visual agnosia for letters, the patient is unable to recognize a letter by static examination (by mere sighting), but can recognize it if the outline of the letter is traced in the air or on a sheet of paper by the examiner's finger at a certain speed, i.e. within an optimum time interval. The same phenomenon has been noted for some geometrical figures or some pictures.

PATIENT'S BEHAVIOUR (VISUAL AGNOSIA)

The electro-oculographic record as the patient was watching the outline of letters traced in the air by the examiner showed very slow and ample deflexions, corresponding to a *grasp of the gaze*. Whenever the patient could not follows a moving target with his eyes or any slowly moving object, sudden abrupt deflections were recorded, corresponding to the *wandering gaze* in search of a significant cue.

Maria S., a right-handed woman, aged 55, after several cerebrovascular strokes, exhibited slight receptive dysphasia. The tendon reflexes were jerky on both sides; there was right–left disorientation. There was a striking wandering of the gaze and great difficulty in fixing the gaze on objects. When she was trying to get hold of some object pointed to, set on the table in front of her, she did so by slow and unco-ordinated movements of her right hand, her eyes turned in another direction (upwards); she visually identified objects or colours lying in that direction. Even if, by chance, she recognized visually some objects that she "grasped" in the central vision, she could not grasp then with her hand since they immediately disappeared from the perceptive visual field. She could not follow with her gaze a moving target. The E.O.G. tracing recorded while the patient was seeking an object with her gaze, showed ample, slow, chaotic waves, which obviously differed from the E.O.G. recorded in the same patient when she quietly looking in front of her.

Gh. N., aged 62, has suffered from several cerebrovascular strokes. He displayed a slight receptive dysphasia; a mild agraphia, unilateral spatial neglect on the left side, a slight motor weakness for his right limbs, left facial paresis of central origin, right–left disorientation. His gaze was found to lack motility, remaining fixed on the midline. When he wanted to grasp an object at the examiner's request, he extended his hand in the wrong direction and only succeded after repeated groping. Just like in the case of the other two patients, although ocular movements to command and automatic movements were well executed, visual pursuit movements were disturbed. When his gaze "fell" on a moving target he could follow it only with jerky movements and only if it moved very slowly and smoothly. The E.O.G. record during the time the patient was trying to follow the moving target shows both normal features (sudden deflections) and pathological ones (more ample, slow deflexions, wherefore we shall call it *intermediate tracing.*) Here is the E.O.G. tracing at rest in the same patient. The clinical-electro-oculographic relationship is evident.

In all the above instances we were dealing with a visual recognition of a given symbol or shape in which the symbol was traced by a moving target and the object recognized when in motion.

By way of example, the first patient of our series was able to recognize only moving objects or persons. He could not identify visually any object on the table, a knife for instance, but he could recognize it in motion.

The same occurred on letter recognition tests. The patient was unable to recognize a letter printed on cardboard but recognized it by watching its outline traced in the air by the examiner. The speed of the moving target is of importance, there being an "optimum time interval" (Botez, 1961) for recognition to take place, comparable to that known in the neurophysiology of the peripheral nerves.

Thus, some patients with visual agnosia can recognize shapes through motion, perhaps on account of the release of some ancestral mechanisms that may be involved compensatorily. This view finds support in three categories of facts. (1) Firstly, it is well known that an excendingly great majority of birds and animals perceive moving things better than static ones. (2) Secondly, one should bear in mind that the prehistoric pictures in the caves represent moving humans and animals: the primitive man was, then, more impressed by the phenomena of nature in motion. (3) Finally, visual perception studies in the newborn infant showed that it perceives in particular the moving things, an electrooculographically demonstrated fact by Dayton and Jones in 1964. These investigators recorded the eyeball movements in normal newborn infants, and the tracings are almost identical to those obtained by Botez *et al.* in 1964 in agnosia patients on reading tests who were watching the examiner's finger tracing the outline of the letter in the air.

In the normal subject ocular motility is no longer playing the determinant role, for the retina and the specific visual pathways are sufficient to a proper sensory integration. This fact was demonstrated by experimental curarization of a normal subject, carried out on one of the authors.

In order to find role of the ocular muscle receptors in the normal subject for the visual recognition of motion, one of us (M.I.B.) asked that he be given an intravenous injection of 100 mg of succinylcholine following the subcutaneuous injection of 5 mg of d-tubocurarine.

Within the short interval (four and a half minutes) of paralysis among other tests, the test of visual recognition of letters was performed by the static method, by the slow motion examination, and by the dynamic method. After the effect of the drug had passed off, the subject named both the letters presented to him on separate pieces of paper and those drawn in the air by the examiner; he also named the pictures as well as the spatial relationships between the objects presented to him during the time of complete ocular paralysis. This may be readily explained in that the retina and the specific visual pathways, as well as the visual cortex, are preserved in the normal curarized subject.

A further problem of particular importance in sensory integration is that of the significant cue.

It was noted in visual agnosia patients that when their gaze fell upon a detail carrying a maximum informational value, they succeded in recognizing the letter, pictures, or objects. The patient developed new algorithms in the recognition of shapes. The importance of the significant cue is revealed also in tactile agnosia. At first the patient fails to recognize the objects by palpation, but subsequently they recognize them falling upon a significant cue. By way of example, the patient is unable to recognize a fork merely grasping it but he will succeed when stung by or feeling its prongs.

The patient fails to recognize the whole as ensemble, but can deduce it on the basis of a significant cue.

The technique of control systems has developed high-performance systems, adaptive and learning systems that adapt themselves to the new conditions created for an optimum operation, and in the case of learning systems the previously gained experience serves to improve the function to an optimum.

In normal subjects the cerebral activity can be compared with that of a superior learning system.

In agnosic patients with a loss of visual or tactile recognition of shapes there is a dissolution in the Jacksonian sense—on account of which the damaged brain operates with more primitive, with ancestral, mechanisms in a permanent effort of adjustment in the cybernetic sense and of compensation, from the clinical neurological point of view. Such disturbances involving the said ancestral and compensatory mechanisms can be compared to the situations realized by adaptive system in new created conditions.

Practically, the patient is able to recognize, after long examination, an object through one of the above mentioned methods, for instance the letter "D" traced in the air by the examiner within an optimum time interval, but he fails to recognize it any longer if he is again shown the printed model or if tracing is done within a shorter time interval.

During the rehabilitation process that may last months or even years, the patient is beginning to re-learn, but he will never exceed the level prior to disease in the field of higher intellectual functions. The cerebral adaptive system is becoming a learning system again, but it is a limited one.

Certainly, the shape recognition process is far more sophisticated, and our intention was to present but two, more evident aspects observed in visual agnosia cases, in a parallel between the biologic compensation systems and technical ones.

References

Botez, M.I. (1961). *Acta Neurol. Scand.*, **37**, 111–28.
Botez, M.I., T. Serbănescu, and I. Vernea (1964). *Neurology*, **14**, 1101–12.
Dayton, G.D. and M.H. Jones (1964). *Neurology*, **14**, 1152–6.
Gibson, J.E. (1963). *Nonlinear Automatic Control.* McGraw-Hill, New York.
Granit, R. (1955). *Receptors and Sensory Perception.* Yale University Press, New Haven, pp. 358–9.

CHAPTER II–9

Stochastic stimulation used to study spinal interneuronal system in man

MILAN R. DIMITRIJEVIĆ, LUDVIK GYERGYEK, and JANEZ K. TRONTELJ

University of Ljubljana
Yugoslavia

Summary

Rhythmic and regular cutaneous noxious stimulation in paraplegic patients with transverse division of the spinal cord has been shown to produce habituation of the responding structures which was ascribed to the interneuronal system of the spinal cord. To elucidate the function of habituation in more detail, stimulation was made random in frequency and/or amplitude. The standard noxious electrical stimulus was generated by a stimulator controlled by a computer, which was programmed by a stochastic function. The energy content of electromyographically detected responses was computed and correlated with stimulus amplitude and frequency functions. The results of the study are expected to contribute to the solution of certain problems in afferent functional stimulation used in external control of paralysed limbs.

INTRODUCTION

Habituation, i.e. diminution and eventual disappearance of the response to repeated stimulation, is a wide-spread phenomenon in the nervous system. It is this process which enables the central nervous system to neglect the constant parameters of the environment and to be selectively alert to changes, which may be more important for survival.

FIGURE 1 Habituation of the flexor reflex on regular repetitive stimulation. X axis: sequence of stimuli. Y axis: mean quantity of E.M.G. responses

A good model of habituation is the so-called flexor reflex in paraplegic patients with transverse division of the spinal cord. Diminution of the responses to repeated cutaneous noxious stimulation and their eventual disappearance in this case is largely the function of the interneuronal system of the divided part of the spinal cord[1,2] (Figure 1).

Fully or partly habituated responses can be transiently recovered by increasing the strength of stimulation. However, if stimulation is very intensive, no habituation will appear.

SUPPRESION OF HABITUATION

The question to be answered by this work was whether or not development of habituation can be prevented just by irregularity in the pattern of stimulation, without increasing the average strength of stimulation.

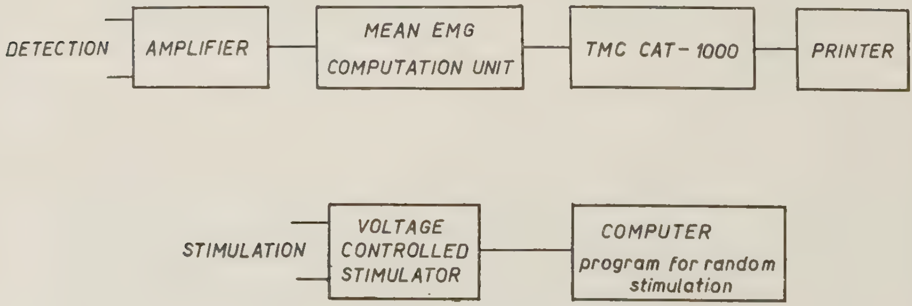


FIGURE 2 Block diagram of the measurement

For this purpose, the pattern of stimulation was made random either in frequency or in amplitude. The experimental arrangement[3,4] is shown schematically in Figure 2.

The results show that it is possible to overcome the process of habituation both by randomization of frequency and of amplitude of stimulation, the latter being more effective than the former (Figure 3 and 4).

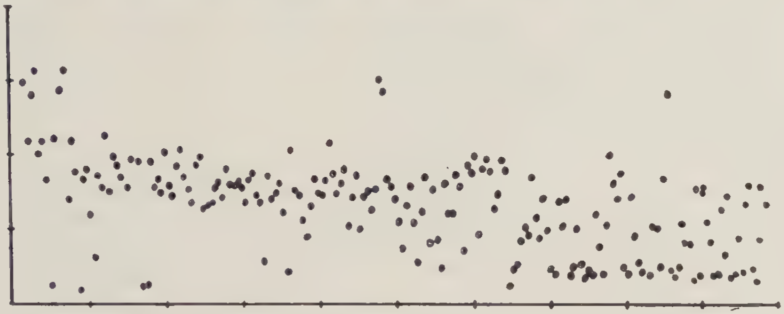


FIGURE 3 Suppression of habituation by random frequency of stimulation. Legend as in Figure 1

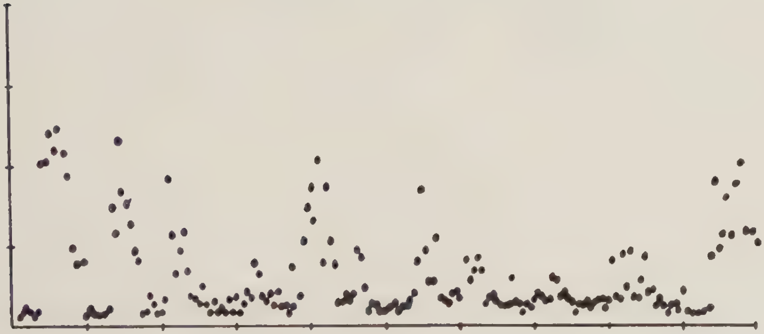


FIGURE 4 Suppression of habituation by random amplitude of stimulation. Legend as in Figure 1

As could be expected, habituation was more effectively suppressed when variance of frequency or amplitude of stimulation was greater; of course this variance must not be excessive if the variance of the responses is not to exceed too much that of regular stimulation.

Another approach to the problem of habituation used in this work was the attempt to compensate for the effects of habituation by changing the amplitude of stimulation by means of a feedback system (Figure 5).

Figure 6 shows the responses to feedback-controlled stimulation in comparison with regular and stochastic stimulation. It can be seen that variance of these responses is greatly reduced. Habituation could thus be very effectively suppressed, although the average stimulation strength was even less than in stochastic and regular stimulation.

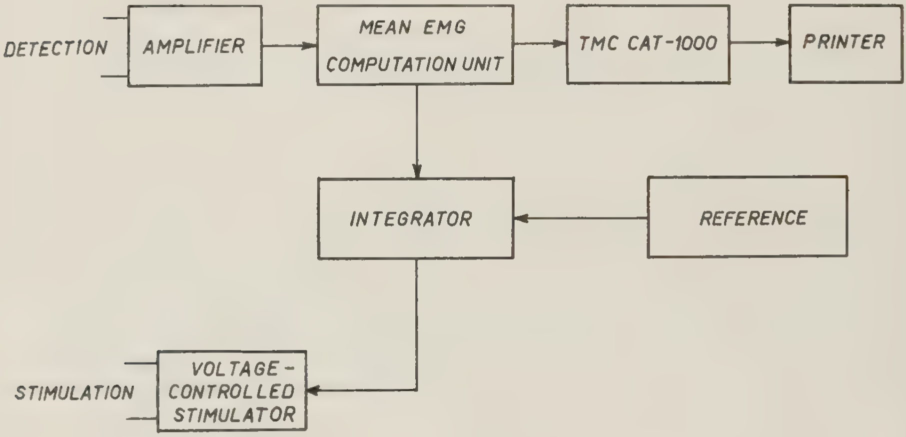


FIGURE 5 Block diagram of feedback system used for dishabituation

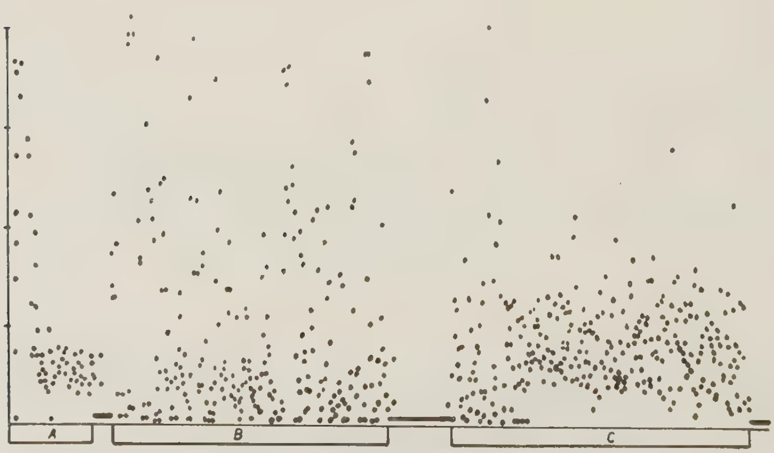


FIGURE 6 A: responses to regular stimulation. B: responses to stimulation stochastic in amplitude. C: responses to feedback-controlled stimulation

This implies that the process of habituation is subject to definitive laws, which, however, are modified by a variety of random and pseudo-random processes. Cross-correlation between the responses and stimuli can be expected to yield some new information of the behaviour of the mechanisms responsible for habituation. These data will be of great interest for controlling engineers working on external control of movement by functional stimulation[5].

References

1. M.R.Dimitrijević and P.W.Nathan, "Studies of spasticity in man. 3. Analysis of reflex activity evoked by noxious cutaneous stimulation", *Brain*, **91**, Part II, 349 (1968).
2. M.R.Dimitrijević and P.W.Nathan, "Studies of spacticity in man. 4. Responses to repetitive cutaneous stimulation", *Brain* (in press).
3. J.Trontelj, L.Trontelj, and J.Trontelj, "A voltage-controlled multi-channel electrical stimulator for programmed afferent functional stimulation", *7th Intern. Conf. Med. Biol. Eng., Stockholm* (1967).
4. J.K.Trontelj and J.V.Trontelj, "Two examples of preparation of biological signals for computation on a computer of average transients", *X. Intern. Automation Instr. Conf., Milan* (1968).
5. M.R.Dimitrijević, F.Gračanin, T.Prevec, and J.Trontelj, "Electronic control of paralysed extremities", *Bio-Med. Eng.*, **3**, 8 (1968).

CHAPTER II–10

Statistical theory and self-organization

A. M. ANDREW
University of Reading, U.K.

Summary

It will be shown that the task performed by certain self-organizing systems described in the literature is essentially the well-known statistical procedure of regression analysis. A comparison will be made of the ways in which these systems and the statistical procedure respond to redundant input information, and the response of animals will also be discussed. A modified form of the statistical procedure will be introduced, suitable for the continuous adjustment of parameters. This is more readily applicable to self-organizing systems than is the standard theory, which assumes "batch processing" of data. The extension of these ideas to systems embodying a threshold element in the output pathway will be considered, and it will be shown that extensions in different ways lead to the unusual perceptron training algorithm and to the forms of analysis due to Sebestyen. The theory will also be extended (by quoting and discussing an earlier published result) to the case where the self-organizing system receives no indication of how its output should have differed to produce a more favourable result. In this case its operation does not correspond to regression analysis. The relevance of the theory to further developments in self-organizing systems will be discussed, with particular reference to multi-layer systems and a new principle termed "significance feedback".

INTRODUCTION

The idea of self-organizing systems has been implicit, if not explicit, in much of what has been said and written under the heading of Cybernetics. Among workers who have discussed it may be mentioned Ashby[1–3], Beurle[4], Turing[5], and Farley and Clark[6]. At least three symposia[7–9] have been devoted to the topic. The reasons for studying it are (a) the desire to understand natural systems, particularly the nervous systems of man and animals, which appear to have powers of self-organization and (b) the possibility of constructing versatile learning automata.

There can be no guarantee that the study of self-organizing systems will achieve either of the desired ends. In the case of the nervous system it is difficult to separate the effects of heredity from those of learning during the lifetime of the individual and it is quite possible that the latter cannot be effective without a considerable amount of pre-organization due to the former. Nevertheless, a number of workers feel that work on self-organizing systems is potentially fruitful, provided the possibility of some degree of pre-organization is kept in mind.

DEFINITION OF SELF-ORGANIZATION

In spite of various attempts at rigorous definition, the idea of a self-organizing system must be partly intuitive. One attempt at definition is in the preface of Yovitz *et al.*[8] where it is stated: "A self-organizing system is a system which changes its basic structure as a function of its experience and environment." One characteristic which is missing from this definition is the idea of purpose; an abandoned motor car which rusts and falls apart is changing its basic structure as a function of its experience and environment, but cannot be termed self-organizing.

There is a difficulty in defining what constitutes a "change in the basic structure" of a system. The writer has sometimes tried to use the criterion that a system will be termed self-organizing if it changes its own internal connections, but should be described as self-optimizing if it only operates to change the values of certain parameters associated with its operation. This distinction cannot be a firm one, since the changing of certain parameters (e.g. gain controls) from zero to a finite value is functionally equivalent to the establishment of new connections. A definition must therefore depend on the intuitive principle that a system will be regarded as self-organizing if the changes which occur in it are more readily described as changes of structure than as parameter-changes.

There is another difficulty in the definition of any terms such as "adaptation", "learning", "self-optimization", or "self-organization" which refer to self-modification of a system. Whether or not a system shows self-modification depends on how much of its history is taken into account as the cause of its responses. Where only a small amount of history is considered the system may appear to be self-modifying although it can certainly be described as a fixed system if the whole of its history is taken into account as causing each response. This point is treated very clearly by Glushkov[10] in terms of automaton theory. He discusses the decomposition of a self-modifying system into a "learning automaton" and an "operative automaton". A system can only show self-modification if the description given of it at any instant is a description of the operative automaton only. Whether such a description seems appropriate to a given system is again a matter of subjective judgement.

Ashby[3] has suggested that almost any large, richly interconnected system will show self-organization, but I have not found his arguments convincing. I believe that he has now modified his views. It will be assumed here that a self-organizing system embodies a special mechanism, like a "perceptron training algorithm" to determine the changes due to experience. If the system is to start from a fairly low level of organization it is necessary for this mechanism to operate on information locally available to it. The information may include some signals which are available throughout the system, such as the measure of "hedony" or degree of goal-achievement, but apart from these the changes made in any part of the system should not depend on an overall view.

The mechanism is of most interest if it operates continuously, rather than by collecting data over an arbitrary sampling interval and subjecting them to analysis. The well-known perceptron training algorithms are acceptable in this respect, but have deficiencies which will be discussed later.

Some of the following discussion will refer to systems which provide only for adjustment of parameters and are therefore not truly self-organizing as discussed above. If a suitable mechanism for parameter adjustment can be found, it is usually not difficult to extend it to allow self-organization. The "Pandemonium" system of Selfridge[11] is an example of a system depending on parameter adjustment, with the added feature that a measure of the "worth" of elements within it is computed and used to determine structural changes.

Other examples are the systems of Foulkes[12] used to determine the statistical structure of sequences of characters, and Roberts' modification of the perceptron[13]. Some other possibilities have been discussed by Andrew[14–17].

PARAMETER-ADJUSTMENT AND REGRESSION ANALYSIS

A self-modifying system must evolve a policy which determines its response as a function of the inputs it receives from its environment. Depending on the nature of the environment it may or may not be necessary to take past values of these inputs into account. If the totality of signals from the environment, including a sufficient amount of their history, is termed a configuration of the environment, then some proposed schemes for self-modifying systems depend on exhaustive classification of the discriminable configurations.

In non-trivial environments the number of discriminable configurations is enormous and any system which relies on exhaustive classification requires a vast amount of information storage and must take a very long time to evolve a policy. It can only respond in a manner which utilises its past experience when it is faced with a configuration it has experienced at least once before. In other words it is incapable of inductive inference.

Where some or all of the inputs and outputs of the system are continuous signals rather than discrete ones, the continuity provides a simple measure of similarity between configurations, and hence the possibility of inductive inference. By processing the information in ways which preserve the continuity, the measure of similarity is automatically exploited. One way of doing this in a self-modifying system is by letting the output signals be computed as polynomial functions of the inputs. The self-modification can then consist of the adjustment of coefficients in the polynomial functions. A "learning filter" working in this way has been described by Gabor[18,19] and a similar scheme was proposed independently, though later, by Andrew[14]. These schemes have much in common with a great deal of other work on self-optimizing control systems.

It is assumed that a measure of the degree of goal-achievement resulting from the system's responses is available to it and is used in determining the modifications. Selfridge[20] has used the attractive term "hedony" for this measure; Wiener[21] has termed it "affective tone".

In some circumstances a self-modifying system may have only the simple hedony measure available to it, but in others the measure may be accompanied by an indication of how the response of the system should have differed to produce a more favorable outcome. This extra indication may be termed "error information" and it is always available when the goal of the system is to generate an output which matches or predicts some quantity which becomes available for comparison. The ways in which a self-modifying system may operate depend very much on whether or not error information is available.

Where precise error information is available, as in the case of prediction or matching, the task of the self-modifying system is precisely that which is carried out by statisticians as Regression Analysis. The method of operation devised by Gabor[18,19] does not assume the availability of error information. However, the tasks he set to his filter for purposes of demonstration were all ones in which error-information was in fact available (as they must be if the filter is to operate from recorded data, as his does). Consequently, Lubbock[22] was able to suggest an alternative form of operation which utilizes the error information and converges much more rapidly. His method is very similar to the standard procedure of regression analysis.

This procedure is presented essentially as follows in standard works, e.g. Weatherburn[23]. If x_1 is the variable whose value is to be estimated from x_2 and x_3, the estimate x_1' of x_1 is computed using the determinantal equation

$$\begin{vmatrix} x_1'/\sigma_1 & x_2/\sigma_2 & x_3/\sigma_3 \\ r_{12} & 1 & r_{32} \\ r_{13} & r_{23} & 1 \end{vmatrix} = 0 \tag{1}$$

where σ_p is the standard deviation of x_p and r_{pq} is the correlation of x_p with x_q. The extension to a larger number of variables is obvious.

Lubbock's method, instead of operating according to Equation (1), pre-processes the inputs, which will now be termed $x_1, x_2, \ldots, x_n$ to derive a set of signals $\theta_0, \theta_1, \theta_2, \ldots, \theta_n$ which are mutually orthogonal. This is done by setting

$$\begin{aligned} \theta_0 &= 1 \\ \theta_1 &= x_1 - a_{10}\theta_0 \\ \theta_2 &= x_2 - a_{21}\theta_1 - a_{20}\theta_0 \end{aligned} \tag{2}$$

and so on

where

$$a_{pq} = \overline{x_p\theta_q}/\overline{(\theta_q)^2} \tag{3}$$

The regression method can then be applied, substituting $\theta_0, \theta_1 \ldots$ for $x_2, x_3 \ldots$ in Equation (1). Application of the method is then simplified because

$$r_{pq} = 0 \quad \text{if} \quad p \neq 1 \quad \text{and} \quad q \neq 1 \tag{4}$$

and the determinantal equation becomes

$$x_1'/\sigma_1 = r_{12}x_2/\sigma_2 + r_{13}x_3/\sigma_3 + \cdots \tag{5}$$

or

$$x_1' = b_0\theta_0 + b_1\theta_1 + \cdots \tag{6}$$

where

$$\begin{aligned} b_p &= r_{1p}/(\sigma_1\sigma_p) \\ &= \overline{x_1\theta_p}/\overline{(\theta_p)^2} \end{aligned} \tag{7}$$

The x_1 appearing in Equation (7) is the true value of the variable to be estimated, whereas that in Equation (2) represents one of the variables on which the estimation is based.

Both the regression method and Lubbock's have been described in the case where only linear functions are computed. Either method is readily extended to allow non-linear functions by introducing further variables x_{n+1}, x_{n+2}, etc., as required, equal to products of the original variables, as

$$\begin{aligned} x_{n+1} &= x_2x_3 \\ x_{n+2} &= x_2x_4 \\ x_{n+3} &= x_2x_3x_4 \end{aligned} \tag{8}$$

CONTINUOUS OPERATION

The standard regression method has been devised for use by statisticians who collect and then process a batch of data. In a self-modifying system it is more satisfactory to compute the required statistics continuously, as in the well-known method (see Andrew[14], Muir[24]) for computing a running estimate m of the mean of a variable x as

$$m_i = (1 - k)\, m_{i-1} + kx_i \tag{9}$$

where m_i is the value of m following the i^{th} observation of x and k is a constant such that $0 < k < 1$. Where $k \ll 1$ this corresponds very closely to exponential smoothing of the values of x using a time constant equal to the time in which $1/k$ values of x are observed.

However, there are difficulties in using such running estimates to implement the regression equation (1). This can be seen by examining the familiar expression for the correlation coefficient

$$r_{pq} = \overline{(x_p - \bar{x}_p)(x_q - \bar{x}_q)} \Big/ \sqrt{\left\{\overline{(x_p - \bar{x}_p)^2}\ \overline{(x_q - \bar{x}_q)^2}\right\}} \tag{10}$$

This could be evaluated continuously by forming running estimates in the manner of Equation (9) for the numerator and each of the mean-squares terms in the denominator. However, each contribution to each of these running estimates has to depend on a running estimate $\bar{x}_p$ or $\bar{x}_q$. Hence the estimate of r_{pq} following the ith samples of x_p and x_q is influenced by the values of these running estimates following the $(i - 1)$th, $(i - 2)$th, etc., samples. The value of r_{pq} formed cannot therefore be said to result from any particular allocation of weights to the set of observations on which it is based and may be misleading if the mean levels fluctuate at a rate which is not slow compared to the effective time-constants of smoothing.

A similar objection applies to the straightforward modification of Lubbock's method to give continuous evaluation. The numerator and denominator of the r.h.s. of Equation (7) can be computed as running estimates, but in turn they depend on other running estimates in the derivation of the θ values.

The regression method can, however, be fairly easily reformulated so as to avoid the need to have running estimates based on running estimates. The modification allows the continuous evaluation of the regression coefficients in such a way that their values at any instant are precisely those which could have been obtained using the standard batch-processing method. The coefficients are computed on the whole history of the system, but with exponential weighting favouring more recent observations.

It is necessary to introduce a slight change of notation at this point. Let y represent the variable to be estimated and y' the estimate formed. Let $x_1, x_2, \ldots, x_n$ be the variables from which the estimate is to be formed, and take

$$x_0 = 1 \tag{11}$$

Let $k_0, k_1, \ldots, k_n$ be the regression coefficients, so that

$$\begin{aligned} y' &= k_0 x_0 + k_1 x_1 + \cdots \\ &= \sum_{p=0}^{n} k_p x_p \end{aligned} \tag{12}$$

If a mean-squares criterion of goodness of fit is used, the aim is to adjust the k_p so as to minimize

$$\sum (y - y')^2 = \sum (y - \sum_{p=0}^{n} k_p x_p)^2 \tag{13}$$

where the outer summation is over the set of observations of y, x_1, x_2, ... on which the regression is based.

Equating derivatives of (13) with respect to the k_p to zero gives the "normal equations" in the form (dividing throughout by the number of observations, so that summations are replaced by means)

$$\begin{aligned} \overline{x_0x_0}k_0 + \overline{x_0x_1}k_1 + \cdots, \overline{x_0x_n}k_n &= \overline{x_0y} \\ \overline{x_1x_0}k_0 + \overline{x_1x_1}k_1 + \cdots, \overline{x_1x_n}k_n &= \overline{x_1y} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot & \\ \overline{x_nx_0}k_0 + \overline{x_nx_1}k_1 + \cdots, \overline{x_nx_n}k_n &= \overline{x_ny} \end{aligned} \tag{14}$$

All the mean values appearing in (14) may be computed as running estimates by the method of Equation (9). This simply means that the regression is based on all observations which have been made, but with greater weight attached to recent ones and an exponential reduction of weight with increasing remoteness. Equations (14) can be solved to obtain values for the regression coefficients. This constitutes a method for their continuous evaluation without the need to let running estimates of values be computed from other running estimates.

CONTINUOUS ADJUSTMENT

In a self-organizing system values of $k_0, k_1, \ldots, k_n$ are being used to produce estimates y' of y at the same time as their values are being adjusted. In that case the regression method may be further modified so that what is computed is not $k_0, k_1, \ldots, k_n$ but $k'_0, k'_1, \ldots, k'_n$ where k'_p is the amount by which k_p must be changed. This is equivalent to fitting a regression function

$$\Sigma\, k'_p x_p \tag{15}$$

to the negative of the error of the estimate

$$(y - y') \tag{16}$$

The k'_p can be estimated from a set of equations similar to (14)

$$\overline{x_0 x_0} k'_0 + \overline{x_0 x_1} k'_1 + \cdots, \overline{x_0 x_n} k'_n = \overline{x_0 (y - y')}$$

$$\overline{x_1 x_0} k'_0 + \overline{x_1 x_1} k'_1 + \cdots, \overline{x_1 x_n} k'_n = \overline{x_1 (y - y')}$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$\overline{x_n x_0} k'_0 + \overline{x_n x_1} k'_1 + \cdots, \overline{x_n x_n} k'_n = \overline{x_n (y - y')} \tag{17}$$

Since the k'_p are the changes to be made in the k_p, it is necessary to suppose that after every m observations (where m may be unity) the values k_p are in fact modified by the addition of the values of k'_p obtained by solving Equations (17). It is then necessary to consider how the running estimates of mean values are affected by this change.

The mean values appearing on the left-hand sides in (17) are unaffected, but since y' has a different meaning before and after the changes, special consideration must be given to the mean values on the right-hand sides.

$$\text{Let } b_q = \overline{x_q (y - y')} \tag{18}$$

be one of the running estimates as it stands before the change of the k_p values. In this, y' refers to a function computed using the k_p values before the change. Let y'' refer to the estimates based on the k_p values after the change, and b'_q the appropriate value of the running estimate after the change.

$$\begin{aligned} b'_q &= \overline{x_q (y - y'')} \\ &= \overline{x_q [y - (y' + \Sigma k'_p x_p)]} \\ &= \overline{x_p (y - y')} - \Sigma k'_p \overline{x_p x_q} \\ &= 0 \end{aligned} \tag{19}$$

since the k'_p are roots of Equations (17) and hence must satisfy the qth equation.

Hence the algorithm which can be used for continuous adjustment of regression equations is as follows. Following the occurrence of a new set of observations of $y, x_1, x_2, \ldots, x_n$ all the mean values in Equations (17) should be updated according to the method of Equations (9). In the case of the values on the right-hand sides, if they were set to zero after the previous set of observations this amounts to setting the value

$$kx_q(y - y') \tag{20}$$

for the qth equation. k is the constant appearing in Equation (9).

If it is desired to adjust the regression coefficients at this stage of operation, the Equations (17) should then be solved for the k'_p, these k'_p added to the respective k_p values, and all the r.h.s. running estimates set to zero. If the k_p values are not changed, the r.h.s. running estimates should not be set to zero, and will be modified by the method of Equation (9) following the next set of observations.

Figure 1 shows some responses of an implementation of the above method. The horizontal scale in each graph represents the number of observations. For each observation, x_1, x_2, and x_3 were assigned values from a pseudo-random generator. For x_2 and x_3 the values were selections from a rectangular distribution over the range -1 to $+1$, but x_1 was from 0 to 2. The "true value" y was computed as

$$y = k_0 + k_1x_1 + k_2x_2 + k_3x_3 \tag{21}$$

where k_0, k_1, k_2, and k_3 are "true values" of the coefficients, initially set at 0, 1, 2, and 3 respectively, but changed as indicated in the subtitles at time zero. The graphs represent the estimates of k_0, k_1, k_2, and k_3 formed by the regression method, with adjustment of the regression coefficients following every observation (i.e. $m = 1$). For each of the runs shown in Figure 1 the settings of the running estimates $\overline{x_px_q}$ were zero at the start of the run at $t = -4$, except for the following:

$$\overline{x_0x_0} = 1 \qquad \overline{x_1x_1} = 4/3$$
$$\overline{x_2x_2} = \overline{x_3x_3} = 1/3$$
$$\overline{x_0x_1} = \overline{x_1x_0} = 1 \tag{22}$$

That is to say, account was taken in these initial settings, of the fact that the mean of x_1 was non-zero but not of the special conditions existing from the start in runs (c), (d), and (e). The value of k used in updating all the running estimates was 0.1, corresponding to a time constant of smoothing of ten on the scale shown.

Figure 2 shows the results of a similar set of runs using an adaptation of Lubbock's method in which the coefficients are adjusted following each observation. In this the time-constant of smoothing for the running estimates

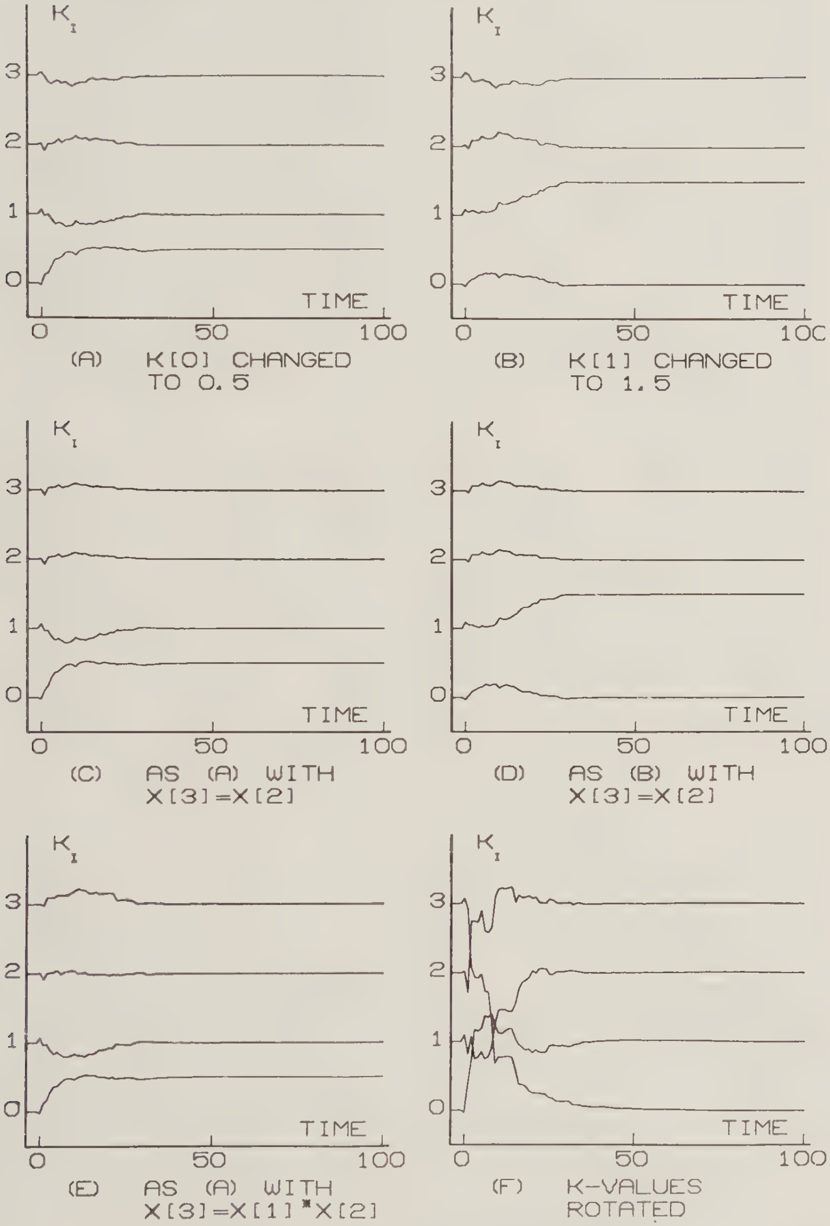


FIGURE 1 Regression. One time constant

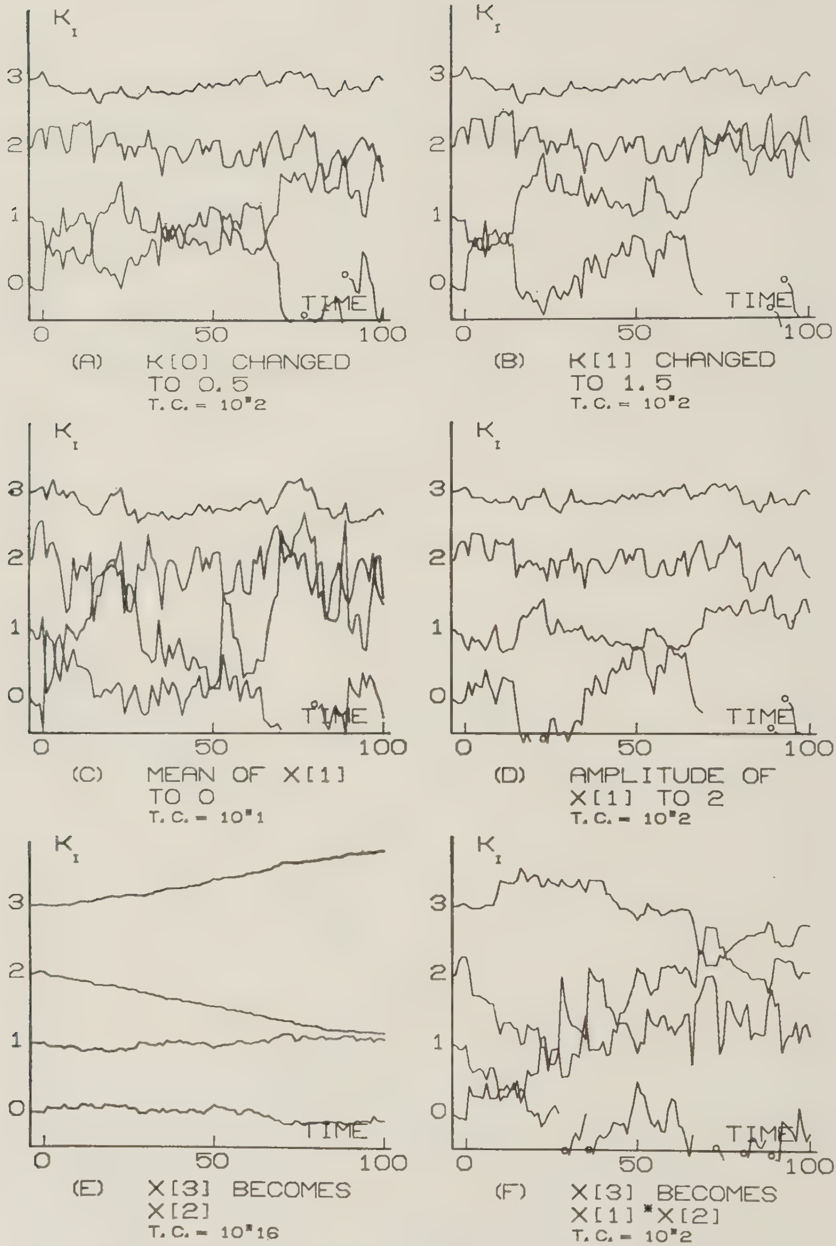


FIGURE 2 Lubbock method with running adjustments

used to compute the regression coefficients approximately according to Equation (7) was again ten, but that used in deriving the θ values according to Equation (3) was a multiple of ten as shown in the subscript of each graph (in which the asterisk takes the place of a multiplication sign). Each of these runs was carried out initially with this time constant at ten, but if any estimated k value deviated from its "true" value by more than 2.0 at any point in the run, or by more than 1.0 at the termination ($t = 100$), the run was repeated with the time constant doubled.

It can be seen that there is interaction between the two processes involving time constants, leading to large fluctuations in the estimates of the k values. The changes in conditions which were made at time zero in runs (c) to (f) are changes which do not affect the k values. These changes are not represented in Figure 1 because under the conditions of the runs in Figure 1 they produced no fluctuations of the estimated values whatsoever. It is clear from Equations (17) that the method of Figure 1 makes no changes in the estimated k values so long as there is no discrepancy between y' and y, and thus there are no changes in the estimates so long as the true values are undisturbed.

It has been stated that the adjustment of the regression coefficients in the modified form of Lubbock's method was approximately according to Equation (7). To be more precise, the ith value of each coefficient was computed as follows:

$$b_{pi} = b_{p(i-1)} + 0.1\,(y - y')\,\theta_i/\overline{\theta_i^2} \tag{23}$$

where b_{pi} is the estimate of b_p following the ith observation.

The coefficients b_p do not correspond to the coefficients k_p used in computing the "correct value" y according to Equation (21). The k_p are coefficients in a linear combination of the x_p, whereas the b_p are coefficients in a linear combination of the θ_p, where the relationship between the x_p and the θ_p is represented by Equation (2).

The estimates of the k_p graphed in Figure 2 are derived from the b_p by the following transformations:

$$k_0 = b_0 - b_1 a_{10} - b_2(a_{20} - a_{21}a_{10}) - b_3(a_{30} - a_{31}a_{10} - a_{32}a_{20} + a_{32}a_{21}a_{10})$$

$$k_1 = b_1 - b_2 a_{21} - b_3(a_{31} - a_{32}a_{21})$$

$$k_2 = b_2 - b_3 a_{32}$$

$$k_3 = b_3 \tag{24}$$

where a_{pq} is as defined by Equation (3).

The method used in obtaining Figure 1, depending on the solution of Equations (17), is the ideal way of forming continuous estimates of the regression coefficients with exponential weighting of observations for recency. The method is quite "robust" in that small variations in it have little effect. For instance, a set of runs exactly like those shown in Figure 1, but in which the time constant used in smoothing $\overline{x_2 x_3}$ $(= \overline{x_3 x_2})$ was double that used in smoothing all the other product terms, gave records indistinguishable from those of Figure 1. Reduction of the multiplying factor k in Equation (20) from 0.1 to 0.08, while 0.1 was used elsewhere in the computation, also failed to produce a perceptible difference. Raising the value of this factor to 0.25 produced the results shown in Figure 3. This is very similar to Figure 1, but has an "underdamped" appearance.

REDUNDANCY

The essential difference between Lubbock's method and the standard regression method lies in the way they deal with the case where the $x_1, x_2, \ldots, x_n$ are not orthogonal over the set of observations. In the context of self-organizing systems the non-orthogonality would be referred to as redundancy among the set of inputs.

Redundancy may be due to one or more inputs being linear combinations of others, or simply being exactly equal. Suppose for instance that $x_p = x_q$ in all observations, where $p \neq q$. If Lubbock's method is used, only one of x_p and x_q comes to be utilized in the estimation of y. Which one is utilized depends on the order in which the variables are treated according to Equation (2). If the attempt is made to apply the standard regression method it is found that the set of simultaneous equations to be solved has no unique solution. Solutions can be found by fixing some of the regression coefficients arbitrarily and solving for the rest; these include solutions which leave one or other of x_p and x_q unused.

Some evidence is available from the work of Lovejoy[25,26] on the behaviour of animals learning to respond to two simultaneous, equivalent stimuli. It appears that they come to respond to one stimulus only, as can be determined by observing the responses when the stimuli are allowed to become non-equivalent. Which stimulus is chosen may be influenced by the previous experience of the animal.

If it is assumed in advance that the inputs will be orthogonal, both Lubbock's method and the regression method can be greatly simplified. In the

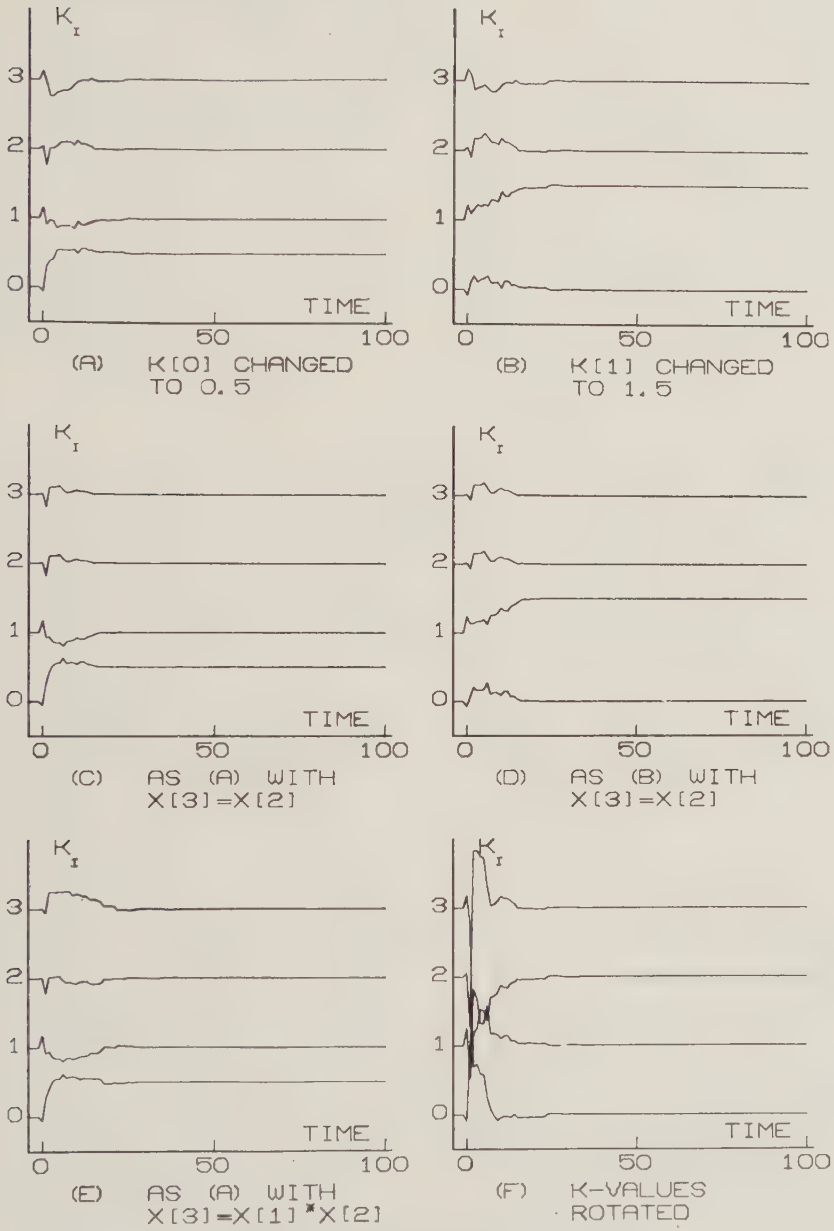


FIGURE 3 Regression. Error product $\times$ 0.25

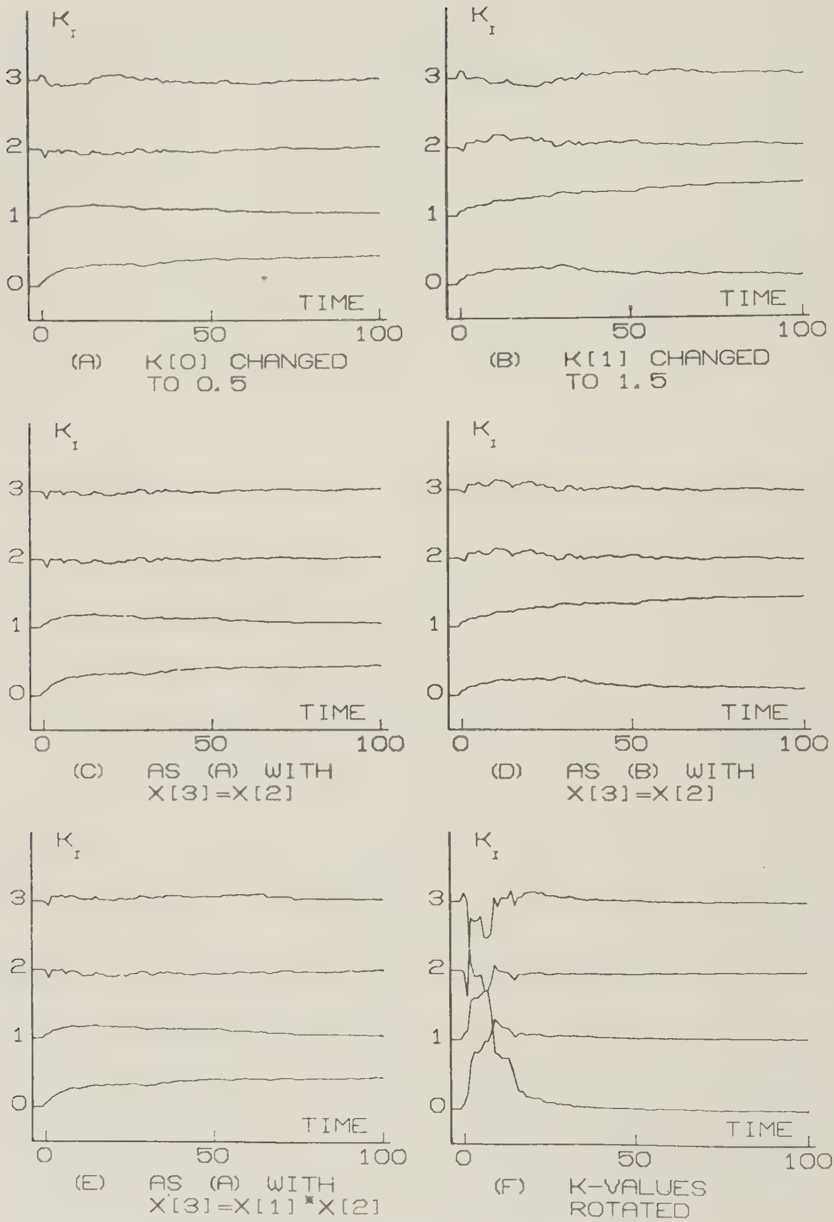


FIGURE 4 Regression. No cross-correlation

Lubbock case it is then permissible to set

$$\theta_q = x_q \tag{25}$$

for all q.

In the regression case all the smoothed products $\overline{x_p x_q}$ for which $p \neq q$ can be set to zero in Equations (14) or (17), and Equations (14) become

$$k_q = \overline{x_q y}/\overline{(x_q)^2} \tag{26}$$

which, in view of Equation (25), is equivalent to Equation (7). Hence, if the assumption of orthogonality is made in advance, the two methods are equivalent.

Similarly, ignoring the cross-products in Equations (17) gives

$$k'_q = \overline{x_q (y - y')}/\overline{(x_q)^2} \tag{27}$$

in which the numerator takes the form of Equation (20) if an adjustment of the k values is made following each observation, and the overall effect on the k values is exactly like the change in b values represented by Equation (23).

Operation according to Equation (27) is more readily implemented in a self-organizing system than operation depending on the solution of a set of simultaneous equations like Equations (17). This follows from the general principle stated earlier that parts of the system should operate on information locally available to them.

Figure 4 shows records from a set of runs like those of Figure 1, but operating according to Equation (27) instead of by solving the set of simultaneous Equations (17). It will be seen that operation is little impaired, even though the mean of x_1 was non-zero, and in runs (c) and (d) the inputs x_2 and x_3 were equal.

It is intuitively obvious, and borne out by these results, that operation according to Equations (17) is more "robust" than operation according to Equations (14). The former is an error-actuated or closed-loop form of operation and therefore can converge on correct values of the regression coefficients even when their cross-correlations are ignored.

However, redundancy cannot always be dismissed so lightly. Suppose there are n of the variables x_p which are equal to one another. Then the component of $(y - y')$ correlated with them is likely to be much larger than that which would be correlated with any one of them if they were mutually orthogonal. Hence, if the changes in the k values are determined by Equation (27), that in each k value may be much greater than it would be in the case of orthogonal inputs. Furthermore, all n of the k values are changed in the same direction by the same amount, further amplifying the total effect by a factor of n. Such amplification must lead to instability for sufficiently large n, so it is not always permissible to ignore the possibility of redundancy. It could possibly be ignored as such if the system embodied a means of reducing the magnitude of the changes in a parameter whenever its value became unstable.

DISCRETE OUTPUT

If a self-organizing system is applied to pattern recognition, what is required as output is not an estimate y' of a continuously variable quantity y, but an indication whether or not the pattern belongs to a particular category. In devices of the "perceptron" type, a quantity is computed in the same way as y' and the output indication is determined by comparison of this with a fixed threshold. A two-valued "correct answer" becomes available for comparison.

One way in which a system may operate in this case is by defining a function y having one fixed value, say unity, whenever the pattern presented belongs to the category, and another fixed value, say minus unity, when it does not. Then if $x_1, x_2, \ldots, x_n$ are the measures of pattern features on which recognition is to be based, the regression method as described can be applied to make y' approximate y as closely as possible. If the decision threshold is set at zero this should lead to correct classifications.

The general principle of adjusting a system to compute a function y' approximating a function y in this way is suggested by Sebestyen[27]. He uses a method of derivation very different from that given here.

The Sebestyen method is an entirely practical one, as he demonstrates, but the requirement that y' should approximate to a two-valued function y as described above is unnecessarily restrictive. All that is important is that every computed value of y' be on the correct side of the threshold. The attempt to make it lie close to $+1$ or -1 probably makes it deviate more than it need from the optimal contour where it crosses the threshold.

An alternative is to let the regression coefficients be altered only when a wrong classification is made, and to let the change be determined only by the comparison of this classification with the correct one. A fairly obvious modification of Equation (27), after substituting from Equation (20) for the numerator, is

$$k'_q = Dcx_q \tag{28}$$

where c is a constant and

$D = +1$ if pattern presented belonged to the category but was classified as not belonging
$= -1$ if pattern did not belong to the category but was classified as belonging (29)

Equation (28) is in fact the usual perceptron training algorithm as presented by, among many others, Nilsson[28]. In his terminology, a vector W, termed the weight vector, corresponds to the set of coefficients $k_1, k_2 \ldots k_n$. A pattern vector Y corresponds to the set of inputs $x_1, x_2 \ldots x_n$. Then if Y_k is the kth pattern to be presented, let W_k and W_{k+1} be the weight vectors before and after its presentation. Let $\mathscr{Y}_1$ be the set of pattern vectors belonging to a category and $\mathscr{Y}_2$ the set which do not. The training algorithm for a perceptron is then stated as follows:

$$\begin{aligned}
W_{k+1} &= W_k \quad \text{if} \quad Y_k \cdot W_k > 0 \quad \text{and} \quad Y_k \in \mathscr{Y}_1 \\
W_{k+1} &= W_k \quad \text{if} \quad Y_k \cdot W_k < 0 \quad \text{and} \quad Y_k \in \mathscr{Y}_2 \\
W_{k+1} &= W_k - cY_k \quad \text{if} \quad Y_k \cdot W_k \geqslant 0 \quad \text{and} \quad Y_k \in \mathscr{Y}_2 \\
W_{k+1} &= W_k + cY_k \quad \text{if} \quad Y_k \cdot W_k \leq 0 \quad \text{and} \quad Y_k \in \mathscr{Y}_1
\end{aligned} \tag{30}$$

corresponding exactly to what was described above.

It has been shown that the above algorithm must converge, subject to certain conditions. Two distinct methods of proof are presented by Nilsson; these are selected from many. What is very remarkable is that the conditions do not include any requirement of orthogonality among the components of Y. The method is independent of redundancy among the inputs.

It is worth noting in passing that the convergence proofs apply also to some variations of the above algorithm in which c is not a constant. The conditions for convergence are that a weight vector exists which will perform the required discrimination, and that the "training sequence" presented to the system is such that the following hold.

(1) Every member of the sequence is a member of $\mathscr{Y}$, the union of $\mathscr{Y}_1$ and $\mathscr{Y}_2$.
(2) Every element of $\mathscr{Y}$ occurs infinitely often in the training sequence.

OPERATION WITHOUT ERROR INFORMATION

The discussion up to now has been entirely of systems having error information available to them. Yet another approach to the study of such systems is that of Donaldson[29] who talks of "error decorrelation". His treatment leads to results essentially similar to those discussed here.

Where error information is not available it is necessary to impose experimental fluctuations either on the parameters to be adjusted or on the control signals from the system to its environment. The latter appears to be preferable.

A system operating without error information may work in either of two ways. It may embody a mathematical representation or "model" of the environment or it may not. Where it does embody a model, it must operate to improve the model as it gains experience. Then the optimal control actions are derived from the model, either by direct computation or by using the model as a high-speed analogue to evaluate possible actions. The alternative is to modify the control policy directly, without the intermediary of a model.

The purpose of the model is to predict the degree of goal achievement, or hedony, resulting from actions of the system in the context of different inputs from the environment. For adjustment of the model, error information is available, even though it is not available to the system as a whole. Consequently, when a system without error information embodies a model, all of the foregoing theory is readily applicable.

It has been shown (Andrew[30]) that a system which adjusts its control policy directly may be mathematically equivalent to one operating by means of a model, provided the adjustment of policy is done in a particular way. Certain simplifying assumptions were made in this treatment, namely that the inputs to the system were mutually orthogonal and that the model was a polynomial of second degree. Also the treatment is in terms of "batch processing" of statistical data rather than "running estimates" as discussed here. Nevertheless the treatment provides at least the framework of a bridge for the carry-over of results from the relatively amenable case of operation with error information to the relatively intractable case of operation without it.

MULTI-LAYER OPERATION

The operating principles for self-organizing systems which have been discussed are only applicable when the parts of the system which undergo changes are parts which contribute directly to the output. In the polynomial-adjusting schemes of Gabor[18,19] and Andrew[14] the polynomial terms embodying the adjustable coefficients are summed to form the output. Similarly, in a perceptron, the connections whose weights are adjusted provide the inputs to a response unit which forms an output of the system. The adjustable elements in these systems are all in a single functional layer.

It is reasonable to suppose that self-organizing systems could be more versatile if the adjustments were not restricted to a single functional layer. It is not easy then to find a suitable "training algorithm". One suggestion has been made by Andrew[31]. The regression method can be re-formulated in terms of the sensitivity of the output of the system to changes at different points within it. Where the adjustable parts contribute directly to the output, this sensitivity is very easily determined. To determine it at points throughout a system, a sub-system of special information pathways has been suggested. These would perform a function to which the name of "significance feedback" has been given. It allows a continuous, automatic sensitivity analysis (Tomovic[32]) of the system as it operates and thus allows the use of a suitable training algorithm in the multi-layer case.

The writer is conducting experiments with multi-layer systems of neuron-like elements embodying significance feedback. They are applied to fairly simple concept-forming tasks of the kinds discussed by Hunt, Marin, and Stone[33]. Preliminary results look interesting but the study is still at an early stage.

CONCLUSIONS

Statisticians and self-organizing systems are both required to operate as best they can on whatever information is available. The main purpose of this paper has been to show how a procedure developed by the former can be adapted for embodiment in the latter. The treatment as it stands is applicable only to systems which (a) have error information available to them and (b) make adjustments in only one functional layer. Ways of overcoming these limitations are discussed.

Acknowledgement

I am indebted to Professor P. B. Fellgett for interest and encouragement and to the Computer Unit of Reading University for an efficient and amiable service.

References

1. W.R.Ashby, "Design for a brain", *Electron. Eng.*, **20**, 379 (1948).
2. W.R.Ashby, *Design for a Brain*, Chapman and Hall, London (1952), (1960).
3. W.R.Ashby, "Principles of the self-organizing system", in *Principles of Self-organization* (Eds. H. Von Foerster and G.W.Zopf), Pergamon Press, Oxford (1962), p.255.
4. R.L.Beurle, "Properties of a mass of cells capable of regenerating pulses", *Phil. Trans. Roy. Soc., London, Ser. B*, **240**, 55 (1956).
5. A.M.Turing, "Intelligent machinery", *Rept. Natl. Phys. Lab., Teddington* (1948).
6. B.G.Farley and W.A.Clark, "Simulation of self-organizing systems by digital computer", *I.R.E. (Inst. Radio Engrs.), Trans. Inform. Theory*, PGIT **4**, 76 (1954).
7. M.C.Yovits and S.Camerson, (Eds.), *Self-organizing Systems*, Pergamon Press, Oxford (1960).
8. M.C.Yovits, E.T.Jacobi, and G.D.Goldstein (Eds.), *Self-organizing Systems*, Spartan Books, Washington (1962).
9. H.Von Foerster and G.W.Zopf (Eds.), *Principles of Self-organization*, Pergamon Press, Oxford (1962).
10. V.M.Glushkov, *Introduction to Cybernetics*, Academic Press, New York (1966).
11. O.G.Selfridge, "Pandemonium: a paradigm for learning", in *Mechanization of Thought Processes*, H.M.S.O., London (1959), p.513.
12. J.D.Foulkes, "A class of machines which determine the statistical structure of a sequence of characters", *I.R.E. (Inst. Radio Engrs.), WESCON Conv. Record*, Part 4, p.66 (1959).
13. L.G.Roberts, "Pattern recognition with an adaptive network", I.R.E. (Inst. Radio Engrs.), *Intern. Conv. Record*, Part 2, p.66 (1960).
14. A.M.Andrew, "Learning machines", in *Mechanization of Thought Processes*, H.M.S.O., London (1959), p.473.
15. A.M.Andrew, "Self-optimizing control mechanisms and some principles for more advanced learning machines", *Automatic and Remote Control (Proc. I.F.A.C. Congr.)*, Butterworth, London (1961), p.636.
16. A.M.Andrew, "An experimental comparison of some algorithms for self-organizing systems", *I.R.E. (Inst. Radio Engrs.), Trans. Inform. Theory*, **IT-8**, No.5, 163, 1962.
17. A.M.Andrew, "Learning systems", in *Automaton Theory and Learning Systems* (Ed. D.J.Stewart), Academic Press, New York (1967), p.107.
18. D.Gabor, "Communication theory and cybernetics", *I.R.E. (Inst. Radio Engrs.), Trans. Circuit Theory*, CT-**1**, No.4, 19 (1954).
19. D.Gabor, W.P.L.Wilby, and R.Woodcock, "A universal non-linear filter, predictor and simulator which optimises itself by a learning process", *Proc. Inst. Elec. Engrs. (London)*, Part B, **108**, 422 (1961).
20. O.G.Selfridge, "Pattern recognition and learning", in *Information Theory* (Ed. E.C. Cherry), Butterworth, London (1956), Paper No.33.
21. N.Wiener, *Cybernetics*, Wiley, New York (1948), p.150.
22. J.K.Lubbock, "A self-optimizing non-linear filter", *Proc. Inst. Elec. Engrs. (London), Part B*, **108**, 439 (1961).

23. C.E. Weatherburn, *A First Course in Mathematical Statistics*, Cambridge University Press, Cambridge (1952), p. 247.
24. A. Muir, "Automatic sales forecasting", *Comput. J.*, **1**, 113 (1958).
25. E. Lovejoy and D.G. Russell, "Suppression of learning about a hard cue by the presence of an easy cue", *Psychon. Sci.*, **8**, 365 (1967).
26. E. Lovejoy, "Monte Carlo simulation of an attention model for discrimination learning", *Meeting of A.I.S.B. Study Group of Brit. Comput. Soc.*, *22nd March* (1967).
27. G.S. Sebestyen, *Decision-making Processes in Pattern Recognition*, Macmillan, New York (1962), p. 61.
28. N.J. Nilsson, *Learning Machines*, McGraw-Hill, New York (1965), Chap. 5.
29. P.E.K. Donaldson, "Error decorrelation: a technique for matching a class of functions", *Proc. 3rd Intern. Conf. Med. Electron., London* (1960), p. 173.
30. A.M. Andrew, "To model or not to model", *Kybernetik*, **3**, 272 (1967).
31. A.M. Andrew, "Significance feedback in neural nets", *Rep. Biol. Comput. Lab., Univ. of Illinois, Urbana* (1965).
32. R. Tomovic, *Sensitivity Analysis of Dynamic Systems*. McGraw-Hill, New York (1963).
33. E.B. Hunt, J. Marin, and P.J. Stone, *Experiments in Induction*, Academic Press, New York (1966).

CHAPTER II–11

*Algebra of the logical deductive intelligent activity**

ANIO O. ARIGONI

Instituto di Fisica

Università di Ferrara Italy

Summary

The functions developed by either living or mechanical devices can be considered to belong to three different sorts: informative detection, attraction, and elaboration. The devices of elaboration are considered almost exclusively from the qualitative point of view; they are in fact subdivided into logic and analogic; in the two derived large classes they are distinguished by several different kinds of functions that the devices are able to carry out. A characteristic, which is common to all the elaborative devices, that peculiarly distinguishes them from the others, is the quantity of intelligent activity they are capable of developing. In the author's views intelligence is, in the present case, understood as a qualification of strictly logic functions, which are carried out by either living or mechanical devices. It is proposed to enumerate the qualities that a device must possess to be able to show intelligence.

INTRODUCTION

One among the most promising and challenging fields of cybernetics is certainly the study of intelligence. But in trying to devote ourselves to it, it is quite natural to pose some questions. What that study really is? What it

* Abridged by excluding some mathematical proofs.

consists of? Yet, what contribution could cybernetics yield, since philosophy has already studied intelligence during millennia?

It is difficult enough to answer the above questions; nevertheless, it is possible to devise a work program to broach this very complex subject. One way consists of analysing quantitatively the different activities operated by three salient interwoven moments of intelligent deductive behaviour*: memorization, discrimination, and decision.

By basic studies it is proposed to establish a procedure for quantifying the decisional intelligent activity[1,2], by considering some fundamental characteristics of the same activity[3,4], as well as some practical implications[5,6]. The epistemological point of view is considered[7] and the study extended to applications of networks of intelligent elements. The object of the study is to investigate how artificial or natural organs are involved in intelligent activity.

THEORETICAL PRELIMINARIES

The first quantification of the intelligent activity had been made by L. Szilard in considering Maxwell's paradox[8]. The activity of the well-known demon, subject of the paradox itself, constituted a valid model to which we referred our studies; it is a decrease of entropy that the demon provokes by its operation, that is, the quantification of its activity of intelligent and logical kind. That decrease, termed "negentropy" by L. Brillouin[9], constitutes the quantity of decision due the corresponding dissipation of information.

The activity of Maxwell's demon is the analogue[11] of a logical elementary device[14]. If an analogy of the operation of a complex intelligent organ were required, it would be necessary to imagine a team of demons working on a mass or masses of gas, depending on how the elementary devices embodging the organ itself are connected among themselves. The basic formulation of the quantity of intelligent activity or elaboration, E, which can be proved to be equivalent[1] is given by the following two equations:

$$E = \text{antilog}_2 (H_A - H_E) \tag{1}$$

$$E = \frac{i_r}{i_i} \tag{2}$$

* Intelligent logical behaviour is manifested by a device implementing a logical function, which by receiving information gives a response consisting of elaborated information in accordance with logical criteria.

In (1) and (2) the symbols i_r and H_A represent respectively the quantity of subjective indecision for recognition based on supplied information, and the negentropy of the information itself; i_i and H_E represent the subjective indecision of interaction of the efferent information with its entropy.

The quality or effectiveness of the elaborated information, that is its degree of response to the aim for which an elaboration is performed, is expressed by an index whose quantity is given by[4]

$$Q = H_A \cdot E \tag{3}$$

Hence

$$Q = (H_A)^2 - (H_A \cdot H_E) \tag{4}$$

The index of effectiveness has the important properties, which are derived from (Equation (4))

a. $H_E = 0 \quad \rightarrow Q = \text{maximum}$

b. $H_E < H_A \rightarrow Q > 0$

c. $H_E = H_A \rightarrow Q = 0$

d. $H_E > H_A \rightarrow Q < 0$

It may be shown that the index Q has semantic qualities relative to the input information.

INTELLIGENT ACTIVITY OPERATED BY ENSEMBLES OF ELEMENTS

The devices hitherto considered were of elementary consistence, that is their structures could not have been subdivided, unless they ceased to be able to behave intelligently. It is now proposed to analyse those whose activity will result from the combination of complex quantities. Further, it is important to include in the study some theoretical elements quite different from those belonging to Information[10]. They are of a semantical kind[12].

In quantifying the elaboration E operated by an elementary device, the logical function it implements does not matter; the said quantification, in fact, depends on the objective data H_A and H_E, as it can be seen by equation 1. In such cases, even if the Equation (2) were utilized, the subjective terms of this are determined by the same objective data[2]. In so far as networks are concerned, the quantity of intelligent activity they can operate, E, depends on three fundamental factors: (1) the quantities E the various devices can operate; (2) the differentiation among the different type of logical functions they are implementing; (3) in what manner they are connected to each other. These theorems can also be proved by a rigorous mathematical analysis.

The corollary concerns information arriving from a complex of elementary intelligent devices, and derives from the theorem of the previous section. It states the following: the entity of the effectiveness index Q depends on three fundamental factors: (1) the entity of the indices of the intermediation information flows; (2) the type differentiation among the logical functions implemented by each device; (3) in what manner the last are connected to each other. These conclusions may be derived by rigorous algebraic means.

INFORMATION EFFECTIVENESS AND ITS ALGEBRA

The conferment of effectiveness on information is the object of applying intelligent activity to primary affering variables[4]; thus, the process of elaboration is characterized by the following dimensional equalities[11,13].

Affering information (bits) not effective	×	Elaboration (undimensional quantity)	×	elaborated information (bits) effective

It appears from previous observations that for elementary devices Q is a direct and exclusive function of H_A and H_E; for complex organs, that is for ensembles of a generical quantity of devices implementing elementary functions, which are variously connected among themselves, Q itself may be correlated to H_A by either E in a suitable way, or to the single $E_1, E_2, \ldots, E_n$.

To generalize, one must develop an algebra with which to find a transfer function that correlates Q to H_A, relative to whatever ensemble of n connected elements; it will consist of a set of formulations for combining different quantities, depending on the factors previously discussed: $E_1, E_2, \ldots, E_n$; $f(D_1), f(D_2), \ldots, f(D_n)$ (connecting ways among the elements*).

In order to simplify the presentation of the paper, the development of the algebra have been confined to ensembles of elements in which there is no feedback.

* The above-described factors do not neglect the important characteristic of non-linearity manifested by certain elementary logical functions, which, as it is clear, are greatly affecting the different H_E factors, depending on the working levels of affering entropies[3].

"Cascaded" Elements (in Series)

The simplest and direct way for connecting two or more elements are those illustrated by Figure 1 in the respective parts (a) and (b). The relative formulation, as previously developed, enable us to derive equations for calculating E and Q which are as follows:

$$E/c = \prod_{i=1}^{n} E_i \tag{5}$$

$$Q/c = H_A \prod_{i=1}^{n} \text{antilog}_2 \left(\frac{Q_i}{H_E^{(i-1)\text{-th}}} \right) \tag{6}$$

where Q/c is the index of effectiveness of information derived from n cascaded elements.

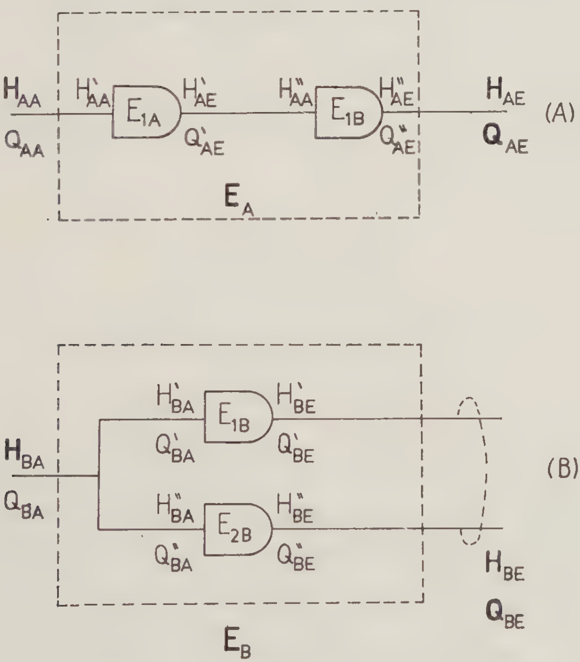


FIGURE 1 Connection between elements (devices)

Parallel Elements

The complex quantity of intelligent activity operated by n parallel connected elements, E/p, may be calculated to be

$$E/p = [\text{antilog}_2 (H'_{BA} \cdot F_1 - H'_{BE})] \times \cdots \times [\text{antilog}_2 (H^{n\text{-th}}_{BA} \cdot F_n - H^{n\text{-th}}_{BE})] \quad (7)$$

$$Q/p = H_{BA} E_B = H_{BA} \left(\frac{\log_2 E_{1B}}{H'_{BA} \cdot F_i} + \cdots + \frac{\log_2 E_{nB}}{H^{n\text{-th}}_{BA} \cdot F_n} \right) \quad (8)$$

where F is a factor less than 1 and proportional to the amount of intelligent activity operated on all the remaining H_{BA} by n different functions E.

Unconnected Elements

This case is considered in Figure 2, and it may be proved that

$$E/c = [\text{antilog}_2 (H'_{CA} - H'_{CE})] \times [\text{antilog}_2 (H''_{CA} + H''_{CE})] = E_{1C} \times E_{2C} \quad (9)$$

$$Q/c = H_{CA} (\log_2 E_C) = (H'_{CA} + H''_{CA}) \cdot (\log_2 E_C) = H_{CA} (\log_2 E_{C1} + \log_2 E_{2C}) \quad (10)$$

are valid.

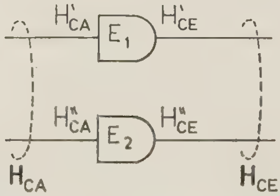


FIGURE 2 Unconnected elements

Mixed Connection of Elements

The variety of the possible connections existing for finite quantities, configurations, and diversity of the implemented elementary functions, may be quite extensive; nevertheless, it may be postulated that any scheme of connection may be reduced to one of the fundamental two, which operate the same quantity of intelligent activity, as in previous sections. To show this, it is necessary to consider the possibility of (1) decomposing ensembles, having more than one afferent variable, in equivalent ensembles, having a unique variable; (2) analysing and synthesizing two or more information flows, that is transforming the scheme shown by Figure 3(A) into that shown either by the part (B) or that (C) one.

Point (1) may be developed by stating the following corollary to the postulate expressed by (1): an ensemble having a number of different variables,

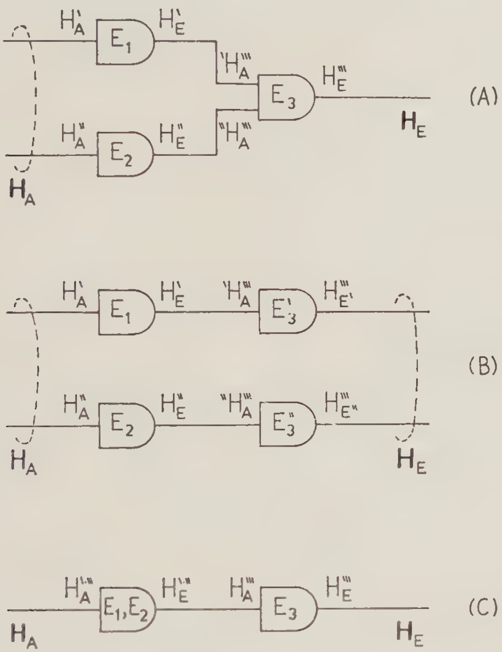


FIGURE 3 Mixed connection of elements

may be subdivided in as many discrete ensembles, every of which including all the intelligent elements and all the afferent variables taking part in the formation of the corresponding output; further, elements and inputs relative to two or more outputs, shall figure, especially modified, in each ensemble. The opposite of this corollary is also true.

Figure 4 illustrates, for example, the concepts described in the above corollary. The schematized ensemble in part (A) of the figure may be decomposed into the two subensembles shown in part (B); vice versa, the latter may be synthesized to form, after suitable simplifications, the first one.

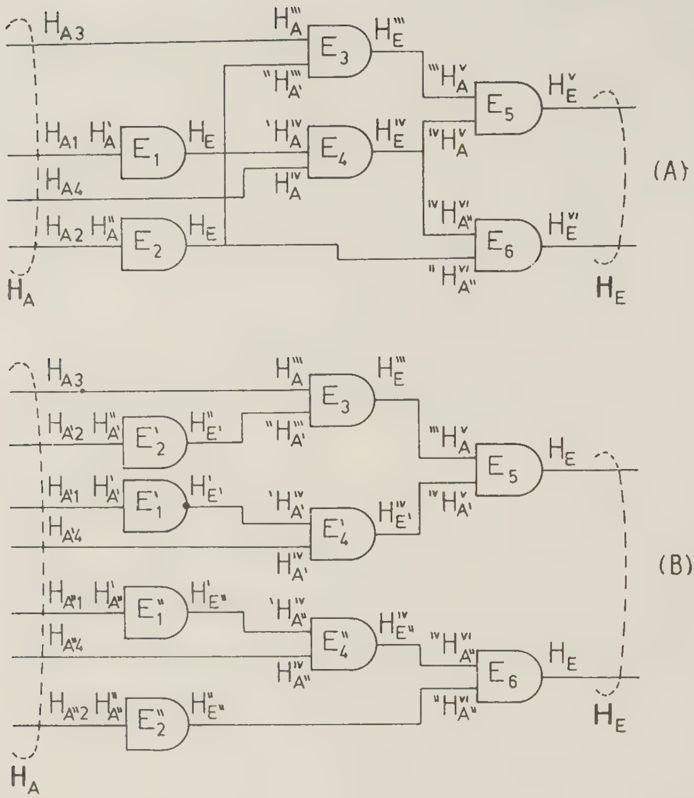


FIGURE 4 Corollary of Equation (8)

CONCLUSIONS

It had been shown what elements to take into consideration for the calculation of the intelligent activity E, operated by a complex of elementary devices, on the input information flows and of the effectiveness of the respective inputs. It was possible to derive a set of equations for the calculations of E and Q relative to ensembles of elements, either connected in series or in parallel. Some postulates and corollaries have been further stated, which constitute the means for transforming any complexity of ensembles of elements into others, whose E and Q were directly computable by the above equations. These two entities may, in fact, be computed on basis of entropies H_A and H_E.

Since any complex may be subdivided into as many others as the number of efferent variables then

$$H_E = H'_E + H''_E + \cdots + H_E^{k\text{-th}} = \sum_{j=1}^{k} H_E^j \tag{11}$$

So far as H_A is concerned, its computation has to take into account the meaning of the different components, contributing to the input of the ensemble, and of those derived from different elaborative elements. Thus, at the inputs for the organ's stages 1, 2, ..., k

$$H_A^i = H_{A'}^i + H_{A''}^i + \cdots + H_{A^{l\text{-th}}}^i = H_A^i \cdot F_1 + H_A^i \cdot F_2 + \cdots + H_A^i \cdot F_l$$
$$= \sum_{g=1}^{l} H_A^i \cdot F_g \tag{12}$$

At the f-th stage, H_A is

$$H_{A_f} = (H'_A + H''_A + \cdots + H_A^{k\text{-th}})_f = \left(\sum_{i=1}^{k} \sum_{g=1}^{l} H_A^i \cdot F_g\right)_f \tag{13}$$

From Equation (1) the total intelligent activity operated by the whole ensemble, results in the following general equation:

$$E = \text{antilog}_2 (H_A - H_E) = \text{antilog}_2 \sum_{f=1}^{q} \left(\sum_{i=1}^{k} \sum_{f=1}^{l} H_A^i \cdot F_g - \sum_{i=1}^{k} H_E^i\right) \tag{14}$$

The relative Q is given by

$$Q = H_A \cdot E = \left(\sum_{i=1}^{k} \sum_{g=1}^{l} H_A^l \cdot F_g\right) \cdot E$$

which, from Equation (4) is

$$Q = \left[\left(\sum_{i=1}^{k} \sum_{g=1}^{l} H_A^i \cdot F_g\right)_1\right]^2 - \left[\left(\sum_{i=1}^{k} \sum_{g=1}^{l} H_A^i \cdot F_g\right)_1 \times \left(\sum_{i=1}^{k} H_E^i\right)_q\right] \tag{15}$$

Thus, the intelligent activity E that any logical deductive organ can operate, and the effectiveness Q of the given responses, may be computed (1) by basing the computations on successive reductions of the complexity, in order to reach the configurations seen in Figure 1 and (2) by considering the entropies respectively contributing to and derived from the ensemble formed by connected elements, as shown in Figure 4.

References

1. A.O.Arigoni, "Quantification of the intelligent activity", *Cybernetica, Intern.* Assoc. *Cybernetics*, **5**, No.1 (1969).
2. A.O.Arigoni, "Subjective indecision for the recognition of messages", *XVII Intern. Congr. Commun., Genova, Italy* (1969).
3. A.O.Arigoni, "Caratteristichen onlineari dell'attività intelligente", *Atti Fond. Giorgio Ronchi Contrib. Ist. Nazl. Ottica*, **24**, No.1, 46–63 (1969).
4. A.O.Arigoni, "Variazioni isoentropiche della quantità di linee afferenti ad un dispositivo di elaborazione", *Ann. Univ.* Ferrara, *Sez. 14 Fis. Sper. Teor.*, **1**, No.4, 37–51 (1969).
5. A.O.Arigoni, "Ottimazione cibernetica delle diagnosi mediche", *Relazione* XXVI *Convegno della Salute, Ferrara* (1969).
6. A.O.Arigoni, "Increasing the communication entropy", *XVII Intern. Congr. Communic., Genova, Italy* (1969).
7. "Considerazioni epistemologiche su un procedimento di quantificazione dell'attività intelligente", (to be published).
8. L.Szilard, "Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen." *Z. Physik*, **53**, 840–50 (1929).
9. L.Brillouin, *La Science et la Theory de l'Information*, Masson et Cie, Paris (1959).
10. P.Guillard, *La Semautique*, Presse Universitaire de France, Paris (1955).
11. A.O.Arigoni, "Effectiveness of the elaborated information" (to be published).
12. A.O.Arigoni, "Descriptive language evidencing the entropy of the carried information." *Ann. Univ. Ferrara, Sez. 14, Fis. Sper. Teor.*, **1**, No.3, 25–36 (1969).
13. A.O.Arigoni, "Sulle dimensioni della probabilità", *Ann. Univ. Ferrara, Sez. 14, Fis. Sper. Teor.*, **1**, No.2, 17–24 (1968).
14. A.O.Arigoni, "Modello fisico proposto per la simulazione dei processi logici deduttivi" *Ann. Univ. Ferrara, Sez. 14. Fis. Sper. Teor.* Suppl. **1, Vol. 1,** 1–40 (1968).

CHAPTER II–12

Models of nerve cell interactions in the cat visual system

G. P. MOORE, J. H. PENALOZA-ROJAS, and B. OKUMURA

Biomedical Engineering Laboratories
University of Southern California
Los Angeles, California, U.S.A.

Summary

Multiple neuronal spike trains have been recorded in the lateral geniculate nucleus of the cat thalamus under pentobarbitol anesthesia. Spikes from each pair of cells recorded were fed into a computing system which was programmed to display autocorrelation histograms for each cell and the cross-correlation histogram for each simultaneously recorded pair. A significant fraction of the cell pairs observed show correlated patterns of firing some of which we attribute to excitatory synaptic connections. Hypotheses about the connections between these pairs, which can be inferred from analysis of the cross correlation, have been tested successfully by subsequent digital computer simulations. The cross-correlation histogram, although derived from purely *extracellular* pulse train timings, can also yield details about the *intracellular* synaptic potentials generated in these cells, and inferences drawn from our records are quite consistent with previously published examples from intracellular records.

Several different patterns of neuronal discharge have been reported for cells in the region of the cat brain known as the lateral geniculate nucleus, an area which is the main relay station between the retina and visual cortex. Several laboratories[1–4] have provided quantitative measures of discharge patterns in these cells by means of *interspike interval histograms* which plot the distribution of elapsed times between impulses, a technique which is now widely used to characterize neuronal activity[5,6].

In our experiments similar statistical measures are being used to process extracellular pulse trains obtained simultaneously from two neurons not only to characterize the temporal characteristics of their discharge but also to test for correlations in their discharge which might be used as indications of possible connections between the observed cells. It is hoped that with sufficient numbers of such observations a general pattern of network organization in this part of the brain may be revealed. In this paper we present some preliminary results to illustrate the techniques we are using to investigate the structure of small neuronal networks.

A typical experimental result is shown in Figure 1. Each channel of data is obtained through a microelectrode inserted into the lateral geniculate region and in these experiments the two recording tips sample activity from cells separated from each other by less than 1 mm. The animals are under pentobarbitol anesthesia and in this condition and under ambient light conditions the spontaneous discharge one observes is known to show a pattern of high-frequency bursts rather randomly distributed in time[2,3,8].

From the record such as that in Figure 1, the time of occurrence of every pulse in each channel is fed into a computer which processes the data to obtain two measures of activity which have been termed the autocorrelation histogram and cross-correlation histogram. The former is computed for each cell in the pair and is a measure of the probability of the occurrence of a pulse in that cell occurring as a function of time τ before or after a pulse in that cell. It is a generalization of the interval histogram for that cell which plots the probability of occurrence of the *next* pulse in that cell after a pulse at $\tau = 0$. (For a discussion of these and other measures, see Moore *et al.*[5]; Perkel *et al.*[6].) Any tendency toward rhythmicity (i.e. preferred intervals of discharge) is clearly revealed in this kind of display.

The cross-correlation histogram obtained for each pair of cells is a measure of the liklihood of occurrence of a pulse in one cell as a function of time τ before or since a pulse in the other cell. The fundamental theorem relevant to this histogram is that if the two cells are independent, i.e. their pulse trains are independent, the cross-correlation histogram will have a constant value, i.e. will be flat. When the cross-correlation function is not flat, functional interaction may be suspected.

In Figure 2 we show a set of correlation histograms for two cells. On the

FIGURE 1 Pulse trains from two simultaneously recorded neurons from the lateral geniculate body of the cat. Pentobarbitol anaesthesia

top the autocorrelation for Cell A, in the middle the autocorrelation for Cell B, and at the bottom the cross-correlation showing the probability of Cell B firing as a function of time since Cell A fires. Both autocorrelations show a central peak in the region of $\tau = 0$ which results from the burst structure of their activity. That is, there is a high probability of a given pulse being

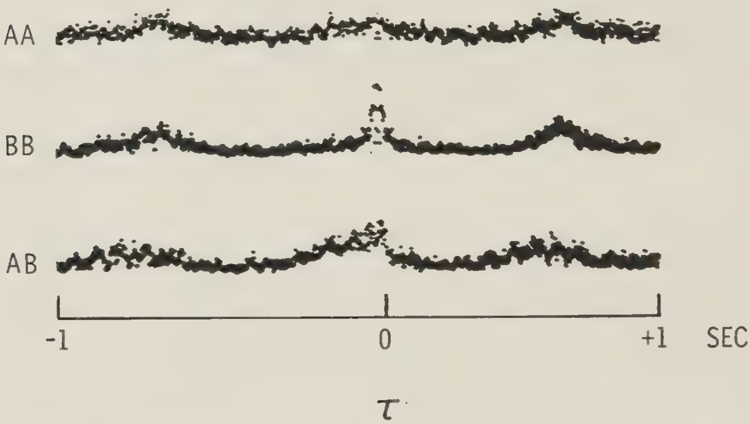


FIGURE 2 Example of autocorrelation and cross-correlation histograms from a pair of simultaneously recorded geniculate neurons. Abcissa: elapsed time from reference neuron discharge. Ordinate: proportional to number of observations of a discharge with that interval from reference discharge

preceded or followed by other pulses at short intervals. The middle trace also shows that there is a slight tendency for the bursts of Cell B to be repeated at intervals of about 500 ms (indicated by the broad peak at $\tau = \pm 500$ ms), a feature shared to a lesser degree by Cell A.

Of particular interest to us, however, is the structure of the cross-correlation histogram which reveals several interesting features. First, the peak immediately to the left of the origin ($\tau = 0$) indicates that Cell B firings tend to precede Cell A firings (all Cell A firings occur at $\tau = 0$ by definition) and have a relatively low probability of following Cell B firings. The reverse cross-correlation (which measures the firing probability of Cell A as a function of time before or since Cell B firings) is the mirror image of the correlogram shown[7], and this means that following a Cell B discharge there is on the average a transient increase in the probability of Cell A firing and this

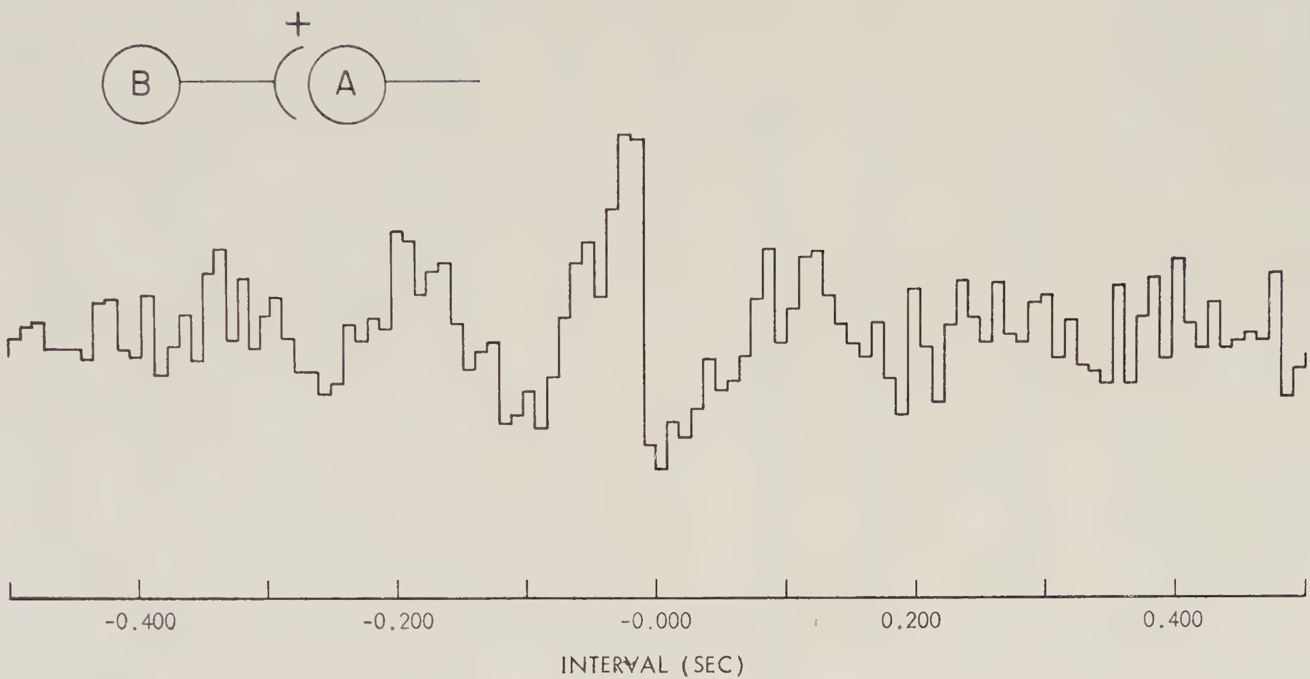


FIGURE 3 Cross-correlation histogram computed for two neurons whose simulated interaction was chosen to match the data of Figure 2. Network model was based on direct monosynaptic excitation between observed cells

transient gradually declines (with a nearly exponential decay) after about 100 ms.

Several explanations of this phenomena can be offered. The most immediate is that Cell B makes a direct excitatory connection with Cell A and that the excitatory synaptic potential generated in A by each Cell B firing raises the probability of Cell A firing throughout the duration of, and in

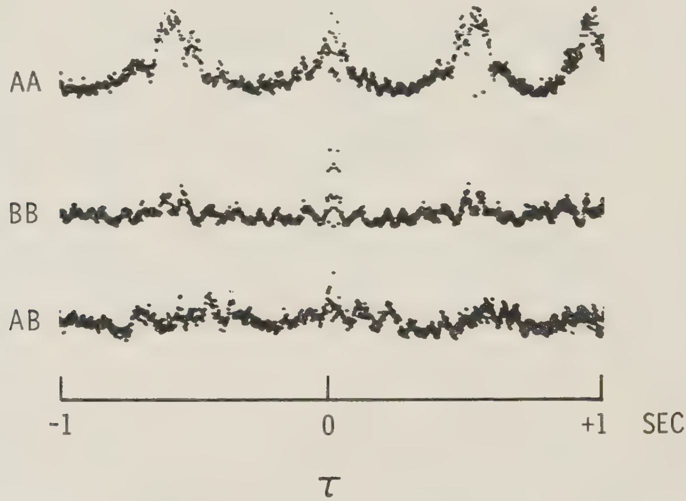


FIGURE 4 Another example of auto- and cross-correlation histograms for two cells. Cross-correlation shows complex structure

rough proportion to the amplitude of, the excitatory potential. This interpretation is consistent with the shape of the central asymmetric peak and the known shape of excitatory potentials in these cells.[9,10]. It is also consistent with the appearance of the paracentral peaks in the cross-correlation histogram at $\tau = \pm 500$ ms, because if Cell B does excite Cell A then A should fire with higher probability not only immediately after Cell B but also at past and future times when Cell B also fires.

Indeed, our hypothesis about the relationship between Cells A and B can be subjected to direct test by means of digital computer simulation using a program which has been developed for that purpose[11]. In each simulation, a cell network is chosen that appears to conform to the known or suspected anatomical relationships, and numerous parameters associated with each cell in the network are assigned values (viz. membrane and threshold potentials, synaptic potential amplitudes and time constants, etc.). In Figure 3 we

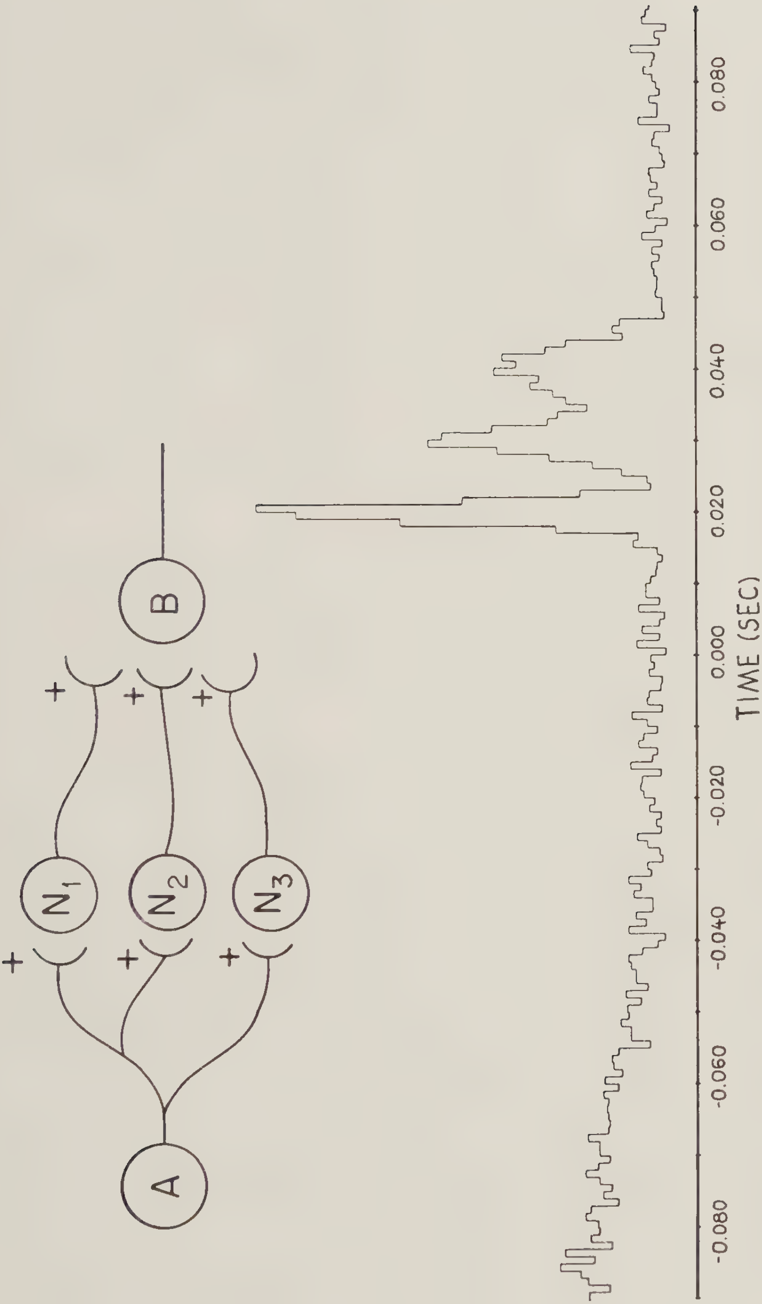


FIGURE 5 Computer simulated network involving multiple interneuron excitation produces cross-correlation histogram similar to that for the neurons shown in Figure 4

show an example of a cross-correlation histogram computed for a simulated network embodying the assumptions made earlier about the relationship between the real cells whose histograms are shown in Figure 2. The central peak results from features of the excitatory connection between the two simulated cells, the other peaks are principally derived from patterns in the presynaptic cell discharge.

A similar but more complex result is shown for another pair of neurons from the cat lateral geniculate (Figure 4). The central peak is reversed in direction, signifying that Cell A excites Cell B, but there is additional structure in the central peak in the form of distinct subpeaks. This is a typical complication which requires additional hypotheses for interpretation. The first possibility is that each Cell A spike can trigger not one but several high-frequency correlated spikes in Cell B, i.e. can trigger a burst in Cell B each pulse of which gives rise to a subpeak. Again, this is consistent with observations reported for these cells by others who have used intracellular electrodes to observe synaptic activity[9,10].

Another alternative explanation is that Cell A exites Cell B via several interneurons with slightly differing conduction delays so that Cell B discharges can result from each Cell A pulse after varying delays. This hypothesis has also been checked by computer simulation, and a typical result is shown in Figure 5.

In each case, computer simulation can be used to support or reject a specific hypothesis and, when successful, to predict further tests or future experimental observations. On the other hand, a given empirical observation can often be explained by several mutually exclusive hypotheses. In the examples given above, we have resorted to known physiological features of the cells being investigated and therefore have formed a hypothesis of excitatory synaptic couplings. Several alternatives, however, have also been investigated. For example, the central peak in Figure 2 could also result from Cells A and B being subjected to *inhibition* from a common source as shown by computer simulation results in Figure 6, and indeed, inhibitory potentials are known to occur in lateral geniculate cells[9,10,12].

The important points to be emphasized from the preliminary results we are reporting here are as follows.

(1) Correlation analysis can be used to detect dependencies in the firing patterns of neurons.

(2) In some cases analysis of the structure of the autocorrelation and cross-correlation histograms of the observed cells can lead to very specific hypo-

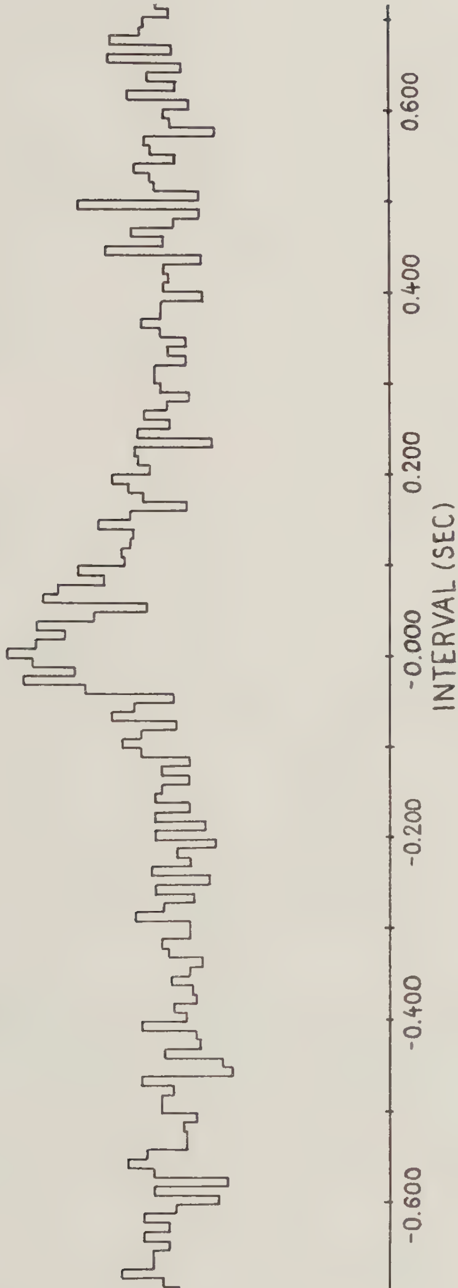


FIGURE 6 Alternative model to account for cross-correlation of Figure 2. Network simulation involved inhibition of Cells A and B from common source instead of direct excitation. Central peak resembles central peak of Figure 2 cross-correlation

theses about connections between the observed cells and these resulting hypotheses can be tested by computer simulation.

(3) In general although a given cross-correlation can be explained by several hypotheses, other hypotheses are clearly incompatible with the observed data (for example, in Figures 2 and 4 the possibility of direct synaptic inhibition is ruled out). Even in the case of alternative hypotheses, each is likely to involve very specific assumptions about cell parameters so that additional observations can be used to validate or reject the model and further experiments can be planned with greater precision.

(4) In any case information about processes normally considered to be internal to the neuron (such as synaptic potentials) can be obtained by purely external measurement.

References

1. R. Corazza, M. Endo, V. Tradardi, and C. Umilta, "Tonic neuronal activity of the central visual pathways during dark adaptation", *Arch. Ital. Biol.*, **106**, 270–82 (1968).
2. A. Herz, O. Creutzfeldt, and J. Fuster, "Statistische Eigenschaften der Neuronaktivitat im ascendierenden visuellen System", *Kybernetik*, **2**, 61–71 (1964).
3. P. O. Bishop, W. R. Levick, and W. O. Williams, "Statistical analysis of the dark discharge of lateral geniculate neurones", *J. Physiol.*, **170**, 598–612 (1964).
4. J. Fuster, A. Herz, and O. Creutzfeldt, "Interval analysis of cell discharge in spontaneous and optically modulated activity in the visual system", *Arch. Ital. Biol.*, **103**, 159–77 (1965).
5. G. P. Moore, D. H. Perkel, and J. P. Segundo, "Statistical analysis and functional interpretation of neuronal spike data", *Ann. Rev. Physiol.*, **28**, 493–522 (1966).
6. D. H. Perkel, G. L. Gerstein, and G. P. Moore, "Neuronal spike trains and stochastic point processes, I: single spike trains", *Biophys. J.*, **7**, 391–418 (1967).
7. D. H. Perkel, G. L. Gerstein, and G. P. Moore, "Neuronal spike trains and stochastic point processes, II: simultaneous spike trains", *Biophys. J.*, **7**, 419–40 (1967).
8. D. H. Hubel and T. N. Wiesel, "Integrative action in the cat's lateral geniculate body", *J. Physiol.*, **155**, 385–98 (1961).
9. J. T. McIlwain and O. D. Creutzfeldt, "Microelectrode study of synaptic excitation and inhibition in the lateral geniculate nucleus of the cat", *J. Neurophysiol.*, **30**, 1–21 (1967).
10. O. D. Creutzfeldt, "Functional synaptic organization in the lateral geniculate body and its implication for information transmission", in *Structure and Function of Inhibitory Neuronal Mechanisms* (Eds. C. von Euler, S. Skoglund and U. Soderberg), Pergamon Press, Oxford (1968).
11. D. Perkel, G. Moore, and J. Segundo, "Continuous-time simulation of ganglion nerve cells in *aplysia*", *Biomedical Sciences Inst.*, Vol. 1 (Ed. F. Alt), ISA–Plenum Press, New York (1963), pp. 347–57.
12. H. Suzuki and E. Kato, "Binocular interaction at cat's lateral geniculate body", *J. Neurophysiol.*, **29**, 909–20 (1966).

CHAPTER II–13

Harmonic analysis in the visual pathway of the higher mammals

J.G. SANTESMASES, F. RUBIO,
L. ARÉS, J. MIRA, and R.H. VERDUZCO

*Instituto de Electricidad y Automática,
University of Madrid, Madrid, Spain*

Summary

The most recent findings of Campbell and coworkers show that in the visual system of some vertebrates a sort of spatial frequency selectivity is present, as well as a selectivity to the orientation of a sinusoidal grating. On these bases, the visual pathway as an optical data processing system is studied. The existence of "spatial tuned networks" is postulated and the equation that accounts for such a network is presented as well as the design of the neural net that can achieve such a function. This net performs spatial integrals and weight factor multiplication, operations widely admitted for such networks. The optimum pattern for testing this selective process is presented, pointing out that orientational selectivity could easily be explained in terms of the spectral analysis that is verified in the visual pathway. If spectral analysis is being made in this field it can be thought that it is also, possible that, under certain conditions, the characteristic operations of a linear system such as spatial filtering, correlation, etc. are carried out. These operations can be formulated by means of integral transforms, whose kernel depends on the particular operation performed. These operations can be supported in the visual pathway by the fact that this has a layered structure which is specially suitable for this activity.

INTRODUCTION

Certain properties which are shown to exist in various mammals (cat, monkey, and man) seem to indicate that in the visual pathway of these beings a spatial harmonic analysis is carried out analogous to that which takes place in the ear in the temporal domain.

This work is an attempt to search a model which manifests this property and which at the same time is able to suggest new neuro and psychological experiments making it possible to show how far the model is valid as well as how to find out the functional structure which promotes this activity.

F. W. Campbell[1] considers that the properties of analysis and transmission of optical data of the visual system of both humans and higher mammals can be studied taking into consideration the fact that this system consists of two parts: the optical system of the eye and the retina–brain system.

Campbell and Green[2] make a quantitative study of the relative role of the optical system and of the nervous system in the detection of gratings–these being fundamental to show these phenomena—for differing spatial frequencies. They reach the conclusion that the threshold contrast—the minimum contrast necessary in perceiving a sinusoidal grating—is imposed by the nervous system more than by the optical. In view of this, we are interested, fundamentally, in the retina–brain system.

The experiments carried out during the last years by Enroth-Cugell and Robson[3], Cooper and Robson[4], Campbell *et al.*[5], and Campbell and Kulikowsky[6] permit us to summarize the properties of the visual pathway related with this topic in the following items.

(a) In the retina of some mammals (cat, monkey, and man) channels exist tuned to different spatial frequencies.

(b) This selectivity is greater in the man than in the monkey and in the later greater than in the cat.

(c) The bandwidth of these channels is roughly smaller than $\frac{1}{8}$ of the spatial frequency at which they are tuned.

(d) This selectivity is detected at the level of the ganglion cells or superior centres of the visual pathway.

(e) At the level of the visual cortex there exists a selectivity response to a determined orientation of a grating.

To summarize, it can be said that, according to the neurophysiological findings (cat and monkey) of the above authors, two important properties of an image have been coded: (1) the information about the orientation of an edge; (2) the extraction of the content in spatial frequencies of the image which is formed on the retina.

Since in the retina–brain system the extraction of the frequency content of a pattern seems to be carried out, there is a real basis for the representation, $F(\omega_x, \omega_y)$, of a certain pattern, $f(x, y)$, in the spatial frequency domain by means of the two-dimensional Fourier transform this representation being not only a way of calculus but having also a certain neurophysiological meaning.

All the functions to be considered will be presented in the two domains: that of spatial variables (x, y) and that of frequencies (ω_x, ω_y); throughout there exists a complete correspondance between both domains. The particular interest at each instant as well as the facility of interpretation of the phenomena in question in one domain or the other will determine on which of the two domains attention is to be paid.

In order to be able to apply this kind of analysis, it is necessary to consider that the processing system of data–retina and superior centres is linear and space invariant. In fact, under certain conditions, these properties are fulfilled.

(a) By spatial invariance the following property is understood: if under an input $f(x, y)$ the system responds with $g(x, y)$, under the input $f(x + a, y + b)$ the response of the system is $g(x + a, y + b)$. Leaving to one side edge effects and considering homogeneous regions of the retina, such as the fovea—or the area centralis in the case of the cat—and the parafovea, it is possible to admit this invariance in the retina.

(b) To consider the retina as a linear system is the strongest hypothesis of the present analysis, being based in research carried out by Maffei[7], where the conclusion is reached that if the modulation of the luminous intensity on the retina is less than 20% its behaviour is linear. The research of Enroth-Cugell and Robson[3] shows that there exist in the retina ganglion cells—cells in their nomenclature—which respond linearly to their stimuli.

Under the conditions stated it is therefore possible to study the data transmission in the retina using as a basis the techniques of the integral transforms.

SPATIAL HARMONIC ANALYSIS

The treatment of these phenomena of spectral analysis by means of integral transforms in which the kernel is a difference one, makes it possible to unify the formulation of the phenomena of lateral interaction[8,9] with those of harmonic analysis.

The layered structure of the retina (Figure 1) as well as that of the lateral geniculate body (Figure 2) can account for the lateral interaction phenomena firstly described by Hartline in the compound eye of the *Limulus*. In

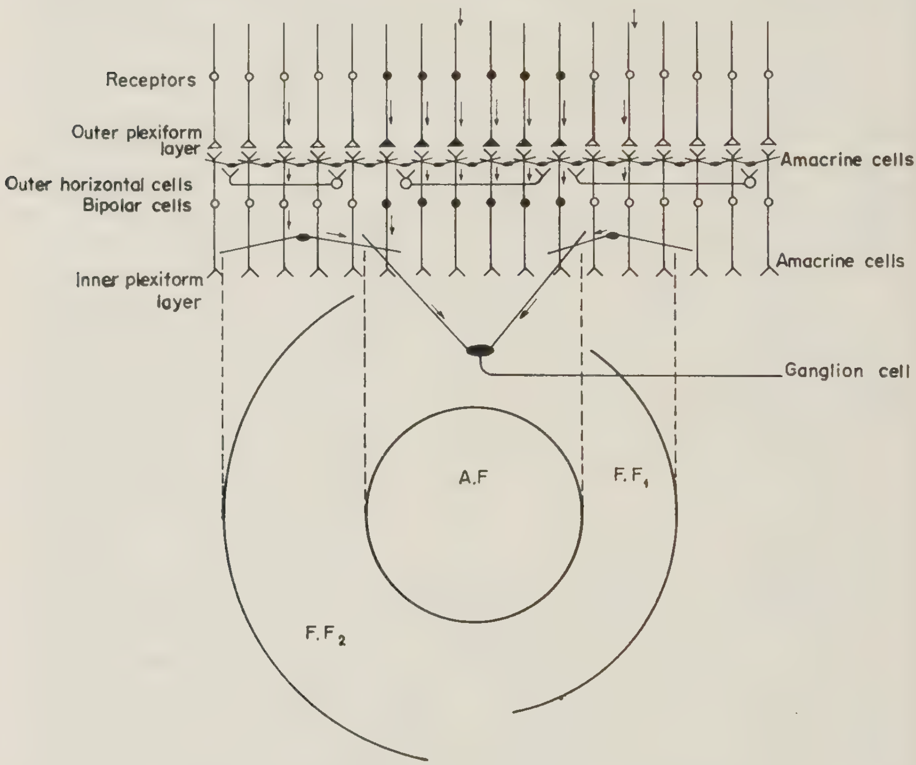


FIGURE 1 Anatomic and functional field of a ganglion cell

the retina of the mammals, the amacrine cells layer—that we admit, following Gallego[10] exist in the outer plexiform layer as well as in the inner one—seems the more adequate structure to be accomplished by these phenomena in them. In the lateral geniculate body the association cells (or short axon cells) can explain this activity.

It has been indicated[8,9] that in the retina of higher mammals the lateral inhibition phenomena can be expressed by means of integral transformations in certain conditions.

In the case of non-recurrent interaction, if $f(\alpha, \beta)$ is the input signal on the element in the coordinate point (α, β) of a coordinate system located in the horizontal plane where such a generic element is located, $g(x, y)$ is the output signal of the element (x, y), and $k_0(x, y; \alpha, \beta)$ is the weighting factor which represents the action of the element (α, β) on the element (x, y), then the response of this element will be

$$g(x, y) = f(x, y) + \iint_R k_0(x, y; \alpha, \beta) f(\alpha, \beta)\, d\alpha\, d\beta \tag{1}$$

This expression can be considered as an integral, if the function to be determined is $g(x, y)$, or as an integral equation, if such a function is $f(x, y)$. The

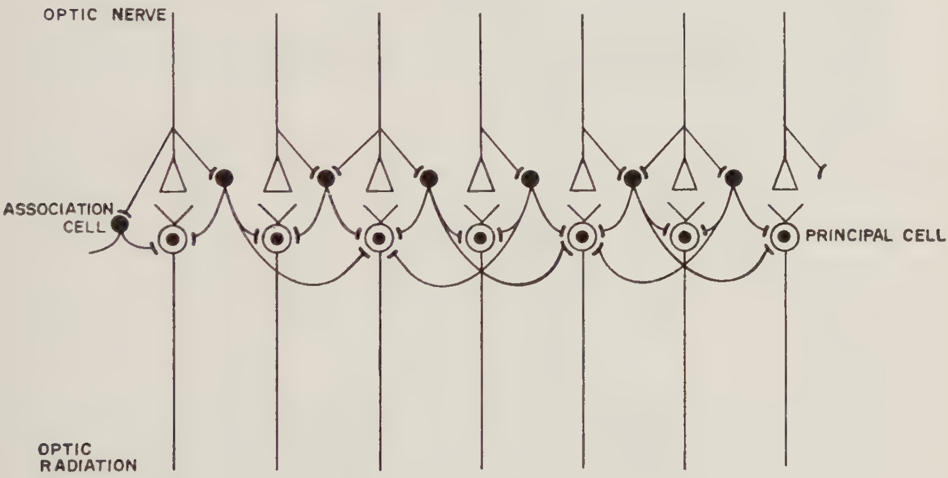


FIGURE 2 The lateral geniculate body

recurrent interaction can be formulated in an analogous way. The most general expression for interaction can be written in the form

$$g(x, y) + \iint_R k(x, y; \alpha, \beta) f(\alpha, \beta)\, d\alpha\, d\beta \tag{2}$$

The expression (2) contains (1) as a particular case in which the weighting factor or kernel is

$$k(x, y; \alpha, \beta) = \delta(x - \alpha, y - \beta) + k_0(x, y; \alpha, \beta)$$

We have suggested in a previous paper[11] different expressions for this kernel, depending upon the transversal layer of the visual pathway to be simulated according to the anatomical data available.

If the system represented in Figure 3 is linear and space invariant, the response of the system to an input, $f(x, y)$ is expressed as a convolution of the input with the response $k(x - \alpha, y - \beta)$ of the system to the impulse function

$$g(x, y) = \int_{-\infty}^{+\infty} \int k(x - \alpha, y - \beta) f(\alpha, \beta) \, d\alpha \, d\beta \tag{3}$$

The above relation may be considered as an interaction between neural elements where the kernel or coefficient of interaction is $k(x - \alpha, y - \beta)$, that is to say, a difference kernel, such as occurs with the coefficients characteristic of lateral interaction.

From now on, we shall restrict ourselves, for the sake of simplicity, to the one-dimensional case. In the simplest case a system tuned at a certain spatial frequency, ω_{0x}, must possess a kernel

$$k(x - \alpha) = A \cos \omega_{0x} (x - \alpha) \tag{4}$$

whose Fourier transform is (Figure 4)

$$A \frac{\pi \, e^{-i\omega_x \alpha}}{2\omega_{0x}} \left\{ \delta\left(\frac{\omega_x}{2\omega_{0x}} + \frac{1}{2}\right) + \delta\left(\frac{\omega_x}{2\omega_{0x}} - \frac{1}{2}\right) \right\} \tag{5}$$

FIGURE 3 Linear and space invariant system

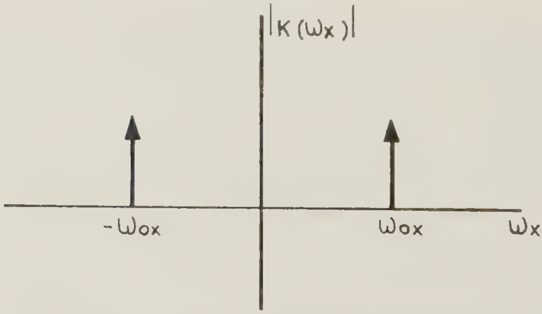


FIGURE 4 Fourier transform (Equation (4))

The above expression admits a simple interpretation in terms of neural nets, for which it is sufficient to assume that the ganglion cell to be found in the position x is subjected to interactive concentric fields, alternately excitatory and inhibitory ones (Figure 5), by means of the neighbouring elements. To admit an interaction of this kind is equivalent to assuming a relation one to

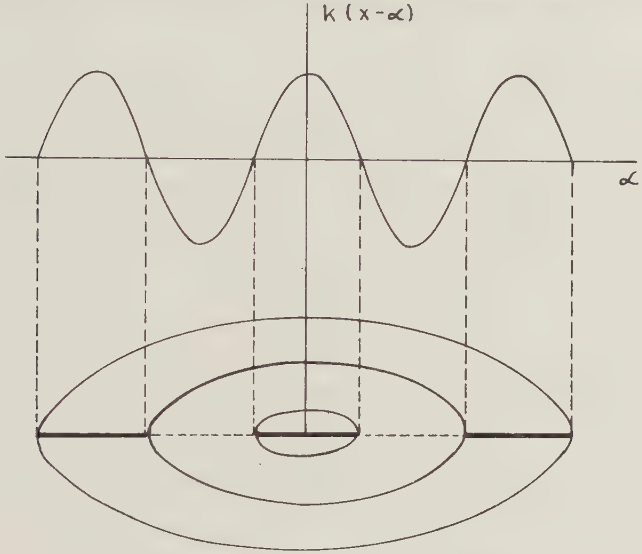


FIGURE 5 Interactive concentric fields

one between receptors and ganglion cells and that all the inputs corresponding to the region of the retina in consideration, the fovea, act on a ganglion cell.

Let us take into account the fact that the cat has no fovea. The equivalent region is the one called "area centralis" in which the relationship between receptors and ganglion cells is not one to one but there exists a convergence of elements. Consequently this interpretation can be admitted in the case of the monkey and of the man but not in that of the cat. This fact—whether the relation is one to one or not—has no importance in order to apply to the conceptual model. Since in the man it is not possible to register at the ganglion cell level, this spatial selectivity has not been shown at this level; however it is possible that this selectivity already exists there, and that there is no need to reach the cortex for showing this. Of course, living being, according to their degree of consciousness, process the information at the higher centres, but we believe that the retina, owing to its special structure, can account for this phenomenon.

To postulate a kernel with these characteristics is equivalent to consider that a given ganglion responds exclusively to one frequency, when in fact it is known that it gives maximum response to one frequency and to a less extent to its neighbouring ones, until reaching a given frequency where there is no response. All these facts suggest a search for another coefficient of inter-

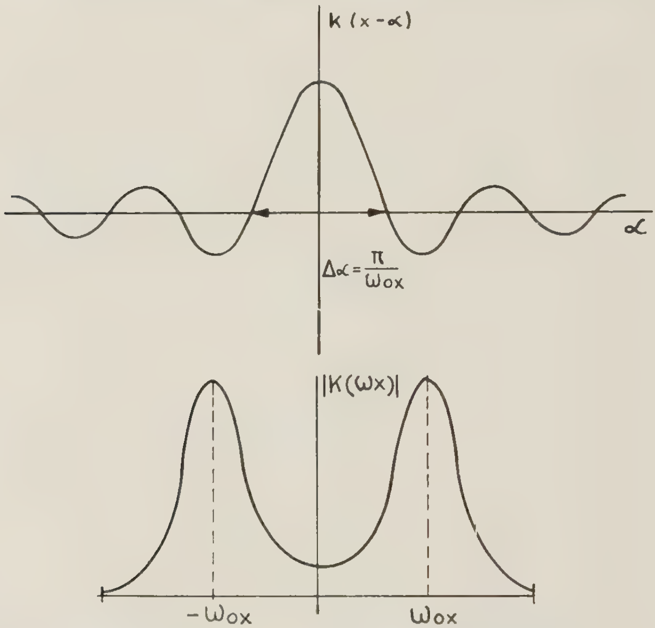


FIGURE 6 Functions of Equations (6) and (7)

action which shows with greater fidelity the neurophysiological findings. Let us consider an interaction coefficient defined by the expression

$$B\,(x - \alpha) = A \cdot \mathrm{e}^{-k\,|x-\alpha|} \cos \omega_{0x}\,(x - \alpha) \tag{6}$$

In the frequency domain this function has a transform

$$\frac{AB\pi}{\omega_{0x}} \left\{ \frac{\mathrm{e}^{-i\omega_x \alpha}}{B^2 + \omega_x^2} * \mathrm{e}^{-i\omega_x \alpha} \left[\delta\left(\frac{{}_0\omega}{2^{x}\omega_x} + \frac{1}{2} \right) + \delta\left(\frac{\omega_x}{2\omega_{0x}} - \frac{1}{2} \right) \right] \right\} \tag{7}$$

where the symbol (*) holds for the convolution integral. Both functions are represented in Figure 6. As can be observed, in the transform domain, a system which presents a transfer function of this kind shows selectivity at the frequency ω_{0x}. It is observed that the smaller its "width", $\Delta\alpha$, the greater is such selectivity.

The ganglion cells which present this selectivity have a receptive field composed of concentric zones which are alternately excitatory and inhibitory, and whose action decreases more or less rapidly with the distance depending on the values of B. The neural net which may account for this property is represented in Figure 7.

The Figure 7(a) represents, schematically, the whole convolution process which accomplishes the spatial selectivity. Two ganglion cells have been represented with their corresponding receptive fields. In this figure the fields are of different sizes and overlapping. In Figure 7(b), the interaction coefficients corresponding to the elements which constitute a generic receptive field are represented with some detail. It can be seen in the figure that the central zone of the receptive field is composed by two elements whose response is multiplied by weighting factors representing the value of the interaction coefficient $k(x - \alpha)$. The output of the elements of the interaction layer are added on α to obtain the response of the corresponding ganglion cell in position x.

Since we have assumed that the system is linear, it is possible to characterize each ganglion cell in the case of the retina by a natural frequency, ω_{0x}, that is intimately related to the size of the receptive field, Figure 6,

$$\Delta\alpha = \frac{\pi}{\omega_{0x}} \tag{8}$$

The ganglion cells with large receptive fields are tuned at smaller frequencies than those with small ones. This result is quite logical since if we suppose a sinusoidal grating in one direction as stimulus, the greater the spatial frequency the narrower are the strips of the grating, so it requires for its detection smaller receptive fields than those needed to detect smaller frequencies.

According to our assumptions—alternatively excitatory and inhibitory fields—we have arrived at the existence of a frequency distribution for each ganglion cell which depends on the proper characteristics of each net. Otherwise, Campbell and his colleagues have experimentally reached a frequency distribution at the ganglion cells under the stimulus of a sinusoidal grating. Then it is possible to think that perhaps a direct relation between the frequencies obtained by us and the frequencies found by the neurophysiologist may exist.

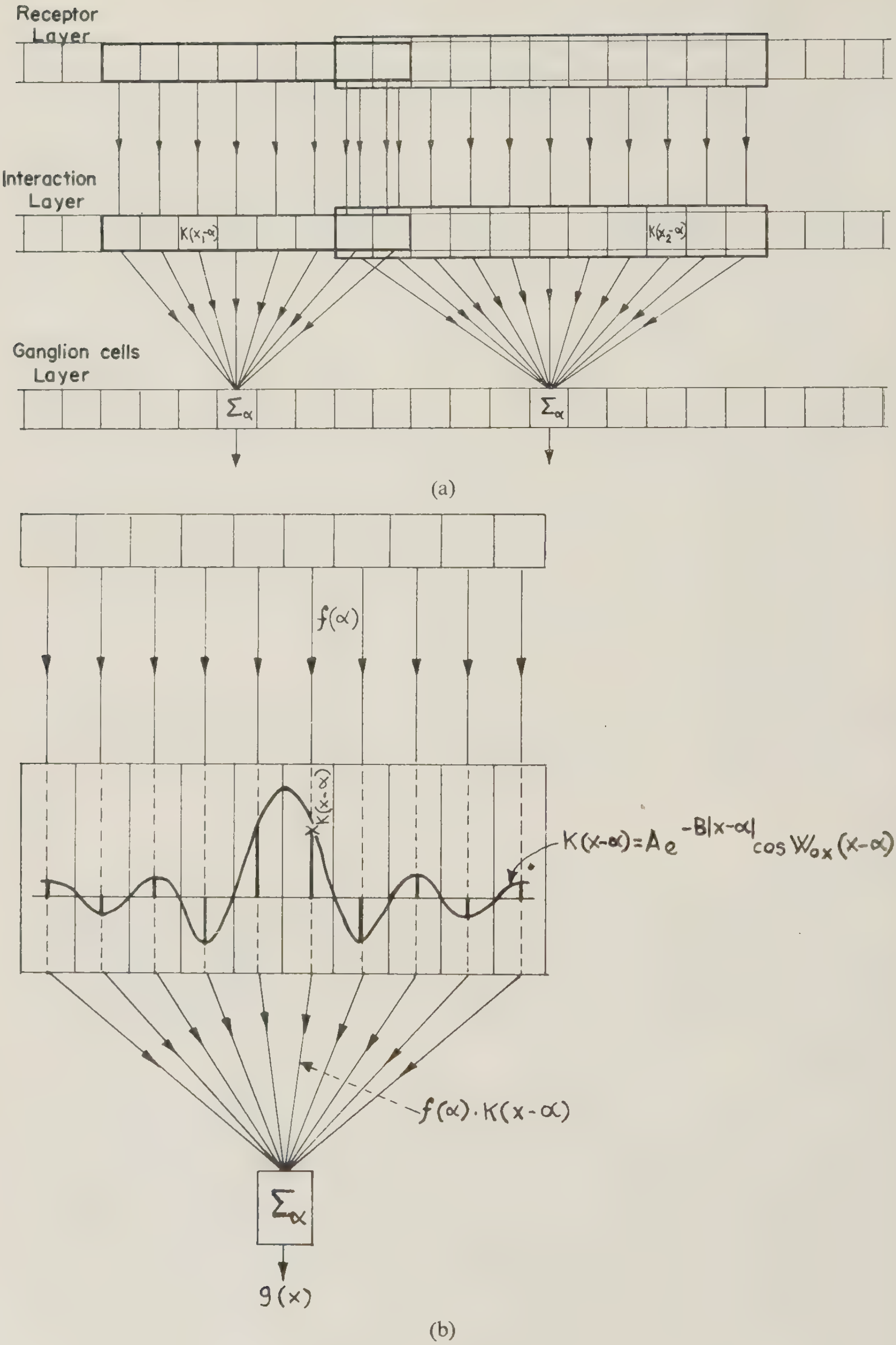


FIGURE 7 The neural net

According to the bases previously considered, it seems that the greater the frequency at which a ganglion cell is tuned, the smaller is its selectivity, i.e. the greater is its bandwidth. This parameter depends on the constant B, that is to say on the smaller or greater distance at which the interactive actions on a ganglion cell are felt.

From our point of view it is possible to explain the harmonic analysis phenomena, assuming that the receptive retinal fields are randomly distributed as far as size is concerned and that there exist a great number of horizontal connections to make possible the assumption done. It sounds rather convincing that these two conditions are fulfilled in the visual pathway of the mammals.

A neurophysiological experiment which can be suggested to prove or to modify conveniently the model consists of measuring the receptive field corresponding to a generic ganglion cell, after this to determine the natural frequency (8) of the net corresponding to this ganglion cell; then to stimulate this ganglion cell with gratings of different frequencies, and to find out the frequency for which the response is maximum and then to look for the relation between the grating frequency giving the maximum response and the frequency of the net in consideration. The existing anatomical and functional knowledge about the retina shows that the sizes of the receptive fields vary between relatively wide limits. According to Gallego[10], in the cat the limits vary in diameter from 450 μ to 80 μ, and the same may occur in the other mammals. This provides, at least, a structure able to support the model described here.

There exists in the retina a set of neural nets of the kind described above; in order to know whether one of them is tuned or not, it is necessary to measure the response of all of them at one instant.

On exciting the system with a periodic signal which is not sinusoidal, the register in all the ganglion cells provide a series of values which must be proportional to the Fourier coefficients of the harmonic series corresponding to the input configuration.

The expression corresponding to (6) will present—in two dimensions—two principal natural frequencies, ω_{0x} and ω_{0y}, with which it is possible to detect the orientation of a grating. Nevertheless the experiments made in the cat and in the monkey, at the level of ganglion cells and of the lateral geniculate body, seem to indicate that this is not so. However, if this orientation is detected at the level of the cortex, how is it possible an information—that seems to have been lost before—can be used in higher centres? Perhaps, the information related with the orientation is coded in a different way that the information related to the frequency? or perhaps, does it travel by a parallel channel?

The most appropriate stimulus to manifest these phenomena seems to be a double periodic grating, $\cos \omega_{0x}x \cdot \cos \omega_{0y}y$, i.e. that whose transform is

$$\frac{\pi^2}{4\omega_{0x}\omega_{0y}}\left\{\delta\left(\frac{\omega_x}{2\omega_{0x}}+\frac{1}{2}, \frac{\omega_y}{2\omega_{0y}}+\frac{1}{2}\right)+\delta\left(\frac{\omega_x}{2\omega_{0x}}-\frac{1}{2}, \frac{\omega_y}{2\omega_{0y}}+\frac{1}{2}\right)\right.$$
$$\left.=\delta\left(\frac{\omega_x}{2\omega_{0x}}+\frac{1}{2}, \frac{\omega_y}{2\omega_{0y}}-\frac{1}{2}\right)+\delta\left(\frac{\omega_x}{2\omega_{0x}}-\frac{1}{2}, \frac{\omega_y}{2\omega_{0y}}-\frac{1}{2}\right)\right\}$$

It is obvious, when the retina accomplishes a spatial harmonic analysis, that the concept of simple pattern must be searched out in the frequency domain not in the spatial variables one. For instance, a luminous spot used as stimulus—$\delta(x, y)$—is not useful at all from the point of view of harmonic analysis, since this function has a Fourier transform which contains all the frequencies with the same amplitude.

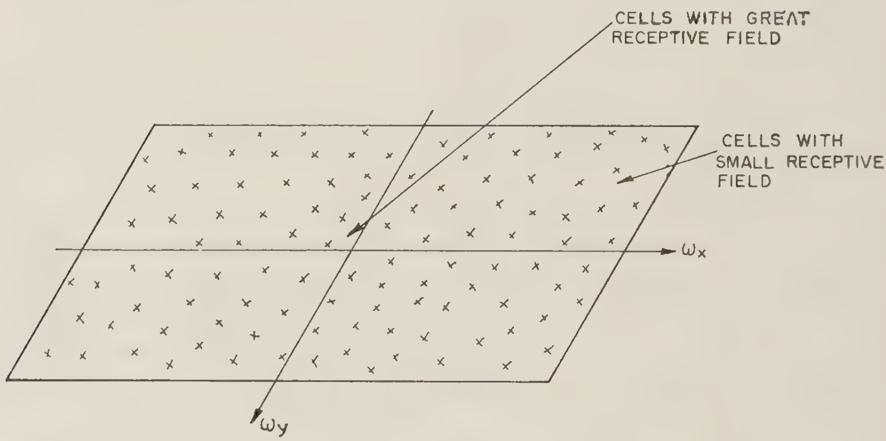


FIGURE 8 Receptive fields of the retina

As it has been said, each of the fibers of the optic track responds with greater intensity to a different spatial frequency. These frequencies are not located in the optic nerve in an ordered way, but they are randomly distributed. If it is supposed there are a mapping from the optic track to the visual cortex, according to the size of the receptive retinal field of the ganglion cells (Figure 8), a real Fourier transform of the input pattern is achieved in the later level. This function could be carried out at the visual cortex level.

SPATIAL FILTERING

It is known that the response of an organism to an external stimulus depends upon two procedures: a *recognition* procedure which identifies or interprets the input with respect to the recognition procedure and a *decision* procedure which determines the appropriate response under the interpretation of the stimulus and the internal needs of the organism. Assuming that the retina carries out a spatial harmonic analysis, the detection procedure may be accomplished by spatial filtering that in fact is a selection process.

There exists an analogy between this process of spatial filtering and the filtering in coherent optical systems. The optical filtering process involves three operations: (1) defocusing or spatially dispersing the input image; (2) modulation of the input image by the filter; (3) focusing or compression which occurs in the process of reconstruction. There exist neurophysiological data pointing out that the two first operations can be carried out in the retina. The last one—reconstruction—is not necessary in the organisms, since the transformed data can be used straighforward by these. This is equivalent to admit that the visual image an organism has of the external medium does not ressemble at all to a photograph in the sense that the incoming signals are coded and its projections to the cortex are distorted ones.

Now the attention will be directed towards the second aspect, since the first one has been already considered. Spectral analysis is an important operation by itself. The retina or the visual system of the mammals acomplishes it to use the data, modifying the content in spatial frequencies of the input, $F(\omega_x, \omega_y)$, by changing the amplitude and/or phase in each frequency, that is to say, by accomplishing an operation of spatial filtering. Following the optical analogy, these filters would act modulating the input signal:

$$R(\omega_x, \omega_y) = H(\omega_x, \omega_y)\, F(\omega_x, \omega_y)$$

where $H(\omega_x, \omega_y)$ is the transfer function of the filter.

The most adequate morphological structures of the retina that can support this phenomenon are horizontal layers, notably the amacrine cells ones, since they wholly cover all the extension of the retina. This does not contradict the fact that the spectral analysis is detected at the ganglion cells level.

The analysis can be carried out in lower levels, but only can be detected at this level since it is the first one in which the neurophysiological data are trustful.

In man, the filtering process, if it exists, would be possibly searched in the L.G.B. or in the corresponding cerebral area.

This spatial filters would be formed by passive elements—which are identified with the continuous horizontal layers of amacrine cells—that is to say, by elements that do not amplify the arriving signals, so that

$$|H(\omega_x, \omega_y)| \leq 1$$

The most elementary filters that can be found are the so called *binary spatial filters.* A binary filter has a transfer function whose modulus is 1 or 0. They are simply masks that allow or not to pass certain frequencies of the image.

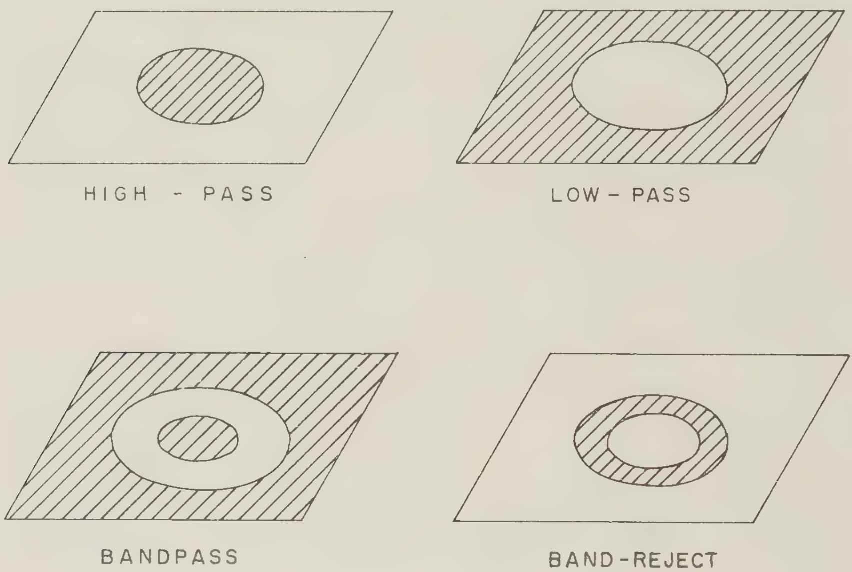


FIGURE 9 Filters

Following the terminology used in filter theory, the following filters can be distinguished: low-pass, high-pass, bandpass and band-reject filters (Figure 9).

If the stimulus is a square-wave grating, its Fourier transform is a sequence of maxima, that is to say, $\sin(k\omega_x)/k\omega_x$ functions located at the zero frequency, the fundamental frequency, and at multiple harmonics of the fundamental frequency. For example, if the mask acts in the frequency domain so that only the zero-frequency term and the first harmonic are passed, the original square-wave grating is converted to a sinusoidal grating having twice the fundamental frequency. As has been said before these filters could be located in the amacrine cells layer or in the dendritic trees for the lower mammals or in L.G.B. for the man.

These filters are the simplest kind of spatial filters, nevertheless they can perform many important operations. They can be useful for detecting extended periodic signals in the presence of random noise. This process can be accomplished in the retina, since owing to the eye movements (high-frequency tremor, slow drift, and saccade) all the signals can be considered as periodic. Further it is known the existence of the so called "neural noise", caused by the spontaneous activity of the cells, the fluctuations of the thresholds and the presence of the neuroglia, that is not a real "insulator" between the conductor chains. These filters would leave to pass the characteristic frequencies, but the signal-to-noise relation is greatly increased, since the noise spectrum is scattered more or less uniformly in the frequency domain.

This kind of filters can also be used to increase the contrast of a pattern by partially or completely eliminating the zero-frequency term in the spectrum. This operation is greatly related with the enhancement of edges and with that of differentiation that can be explained also by means of the lateral inhibition phenomena. These processes on the sensorial data being admitted by all the neurophysiologists.

In the retina it is also possible to admit other linear transformation, such as autocorrelation, cross-correlation, etc. Such operations can account for more realistic spatial filtering processes that those developed previously. Likewise the pattern recognition phenomena detected by several neurophysiologists at ganglion cells level in the lower animals can be explained assuming that the retina performs the linear operations mentioned above.

Acknowledgements

Many of the ideas contained herein were stimulated by conversations with Prof. R. Moreno Díaz of the University of Zaragoza (Spain). Three of the authors have worked with the support of an IBM S.A.E. grant via the Centro de Cálculo de la Universitad de Madrid.

References

1. F.W. Campbell, "The human eye as an optical filter", *Proc. I.E.E.E. (Inst. Electron. Engrs.)*, No. 6 (1968).
2. F.W. Campbell and D.G. Green, "Optical and retinal factors affecting visual resolution", *J. Physiol.*, **181**, 576–93 (1965).
3. Ch. Enroth-Cugell and J.G. Robson, "The contrast sensitivity of retinal ganglion cells of the cat", *J. Physiol.*, **187** (1966).
4. G.F. Cooper and J.G. Robson, "Successive transformations of spatial information in the visual system", *Symp. on Pattern Recognition*, *Teddington* (1968).
5. F.W. Campbell, B.G. Cleeland, G.F. Cooper, and Ch. Enroth-Cugell, "The angular selectivity of cortical cells to moving gratings", *J. Physiol.*, **198** (1968).
6. F.W. Campbell and J.J. Kulikowski, "Orientational selectivity of the human visual system", *J. Physiol.*, **187** (1966).
7. L. Maffei, "Electrical activity at the different layers of the vertebrate retina", *Lecture, School of Phys., Enrico Fermi, Varenna*, Italy (1968).
8. J.G. Santesmases and F. Rubio, "Modelo de inhibición lateral en el ojo compuesto del Límulus", *Annales Real Soc. Espan. Fis. Quim.* Tomo LXIV, 243–50 (1968).
9. F. Rubio, "Modelos de procesos retinales: Simulación electrónica y en calculadoras", *Tesis, Universidad de Madrid* (1969).
10. A. Gallego, "Connexions transversales au niveau des couches plexiformes de la rétine", *Actualites Neurophysiol.*, **6**[e] *Sér.* (1965).
11. R. Moreno et al, "Applicación de las transformaciones integrales al proceso de datos en la retina" *Revista de Automática,* **n. 5** (1969).

CHAPTER II–14

A theory of size and intensity invariance and the neural origin of visual illusions in the brain

W.K. TAYLOR

University College, London

Summary

When we learn to recognize an object the size of the retinal image is often restricted to a certain range during the training phase but may be many times larger or smaller when it is recognized on future occasions. In addition, the illumination of the object may differ by many orders of magnitude between the training and recall presentations. It is suggested that specialized neural networks in the retina and brain preprocess the widely differing patterns of retinal excitation produced by the same object at different distances and intensities, but the same orientation, so that the final neural pattern supplied to the memory areas of the brain is invariant. Thus the memory areas responsible for recognizing one view a particular car, for example, would receive the same processed signals during large variations of street lighting and car distance. The distance and brightness of a familiar object can, however, be estimated from the size and intensity of its retinal image so that this information must be stored separately. The memory capacity required for these two parameters is insignificant, however, when compared with that required to recognize complex shapes. Theoretical models of neural preprocessing networks posessing the desired properties are represented in a signal-flow-graph notation and their properties are investigated quantitatively. It is shown that their function partially resembles that of an automatic zoom-lens that increases or decreases effective magnification as an object receeds or approaches, so that its image size and brightness on a television screen remain constant, the magnification and intensity being indicated on separate dials. The theory leads to explanations of visual illusions in terms of neural signals. It is shown, for example, that the two equal length lines of the Müller–Lyer illusion produce measureably different signals in the size-estimating area of the neural model.

When we are shown an unfamiliar object at a certain distance and told its name it is often possible to recall the name when the object is seen on a future occasion at a different distance and a different level of illumination. We may, for example, be introduced to someone at armslength or handshake distance in a dimly lit room. The face may still be recognized, however, when it is seen later at a much greater distance in sunlight. The retinal image on each occasion falls on quite a different set of receptors so that a simple correlation with a stored representation of the receptors activated on the first or learning presentation is not possible. At the same time, however, we know that the object is further away than when it was first presented and the distance can be estimated by comparing the new retinal image size with the size that was recorded when the object was at armslength. This is additional information to that gained by accommodation of the eye or binocular fusion and may conflict with the latter in visual illusions.

Experiments with machine learning and recognition of faces on the UCLMII learning machine[1–3] have shown that approximately 20,000 bits of memory are sufficient to recognize 10 faces providing they do not change in size or position by more than a few percent of their training values. There is little evidence to support the idea that we automatically recognise initially unfamiliar complex objects in any orientation after training in only one orientation, although on some occasions this may be possible by recognition of parts, texture, or colour, etc. If the problem is limited to that of size and intensity invariance then using a simple correlation technique it is necessary to increase the memory capacity by an order of magnitude if a significant range of object sizes are to be stored and to repeat the training procedure for a number of object sizes distributed throughout the range. A more practical solution is to use a single "shape" memory for all possible sizes and to transform the different sizes of retinal image to one standard size that can be supplied to the shape memory, at the same time storing the comparatively small amount of extra information about the transformation that was required to do this. The suggested operation is similar to that performed by a zoom lens on a television camera. The lens can be adjusted by a single parameter to keep the image of an actor's face covering the camera tube face as his distance from the camera varies over a wide range. In the absence of any such mechanism in the optics of the eye it is suggested that there is an inborn neural network that rapidly performs part of the necessary operation in conjunction with the stored memory of the object established on previous presentations.

One thing that the brain has in abundance is parallel connections, as evidenced by the large volume of white matter. This paralleling of millions of relatively low capacity information channels results in certain types of calculation being performed at a much higher speed than is possible in the relatively small number of fast channels of conventional computers. By employing integrated micro-circuit techniques, however, there is a strong possibility of exceeding the parallel information processing capacity of the brain in the foreseeable future.

One clue to the possible size invariance soluti on adoped in the brain is the poor memory for absolute size which is about $\pm 5\%$ for all sizes. In one experiment, for example, experienced draughtsmen were asked to draw lines of various commonly used lengths such as $\frac{1}{8}$ in., $\frac{1}{4}$ in., $\frac{1}{2}$ in., 1 in., 2 in., 4 in., and 8 in. without using a graduated ruler, and the mean modulus of error was approximately constant at 1 part in 20 for all sizes. This suggests that the memory for all sizes of object has a quantization error of 1 part in 20. If the retina is to be well matched to this memory there is no point in retaining an accuracy of $\pm$ one or two receptors from the entire retina of approximately 10^8 receptors. It is well known that resolution is reduced with distance from the fovea due to the summation of large numbers of receptors at single ganglion cells and it is suggested that this summation process is continued in the brain until only a relatively small number of independent channels finally reach the shape memory, and that only a small proportion of these are actually activated by a visual scene. The memory would thus receive a coarsely quantized sample of the transformed or preprocessed retinal image that nevertheless contains sufficient information to select the name of the object from the memory or to initiate other selective or motor activity. This severe reduction of redundancy does not necessarily mean that small significant details like the beauty spot on a cheek are discarded since it has been shown that a detail filter network[4–6] will allow such discontinuities to pass through whilst suppressing the smooth skin except at its boundary.

The important question is how to carry out a summation process that will give equal resolution for all distances and hence for all retinal image sizes of an object. A limit is obviously reached as an object like an aeroplane becomes so far away that it only covers a few retinal cones. The number of receptors required by the UCLMII machine for face recognition was 100, arranged in a 10×10 matrix but 100 cones were not sufficient for identification of the same faces by human observers. The minimum number of cones required for the recognition of reasonably complex shapes has been estimated to be in the region of 400 and this is assumed to be the required number of inputs to the shape memory. It will be denoted in general by N.

The cones of the retina are arranged at random with the constraint that there is approximately linear density in all directions. (Close packing of the circular photocells in the learning machine UCLMII results in a non-uniform linear density which is more sensitive to edges that coincide with the rows.) If the mean density of receptors per unit area is ϱ then the size of a circular area containing N cones has a diameter d_1 given by

$$\varrho\pi\left(\frac{d_1}{2}\right)^2 = N \tag{1}$$

Assuming rods and cones to have the same diameter and constant density ϱ the number of receptors in a retina of overall diameter D (assuming an equivalent plane retina) is approximately

$$10^8 = \varrho\pi\left(\frac{D}{2}\right)^2 \tag{2}$$

and if D is taken to be 4 cm then

$$\varrho = \frac{4\times 10^8}{\pi\times 16} = \frac{10^8}{4\pi} \quad \text{receptors/cm}^2$$

giving the linear distance occupied by one receptor as

$$\frac{2\sqrt{\pi}}{10^4}\times 10^4 = 2\sqrt{\pi} \quad \mu\text{m}$$

which is the right order of magnitude.

We now assume that the receptors are grouped into S subretinas of different overall diameter d_n which are formed by summating receptor signals into N independent uniformly distributed groups, each containing a summation of signals from the same number of receptors given by the relationship

$$\frac{\varrho\pi}{N}\left(\frac{d_n}{2}\right)^2 = \left(\frac{d_n}{d_1}\right)^2 \tag{3}$$

If the diameter of the nth subretina is d_n we have the problem of choosing the values of $d_2, d_3, \ldots, d_n, \ldots, d_{S-1}$ for the $S-2$ subretinas of size intermediate between d_1 and d_S. It seems reasonable to suppose that central vision should be given more representation than peripheral vision and an exponential distribution of the form

$$d_n = d_1 c^{n-1} \tag{4}$$

is assumed. Knowing d_S we find

$$c = \left(\frac{d_S}{d_1}\right)^{1/S-1}$$

As an example let $S = 3$, $d_S = 27$ and $d_1 = 3$. This gives $c = (27/3)^{1/2} = 3$, $d_1 = 3$, $d_2 = 9$ and $d_3 = 27$. The three subretinas are shown separately and superimposed in Figures 1a and 1b respectively.

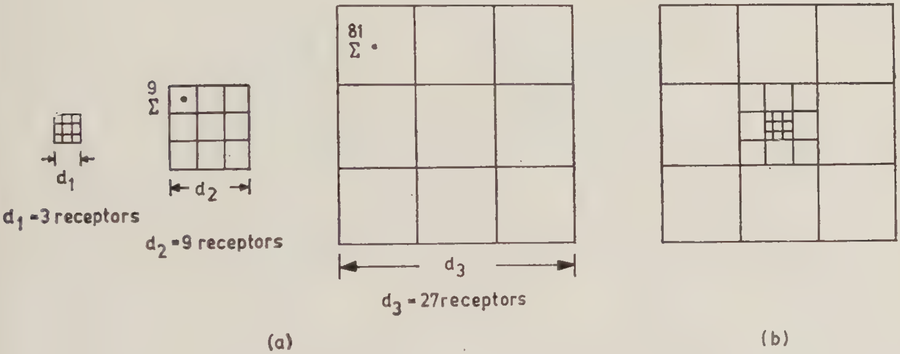


FIGURE 1 Subretinas

We require the maximum signal, obtained by summation over each square, to be the same for all subretina sizes. It is thus necessary to normalize the sum by dividing it by the number of receptors that contribute to it. In other words, for the nth subretina we form

$$\frac{1}{\varrho\pi\,(d_{n-1}/2)^2} \sum_{i=1}^{\varrho\pi\,(d_{n-1}/2)^2} x_i \qquad (n > 1) \tag{6}$$

For the example of Figure 1 the elements are square, and if d_n is measured in receptor units we have in general a signal

$$\frac{1}{(d_{n-1})^2} \sum_{i=1}^{(d_{n-1})^2} x_i \tag{7}$$

which gives $1/9 \sum_{1}^{9} x_i$ for the second or medium size subretina ($n = 2$) and $1/81 \sum_{i=1}^{81} x_i$ for the large subretina. Thus for unit retinal element signals each subretina gives unit signals.

The way in which this may come about in the nervous system is by the physical limitation of neuron area available for synaptic contact. Thus a ganglion cell receiving signals from 81 receptors will only have 1/81 of its area available for each receptor whereas a ganglion cell with the same area

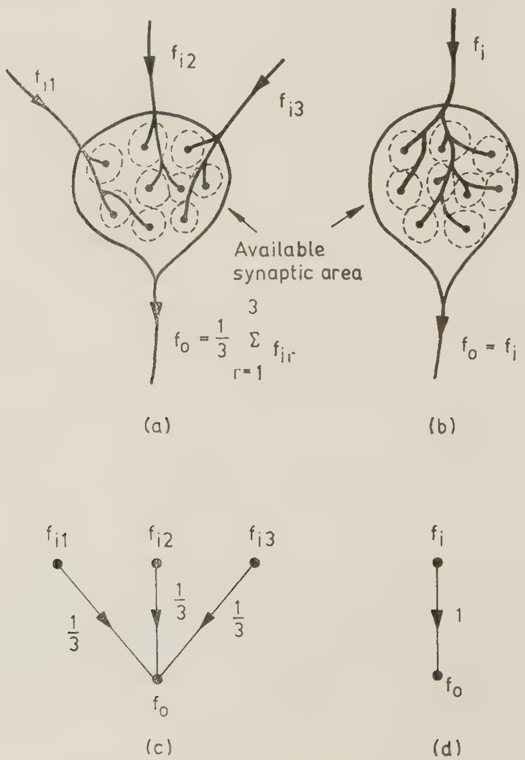


FIGURE 2 Branching

but only 9 afferent fibres will accommodate 9 synaptic endings per fibre, giving each a gain of $9/81 = 1/9$, etc. With larger numbers the correct weight could thus be established statistically by random fibre growth followed by branching and synaptic multiplication until the available cell area is used up. Figure 2 illustrates this effect for the simple case of a neuron with nine units of area available for synaptic contact. In (a) the neuron has three afferent fibres each of which can make three synapses before the neuron is full but in

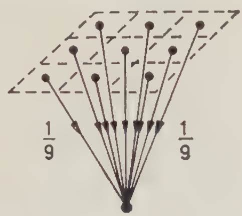


FIGURE 3 Outputs of subretinas

(b) the branching of a single fibre continues until there are nine synapses for the single afferent. Any additional branching could be eliminated eventually during development if only branches that find a vacant site and "take root" survive in the long term.

This hypothesis also ensures that a neuron cannot be overdriven by synaptic bombardment, even if all endings are excitatory and all the afferent frequencies have a maximum value of say 500 impulses/second. In Figure 2, for example, there are nine synapses and each is therefore assumed to have a weight of 1/(No. of synapses) = 1/9. The resultant weight per afferent axon in (a) is $3/9 = 1/3$ and in (b) $9/9 = 1$, so that $f_0 = 1/3 \sum_1^3 f_{ir}$ and $f_0 = 1/9 \sum_1^9 f_i = f_i$ respectively. Most neurons in the brain have several thousand synapses but the same principle can be extended to any number, the smaller weight of each synapse being compensated by greater numbers so that all active neurons fire at frequencies within the permissible dynamic range of zero to approximately 500 impulses/second. This high figure is rarely attained under normal conditions since inhibition also plays a big part in restricting the frequency range.

The signal flow graph (S.F.G.) notation shown in Figure 2(c) and (d) provides a convenient shorthand for neurons and their synapses since the nodes represent the algebraic summation of the synaptic depolarizing and hyperpolarizing currents and the branch multipliers represent the synaptic

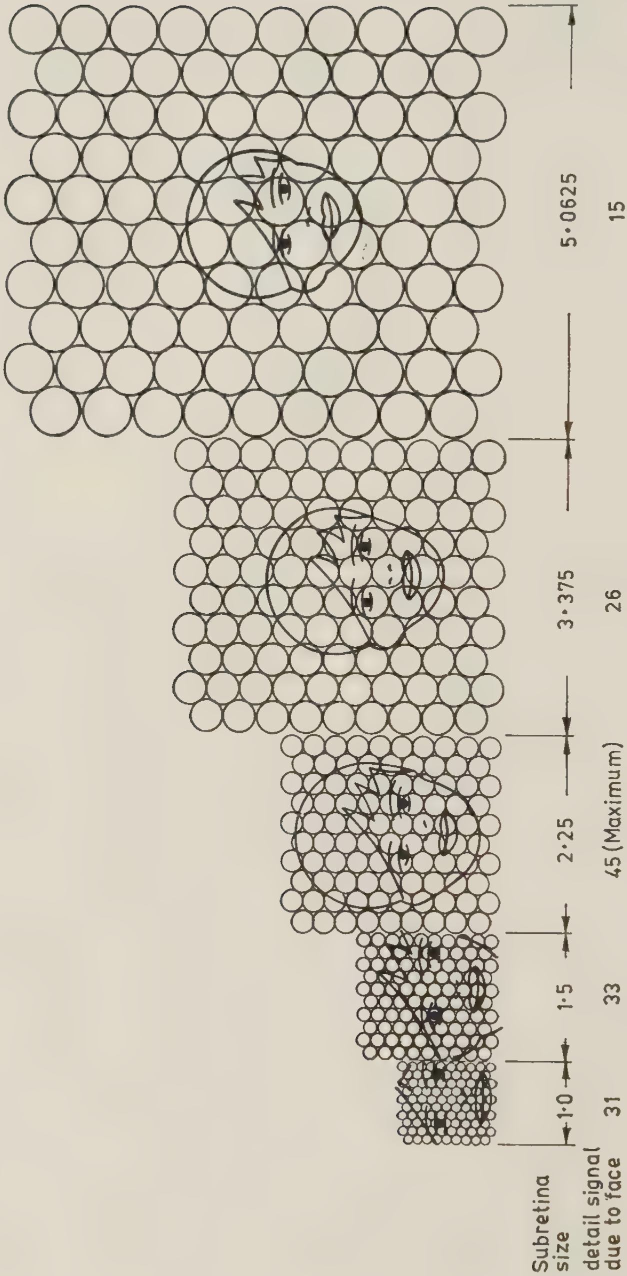

Subretina size
detail signal due to face
1·0
31
1·5
33
2·25
45 (Maximum)
3·375
26
5·0625
15

FIGURE 4a

weights with $+$sign for excitation and $-$sign for inhibition. The nine outputs from the middle size sub-retina in Figure 1(a) may each be represented as shown in Figure 3.

The learning machine UCLMII has a matrix of 100 photomultipliers arranged in rows as illustrated in Figure 4. If this machine were required to recognize objects of varying size automatically and instantaneously, thus

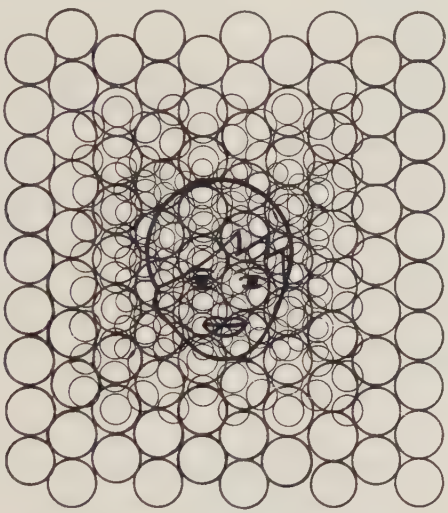


FIGURE 4b Dominance of a subretina

ruling out the possibility of a conventional zoom lens, it would require many more receptors but no additional memory capacity since all the subretina outputs would converge onto the original memory. This would cause complete confusion if it were not possible to arrange for just one of the subretinas to gain control of the final common path at any instant of time.

When five subretinas, each containing 100 receptor elements, are employed the individual subretinas and composite retina for $c = 1.5$ is as shown in Figure 4. The superimposed outline of a face serves to illustrate the way in which one of the subretinas becomes dominant.

SUBRETINA DOMINANCE

The constant and slowly changing areas of a visual scene can be eliminated by a detail filter[4-6] that has a S.F.G. of the form shown in Figure 5 where each output is equal to the input directly above $-1/6$ (sum of six surrounding inputs). The resultant is clearly zero for uniform illumination of any intensity but non-zero at edges and corners, etc. In the nervous system negative nerve impulse frequencies are not possible and this can be simulated by including the diode D.

When the input image arises from a solid triangular object the corners give larger signals than the edges and the central areas give zero detail-filter output. Faces tend to be converted to line drawings and a measure of the quantity of detail in an object is the total length of edge in the filtered image measured in terms of the number of significant output signals. If the intensity is normalized the sum of a subretina detail-filter output is proportional to the detail seen by that subretina. It will now be evident that the detail output of any subretina has a maximum when the object image is approximately the same size as the subretina since a smaller image only covers a small percentage of the receptors, and in the limit only one receptor.

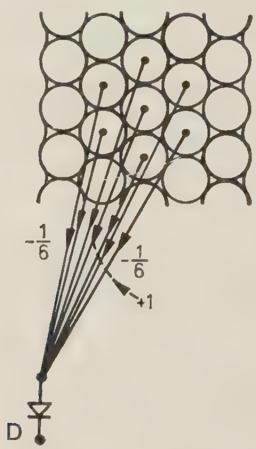


FIGURE 5 Detail filter

The subretina can, moreover, only detect a small section of an enlarged image and in the limit will, at best, only see a single piece of boundary. The detail signals are passed through a maximum amplitude[4-7] filter so that only the largest emerges. This is arranged to act as a gating signal for the subretina producing it so that only the best matched subretina has access to the memory and other cortical areas, all other subretinas being inhibited as illustrated in Figure 6. The size information simply denotes which subretina has control but it can be associated with distance information such as binocular fusion, accommodation, hand position at touching distance or number of walking paces required to reach the object. A machine or brain working

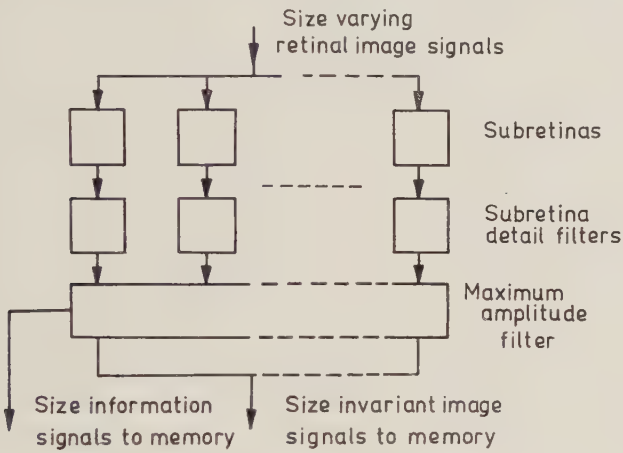


FIGURE 6 Inhibition of subretinas

on these principles could, for example, store the following information. The face of Miss A matches subretina 5 (Figure 4) when Miss A is at armslength. If Miss A is recognized on some future occasion on subretina 2, then her distance can be predicted as $1.5^4/1.5' = 1.5^3 = 3.375$ armslengths or in general if subretina p is used at a reference distance r and subretina q is used at a subsequent meeting, then the estimated distance on the second occasion is

$$r_e = rc^{p-q} \tag{8}$$

VISUAL ILLUSIONS EXPLAINED BY THE THEORY

The theory leads to simple signal-processing interpretations of a number of visual illusions that have long puzzled psychologists. Figure 7, for example, shows the well known Müller–Lyer illusion, an "explanation" for which is sometimes given in terms of the corners of a room or building, the upper figure being an inside corner of a room and the lower figure the outside corner of a building. This interpretation can be disproved, however, by placing circles on the ends, as shown in (b), and apart from this the illusion persists for different orientations of the figures. The theory of size invariance presented here interprets the illusion as a change in the size of subretina producing maximum detail as the point of fixation goes from one central line to the other. In the upper figure the detail extends over a greater distance than it does in the lower figure so that the upper figure gains control of a larger subretina. In making the size estimate it is necessary to store the information from one figure and to compare it with the information derived from the

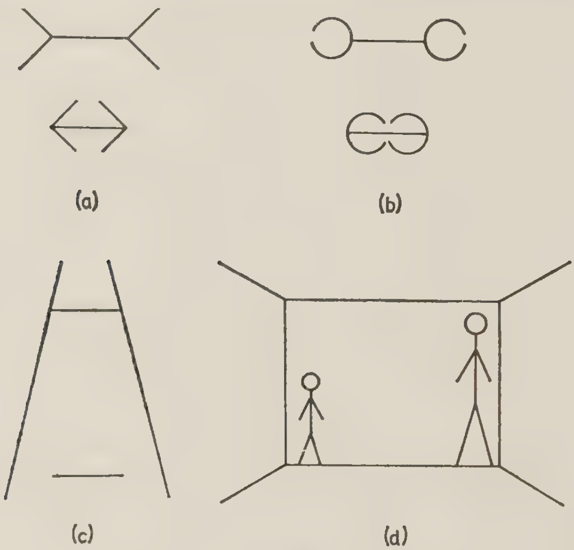


FIGURE 7 Müller–Lyer illusion

second figure since the two central lines are not sufficiently close to allow simultaneous comparison. The subretina size switching automatically accompanies the fixation changes and the subretina selection signals carry the illusion information.

Illusion (c) it often attributed to the perspective effect of railway lines, the lower line being a short sleeper. When viewed at right angles the tapering lines could represent a wall in perspective. In the absence of the sleepers the final common retina has the same image as the fixation point moves along the lines but the subretina size changes to match the line separation. The retinal image is interpreted as that of an object of fixed size, i.e. the separation of the rails, or height of the wall, at varying distance, thus over-ruling the constant accommodation and binocular fusion signals by stronger pattern weighting.

An object such as the sleeper, superimposed on the strongly weighted familiar tapered line pattern that is invariant with fixation, thus appears to vary in relative size as the point of fixation changes from one sleeper to the other.

If the lines in (c) are made parallel the illusion disappears but any reduction in the actual length of one sleeper appears on the final-common retina and is thus measurable as a change in the pattern signals. This is also the situation in the distorting room illusion shown in (*d*) in which the same man, or indeed any object known to be of fixed size, appears to change size as he moves horizontally on the baseline. He is of course walking towards or away from the observer at the same time and this would normally be allowed for without the room or with normal perspective and interpreted as a distance change, the final common retina image remaining size invariant and the appropriate subretinas being selected. The strong detail of the horizontal edges at floor and ceiling of the distorting room form a strong image of constant size on the retina which tends to gain constant control of the best-matched subretina, which is larger than that for the man alone. Thus instead of the final retina having a size invariant man, it has a picture of a man that actually shrinks as he moves to a greater distance and for any given set of system parameters, which include the memory weight, the apparent size change is predictable. The effect is reduced by the fact that the extent of the man's detail is variable. We can conclude that the memory of known shapes and their image size at a known distance will normally give a good estimate of their distance at any other image size except when two well-known shapes appear together and one is not in the correct proportion to the other. There is then conflict and when one gains control of a subretina the apparent size of the other will change. In the case of a faint after image that is projected onto familiar objects the latter always gain control and the apparent size of the after image can thus change by enormous amounts as it rests first on a finger tip and then on a distant house.

References

1. W.K. Taylor, "Learning characteristics of a trainable pattern recognition machine", I.E.E. Conf. *Publ.*, No. 44 (1968).
2. W.K. Taylor, "Machine learning and recognition of faces", *Electron. Letters*, **3**, No. 9, 436–7 (1967).
3. W.K. Taylor, "Machines that learn", *Sci. J.*, 102–6.
4. W.K. Taylor, "Pattern recognition by means of automatic analogue apparatus", *Proc. Inst. Electron. Engrs. (London), Part B*, **106**, 26, 198–209 (1959).
5. W.K. Taylor, "Cortico-thalamic organization and memory", *Proc. Roy. Soc. (London), Ser. B*, Part 976, **159**, 466–78 (1964).
6. W.K. Taylor, "A model of learning mechanism in the brain", in *Cybernetics of the Nervous System* (Eds. N. Wiener and Schade), Elsevier, Amsterdam (1965), pp. 369–97.
7. W.K. Taylor, "Neuronal models of pattern recognition and learning mechanisms in the brain", *A.G.A.R.D. Bionics Symp., Brussels, 1968* (in press).

CHAPTER II–15

Biohology—a new model of information processing

P. GREGUSS

New York Medical College, New York, U.S.A.

and

Ultrasonic Laboratory, RSRI, Búdapest, Hungary

Summary

Recently I have shown that the striking abilities of echolocating animals can be interpreted on the basis of holography, since when emitting the signal-bearing wave, the part of the brain that emits the wave sends this information to a second part of the brain, where the signal-bearing wave (the information) is received and processed. The information sent by the first part of the brain to the second is the coherent reference background. It may be shown that bioholography is a fundamental principle of biological information processing. It can, however, present itself in different forms and may be masked so that it is difficult to recognize it. This is attributed to the fact that unfortunately the transmission of informations attached to the phase of the stimulus wave has not up to now been studied directly in the biological information processing. The nervous system, however, can be regarded as a transfer function existing between a read-in function and a read-out function. On this basis and taking into account the quasi-simple rectangular geometric pattern of the neuron network, the steady presence of a coherent background can be made probable. With the aid of the bioholographic storage and retrieval of information in the brain, one may study the problem of "mind-brain" dualism.

INTRODUCTION

We have shown recently that the striking abilities of bats in pattern recognition can be interpreted on the ground of biological information-processing mechanism based on a holographic principle, i.e. during the processing of the information carried by the reflected ultrasonic waves a coherent background is used. When the bat emits an ultrasonic impulse of a certain frequency, at the same time the part of his brain which ordered the pulse to be

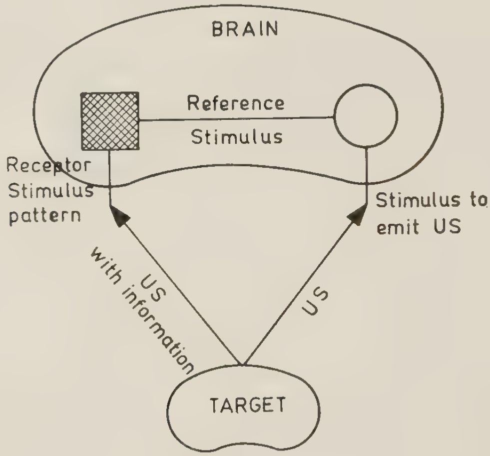


FIGURE 1 Processing of information by the bat

emitted sends a stimulus pattern on this to a second part of the brain, where the echo (the target information) is received and processed (Figure 1). Furthermore, every time when background noise increases the intensity of this, the stimulus pattern increases too without increasing the intensity of the emitted pulses. This is, however, in good agreement with the bioholographic conception, since—as shown is Figure 2—in this way the information carrying term can be amplified. If, for example, A is the amplitude of the reference stimulus—"the bat informs himself of the emission of the ultrasonic impulse"—and S the signal amplitude the resultant amplitude is

$$(R^2 + S^2 + 2RS\cos\varphi)^{1/2}$$

where φ is the phase angle between the reference stimulus and the signal bearing stimulus (phase information). R^2 is uniform, the second term S^2 negligibly small. The information is carried by the last term and is, indeed, amplified by the factor $2R$. In other words, if the bat due to the increasing backnoise increases the intensity of its reference stimulus the information carrying term which the bat is recording will be amplified by twice this amount, and so it is easy to understand that why the process signals may be as much as 2,000 times fainter than the background noise.

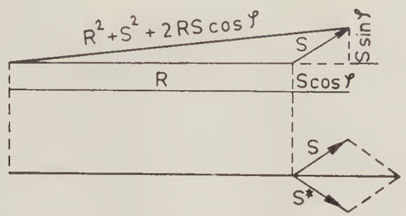


FIGURE 2 Amplification of information

The scope of this paper is to show that the holographic principle is not restricted to the information processing mechanism of bats, but it may be valid also for other echolocating animals, even in the case when the information carrier wave is independent of the processor, e.g. in visual perception when light waves are carrying the information.

ACTIVE INFORMATION PROCESSING

In the cases when the information carrier waves to be processed are generated by one or another way by the animal itself we are speaking of active information processing. The anatomical structure which may back the concept of the bioholographic information processing mechanism in these cases seems to have been found also in dolphins, as J.J. Dreher reports in one of his works[1]. These echolocating animals, it is known, are perceiving their emitted and from a target reflected ultrasonic impulses not with their ears as one would expect but with a structure on the forehead called "melon", which position is such that at least a part of the emitted ultrasonic impulse strikes it creating so a reference background. The histological examination has shown that this structure has a highly regular lay-out, and if we are comparing it with a photographic plate as wave sensor, its line/mm resolution appears to be in the order of 100 lines/mm. This, however, is about one order of magnitude less than needed for light holography, but if we take into account that there are using wavelengths several magnitudes longer than that of the light this resolution capability is more than enough for these animals.

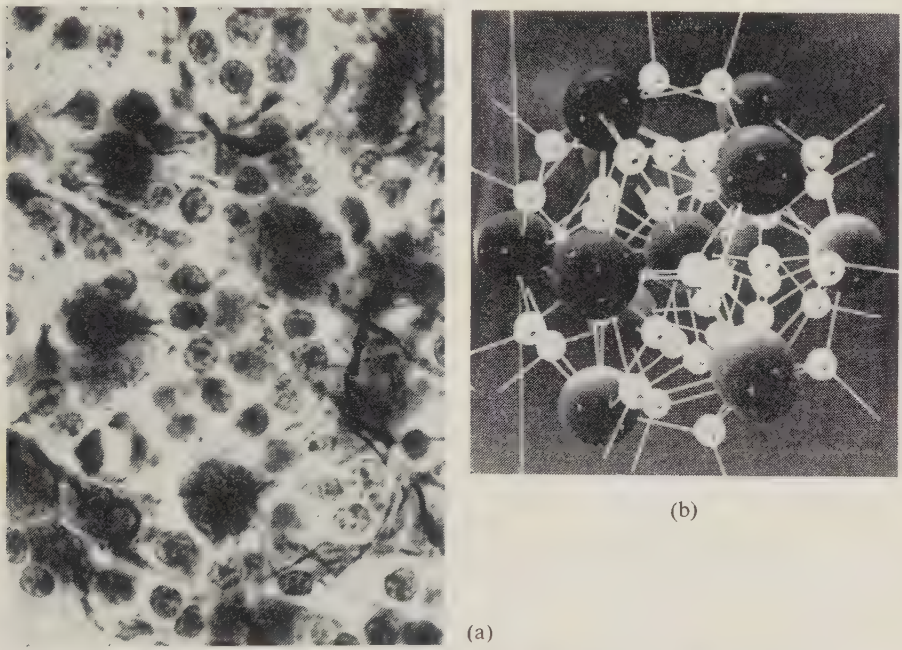
Another symptom, which refers to a bioholographic information processing, is the observation that the ultrasonic impulse emitted by most of these species is not a single but a double impulse. It is, for instance, true that in a hologram the whole information pattern is recorded, but by the wavefront reconstruction the ($S \sin \varphi$) part of it is lost. This, however, as shown by Gabor[2], can be avoided if two holograms are taken from the information pattern but with reference background delayed $\lambda/4$ from each other, and then the reconstruction of the information pattern is performed from the two holograms at the same time. In the double impulse of the porpoises the delay between the part impulses correspond to $\lambda/4$, i.e. the animal may record two holograms from the same information pattern with a reference background delayed by $\lambda/4$ and so by decoding he receives the whole information pattern, indeed.

PASSIVE INFORMATION PROCESSING

We think it is not by chance that the first experimental and histological references to a bioholographic information processing mechanism has been found in bats, porpoises, and fishes[3], i.e. in animals living in media which can be described as a real three-dimensional environment, since there the degree of freedom of movement is more than on a surface. Movement namely presumes processing of phase information. In connection with this we should like to refer to the life of salamanders. Nine months of the year these amphibia are living and moving on solid surface, but during their mating period they are living mainly in water. And what happens? As long as they are not spending the best part of their lives in water their sideline system—the organ which allows a holographic information processing of the informations carried by the waves generated during swimming—is totally degenerated but after returning in an environment in which movement requires more phase information processing, i.e. during the mating period of three months, the sideline system redevelops, but degenerates again after the animal leaves water.

The question, therefore, arises: can we suggest a holographic information processing mechanism also in the cases, where the information carrier wave is independent of the processor? Before going into details we have to point out that a living organism, as an information processing unit, can always be treated as an optical system, i.e. the biomechanism of the information processing can be described principally by the same mathematical methods as it is customary in electron optics and in communication theory. This, however, means that the information processing mechanism of a living organism can be treated as a sequence of Fourier operations. This, on the other side, refers to a neural network which is capable of performing Fourier operations.

Our belief is that such a neural network exists and this belief is based on the recent findings of J. Szentagothai *et al.* of the Medical University of



(a) (b)

FIGURE 3 Electron-microscopy study of granule cells (cerebral cortex) and its 3D model after A. Pellionisz

Budapest[4]. Szentagothai pointed out that the cerebral cortex shows a remarkable regularity, nearly that of a crystalline structure. He also draw the attention to the very striking phenomena that the ratio between the climbing fibres and the Purkinje cells is 1 : 1 and that the climbing fibres can be regarded as units reading out the background excitation or inhibition level of the Purkinje cells, which were excited by the input of the moss fibres. At the same time A. Pellionisz has shown through electron microscopic studies that each granule cell can be regarded as a tetrahedron having one-one dendrite at the four vertices (Figure 3). This, however, refers to a structure similar to hybrid junctions used in Butler matrices, which has been applied by L. Rudolph[5] to data processing problems thus making possible the performance of Fourier transforms with microwaves in nanosecond rates. If we now consider that microwaves are propagating with a velocity in the order of 10^{10} cm/s and to perform a Fourier transform this network needs 10^{-9} s and then assume the neural network uses this technique, and knowing that the propagation velocity of a stimulus is between 10^2–10^4 cm/s, we get 10^{-1}–10^{-3} s rates which is in good agreement with the neural activity rates found experimentally. Hence there is reason to believe that the bioholographic concept

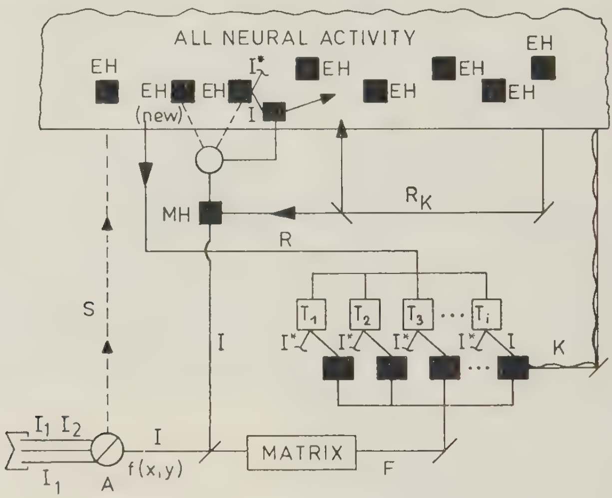


FIGURE 4 Model of information processing

is not restricted only to the case where the signal bearing wave is produced by the living being itself. This belief is also backed by the findings of I. N. Sokolov[6], who showed that the sensory input is always somewhere matched against a comparator before being processed further in the central nervous system. According to our assumption the comparator may be a stimulus pattern issuing from a given part of the whole neural activity.

Though our model does not claim morphological equivalents, its mechanism could lead to neurophysiological or experimental psychological studies, starting from the idea that a good deal of impressions (information) of the external world must be eliminated during information processing, otherwise our memory would be unable to store the entire flow of sensory information. So as can be seen from Figure 4 a function group A acts as a gate at the beginning of the information processing and lets only a stimulus pattern through which can be described by the function $f(x, y)$. (x and y are the characteristic coordinates of an information pattern.) At the same time this information pattern induces an arbitrary stimulus pattern R from the all neural activity and this serves as a reference background. Due to the holographic principle, under the effect of these reference stimuli a series of information elements are reconstructed from a finite number of "model elements" stored in the holographic form T_i. In the meantime the Fourier transform F of the information pattern $f(x, y)$, which passed through the function group A, is formed in a group of neurons and this can convolute with the information elements I_i reconstructed by R. As a result of this convolution a series of stimuli arises, which by further interference gives birth to a "code stimulus", K now representing the original information pattern $f(x, y)$. This code stimulus then generates a specific reference stimulus, called by us code stimulus R_K from the all neural activity to which all engrams (EH), memory traces, i.e. information patterns stored in holographic form, belong.

At this stage of information processing two possibilities arise: either there is an engram, which was formed with the specific reference background identical with the specific reference stimulus R_K, or not. In the first case —owing to the holographic principle—the original information pattern is reconstructed in the brain, i.e. the recall is performed. In the second case, however, the specific reference stimulus R_K interacts with the original information pattern $f(x, y)$ recording it in a holographic form (MH), which then will be a part of the all neural activity of the brain. This stage of information processing is identical with the formation of the "short-term" memory, which after several repetitions may change to "long-term" memory.

The next question which arises is: what is the material realization of the information pattern stored in holographic form. In this connection it should be mentioned that in the recent years it was assumed that in the process of learning the structure of RNA changes and in this way it may be a component of memory. It is well known, however, that the messenger or sRNA molecules in the cell (also in a brain cell) have not a long life time and, therefore, cannot serve as the basis of a really long-term memory. So we are forced to look after some memory mechanism which is based on the structural changes in DNA during the information pattern recording process.

It has been shown by J. Ladik of the Central Research Institute of Chemistry of the Hungarian Academy of Sciences that it is possible that the stimulus pattern, which is formed according to our model by the interference of a specific reference background elicited by the code stimulus pattern K and the original information pattern induced stimulus pattern, yields a change in the base sequence of DNA thus forming an engram, a memory trace of the original information pattern[7]. The change of the original information of these DNA molecules will remain in these DNA molecules for a long time and therefore they really can serve as an information storage. Of course, in this case the long-term memory can function only statistically. This means in the process of recording the stimulus pattern which contains in a holographic form the original information pattern—i.e. both the amplitude bound and phase bound informations—acts on a large number of brain cells. Some parts of the stimulus pattern, however, will not cause any change in the base sequence of the DNA molecules of these cells but at a fraction of cells there will be changes in the information stored in their DNA molecules. Since, however, in holograms as in a recording medium *every element* of the input information pattern is *entirely* distributed, the statistical character of this recording process does not produce a disturbance. Not even if we assume that most probably again the majority of these changes will occur in such region of the DNA molecules for which the information can not be recalled by the specific background stimulus pattern R_K. *Some* changes of the base sequence in *some* specific regions of *some* DNA molecules in a certain group of brain cells occur, however, and are specific in the sense that they do correspond to the storage of the information pattern processed. This model of biological information processing and storage is all the more attractive because it is consonant with concepts, which are viewing the brain as a statistically organized system that displays certain classes of orderly behaviour[8].

References

1. J. J. Dreher, *Acoustical Holography*, Plenum Press, New York (1968).
2. D. Gabor, *Proc. Roy. Soc. (London), Ser. A*, **197**, 454 (1949).
3. P. Greguss, *Fizikai Szemle*, **19**, 225–234 (1969).
4. J. Szentagothai, *Proc. I.E.E.E. (Inst. Elec. Electron. Engrs.)*, **56**, 960 (1968).
5. L. Rudolph and P. Blackwell, *Conf. System Sci., Honolulu* (1968).
6. I. N. Sokolov, *Neuronal Models and the Orienting Reflex, The Central Nervous System and Behavior*, M.A.B. Brazier, New York (1960).
7. J. Ladik and P. Greguss, *Intern. Symp. Biol. Memory, Tihany* (1969).
8. P. Greguss, *Nature*, **219**, 482 (1968).

CHAPTER II–16

A neurocybernetical approach to the problem of language

E. NICOLAU and C. BĂLĂCEANU
Bucarest University, Bucarest, Rumania

Summary

Language is the most important human communication system. From the cybernetical point of view the language is a formal system enabling a system A to communicate with a system B. The case B = A is also possible. Information is carried not only by some static elements, but also by some kinematic and dynamic factors. Our neurocybernetical models are able to explain the normal as well as the pathological aspects of language—such as aphasia and disartria. We consider the language as the result (output) of the speaking mechanism of the brain. We are developing some neurocybernetical models for the generation of sentences, for the interpretation of sounds, etc. Models for pattern extraction and recognition, and for decision making devices enable us to explain the interrelations between thinking and speaking. We are also studying the correlations between these decision making devices, thinking, and affection.

INTRODUCTION

Communication takes place if two systems are interconnected so that the output signals of one become input signals for the other. Communication may take place if at least one of the first system's output signals is accepted as an input signal for the other and if this signal bears information.

In order to establish communication, it is necessary that at least the following elements be available: a source of information, a channel, a receiver—these forming the physical equipment. Moreover, there should be a flow of information from the source to the receiver— a flow which is processed in each element of the system by means of certain algorithms.

Speech is the means by which a transfer of information from the source to the receiver is achieved. In any communication we should distinguish at least three aspects: the energetic or substantial aspect; the statistical aspect; the semantic aspect.

In respect of the first aspect, communication is ensured in speech by the energy of acoustic waves. Experiments have shown that in spoken communications acoustic energy has a certain distribution in groups called formants. The formants in their turn are grouped in phonema (vowels or consonants). From the mathematical point of view the vowels are oscillations having the form

$$f(t) = a(t) \cdot v(t)$$

where "a" is an amplitude variable in time and "v" a periodical function. The above representation explains the instantaneous vowel spectrum. It is possible that a similar representation also explains the aspect of consonants considered as transient phenomena. The spectral analysis of spoken language is not essential for a theory of the language, but it plays a fundamental part for the theory and realization of the systems of recognition of acoustical enunciations.

In respect of the statistical aspect, certain units of the language are initially defined. Formants have not been considered for the spoken language, as graphic elements composing the written letters have not been considered either. These are the subatomic components of the language.

Elementary units of the language are phonema for speech and letters for writing. They represent atomic units. Words represent molecular units and some typical expressions (as good-day, good-bye, etc.) may be considered as macromolecular units.

The statistical study of the language generally refers to the distribution of atomic and molecular units. Thus, for the distribution of phonema and letters the following experimental law is confirmed (Nicolau):

$$f_n = A \cdot 2^{-kn}$$

where f_n is the frequency of occurrence of the n rank letter or phonema; n represents the rank of the letter or phonema in their decreasing hierarchy determined in terms of the frequency of occurrence; and A and k are constants.

Other laws have also been confirmed for words; among them, the best known is that of Zipf:

$$f_n = An^{-k}$$

where f is the frequency of occurrence of the words in this instance.

Also in the context of the statistical aspect of human language entropies of the first and second order have been computed for letters and phonema (Nicolau; Pelikan). The researches of Voinescu *et al.* have shown that the statistical structure of the language (law of Nicolau, entropies) is not significantly altered in the case of aphasia. It appears, therefore, that this statistical structure is determined by the mode of extension organization of the human language.

THE SEMANTIC ASPECT

The semantic aspect is the most complex one. Notwith standing the statistical character of the language, each element of a sentence has a well-defined character so that it can be identified by means of a deterministic process.

In the context of the semantic organization of language we should specify that the number of units used, be they either atomic (letters, phonema, elementary symbols) or molecular (words), belong to a finite repertory (alphabet for the former, lexicon for the latter). Any enunciation or message is a result of the concatenation of several units. Atomic units thus generate the words; molecular units the propositions. Finally, any proposition is an organized succession of phonema (letters) or words belonging to a stock, to a thesaurus, which implies the existence of a long lasting memory.

Out of a collection of units (alphabet, lexicon), it is possible to obtain an infinite number of linear combinations by concatenation. The set of these may be expressed, as we know, by a tree-graph (Chomsky and Miller). The human language does not use them all. A collection of (grammatical) rules delimits a subset (subgraph). Each natural or artificial tongue is expressed by a subgraph corresponding to the respective tongue's own collection of rules. According to these rules two values be assigned to the sequencies of phonema; with or without meaning. The successions of meaningful linguistic units, therefore, obtained by using the rules of a tongue are called syntagma. Any biological or technical system communicating informations generates syntagma.

The alphabet, dictionary, and collection of rules permitting the generation of syntagma belong to the linguistic or syntagmatic level of communication, an instrument known under the name of tongue. The tongue can thus be modelled by means of an automaton of Turing machine type or of a "push-down system".

LANGUAGE AND COMMUNICATION

For language to serve as communication, it has to ensure a transfer of informations from source to receiver and influence the latter's thought or behaviour. Information is communicated about real or imaginary objects and phenomena, about feelings, etc. A formal analysis of the language in this respect has been started by Carnap and Bar Hillel.

The elements about which we communicate form a set (E). The tongue offers the set of words (M) which are the elements or monema of the language (Bresson).

The relations between E and M (between signified and significant) constitutes the corner stone of the semantic study of the language (F. de Saussure; Ogden and Richard; Quine; Roussel) and represents one of the most complex aspects of coding (the signification spheres of K. Goldstein).

The generation of a sentence implies the selection of words from the thesaurus and their grouping in a more or less long syntagmatic structure (of propositional order). This organization is brought about on basis of some algorithms, which assign to each word a certain place in the proposition according to the meaning and to its part in the respective syntagma. Likewise, algorithms command certain changes in the words (flexional) as well as the addition of some groups of phonema (prefix, etc.), or of some words which play the part of recognition indices or of logic operators (conjunctions, etc.). The structural analysis of a proposition may be expressed on ground of a graph-tree, which is the expression of syntactic constraints of the language (Sherry-Bossert; Lecerf and Ihm). These syntactic interrelations can also be represented by a network (Reich; Lamb).

The decoding of a sentence implies its fragmentation (segmentation, on basis of pauses or key words), then the recognition of words and the part they play in the proposition. This operation also implies a succession of operators which require an operational memory (which determines the maximum possible length of syntagma) and an algorithmization (Ida Rhodes). Modern automation offers a series of models and algorithms for the generation of syntagma and for their decoding (Nicolau and Bălăceanu).

We pass over simple systems which generate aleatory successions of phonema (letters) or words. Likewise we do not insist on the systems which, at a given signal, generate syntagma or successions of syntagma ready elaborated and recorded in the memory (reciting). The theatre involves a succession of such messages (recited by two or several automata), where the end of a message represents the starting signal on the next one. The use of courtesy formulae implies a selection of short recitings (therefore a strategy this time) in which, next to the releasing signal, an internal state of the automaton also intervenes (as in the case of the Mealy automaton).

A more complex class is represented by the translation machines. Here the problem consists in transposing a message from a tongue L_k in a tongue L_j according to the relation

$$S_j = T_j^k S_k$$

where S_j and S_k are the syntagma corresponding to the two tongues and T_j^k is the translation operator. Such an operation implies the construction of a sentence S_j, using the words and rules of tongue L_j but taking into account on basis of the dictionary and of the grammar of tongue L_k the input sentence S_k. A (simpler) translation variant is the transposition of a sentence from one symbolization system into another, but using the same tongue (for instance reading aloud a written text or writing at someone's dictation).

LANGUAGE ORGANIGRAM

However, the human language implies a much more complex mechanism. The generation of propositions as well as the decoding of syntagma does not only develop on the syntagmatic level. We mean the permanent intervention of paradigmatic relations between the sentence and that to which the sentence refers. These relations form a complex network which can be partially explored by psycholinguistic experiments (Herwes; Rosenzweig; etc.). It happens rarely that the human speech device is confronted with the ecolalic reproduction of messages received.

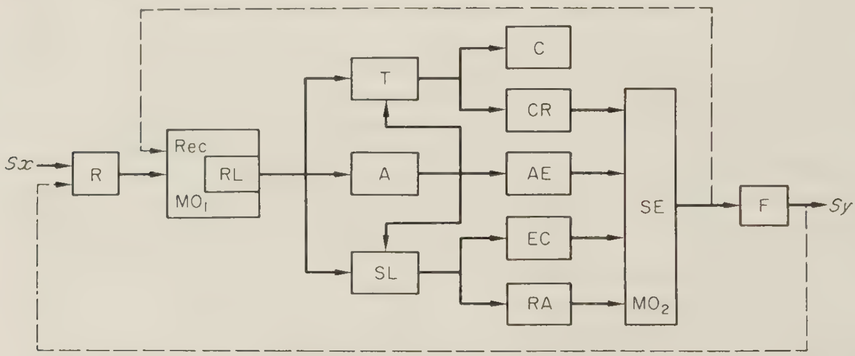
The operation of translation is also not very well known. The human speech device is generally confronted under the pragmatic aspect with the necessity of communicating something about something. In this instance, to the propositional syntagma is also attached a value of truth conferred by logical chrysippian functions (they are bivalent : true or false) or lukaszievician (polyvalent) functions. This sets the complex problem of relations language—thought.

Likewise, language also reflects the internal state of the system it generates. On the psychological level this state constitutes the affect. In the structure of any human enunciation we also find the affect participating:

$$m = m_i + m_a$$

where "m" is the message considered; m_i is the purely semantic message (capable of optimization on logical criteria), and m_a is the additional affection factor (often responsible of optimization in terms of aesthetic and emotional criteria).

The above theoretical data, as well as the analysis of some cases of ecolalic or transcortical aphasia (Bălăceanu and Rosianu) permit the elaboration of the following language organigram.



Input signals S_x are detected by receptors (R) and recognized (Rec). If the signals are linguistic, the operation of recognition is specialized (R.L.). A relatively simple logical selection (S.L.) may ensure the ecolalic recitation (E.C.) of the messages received or the automatic recitation (R.A.). The logical thought system (T) ensures correct answers (C.R.) or adequate behaviour actions (C). Finally, the effective tonalizers (A) may generate some affective expressions (A.E.) (for example: exclamations) and modulate the sentences. It is evident that expression by means of speech, writing or gestures is carried out by an execution servosystem (S.E.), which disposes of a stock of programs and which due to effectors (E) (in this instance the muscles) generates the output signals S_y. Various feedbacks ensure the system operation. Likewise, account should be kept of systems of operational or immediate memory (MO_1, MO_2), as well as a long lasting memory (L.M.).

In the present paper we have presented but a few of the neurocybernetic and psychocybernetic aspects implied in language.

CHAPTER II–17

*Properties of the short-time spectral analysis performed by the peripheral auditory system**

G. GAMBARDELLA

Università di Genova, Italy

Summary

It is well known that the basilar membrane behaves as a mechanical frequency analyser. A mathematical tool (which we call the "generalized spectral analysis integral") is developed for describing such a spectral analysis in a suitable range of frequencies and amplitudes and on the basis of a mathematical model of the basilar membrane due to J.L. Flanagan. By means of this integral it is shown that the form of the "short-time spectrum" ideally computed by the auditory system is invariant under linear time scaling of the input signal, which is not the case for most artificial spectral analysers. Indeed, in order to possess the "form-invariance property", the "weighting function" of the spectral analyser must belong to the class

$$w(\omega, t) = v(\omega t)\, t^b, \quad t > 0, \quad \omega > 0$$

where $v(\omega t)$ is any real function of the product ωt, b is a real constant and the whole function is absolutely integrable for any fixed ω. The properties of the spectral analysers defined by such a class are extensively considered and discussed.

* This work has been sponsored in part by the "Consiglio Nazionale delle Ricerche".

INTRODUCTION

The plan of the present work is as follows. First, we shall briefly describe the mechanical functioning of the human ear on the basis of the existing data and literature. Emphasis is put on the basilar membrane, which is that part of the system where a mechanical frequency analysis of the stimulus (sound) takes place. In adopting a mathematical model of the basilar membrane (developed by J.L. Flanagan[1]), a brief discussion is given about various assumptions on which it is based. Next, a mathematical tool is developed for adequately describing (on the basis of the mentioned model) the short-time spectral analysis ideally performed by the peripheral auditory system. This mathematical tool (which we shall call the "generalized spectral analysis integral") is capable of describing a very wide class of spectral analysers besides and including the model of the auditory system. It also covers, as a particular case, the conventional "short-time Fourier integral" and the well-known means of implementing it. Therefore its interest appears to be quite general, especially in the field of filtering and communication theory.

Finally, it is shown that the short-time spectral analysis ideally performed by the auditory system possesses the "form invariance property" (under linear changes in the time scale of the input signals). The underlying theoretical result we have found has an interest which goes beyond the case of the auditory system, since a very general condition has been derived for a spectral analyser to possess the Form Invariance Property. A brief discussion of the properties of this class of analysers is then made, and an inversion formula appropriate to them is given for recovering the processed signal. Finally, some foreseeable practical uses of these theoretical results are mentioned.

MECHANICAL PROCESSING OF THE SOUNDS IN THE EAR

In the hearing process, in man, three different parts can be schematically distinguished: (1) a purely mechanical stage in which the motion of the external air particles (the "sound") is transformed into the vibrations of the basilar membrane and the annexed parts of the "cochlear partition"; (2) the transduction process by means of which such mechanical vibrations induce or influence the discharge of nerve impulses in the cochlear nerve; (3) the neural processes which are associated with the acoustic stimulus and its perception. Our interest, in the present work, will be focused on part (1), that is the mechanics of the ear. It is the only part which has already been put on a rather firm basis, thanks principally to von Békésy's experiments[2,3].

Furthermore, our point of view about the mechanics of the ear will be mainly a functional one, aiming for a description of the peripheral ear in terms of systems and signal processing theory. A schema of the ear, containing only its essential attributes from our point of view, is shown in Figure 1.

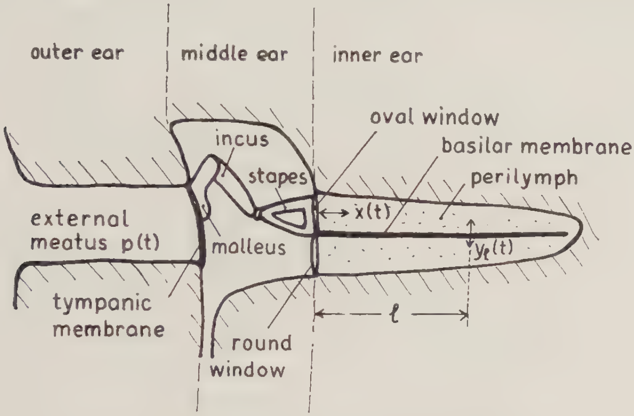


FIGURE 1 Crude schema of the peripheral ear

In such a schema the cochlea appears uncoiled and, of the different parts of the cochlear partition, only the basilar membrane is shown.

Very briefly and qualitatively, the mechanical operation displayed by the ear is as follows.

The sound energy impinging on the outer ear enters the external meatus which acts like a resonant cavity. As a consequence a minor amplification effect takes place around the frequency of 3000 Hz (corresponding to the first vibration mode of the cavity) before the tympanic membrane is excited by the sound energy.

The vibrations of the tympanic membrane are then transmitted via the ossicles chain to the stapes footplate (oval window) which moves like a piston on the fluid (perilymph) of the cochlea. It is worth mentioning that the middle ear (including the tympanic membrane) provides the impedance matching between the air medium of the outer ear and the liquid medium of the inner ear. It also provides a protection against very loud sounds for the more delicate inner ear structure.

The cochlea can be thought of as a tube filled with liquid (the perilymph) and divided longitudinally into two parts by an elastic strip (the cochlear partition); the walls of the tube are rigid (bone) except for one of the two terminal faces which is divided (by the cochlear partition) into two sections: the oval window coincident with the stapes foot-plate, and the round window which is in air contact on the external side.

Since the perilymph is practically incompressible, any displacement due to the stapes is compensated by a corresponding movement of the round

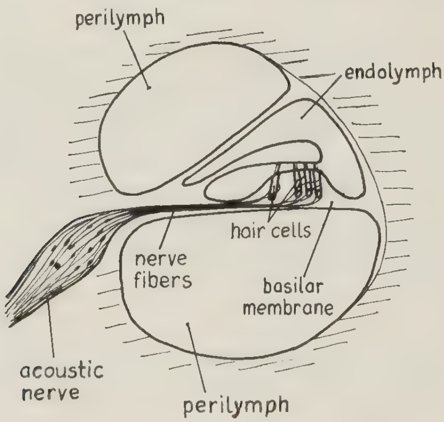


FIGURE 2 Crude schema of a section of the cochlea

window. The cochlear partition is itself filled with a liquid (the endolymph), as shown in Figure 2, and has two membranes; one of them, the basilar membrane, is far more important since it supports the "Organ of Corti", where the sensorial cells and the first nerve fibres are situated.

The section of the basilar membrane, as well as its elastic properties, gradually change all along its longitudinal dimension ($\sim$35 mm). As a consequence, if the frequency of a sinusoidal steady stimulus applied to the stapes goes from the low to the high values of the audible range ($\sim$20 Hz $\div$ $\sim$20 kHz), the region of the basilar membrane undergoing the maximum displacement moves correspondingly from the apex toward the base (see Figure 3). Furthermore the frequency response of each membrane point resembles a broad resonance curve whose peak frequency decreases with increasing distances from the base.

(It is essentially these phenomena which support the theory of the basilar membrane acting as a frequency analyser of limited resolution as first outlined by Helmholtz more than a century ago[4].)

From the qualitative approach used so far we shall pass now to a mathematical model of the mechanical operation of the ear. Of course in such an attempt one proceeds by abstraction, seeking to preserve in the model what are supposedly the essential features and discarding details of the physical structure. Our major assumption is that both the middle ear and the basilar membrane are time-invariant and linear systems within the frequency and amplitude ranges of normal speech.

Such a restriction of the stimulus variability may seem quite strong, but in fact it corresponds to what we think is the most frequent and important use that we make of our auditory system, namely speech processing. In

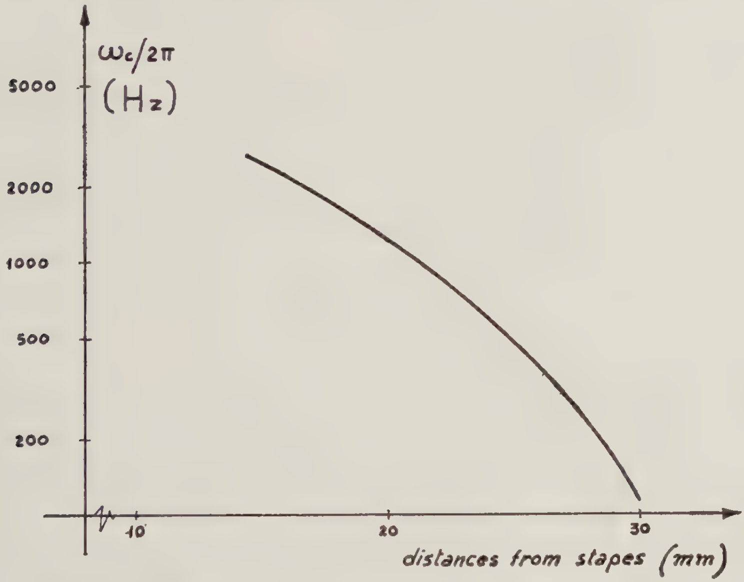


FIGURE 3 Stimulus frequency versus region of maximum displacement in the basilar membrane. (Adapted from Békésy[2], p. 442)

quantitative terms the restriction corresponds to a frequency range of about 100 Hz–5 kHz and an amplitude level not exceeding 100 dB above the standard reference level[5] ($2 \cdot 10^{-4}$ dyne/cm^2).

The input and the output of the middle ear considered as a system are assumed to be respectively the sound pressure $p(t)$ at the tympanic membrane and the stapes displacement $x(t)$ (see Figure 1). The latter constitutes also the input to the basilar membrane system, whose output consists in the transversal displacements ($y_l(t)$) of the basilar membrane itself at the different distances (l) from the base. Obviously the basilar membrane is modeled by a multiple-output system or a collection of systems all having the same input.

We shall now briefly discuss the time-invariance property assumed for our models.

The only physiological evidence against it seems to consist in the adaptive behavior of two small muscles, the "tensor tympani" and the "stapedius", in the middle ear[6]. The effect of this adaptive behavior, however, in our framework appears to be both relatively small and related to particular circumstances which we can neglect here. Therefore, we shall ignore it in our idealization.

As far as linearity is concerned, all the available physiological evidence largely supports it, at least within our assumed range of variability of the stimulus[2,7-11]. In particular, the most recent physiological experiments (some of which were performed with a technique exploiting the Mössbauer effect), have not shown any significant non-linearities, although displacements as small as a few tens of Ångstroms have been measured[7-9,11].

It seems worth emphasizing that the part of the auditory system which we are considering at this point ends at the basilar membrane and does not include the other parts of the cochlear partition, where apparently some non-linear effects do take place.

One possible objection to the linearity assumption is the lack of physiological evidence at low stimulus levels (say below 80 dB) or—which might be equivalent—the lack of measurements of the shape of the response (harmonics) when a high level stimulus (a tone) is employed. This scarcity of experimental data stems from the difficulty in measuring extremely small movements, of the order of 1 Å or even less[12]. If, for example, the stimulus is a 1000 Hz tone at threshold intensity ($\sim 2 \cdot 10^{-4}$ dyne/cm^2), the corresponding movement of the tympanic membrane is 10^{-1} Å; furthermore, extrapolating linearly from tha data at higher intensities, one finds that with the same stimulus the displacement of the stapes and of the Basilar Membrane would be of the order of 10^{-3} Å!

It has not been proved yet whether such displacements are perceptually significant or not. If they are, then[12] the action exciting the transducers (hair cells) can hardly be "shearing forces", at least at this level of displacements. In fact, physiologists have already started looking for evidence of other means of exciting transducers than shearing forces[12–14].

Finally even if non-linear phenomena do exist primarily at low amplitudes of the stimulus—speech in our case—their effect on the signal output could be ideally assimilated with the effects due to the background noise to which we are normally exposed while listening to people (including ourselves)[5,15,16].

Based on the above, we shall assume linearity in both the middle ear and the basilar membrane.

The next step in attempting to construct a model of the middle ear and the basilar membrane, is to provide the characteristic functions (or the impulse responses) of the time-invariant linear systems which we have decided to use. To this end we shall use Flanagan's[1] expressions, which are consistent with the physiological evidence cited above. These expressions show that the middle ear acts as a low pass filter (with cut-off frequency around 1500 Hz), whereas the basilar membrane acts as a distributed system, that is, a collection of band-pass filters each corresponding to a different distance l along the basilar membrane (see Figure 1). Since there exists a one to one correspondence between the filters center-frequencies ω_c and distances l along the basilar membrane, either parameter can be used in describing these filters. With this in mind, the entire middle ear–basilar membrane system will be modeled by a (continuous) set of band-pass filters with impulse responses, as given by Flanagan:

$$h(\omega_c, t) = (\omega_c t)^2 \, e^{-\omega_c t/2} \sin \omega_c t \tag{1}$$

On the basis of this expression we shall now study the mathematical properties of the corresponding spectral analysis. That one is dealing with a spectral analysis (in its more general sense) is beyond doubt, since the output signal of every band-pass filter has energy (or power), which somehow corresponds to the energy of the input signal around the center frequency ω_c of the filter*.

Next we shall see what the relationship is between the spectral analysis performed by the basilar membrane system (as we shall call the above model) and that usually described by means of the "short-time fourier integral" and performed by other physical (artificial) devices.

* In order to approximately evaluate this energy (on a short-time basis) a simple rectifier in series with an integrator at the output of each filter would suffice, and indeed a similar operation, although not yet well understood, seems to take place in the auditory system[17,19].

CONVENTIONAL SHORT-TIME SPECTRAL ANALYSIS

The classical Fourier integral is not adequate as a mathematical tool for describing the spectral analysis performed by any real mechanism, such as speech analysers. On the other hand, the "short-time Fourier integral" seems to be quite appropriate and we shall briefly consider it here.

If $w(t)$ is a "weighting function", real and identically zero for $t < 0$, the "short-time Fourier integral" of a signal $f(t)$ is defined as[1]

$$F(\omega, t_0) = \int_{-\infty}^{+\infty} f(x)\, w(t_0 - x)\, e^{-j\omega x}\, dx \qquad (2)$$

Of course both $f(t)$ and $w(t)$ must be such that the above integral exists. For our purposes $w(t)$ can be assumed to be the impulse time response of a realizable and stable low-pass filter. As an example, $w(t)$ might be taken to be the impulse response of a simple *RC* integrator: $w(t) = e^{-\alpha t}$, $t \geqslant 0$, $\alpha > 0$; $= 0$, $t < 0$. Furthermore, consistent with our physical interests in the present work and the assumptions of the previous section, $f(t)$ will be assumed to be a speech signal, hence a bounded and temporally limited function. Last, t_0 is a real parameter representing the "present" (or the observation) instant of time, and it takes, of course, finite values.

The above assumptions on both $w(t)$ and $f(t)$ are largely sufficient for ensuring the convergence of integral (2).

As is apparent from its very definition, $F(\omega, t_0)$ is the classical Fourier integral of the function $f(x)\, w(t_0 - x)$, which consists of the signal $f(x)$ taken until the present time t_0 and weighted in its past values according to the function $w(t)$. This is why $w(t)$ is called the "weighting function"; it is also called the "window", since its action corresponds to "seeing" the actual signal through a window or a screen emphasizing that part of the signal proximal to t_0. It is clear that $F(\omega, t_0)$ contains some degree of arbitrariness with respect to $f(t)$, consisting in the choice of $w(t)$. Equivalently this degree of arbitrariness may be thought of as corresponding to the choice of the particular device (or system) implementing the short-time spectral analysis of $f(t)$. It was stated above, in fact, that $F(\omega, t_0)$ is an adequate mathematical means for describing the analysis performed by real spectral analysers. One such a correspondence between $F(\omega, t_0)$ and "devices" can be quickly seen if we take, from (2), the "short-time amplitude spectrum" $|F(\omega, t_0)|$ and write it (through a simple change of variable) in the following way:

$$|F(\omega_c, t_0)| = \left| \int_0^\infty f(t_0 - y)\, w(y) \cos \omega_c y \, \mathrm{d}y + j \int_0^\infty f(t_0 - y)\, w(y) \sin \omega_c y \, \mathrm{d}y \right| \tag{3}$$

The two integrals in the above formula correspond physically to the outputs of two realizable filters; in fact, they are the convolutions between the signal $f(t)$ and the impulse responses, namely $w(t) \cos \omega_c t$ and $w(t) \cdot \sin \omega_c t$, of two Fourier-complementary band-pass filters.

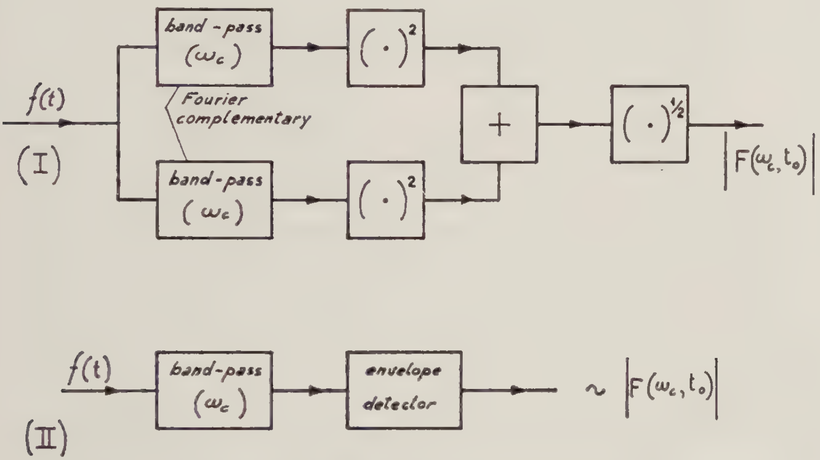


FIGURE 4 Methods for measuring (I) or approximating (II) the short-time amplitude spectrum $|F(\omega_c, t_0)|$

The symbol ω_c still represents the frequency—like the symbol ω in (2)—but it has been introduced for stressing its role of centre frequency in the band-pass filters mentioned above.

The outputs of the two filters at each instant t_0 must of course be squared, added, and the square root of the sum taken in order to obtain the short-time spectrum (see Figure 4). In practice, an approximation of $|F(\omega_c, t_0)|$ can be obtained by taking the time envelope of either band-pass filter (see Figure 4) as anticipated in the previous section (see Footnote on p. 453). The larger the centre frequency ω_c is compared with the bandwidth of the filter, the better is the approximation.

Of course as many band-pass filters (or pairs of them) are necessary as the number of frequency values for which the short-time spectrum is required.

If ω_c is allowed to vary continuously in $(0, \infty)$, then for the implementation of the short-time spectrum $|F(\omega_c, t_0)|$ in the above sense we may consider a distributed system of band-pass filters.

As the impulse time responses clearly show, the different band-pass filters are all formally generated by the same low-pass filter with impulse response $w(t)$. In particular their transfer functions $W(\omega, \omega_c)$ are all a replica of the

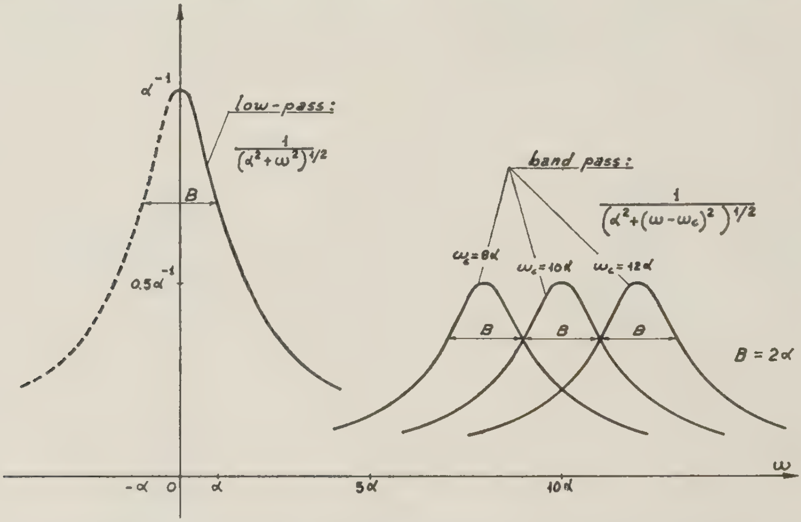


FIGURE 5 Transfer function amplitudes of several band-pass filters and the generating low-pass filter corresponding to the weighting function $w(t) = e^{-\alpha t}, t > 0$

same low-pass characteristic $W(\omega)$ (the transform of $w(t)$) essentially shifted around the different centre frequencies ω_c.

In the case of the example considered before ($w(t) = e^{-\alpha t}, t \geqslant 0$), we have

$$W(\omega) = 1/(\alpha + j\omega) \tag{4}$$

whereas the transfer function of the filter with impulse response $w(t) \cos \omega_c t$ is

$$W(\omega, \omega_c) = \frac{1}{2}\left[\frac{1}{\alpha + j(\omega + \omega_c)} + \frac{1}{\alpha + j(\omega - \omega_c)}\right]$$

If $\alpha \ll \omega_c$, we may further write (for positive ω)

$$|W(\omega, \omega_c)| \simeq \frac{1}{2}\left[\frac{1}{\alpha^2 + (\omega - \omega_c)^2}\right]^{1/2} \tag{5}$$

A diagram of this expression for a few values of ω_c/α is shown in Figure 5 together with $|W(\omega)|$ as deducible from (4). It is evident that what essentially changes from one filter to the other is the centre frequency. The bandwidth of the filters is always 2α (for $\alpha \ll \omega_c$), the same, of course, as that of the generating low-pass filter. The "frequency resolution"* of the analyser depends on this bandwidth: the smaller the bandwidth the better the frequency resolution, and, according to the "uncertainty principle", the worse the time resolution. Of course the right compromise as far as the resolution is concerned, depends upon both the signal and the use to be made of the short-time spectral analysis. In the case of the spectral analysers described by the short-time Fourier integral (2), we may further say that their resolution is uniform at all frequencies.

GENERALIZED SHORT-TIME SPECTRAL ANALYSIS

What we have essentially shown in the previous section is that the short-time Fourier integral (2) is an adequate mathematical tool for describing the spectral analysis performed by a set of band-pass filters centered at different frequencies but otherwise having the same transfer function. No less interesting from the physical point of view, however, is the case of spectral analysis performed by band-pass filters differing from each other in more respects than just their centre frequency. In particular we would like to consider the case of the band-pass filters formally generated by different low-pass filters, i.e. having an impulse response of the kind:

$$w(\omega_c, t) \cos \omega_c t \tag{6}$$

This is easily seen to apply to the basilar membrane system (see (1)) and to some artificial spectral analysers with non-uniform frequency resolution.

Besides their evident physical interest, such systems (defined by (6)) seem to be easily handled mathematically. Although, in fact, the short-time Fourier integral is no longer adequate, a slight (from the formal point of view) modification of it (namely substituting $w(\omega, t)$ for $w(t)$) seems to be adequate. We shall call such a modified formula the "generalized spectral

* It is worth pointing out that whenever we mention the "frequency resolution" of spectral analysers, a rigorous approach would imply the consideration of the noise level and its frequency distribution in the range of interest (not necessarily uniform, as we are implicitly assuming for simplicity's sake).

analysis integral", since the short-time fourier integral (2) becomes a particular case of it. Formally we have

$$F_g(\omega_c, t_0) \triangleq \int_{-\infty}^{+\infty} f(x)\, w(\omega_c, t_0 - x)\, e^{-j\omega_c x}\, dx \tag{7}$$

Convergence is ensured, as in (2), by our assumptions on both $f(t)$ and $w(\omega_c, t)$. The latter, as implied by (6), is assumed, for every ω_c, to be the impulse time response of a stable realizable low-pass filter, while $f(t)$ is still a continuous and temporally limited signal (such as a sample of speech).

The correspondence (in the sense anticipated above) between $F_g(\omega_c, t_0)$ and a set (or a distributed system) of band-pass filters, can be made more evident if we take the magnitude of (7), as we did for integral (2) in the previous section:

$$|F_g(\omega_c, t_0)| = \left| \int_0^{\infty} f(t_0 - y)\, w(\omega_c, y) \cos \omega_c y\, dy + j \int_0^{\infty} f(t_0 - y)\, w(\omega_c, y) \sin \omega_c y\, dy \right| \tag{8}$$

Here, in fact, one can easily recognize in the two convolution integrals the outputs of two complementary band-pass filters, with impulse responses $w(\omega_c, t) \cos \omega_c t$ and $w(\omega_c, t) \sin \omega_c t$. As an example, let us take

$$w(\omega_c, t) = \begin{cases} e^{-k\omega_c t}, & t \geqslant 0 \\ 0\quad, & t < 0 \end{cases}$$

with k a positive constant, e.g. $k = 0.1$. Considered as a "weighting function" or a time "window", $e^{-k\omega_c t}$ appears to operate in different ways upon the different spectral components of the signal. The higher the frequency of the latter, the shorter the window through which they are "seen". Perhaps the parallel with the corresponding example of the previous section is clearer in the frequency domain. Here the band-pass filters with impulse response $w(\omega_c, t) \cos \omega_c t$ have a transfer function which for positive ω and $k \ll 1$ (as is the case if $k = 0.1$), can be written

$$|W(\omega, \omega_c)| \simeq \frac{1}{2} \left[\frac{1}{(k\omega_c)^2 + (\omega - \omega_c)^2} \right]^{1/2}$$

Such transfer functions, as also shown in Figure 6, differ from each other not only in their centre frequencies, but in their amplitudes and bandwidths as well. In particular we see that the bandwidth of the filters is proportional to the centre frequencies ($B = 2k\omega_c$), that is, the filters have a constant Q factor. The frequency resolution of the corresponding spectral analyser is therefore inversely proportional to the frequency and thus uniform in percentage. This property also holds in the case of the basilar membrane

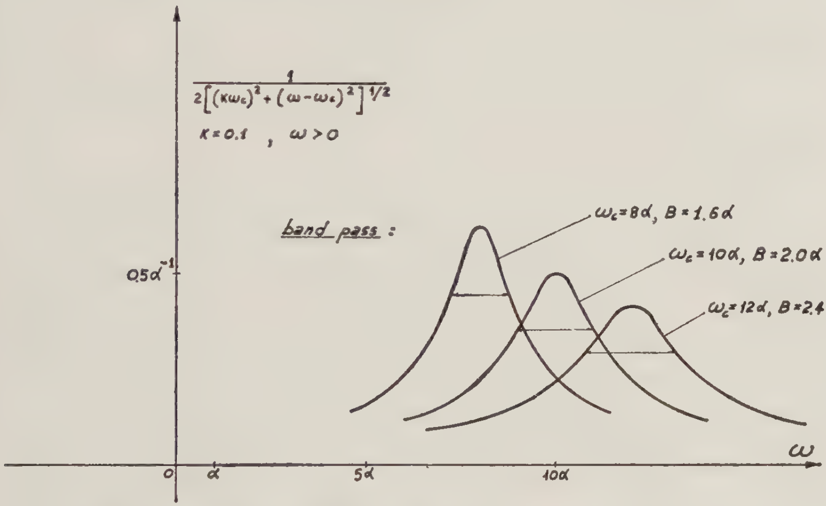


FIGURE 6 Transfer function amplitudes of several band-pass filters corresponding to the "weighting function" $w(\omega_c, t) = e^{-k\omega_c t}$, $t > 0$, $k = 0.1$. The centre frequencies ω_c have been chose equal to those used in the example of Figure 5

system, as implied by (1), and it corresponds to a well-known psychoacoustical effect.

We are now, perhaps, in a better situation for understanding the kind of signal processing represented by $F_g(\omega_c, t_0)$ (7) as compared to $F(\omega, t_0)$ (2). In the latter case $F(\omega, t_0)$ is, as already mentioned, the classical Fourier integral of the weighted signal $f(x)\,w(t_0 - x)$. Consequently $|F(\omega, t_0)|^2$ is a spectral energy density which represents an estimate of the signal spectral energy over a time spread proximal to t_0 and depending on the "length" of the time window $w(t)$.

In the case of (7), $F_g(\omega_c, t_0)$ has, for each ω_c, a value (a magnitude and a phase) equal to the value which the Fourier integral of the weighted signal $f(x)\, w(\omega_c, t_0 - x)$ would have for $\omega = \omega_c$. Formally we may write

$$F_g(\omega_c, t_0) = [F(\omega, \omega_c, t_0)]_{\omega=\omega_c} \tag{9}$$

Therefore $|F_g(\omega_c, t_0)|^2$ too, considered as a continuous function of ω_c, represents an estimate of the signal spectral energy over a time spread proximal to t_0; such a time spread, however, is now different for the different harmonic components of the signal, according to the "lengths" (one for every ω_c) of the "weighting function" $w(\omega_c, t)$.

With this comment in mind and on the basis of the entire previous discussion, we believe that integral (7) can be legitimately called a "spectral analysis integral".

From the mathematical point of view, however, some further comments on integral (7) are in order. It has already been stated (see (9)) that if $w(\omega_c, t)$ is a continuous function of ω_c, $F_g(\omega_c, t_0)$ too is a continuous (complex) function of ω_c. We may now ask ourselves if, from the knowledge of both $w(\omega_c, t)$ and $F_g(\omega_c, t_0)$, the signal $f(t)$ can be recovered as can be done in the case of integral (2) where indeed the problem is easily solved, since the inverse Fourier transform of $F(\omega, t_0)$ gives $f(x)\, w(t_0 - x)$, from which $f(t)$ can be recovered (of course from t_0 back, in the time range where $w(t_0 - x) \neq 0$). In the case of $F_g(\omega_c, t_0)$ a similar general solution does not exist. It seems necessary in this regard to know something more about the "weighting function" $w(\omega_c, t)$ (which corresponds to specifying the kernel of the Fredholm integral equation of the first kind represented by (7)[17]). For one class of "weighting functions" of great physical interest, however, an inversion formula of (7) can easily be found, at least formally, and we shall mention it here. It is the class of $w(\omega_c, t)$ which are functions of the product $\omega_c t$. Their physical interest comes from the fact that the "weighting function" of the basilar membrane system belongs to it (see (1)), and from other properties to be discussed in the next section.

In order to derive in this case the mentioned inversion formula of (7), one can make use of the "Mellin transforms"[20] which are defined as follows:

$$F_m(s) = \int_0^\infty f(x)\, x^{s-1}\, dx \tag{10}$$

$$f(x) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F_m(s)\, x^{-s}\, ds \tag{11}$$

It has been shown that such transforms can easily be derived from the Fourier transforms by means of a simple change of variable. Before applying them to our case, let us write equation (7) in the more convenient form

$$e^{j\omega_c t_0} F_g(\omega_c, t_0) = \int_0^\infty f(t_0 - y)\, w(\omega_c y)\, e^{j\omega_c y}\, dy \tag{12}$$

equivalent to the more compact form

$$g(z) = \int_0^\infty h(y)\, k(zy) dy \tag{13}$$

where

$$\begin{cases} z \equiv \omega_c, \quad g(z) = e^{jzt_0} F_g(z, t_0) \\ h(y) = f(t_0 - y) \\ k(zy) = w(\omega_c y)\, e^{j\omega_c y} \end{cases}$$

Taking the Mellin transform of both terms in (13), one ultimately obtains the formal result

$$H_m(s) = G_m(1 - s)/K(1 - s) \tag{14}$$

from which, by means of the inversion formula (11), the function $h(y) = f(t_0 - y)$ can be formally recovered.

THE FORM INVARIANCE PROPERTY

We have already remarked that the classical Fourier integral $F(\omega)$ is not adequate for handling mathematically the spectral analysis of a signal $f(t)$ performed by real mechanisms. For coping with these we might enstead turn (as has been shown in the previous sections) either to the "short-time Fourier integral" $F(\omega, t_0)$ (2) or, more generally, to the "generalized spectral analysis integral" $F_g(\omega_c, t_0)$ (7). The dismissed classical Fourier integral, however, has some nice (or at least well-stated) properties[21] which we do not know to be shared by the "new integrals" $F(\omega, t_0)$ and $F_g(\omega_c, t_0)$. Among such properties one which seemed to us particularly worth investigating, is the "form invariance property" (F.I.P.). By this we mean the well-known fact that to a linear time scaling of $f(t)$ there corresponds a linear frequency scaling of $F(\omega)$ (by an inverse factor) and vice versa. Now the question is: does the same property hold in the case of real spectral analysers? Or, more rigorously, under what conditions (if any) on the "weighting functions" does the F.I.P. hold in the case of $F(\omega, t_0)$ or $F_g(\omega_c, t_0)$? The interest of such a problem in communication theory (particularly speech processing) is apparent and we shall briefly comment on it later on.*

Let us first consider $F(\omega, t_0)$. Its expression for the scaled signal $f(kt)$ (a "speeded up" or "slowed down" $f(t)$) is

$$F_K(\omega, t_0) = \int_{-\infty}^{+\infty} f(kx)\, w(t_0 - x)\, e^{-j\omega x}\, dx \tag{15}$$

We say that the F.I.P. holds in this case if $F_k(\omega, t_0)$ differs from $F(\omega, t_0)$ simply by a linear scale change in the variables ω and t_0 and a multiplicative factor which can be a function of k. Otherwise stated, the F.I.P. holds if a solution exists of the equation

$$F_K(w, t_0) = \gamma F(\beta\omega, \alpha t_0) \tag{16}$$

where α, β, γ are real functions of k.

If the computations are carried through[22], it is possible to show that the necessary and sufficient condition for equation (16) to be satisfied is that the weighting function $w(t)$ belong to the class

$$w(t) = at^b, \quad t > 0 \tag{17}$$

where a, b are real constants.

We observe that this class of functions does not satisfy all the requirements imposed on $w(t)$ by our assumptions. One requirement was in fact the stability condition

$$\int_0^{\infty} |w(t)|\, dt < \infty$$

which obviously is not satisfied if $w(t)$ is a power function of t. In other words we cannot hope to find (or to design) any spectral analyser which is mathematically described by integral (2) and at the same time possesses the F.I.P. Most conventional spectral analysers in fact do not possess the F.I.P.

For the "generalized spectral analysis integral" $F_g(\omega_c, t_0)$ an equation similar to (16) can be written:

$$F_{g,K}(\omega_c, t_0) = \gamma' F_g(\beta'\omega_c, \alpha' t_0) \tag{18}$$

* Were our interest confined to "amplitude spectrum" analysers, the investigation of the F.I.P. should be made using the magnitudes $|F(\omega, t_0)|$, $|F_g(\omega_c, t_0)|$. It is easy to show, however, that the results in this case would not be different from those that we are going to find for $F(\omega, t_0)$ and $F_g(\omega_c, t_0)$.

The necessary and sufficient condition for Equation (18) to be satisfied can then be shown to be[24]

$$w(\omega_c, t) = v(t\omega_c)\, t^b, \quad t > 0 \tag{19}$$

where $v(t\omega_c)$ is an arbitrary real function of $(t\omega_c)$ and b a real constant. A "weighting function" may well belong to the class defined by (19) and still satisfy all the requirements imposed on it by our assumptions, including the stability condition.

For example the "weighting function" previously considered $e^{-k\omega_c t}$, $t > 0$, is one which satisfies both the stability condition and Equation (19). One further example, and indeed quite an interesting one, is the "weighting function" associated with the basilar membrane system: $(\omega_c t)^2 \exp(-\omega_c t/2)$ (see (1)).

We do not know, at present, if the F.I.P. is actually exploited by the auditory system for the perception of sounds; nevertheless the possibility that it might be seems interesting.

In order to gain a deeper and more general insight into the class of spectral analysers defined by Equation (19) it seems worth looking at the transfer functions of their corresponding band-pass filters. It turns out, in fact, that these characteristic functions always have a shape which is independent of the centre frequencies ω_c and a bandwidth proportional to ω_c. This can easily be shown by taking the Fourier transform of (19), which represents, for each ω_c, the transfer function of the equivalent low-pass filter:

$$\omega_c^{-(b+1)}\, W(\omega/\omega_c) \tag{20}$$

where $W(\omega)$ is the Fourier transform of the low-pass impulse response $\tau^b v(\tau)$, $\tau > 0$. If the bandwidth of (20) is small compared to ω_c, the transfer functions of the corresponding band-pass filters can finally be written

$$\omega_c^{-(b+1)}\, W[(\omega - \omega_c)/\omega_c] \tag{21}$$

which shows that the bandwidth is in fact proportional to the centre frequency ω_c (that is, the Q factor is constant), whereas the shape remains invariant: a change in ω_c only produces a shift and a scaling on the ω axis together with a scaling of the amplitudes by $\omega_c^{-(b+1)}$.

Conversely it can be shown that whenever a spectral analyser has band-pass filters with transfer-functions characterized by the same shape, the same Q factor and by an amplitude factor which is a power function of the centre frequency, then the "weighting function" associated with such an analyser belongs to the class (19) and the F.I.P. holds.

Finally, for the "generalized spectral analysis integral" computed by means of a "weighting function" satisfying Equation (19) there exists an inversion formula, as shown above (see previous section).

From the practical point of view Equation (19) seems to have very interesting applications, particularly in the field of time (or frequency) compression techniques, and the so-called "vocoders" in general. It will suffice to mention here the fact that, if a spectral analyser satisfies Equation (19) and does not work in "real time" (such as some "sound spectrographs"[1]) then the total amount of time required for processing a signal $f(t)$ can be reduced (by "speeding up" the signal) without changing the form of the wanted spectrum.

It is beyond the scope of this paper to discuss further applications of the F.I.P.

Acknowledgments

I feel deeply indebted to J. L. Flanagan, whose work, in addition to being instructive, has been used abundantly in this paper.

Further thanks go my friends Richard D'Ari and Bruno Preziosi for the helpful discussions I had with them on various points in this paper.

References

1. J.L. Flanagan, *Speech Analysis, Synthesis and Perception*, Academic Press, New York (1965), Chaps. IV, V, p. 126.
2. G. von Békésy, *Experiments in Hearing*, McGraw-Hill, New York (1960).
3. I.C. Whitfield, *The Auditory Pathway*, E. Arnold, London (1968).
4. H. Helmholtz, *On the Sensations of Tone*, Dover Publications, New York (1954).
5. H. Fletcher, *Speech and Hearing*, Van Nostrand, Princeton (1929).
6. A.L. Towe, "Audition and the auditory pathway", in *Physiology and Biophysics* (Ed. T.C. Ruch and H.D. Patton), Saunders, London (1965).
7. M. Rubinstein and H. Fischler *et al.*, "Measurement of stapedial-footplate displacements during transmission of sound through the middle ear", *J. Acoust. Soc. Am.*, **40**, No. 6, 1420–6 (1966).
8. H. Fischler and M. Rubinstein *et al.*, "Dynamic response of middle ear structures", *J. Acoust. Soc. Am.*, **41**, No. 5, 1220–31 (1967).
9. P. Gilad and M. Rubinstein *et al.*, "Application of the Mössbauer Method to Ear Vibrations", *J. Acoust. Soc. Am.*, **41**, No. 5, 1232–6 (1967).
10. J.J. Guinan and W.T. Peake, "Middle ear characteristics of anesthetized cats", *J. Acoust. Soc. Am.*, **41**, No. 5, 1237–61 (1967).

11. B. M. Johnstone and A. T. F. Boyle, "Basilar membrane vibration examined with the Mössbauer Technique", *Science*, **158**, 389–90 (1967).
12. M. Lawrence, "Energy conversion in the peripheral ear", in *Sensorineural Hearing Processes and Disorders, H. Ford Hosp. Intern. Symp.*, Churchill, London (1967).
13. N.Y. Kiang, "A survey of recent developments in the study of auditory physiology", *Ann. Otol. Rhinol. Laryngol.*, **4** (1968).
14. H. Davis, "Mechanisms of the inner ear", *Ann. Otol. Rhinol. Laryngol.*, **4** (1968).
15. J. C. Licklider and G. A. Miller, "The perception of speech", in *Handbook of Experimental Psychology* (Ed. S. S. Stevens), Wiley, New York (1951).
16. E. G. Richardson, *Technical Aspects of Sound*, Elsevier, Amsterdam (1953).
17. H. Davis, "A mechano-electrical theory of cochlear action", *Ann. Otol. Rhinol. Laryngol.*, **67** (1958).
18. W. M. Siebert, "Stimulus transformations in the peripheral auditory system", in *Recognizing Patterns* (Ed. A. Kolers and M. Eden), M.I.T. Press, Boston (1968).
19. K. Yosida, *Lectures on Differential and Integral Equations*, Intersciences, New York (1960).
20. E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, Oxford University Press, Oxford (1937), pp. 7, 315.
21. A. Papoulis, *The Fourier Integral and its Applications*, McGraw-Hill, New York (1962).
22. G. Gambardella, "Time scaling and short-time spectral analysis", *J. Acoust. Soc. Am.*, **44**, No. 6, 1745–7 (1968).

CHAPTER II–18

Study of certain clinical disorders related to the biosynthesis and regulation of steroid hormones by computer simulation

DAVID M. AUSLANDER

University of California, Berkeley, California, U.S.A.

and DINESH C. SHARMA

University of Michigan, Ann Arbor, Michigan, U.S.A.

Summary

The activities of several physiological systems are co-ordinated by the neuro-endocrine system which itself is composed of several subsystems. Hormone production and secretion within an endocrine organ are subject to passive regulation, as a consequence of their sequential organization, as well as to feedback regulation and other inputs. In this paper, the inter-glandular and intra-glandular modulation of adrenocortical hormone biosynthesis and secretion under physiological and certain pathological conditions have been studied and simulated. The simulation is a numerical digital computer model with a structure which is, to some degree, isomorphic to the structure of the original biochemical system. The model can be stimulated, changed, and perturbed to predict how the original system would respond under analogous conditions. Particular attention has been paid to a class of clinically important pathological conditions. Biochemically based hypotheses for the structure of the adrenocortical system and enzymatic rate equations, including the effects of inhibition and stimulation, form the basis for the model. A step-wise, quasi-linear integration method is used to calculate system responses.

INTRODUCTION

An organism, although it always arises from another organism, is basically a chemical system. In its operation, which includes its temporal course, maintenance, and reproduction, it appears to rely on the same physical–chemical laws on which an inanimate mechanical–chemical system relies for its performance. A hierarchy of control systems organized at a succession of levels insures proper functioning of the living system. The control at the higher level establishes constraints and integrates the functions of the subsystems of which it is comprised. The ability of the body to regulate its physical and chemical variables within close limits despite wide variations in physiological activities and environmental conditions is dependent to a large extent on the neuro-endocrine system. Its proper operation is necessary for the development of the zygote (genotype) into the adult (phenotype), for performance of the body processes throughout the life span and in reproduction. Disorders of the neuro-endocrine system have dramatic manifestations.

In modelling the biosynthetic process in the adrenal cortex we have attempted to summarize present understanding of this portion of the neuro-endocrine system in a precise quantitative form, hoping to obtain leads to further inquiry and plausible answers to questions of current interest. The purpose of the model is to act as a stand-in for the actual system so that quantitative results can be generated by simulation for experiments that are impractical, difficult, or too expensive to carry out *in vivo*.

The Neuro-endocrine System

The neuro-endocrine system, while co-ordinating the activities of various physiological systems, is itself composed of several hormonal systems, which in turn involve the nervous system, the endocrine glands, the transport system, the target organs, and finally the organs responsible for the catabolism and elimination of the hormones and their inactive products of metabolism. Despite the limited understanding about the mechanism(s) of hormone action, the physiological effects of hormones, their roles as regulators of various metabolic processes and their efficacy in endocrine disorders are well documented. A vast body of information about their chemical structure, biosynthetic pathways, physiological actions and pathological aberrations due to a lack or excess of several natural as well as tailor-made molecules with specific properties has also accumulated.

The Steroid Hormones

Although the hormones with known chemical structures fall into three broad categories, viz. small molecular weight amines, polypeptides or proteins, and steroids, the major developments from a practical point of view have been limited to the group of steroid hormones. Their structures are remarkably similar and they are synthesized primarily in the adrenal cortex, testes, ovaries, and the placenta of pregnancy. The hormones, as well as some of the intermediate precursors secreted by these endocrine organs into the circulatory system are transported in free and bound form to the target tissue. The catabolic processes, although not limited to any specific tissue, are mainly accomplished in the liver.

In humans, physiologically, the main hormones secreted by the adrenal cortex are cortisol and aldosterone, although it is capable of secreting "adrenal androgens" and estrogens also. Ordinarily this contribution is modest as compared to the gonads, but under certain circumstances the secretion of sex hormones by the adrenal cortex is of clinical significance. The testes mainly secrete androgens, particularly testosterone; ovaries and the placenta mainly secrete estrogens and progesterone. The pathways for the biosynthesis of corticoids, androgens, estrogens, and progesterone are interrelated, and a central scheme for their biosynthesis can be drawn; the specialized pathways can be added to this central scheme[1]. The common biosynthetic pathways shared by the adrenal cortex and the gonads probably reflect the common embryologic origin of these endocrine organs in the urogenital ridge[2].

SECRETION OF STEROIDS FROM THE ADRENALS BY HUMANS

Extensive studies indicate great complexity of the adrenocortical system. It is presumed that the ultimate objective of *in vivo* and *in vitro* studies is the extrapolation of the findings to the human body for a better understanding of the system under physiological as well as various pathological conditions. However, experimental studies, despite the ingenuity of the worker, often face serious limitations. In addition to the well-recognized constraints imposed by the choice of the subjects, their positions on the temporal course of development and the choice of the time domain selected for the study, the large number of state variables and their non-linear responses to the experimental conditions (inputs) severely limit the amount of information that can be obtained from any one experiment. Changes in some of the state variables may be observable, but technical difficulties often make measurement impractical. Also, it is difficult if not impossible, to observe the time course of a particular variable for an extended period without disturbing the rest of the system. The limitations of chemical analysis combined with the need for sequential sampling without perturbing the rest of the system make irreconcilable demands on the experimental investigation of certain phases of the adrenocortical system.

The development of a mathematical model on the basis of available information is an alternative to direct experimentation. This is necessarily an approximation procedure since the functional significance of the various components are only partially known and their interrelations are incompletely understood. Also, only a few of the kinetic parameters of the enzymic reactions have been studied. The mathematical construct described below attempts to incorporate the quantitative information available from experimental studies in this and other laboratories.

Digital Computer Model of the Adrenocortical Hormone System

The digital computer model is a numerical representation of the adrenocortical system. The usefulness of such a model is measured by two important properties: (1) the accuracy of the results, the degree to which they match the observed behaviour of the original system, and (2) its rationality —the degree to which the structure of the model matches the structure of the original system. This second property, while not quite as important as the first, is vital for the clinical or laboratory uses of the model. In the model described here the form and structure are derived directly from biochemical descriptions of the adrenocortical system (see Figure 1), and all of the parameters used in the model have direct biochemical significance.

One of the most interesting characteristics of this system is its behavior as a feedback control system to regulate the levels of various body chemicals. The initial temptation for an automatic controls-oriented analyst is to structure the model as a classical feedback control system. This temptation should be resisted, however, since examination of the probable biochemical mechanisms shows that the control is accomplished by means of parametric coupling—the concentration of one chemical affecting the production of another through inhibition or stimulation, a non-linear phenomenon.

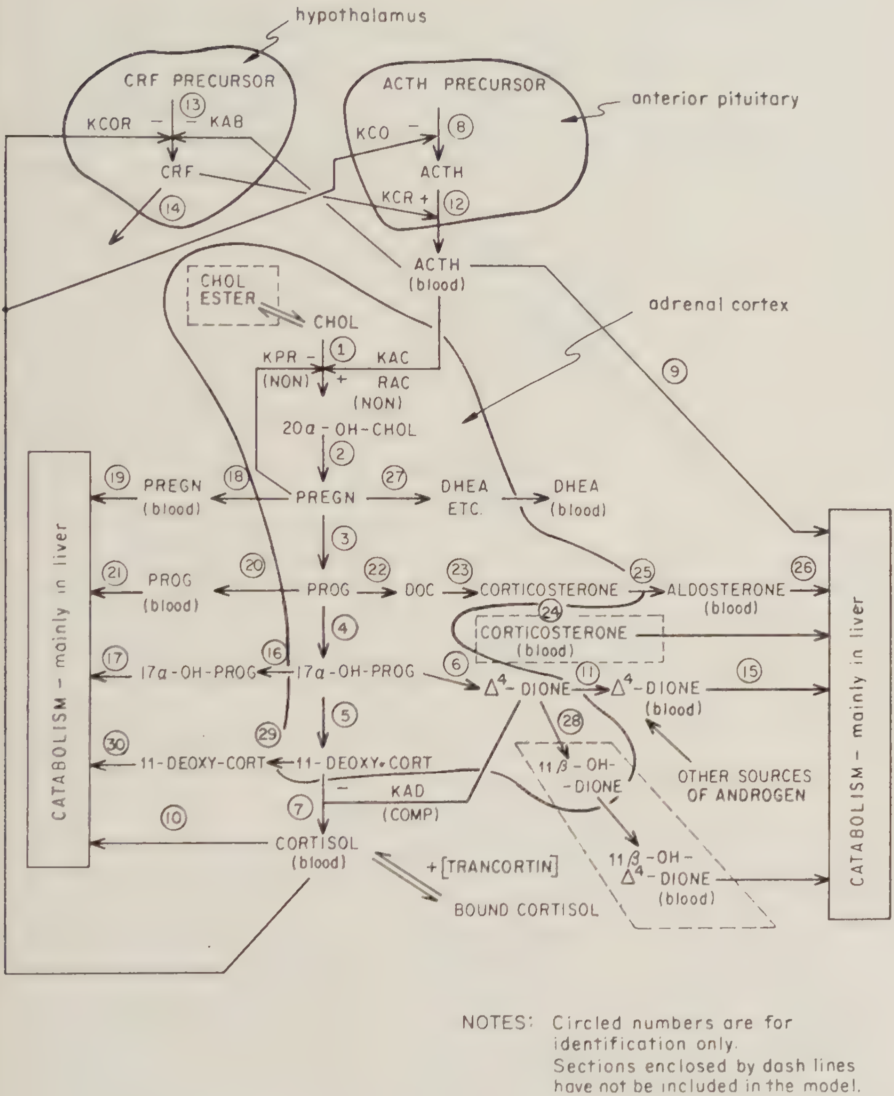


FIGURE 1 Flow sheet for adrenocortical system

Numerical Formulation

The computation of the time response of the model is based on a time interval short enough so that none of the production rates change by a significant amount during that time. The general order of computation, shown in Figure 2, is first to compute the production rate based on the values of the state variables (substrate concentrations) at the beginning of the time inter-

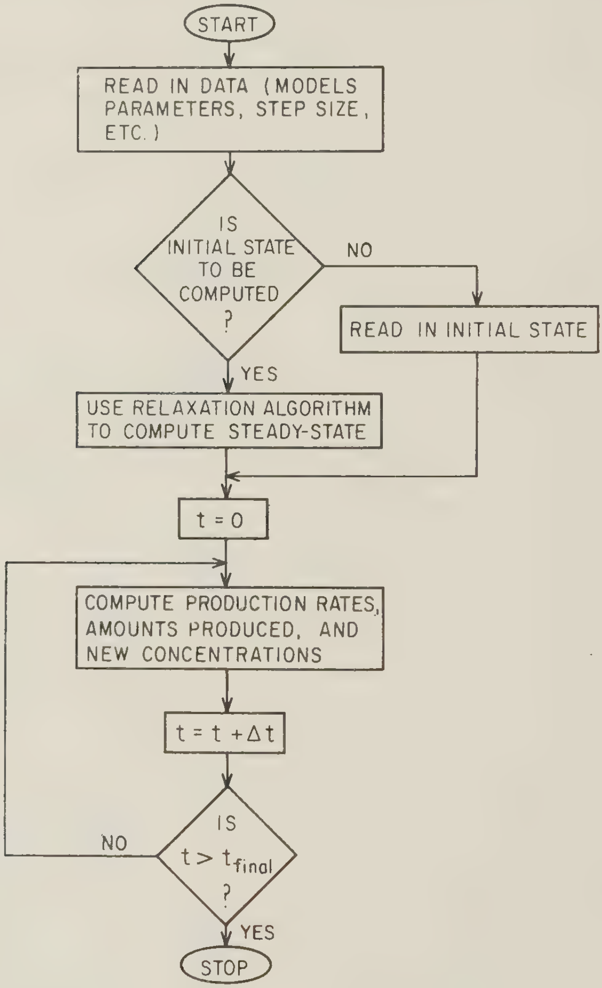


FIGURE 2 Flow chart of overall program operation

val, next to compute the amount of chemical produced in the time interval, and then to update the concentration. The process is then repeated for all of the chemical substances. This method is essentially Euler's method for integrating ordinary differential equations 3 and bases the integration over each time step entirely on values of the substrate concentrations at the beginning of the step. Other methods offer advantages in maximum allowable step size for a given accuracy, but Euler's method is by far the easiest to program and to use. If computing costs do not become excessive, there seems to be no advantage in using any of the more complex methods. A further point to note is that the maximum step size that can be used is limited by the speed of the fastest reaction in the system, regardless of the speed of any of the others or other accuracy considerations.

Each reaction is governed by the biosynthetic rate equation, reproduced below for the case of a substrate A forming a product B:

$$\left(\frac{\mathrm{d}B}{\mathrm{d}t}\right)_{\mathrm{prod}} = V_{\mathrm{max}}a/(a + K_m)$$

where lower case letters refer to concentrations, capital letters refer to actual quantities present, and no inhibitors or stimulants are present. The subscript "prod" indicates that this is the change in B due only to production. For the purposes of numerical computation, the production rate is assumed to be constant over a small time increment, Δt, and the quantity of B produced during the time period is

$$\Delta B_{\mathrm{prod}} = (V_{\mathrm{max}}a/(a + K_m))\,\Delta t$$

The amount of B lost is described by a similar relation (for example, B reacting to produce C) and the *net* change in B is then computed. Any number of sources or destructive paths can be considered in the same manner. After the total quantity of B present has been updated to include the change over the time increment, a new value of concentration is computed by dividing the quantity by the pool volume (a separate pool volume is used for each substance to account for different fluid volumes in various compartments). The basic biosynthetic rate equation is non-linear (because of the substrate concentration in the denominator) even before consideration of inhibitors or stimulants. At "normal" operating levels, however, K_m is much greater than the substrate concentrations permitting some of the equations to be linearized. Since this has the effect of suppressing the saturating characteristics of the reactions, which often appear in clinical conditions or in case of infusions or injections or large quantities of ACTH, linearization is not recommended. For reactions whose rates are modulated by inhibitors or stimulants, the appropriate forms of the biosynthetic rate equation are

$$\left(\frac{\mathrm{d}B}{\mathrm{d}t}\right)_{\text{prod}} = \frac{V_{\max}a}{(a + K_m)(1 + j/K_j)}$$

for non-competitive inhibition by the species J, where j is the concentration of J, K_j is the inhibition constant, and a is the substrate concentration. For competitive inhibition, the rate equation is

$$\left(\frac{\mathrm{d}B}{\mathrm{d}t}\right)_{\text{prod}} = \frac{V_{\max}a}{(a + K_m(1 + k/K_k))}$$

where k is the concentration of the competitive inhibitor K and K_k is the inhibition constant. The more general forms of these equations are written for partially competitive or partially non-competitive reaction rate modulation and can account for inhibition or stimulation. For partially non-competitive modulation the rate equation is

$$\left(\frac{\mathrm{d}B}{\mathrm{d}t}\right)_{\text{prod}} = \frac{V_{\max}a(1 + Rj/K_j)}{(a + K_m)(1 + j/K_j)}$$

The parameter R controls the nature of the modulation: if R is greater than one the substance J will stimulate the reaction, if R is less than one it will inhibit the reaction, and if R is equal to one the concentration of J will have no effect on the reaction. If R is equal to zero, the equation reduces to the case of pure non-competitive inhibition, given above. A similar equation exists for partial competitive modulation. In the computational model the production rates are again approximated by averages over a small time increment; the concentrations of the modulating substances at the start of the time period are used to compute the production rate for that time period.

Hormones produced in the adrenal get into the blood by diffusion. In order to accurately represent the system dynamics it is necessary to designate state variables (quantities which fix the dynamic state of the system at any instant) for the concentrations within the adrenal and for the concentrations in the blood. For the purposes of this study, the diffusion rates were assumed to be proportional to the differences in concentration from the adrenal to the blood. Where possible, the diffusion rate coefficients were chosen on the basis of experimentally determined secretion rates. In order to account for the blood transport time, a time delay of two minutes was introduced between the secretion of cortisol into the blood and its appearance at the hypothalamus or pituitary.

The binding phenomenon has a strong effect on the dynamic response of hormone systems. In this system, the binding of cortisol to transcortin is the most significant binding reaction since cortisol is the main element in the feedback regulating loop to ACTH. For the sake of simplicity, other binding reactions were not included. The rate at which a hormone A is bound to B (or unbound) is given by

$$\left(\frac{\mathrm{d}A}{\mathrm{d}t}\right)_{\text{binding}} = k_1\,(a \cdot b) - k_2 a$$

where, again, lower case letters represent concentrations and capitals represent total quantities present. Note that in this case the reaction can go either way—the production rate can be either positive or negative—depending on the relative magnitudes of the two terms in the rate equation. Separate state variables must be defined for free A, free B, and for the bound compound. For numerical analysis, the equation is treated exactly like the biosynthetic rate equations.

The scheme of computation, then, consists of the approximate integration of a set of differential equations, using one first-order equation for each state variable (or identifiable quantity). For the adrenocortical system this amounts to about thirty differential equations, each of which requires an initial condition (concentration at $t = 0$) to start the computation. A stable equilibrium state is the most logical starting point for dynamic studies; this is a state for which all concentrations are constant with time, and thus for each substance, the net production rate must equal the net destruction rate. By equating production rates and destruction rates, a set of non-linear algebraic equations is generated whose solution gives the desired steady-state condition. An iterative, relaxation type solution is used to compute the steady-state condition and thus the initial state for the dynamic solution. The iterative solution converges very rapidly for most cases, but with some combinations of parameter values it does not converge at all. When this occurs, an alternative method for generating the starting conditions is to give the system zero initial states, on all the state variables and then simply to permit the dynamic solution to run for a long enough period of time so that transients die out.

Parameter Determination

This model, with approximately thirty different equations, contains about 75 independent parameters which must be fixed in order to compute a time response. One property of a numerical simulation model is that no "unknown" or "undetermined" parameters can exist: since the computation is numerical, values must be found, in one way or another, for every parameter before any results can be produced. The major parameters are the V_{max}'s and K_m's for each reaction, the inhibition and stimulation constants, and the diffusion rate coefficients. The values and units for all of the parameters used in the stimulation are shown in Tables 1 and 2. As a start, whenever experimentally determined values were indicated in the literature, these were used in the model. The parameters for which no reliable values could be found were estimated to obtain a complete preliminary working model. If all or nearly all of the model parameters could be fixed by employing experimental or theoretical values, obtained from the literature, the model for the normal system could be tested at this point by comparing its predicted response to a series of disturbances with the observed response of the actual system to the same disturbances. In this case, however, not nearly enough data were available, so remaining parameters were determined through the parameter identification process below.

The problematical nature of parameter determination raises an important question in the philosophy of modeling: to what degree can a model with so many arbitrarily fixed parameters be significant? In an absolute sense a model is no better than our knowledge of each of the parts that make up the model. One can never have complete confidence in a model anymore than one can ever obtain exact solutions to any problems involving physical systems; judgement must therefore be used in determining the ranges of validity a model is likely to have and the amount of confidence that should be placed on the predictions it makes. In order to insure a level of significance beyond that of pure curve fitting, the structure of the adrenocortical system model (i.e. the chemical flow diagram for the system) and the forms of the rate equations were entirely fixed from biochemically supported hypotheses. These put very strong contraints on the nature of responses that could be obtained from the model, regardless of the choice of parameter values. Further quantitative discussion of the effects of parameter variation on the responses is given in the section on parameter sensitivity.

Experimentally obtained information pertaining to the adrenocortical system performance was used as the primary input for parameter identification. This information was primarily steady-state in nature, concentrations, secretion rates, half-lives, etc., so the major goal of the identification was to match the steady-state equilibrium state of the model to all known measurements. Excellent agreement was obtained; the only significant exception, the aldosterone, occurred because a known binding was not included in the model. Unfortunately, there is not enough data available to fix all of the model parameters. One particular problem is that in the normal state the reactions are operating well below their saturation levels. Thus only the ratio V_{max}/K_m is important in determining production rates, because the value of K_m is much larger than the substrate concentration value.

In all of the studies that follow, the model was used in exactly the form determined by the parameter identification studies; no changes were made except those indicated by hypothesis being examined. The parameter values for the normal model have been summarized in Table 1 and 2.

TABLE 1 V_{max} and K_m Values used in the model

Reaction No.	V_{max} M/min	K_m M/l	Reaction No.	V_{max} M/min	K_m M/l
1	4.3×10^{-8}	0.57×10^{-5}	17	0.2×10^{-3}	0.74×10^{-4}
2	0.6×10^{-5}	0.54×10^{-5}	18	0.15	–
3	1.1×10^{-4}	1.7×10^{-4}	19	0.56×10^{-4}	0.74×10^{-4}
4	0.65×10^{-5}	0.74×10^{-4}	20	0.25×10^{-2}	–
5	2.94×10^{-6}	3.03×10^{-5}	21	0.124×10^{-4}	0.74×10^{-4}
6	0.15×10^{-5}	0.6×10^{-4}	22	0.4×10^{-6}	0.6×10^{-4}
7	2.5×10^{-7}	2.08×10^{-5}	23	0.3×10^{-6}	0.85×10^{-5}
8	0.17×10^{-4}	0.74×10^{-4}	24	0.3	–
9	0.6×10^{-4}	0.74×10^{-4}	25	0.2×10^{-8}	0.8×10^{-7}
10	0.2×10^{-3}	0.74×10^{-4}	26	0.153×10^{-4}	0.74×10^{-4}
11	0.020	–	27	0.9×10^{-6}	0.74×10^{-4}
12	0.1×10^{-4}	–	28	0.73×10^{-6}	0.19×10^{-4}
13	0.8×10^{-4}	0.74×10^{-4}	29	0.12E-3	–
14	0.5×10^{-2}	–	30	0.436E-5	0.74×10^{-4}
15	0.12×10^{-3}	0.74×10^{-4}			
16	0.25×10^{-1}	0.3×10^{-4}			

A great deal of work remains to be done, both experimentally and analytically, in order to determine all the parameters for this model. The general agreement between simulated and measured steady-state conditions and the results of the dynamic and clinical studies which follow indicate that the model has the potential of becoming a highly accurate representation of the adrenocortical system as more basic parameters become known.

TABLE 2 Inhibition/Stimulation modulation coefficients (normal case)

Reaction affected	Symbol (see Figure 1)	Value
1	KAC	3.8 m un/l
1	RAC	10.0
1	KPR	0.15×10^{-6} M/l
7	KAD	0.5×10^{-4} M/l
13	KCOR	0.5×10^{-9} M/l
13	KAB	0.947 m un/l
8	KCO	0.25×10^{-7} M/l

Programming and Computer Considerations

The simulation program is written in Fortran IV and runs on a CDC 6400 computer*. Output is obtained in graphical and tabular form. The cost of obtaining information from a simulation program is related most directly to its computer running time characteristics; this program requires approximately twenty seconds of computer time to simulate one hour of system time. The amount of computer time required for the simulation of other biochemical systems similar in nature to the adrenocortical system will scale linearly with the number of state variables.

SIMULATION OF EXPERIMENTAL AND CLINICAL STUDIES

The utility of a model can only be assessed with reference to a larger context. The adrenocortical system is described by certain observable physical and chemical quantities and their interrelations. Its analog consists of certain mathematical functions and the variables contained therein. If a model were developed to reproduce the behavior of the adrenal cortex based only on two variables—the input and the output—within specified limits, its degree of isomorphism with the prototype would be low. For certain purposes,

* University of California, Berkeley, Computing Center.

such a model[4] might be quite adequate, particularly if one wished to study other parts of a larger system in which the prototype is only a part. But, because the enzymatic reactions constituting the sequence for the biosynthesis of the corticoids are the determining factors in most cases of clinical disorders of the adrenocortical system, a high degree of isomorphism is desired. These disorders are not rare and their study is of immediate interest.

Consequently, available experimental information on the biosynthetic pathways was given a major consideration in constructing the model presented above. Simulation of certain clinical disorders, based on the hypotheses presented by various investigators, has been attempted in the following section to develop an interaction between the mathematical construct and experimental studies.

Effect of ACTH and Glucocorticoid Administration

The mechanism of ACTH action(s) is debatable, but its effects on the adrenal cortex have been studied in detail. These include stimulation of the biosyntheses of corticosteroids, accelerated phosphate turnover and a decrease in the cholesteral and ascorbic acid concentrations. Cortisol is the major glucocorticoid secreted by the normal human adrenal cortex. Its secretion under basal conditions as well as in response to stress is dependent upon ACTH.

Infusions of ACTH as low as 0.05 units per hour, i.v. increase cortisol secretion[5]. The maximum response has been obtained by infusing 0.4 units per hour[6]. However, most investigations assure maximum response by administering 25–30 I.U. infused i.v. over 4 to 8 hours or 40 I.U. given intramuscular per day. Eik-Nes *et al.*[7] observed that i.v. infusion of 25 I.U. in 6 hours gave a maximal rate of increase in plasma. 17α-hydroxycorticosteroids, although the effect did not seem to be much different when the period of infusion was 4 hours.

The effect of maximal stimulation with ACTH in normal subjects and in patients with various diseases has been studied[8]. The mean response to ACTH infusion in normals was an increase in plasma concentration of 17α-hydroxycorticosteroids from 7 μg per 100 ml to a range of 32–51 μg per 100 ml; a similar increase has been reported by Nugent and Mayes[9]. In response to ACTH there occurs an initial brisk rise in plasma levels to more than twice the resting concentration at 30 min[10], but cortisol concentration can show a slow rise even as late as 8 hours[11]. A maximum production rate of cortisol, 150 mg/day was found to increase the body pool from 2 mg to 10 mg[12].

Stimulation due to clinical shock in patients undergoing surgery raised the corticoid levels in plasma to 45–70 μg per 100 ml[13,14]. A range of 33–43 μg per 100 ml in Cushing's syndrome[15] and 55–71 μg per 100 ml in adrenocorticol carcinoma[16] have also been reported.

The effect of infusion of dexamethasone and cortisol has been studied by several investigators, including Nugent *et al.*[8] and James *et al.*[17]. In the latter study infusion of 1 mg of dexamethasone per hour led to a prompt exponential fall in the plasma cortisol concentration; after 3 hours the concentration was 23 to 27% of the initial value.

ACTH stimulates the secretion of aldosterone also, but the amount required to increase aldosterone secretion is greater than that required for maximal cortisol secretion. ACTH infusion (25 I.U. in 8 hours) in 8 to 18-year old normals increased urinary excretion of Porter-Silber chromogens from 4–12 mg to 20–60 mg per day; however, the increase in aldosterone excretion ranged from 5–20 μg to 13–62 μg per day[18]. Plasma concentration of aldosterone in young males is usually about 4.7 mμg per 100 ml[19]. In Cushing's syndrome due to hyperplasia the concentration rose to about 12 mμg per 100 ml[20].

The rate of aldosterone secretion is not directly affected by the concentration of aldosterone in peripheral blood as in the case of cortisol[21]. It is indirectly controlled by the blood volume[22]. Also, a potent stimulus for aldosterone secretion is the restriction of sodium chloride intake. Secretion of aldosterone in young normals is about 80 μg per day[19]; in Cushing's syndrome due to hyperplasia the secrctory rate rose to 100–200 μg per day[20].

Infusing ACTH or increasing its level by blocking cortisol production also has an effect on the secretion of other corticosteroids. Corticosterone concentrations in plasma maintain a constant relationship to those of cortisol throughout 24 hours with a maximum level of about 1.5 μg per 100 ml[23]. In normal individuals Brorson[24] reported concentrations of 10.2 μg of unconjugated cortisol and of 1.2 μg per 100 ml of corticosterone. Infusion of 25 I.U. of ACTH in 8 hours raised the corticosterone levels to a range of 7.5–10.0 μg per 100 ml[25]. The blood production rate of corticosterone in normals has been found to be in the range of 0.9–4.4 mg per day[12,26]; in the Cushing's syndrome the production rate was 9.0 mg per day.

The normal secretion rate of 11-deoxycorticosterone (DOC) is 0.4 to 0.6 mg per day[27]. Under the influence of metopyrone, when the adrenal is presumably stimulated maximally the secretion rate of DOC rose 125 fold[28]. Concentration of 11-deoxycortisol in plasma is normally 0–2.3 μg per 100 ml[29,30]. On infusion of metopyrone, the concentration rose to 12 μg per 100 ml, and the secretion rate rose to 230 mg per day[31].

Simulation of ACTH Administration and Suppression of the Adrenal Cortex by Cortisol Infusion

The simulated effects of the infusion and injection of ACTH and of the infusion of cortisol on aldosterone, ACTH, and cortisol blood concentrations are shown in Figure 3 to 6. The major qualitative characteristics of these simulated responses agree with the experimental observations. For large-scale infusion of ACTH, the rate of infusion appears to have little effect; the final values and rates of rise for the cortisol and aldosterone concentrations are essentially identical for infusion of 25 I.U. of ACTH over two hours and over six hours, Figures 3 and 4. This is a graphic demonstration of the system's basic non-linearity, since a linear system would be expected to respond in proportion to the strength of the stimulus. Comparison with the experimental date above shows that the effect of ACTH at maximum stimulation is weaker for the model than has been observed.

Simulated injection of ACTH is shown in Figure 5. The response again exhibits the non-linear saturation phenomenon, since the response to a 10 I.U. injection is only slightly less than the response to a 25 I.U. injection. ACTH injection, as simulated, has only a limited effect: its half-life is so much shorter than that of cortisol or aldosterone that the added ACTH in the blood decays to a very low concentration before it has a chance to have much effect (note that the total time shown in the figure is only a little more than an hour and a half). Simulation of cortisol infusion, Figure 6, displays the characteristic fall in ACTH and aldosterone concentrations as the cortisol is administered.

Certain Disorders of the Adrenal Cortex and their Simulation Studies

In addition to cortisol and aldosterone, the adrenal cortex secretes several other steroids, including "adrenal androgens", corticosterone and progesterone. Secretion of other steroids like 17α-hydroxyprogesterone, 11-deoxycorticosterone and 11-deoxycortisol can be excessive under abnormal cir-

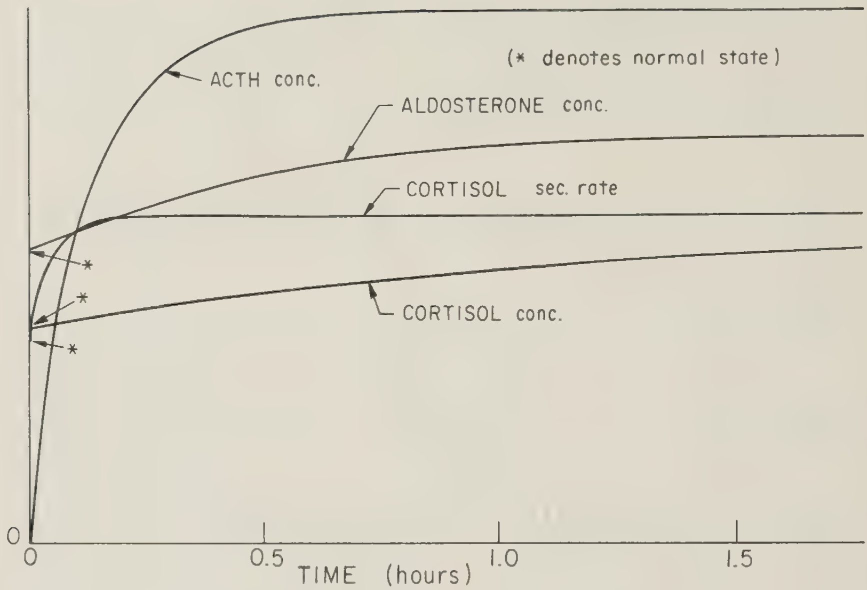


FIGURE 3 ACTH infusion–25 I.U. in 2 hours

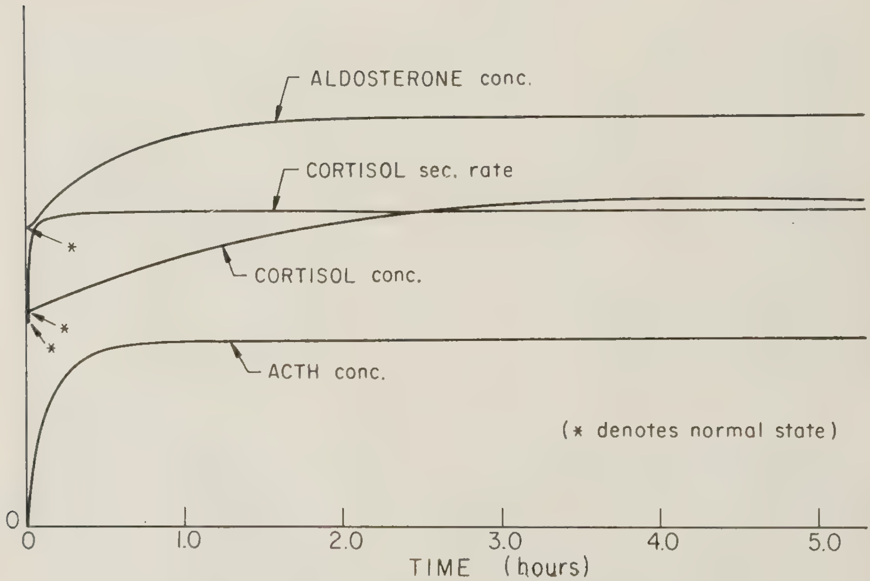


FIGURE 4 ACTH infusion–25 I.U. infused in 6 hours

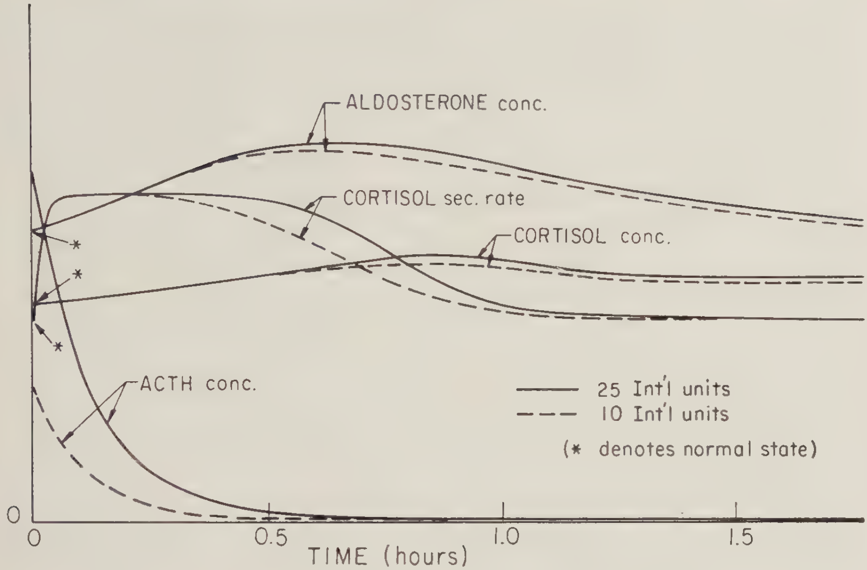


FIGURE 5 ACTH injection

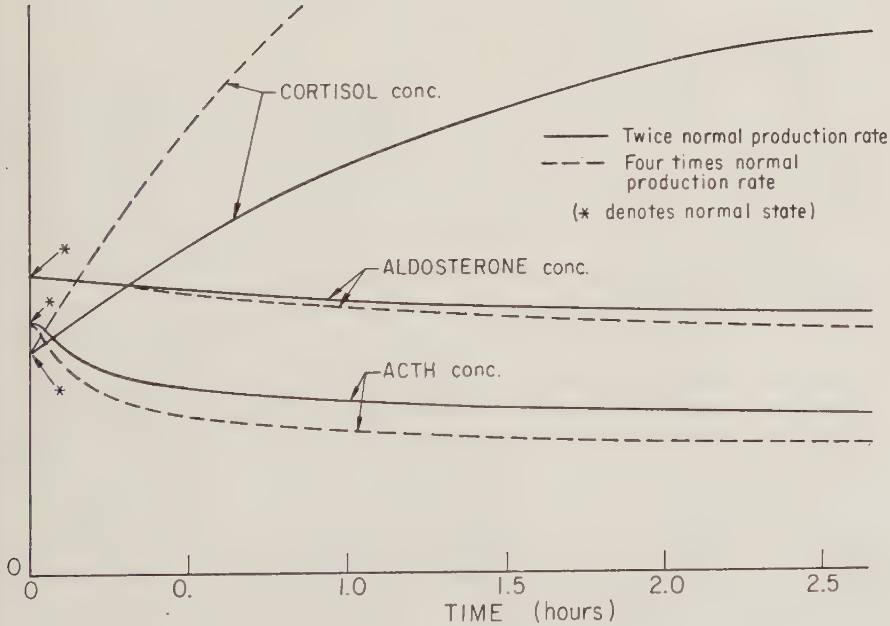


FIGURE 6 Cortisol infusion

cumstances. Secretion of estrogens by the adrenals is not significant except in patients with feminizing tumours of the adrenal cortex[32].

A large number of disorders of the adrenocortical system are a consequence of abnormal secretion of steroids by the adrenals. The primary cause of the disorder may be located either within the adrenal itself or in some other component of the system, e.g. in the hypothalamus or the anterior pituitary. In cases of primary or secondary hyper-activity of the adrenal, the syndrome produced depends upon the rates of secretion of glucocorticoids, mineralocortoids and the sex hormones. The abnormal secretion of these hormones may be due to a disturbance in the normal sequence of corticosteroid biosynthesis in non-tumorous adrenals. Most frequently this impairment is due to a relative deficiency of one of the enzymes necessary for corticosteroid biosynthesis. This congenital defect is due to a recessive autosomal gene for which clinical manifestations appear in the homozygous state. The genetic defect breeds truly in families, and no difference in the sex ratio has been noticed[33,34,35]. Several forms of this disorder known as "congenital adrenal hyperplasia" are now recognized. The clinical manifestations in these variants depend upon the site of the metabolic block in the biosynthetic pathway, the severity of the defect, the sex of the subject, and the time of onset of the disorder.

Most frequently there is a partial impairment in the biosynthesis of cortisol. Hypersecretion of ACTH in an attempt to maintain homeostasis with respect to cortisol results in increased adrenocortisol activity and consequent hyperplasia. The increased concentration of the intermediates immediately before the metabolic block may maintain a near normal cortisol secretion rate; these intermediates are released into the blood also, and their characteristic metabolites are excreted in the urine in excessive amounts. The clinical manifestations may arise from a relative deficiency of cortisol or aldosterone and from an excessive secretion of other steroids. The age at which the disease is manifest appears to be determined by the extent and site of the betabolic defect. The following variants of congenital adrenal hyperplasia have been observed.

(1) Lipoid adrenal hyperplasia due to a defect in the biosynthesis of pregnenolone from cholestorol.

(2) Defect in the oxidation of pregnenolone to progesterone due to a deficiency of 3 β-hydroxysteroid dehydrogenase.

(3) 17 α-hydroxylase deficiency.

(4) C-21 hydroxylase deficiency. Two types of this disorder, viz. salt-losing

and non salt-losing. In salt-losing patients biosynthesis of aldosterone is also affected. In non salt-losing patients only the biosynthesis of cortisol is affected. Three different hypotheses have been advanced to explain this difference.

(5) 11 β-hydroxylase deficiency. Two kinds of this disorder, viz. hypertensive type (probably due to an excessive secretion of 11-deoxycorticosterone) and the normotensive type have been reported.

(6) Defect in aldosterone biosynthesis only.

Genetic defect in the conversion of 18-hydroxycortiosterone to aldosterone has been specifically reported[36]. A defect in C-18 hydroxylation of cortiosterone has also been suggested[37].

In an earlier section on secretion of steroids these metabolic defects and the enzymes involved have been discussed to some extent. In the following section, effects of a partial or almost total deficiency of these enzymes on the secretion of various steroids and their concentrations in the blood are studied by simulating these metabolic blocks of various enzyme reactions. In simulating a partial block the maximum velocity of the enzyme reaction has been considered to be one-tenth of normal; a maximum velocity of one-thousandth of normal has been used to simulate the conditions of a total block. All other parameters have been kept the same as in the normal case. In the case of the C-21 hydroxylation defect, two hypotheses have been simulated: that the non-salt-losing variant results from a less severe block of C-21 hydroxylation and that the hydroxylation of both progesterone and 17α-hydroxy-progesterone is blocked. Two situations have been considered: (a) when, as compared to the normal, hydroxylation of progesterone is 1/10 and that of the substrates 17α-hydroxyprogesterone is 1/1000, and (b) when the comparative blocks in the hydroxylation of these substrates have been reversed.

Simulation of Liver Disorder Affecting Cortisol Metabolism

The liver is the organ primarily responsible for the metabolism of steroid hormones[38]. Because of this role, it might be expected that this organ could indirectly influence the synthesis of corticosteroids by the adrenals. Plasma 17-hydroxycorticosteroid levels have been found to be normal in cirrhosis of the liver, but infused cortisol disappears at a decreased rate in patients with hepatitis or cirrhosis[109]. In patients with liver disease, Peterson[40] observed normal plasma levels of cortisol and corticosterone but found urinary corticoids and 17-ketosteroids to be below the normal level. Although several other infused steroids were metabolized at a normal rate, cortisol metabolism was metabolized at a diminished rate. The rate of the biosynthesis of cortisol and corticosterone was also markedly reduced in patients with cirrhosis. It was suggested that a homeostatic mechanism through the liver–pituitary–adrenals leads to a decreased synthesis of cortisol and corticosterone in patients with liver disease in whom the rate of cortisol catabolism is impaired.

While the secretion of cortisol, corticosterone and adrenal androgens is frequently suppressed in liver disease, aldosterone production appears to remain normal[41]. The half-life of cortisol has been found to range from 190–330 min (normal about 90 min) and that of corticosterone from about 84–102 min (normal about 20 min) in patients with liver disease. The effects of increased half-life of cortisol has been simulated and the results are shown in Table 3.

TABLE 3 Simulation of change in Cortisol half-life

	Normal ($t_{1/2}$ = 70 min)		Increased half-life ($t_{1/2}$ = 140 min)	
	Blood conc. per litre	Secretion rate per day	Blood conc. per litre	Secretion rate per day
ACTH	0.3 m un	3.5 IU	0.221 m un	2.59 I.U.
Cortisol	0.562×10^{-6} M	22.8 mg	0.990×10^{-6} M	20.7 mg
Aldosterone	0.187×10^{-8} M	0.166 mg	0.171×10^{-8} M	0.153 mg
Δ^4-dione	0.355×10^{-8} M	2.05 mg	0.328×10^{-8} M	1.86 mg
11β-OH-Δ^4-dione	–	3.98 mg	–	3.53 mg
Corticosterone (in adrenal)	0.191×10^{-7} M	–	0.173×10^{-7}	–

Parameter Sensitivity Studies

One important task a computer model of a system can accomplish is to determine the relative influence of various system parameters on particular system properties. Parameter sensitivity, as this is called, affects model construction, parameter identification procedures, and very significantly, design of experiments to be carried out on the actual system. In general, the purpose of an experiment is to infer the value of one or more parameters from the measurement of other quantities, usually state variables. Since it is usually impossible to measure all of the state variables, a parameter sensitivity study carried out on a model can indicate to the experimenter which measurements have the strongest bearing on the parameter he is looking for. Such a study can also demonstrate the need for further experimentation, for example, when a parameter which has a strong influence on the system output is not well known.

The changes in steady-state concentrations of some of the important hormones due to changes in various parameters are shown in Table 4. It is

TABLE 4 Parameter sensitivity studies summary of results—steady-state case

Parameter	Change from normal	Changes in steady-state concentration % (Percent)			
	%	Cortisol	ACTH	Aldosterone	Δ^4-dione
KAC	−50	+18	−7.4	+14	+15
KAC	+100	−15	+8.4	−16	−16
KPR	−50	−18	+9.5	−14	−16
KPR	+100	+21	−8.5	+15	+16
V_4	−50	−6.0	+3.2	+59	−6.1
V_4	+100	+2.7	−1.5	−43	+3.7
V_5	+100	+18	−6.3	−2.7	−34
V_7	50	0	0	0	0
	+100	0	0	0	0

interesting that the table reveals one aspect of the basic regulatory mechanism: the rises or falls in cortisol concentration caused by the parameter changes are partially offset by opposite changes in the ACTH concentration. The last two rows of the table imply that the value of V_7 has no effect at all on the steady-state solution. This is because the adrenal concentration of 11-deoxy-cortisol (which diffuses into the blood very slowly) rises (or falls) just enough to maintain the cortisol concentration level. Findings of this sort are good examples of the interaction model building and experimentation—in this case better fixes could be obtained on many of the system parameters if more data were available on hormone concentrations in the adrenal.

Two dynamic studies were also made as part of the parameter sensitivity study. In Figure 7, the effect of a change in KPR is shown. As expected, there are changes in both the initial state and in the dynamics. In the second

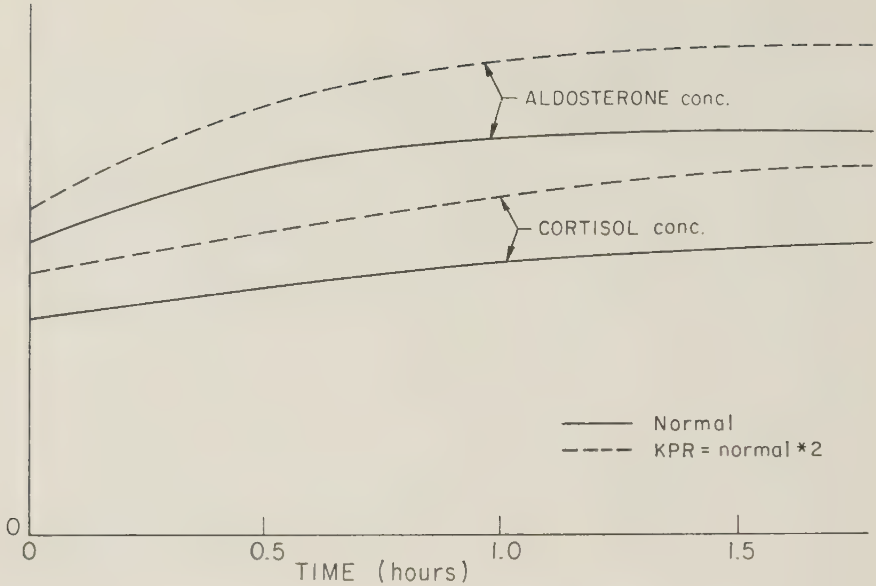


FIGURE 7 Parameter sensitivity studies; ACTH infusion; 25 I.U. per 2 hours

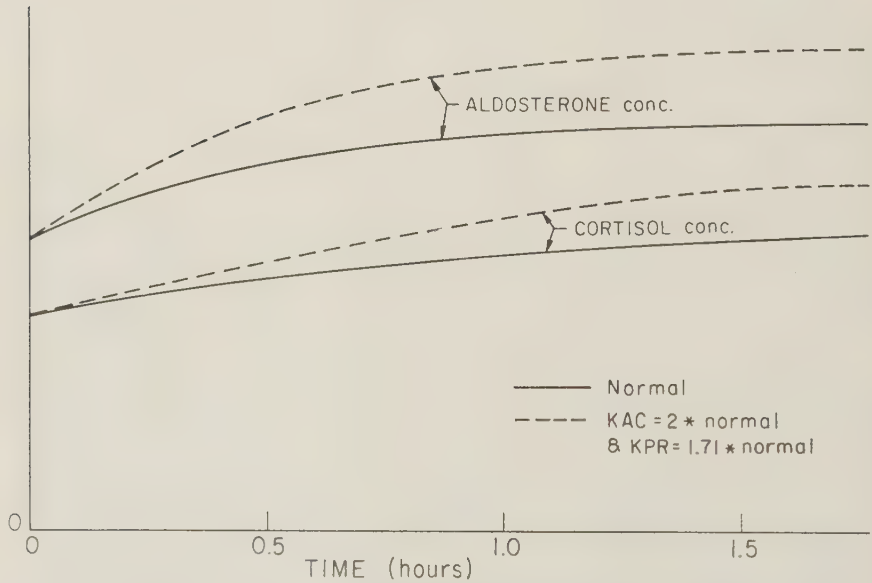


FIGURE 8 Parameter sensitivity studies; I.U. ACTH infusion; 25 I.U. over 2 hours

case, Figure 8, KPR and KAC were changed simultaneously in such a way that the initial state did not change at all. The resulting purely dynamic change in the system response could not have been detected from steady-state studies. As noted earlier, changes of a purely dynamic nature have not yet been included in the parameter identification.

Acknowledgments

This investigation was supported in part by Biomedical Sciences Support Grant FR-7006 from the General Research Support Branch, Division of Research Resources, Bureau of Health, Professions, Education and Manpower Training, National Institute of Health, and by the Special Fellowship Grant in Bioengineering, 1-F03-AM41138-01 from the Institute of Arthritis and Metabolic Diseases, Endocrinology Program, of U.S. Public Health Service.

The paper is an abridged version of the original submitted.

References

1. D.C. Sharma and R.I. Dorfman, *A Generalized Outline of the Metabolism of Steroid Hormones, Biosynthesis and Catabolism*, Holden-Day, San Francisco (1966).
2. D.C. Sharma and J.L. Gabrilove, *Comp. Biochem. Phys.* (in press, 1969).
3. R.W. Southworth and S.L. Deleeuw, *Digital Computation and Numerical Methods*, McGraw-Hill, New York (1965).
4. F.E.Yates, R.D.Brennan and J.Urguhart, *Federation Proc.*, **28**, 71 (1969).
5. G.W. Liddle, D.P. Island, and C.K. Meador, *Recent Progr. Hormone Res.*, **18**, 125 (1962).
6. J.Landon, V.H.T.James, R.J.Cryer, V.Wynn, and A.W.Frankland, *J. Clin. Endocrinol. Metab.*, **24**, 1206 (1964).
7. K.Eik-Nes, A.A.Sandberg, D.H.Nelson, F.H.Tyler, and L.T.Samuels, *J. Clin. Invest.*, **33**, 1502 (1954).
8. C.A.Nugent, W.D.MacDiarmid, A.R.Nelson, and F.H.Tyler, *J. Clin. Endocrinol. Metab.*, **23**, 684 (1963).
9. C.A.Nugent and D.M.Mayes, *J. Clin. Endocrinol. Metab.*, **26**, 1116 (1966).
10. J.Landon, V.Wynn, V.H.T.James, and J.B.Wood, *J. Clin. Endocrinol. Metab.*, **25**, 602 (1965).
11. E.A.Cameron and J.R.Kilborn, *Clin. Chim. Acta*, **10**, 308 (1964).
12. R.E.Peterson, *Recent Progr. Hormone Res.*, **15**, 231 (1959).
13. Y.Einerth, P.Hedner, and O.Wilander, *Acta Chir. Scand.*, **130**, 411 (1965).
14. D.Mattingsly, and C.Tyler, *Proc. Roy. Soc. Med.*, **58**, 1010 (1965).
15. W.Futterweit, D.T.Kreiger, and J.L.Gabrilove, *J. Clin. Endocrinol. Metab.*, **22**, 364 (1962).
16. F.C.Pavlatos, R.P.Smilo, and P.H.Forsham, *J. Am. Med. Assoc.*, **193**, 720 (1965).
17. V.H.T.James, J.Landon, and V.Wynn, *J. Endocrinol.*, **33**, 515 (1965).

18. P.E.Salton, Jr. and E.G.Biglieri, *J. Clin. Endocrinol. Metab.*, **27**, 37 (1967).
19. C.Flood, C.Gherondache, G.Pincus, J.F.Tait, S.A.S.Tait, and S.Willoughby, *J. Clin. Invest.*, **46**, 900 (1967).
20. T.Kono, T.Yoshimi, and T.Miyake, in *Steroid Dynamics, Proc. Symp. on the Dynamics of Steroid Hormones* (Eds. G.Pincus, T.Nakao, and J.F.Tait), *Tokyo, 1965*, Academic Press, New York (1965), p.429.
21. J.R.Blair-West, J.P.Coghlan, D.A.Denton, J.R.Goding, M.Wintour, and R.D. Wrigh, *Endocrinology*, **77**, 501 (1965).
22. F.C.Bartter, G.W.Liddle, L.E.Duncan, Jr., J.K.Barber, and C.Delea, *J. Clin. Invest.*, **35**, 1306 (1956).
23. R.E.Peterson, *J. Clin. Endocrinol. Metab.*, **17**, 1150 (1957).
24. I.Brorson, *Acta Chir. Scand.*, **127**, 162 (1964).
25. R.E.Peterson, *J. Biol. Chem.*, **225**, 25 (1957).
26. E.G.Biglieri, S.Hane, P.E.Salton Jr., and P.H.Forsham, *J. Clin. Invest.*, **42**, 516 (1963).
27. M.G.Crane and J.J.Harris, *J. Clin. Endocrinol. Metab.*, **27**, 1135 (1966).
28. J.J.Harris, C.Hoegel, and M.G.Crane, *J. Clin. Endocrinol. Metab.*, **27**, 106–13 (1967).
29. J.Ditzel, P.F.Hausen, and N.Riskoer, *Acta Endocrinol.*, **45**, 171 (1964).
30. S.H.Waxman, D.F.Tippit, and V.C.Kelley, *J. Clin. Endocrinol. Metab.*, **21**, 943 (1961).
31. J.E.Plager, K.G.Schmidt, and W.J.Staubitz, *J. Clin. Invest.*, **43**, 1066 (1964).
32. J.L.Gabrilove, D.C.Sharma, H.H.Wotiz, and R.I.Dorfman, *Medicine*, **44**, 37 (1965).
33. B.Childs, M.M.Grumbach, and J.J.VanWyk, *J. Clin. Invest.*, **35**, 213 (1956).
34. A.Prader and G.J.P.A.Anders, *Helv. Paediat. Acta*, **17**, 285 (1962).
35. H.K.A.Visser and W.S.Cost, *Acta Endocrinol.*, **47**, 589 (1964).
36. S.Ulick, E.Gautier, K.K.Vetter, J.R.Markello, S.Yaffe, and C.U.Lowe, *J. Clin. Endocrinol. Metab.*, **24**, 669 (1964).
37. H.K.A.Visser, in *Basic Concepts of Inborn Errors and Defects of Steroid Biosynthesis* (Eds. K.S.Holt and D.N.Raine), Livingston, London (1966), p.34.
38. L.T.Samuels and C.D.West, *Vitamins Hormones*, **10**, 251 (1952).
39. E.Englert Jr., H.Brocon, S.Wallach, and E.L.Simmons, *J. Clin. Endocrinol. Metab.*, **17**, 1395 (1957).
40. R.E.Peterson, *J. Clin. Invest.*, **39**, 320 (1960).
41. I.Dyrenfurth, C.H.Stacey, J.C.Beck, and E.H.Venning, *Metab. Clin. Exp.*, **6**, 544 (1957).

CHAPTER II–19

Disorders of effector mechanism at the level of humoral biocybernetical systems in cerebral infarction

C. BĂLĂCEANU
Bucarest University, Bucarest, Rumania
and E. STOICA
Institute of Neurology, Academy of Sciences, Rumania

Summary

The authors describe the cybernetical model of the homeostatic systems for blood coagulability and blood sugar level, emphasizing that the effector mechanism is common for both systems. The behaviour of these systems in a control group, and in patients with cerebral infarction, haemorrhage, or cerebral circulatory insufficiency was studied. Blood sugar and blood coagulability variations induced by administration of 1 g of tolbutamide was determined. These humoral parameters after a typical evolution show a tendency to reach their initial value, except in patients with cerebral infarction in which stationary error appear (long lasting hypoglycaemia and hypercoagulability). The alteration observed in the latter group may be conditioned by the impairment of the effector mechanism. The hypothesis was demonstrated by determining the elimination of VMA (3, methoxy—4, hydroxy mandelic acid), the main metabolite of adrenergic messenger (the output message). In patients with cerebral haemorrhage and cerebral circulatory insufficiency as well as in control group there was an increase of VMA elimination during the transitory state induced by tolbutamide administration, whereas in patients with cerebral infarction a reduction or no alteration in the elimination of this metabolite were observed. The authors pointed out that the effector mechanism disorders of the blood sugar and blood clotting systems may be involved in the onset of the cerebral infarction.

INTRODUCTION

The homeostatic function of the central nervous system consists in keeping constant, or relatively constant, certain physiological parameters, such as the concentration of certain substances in the fluids of the organism, the temperature, the blood pressure, etc. The number of parameters which have to be regulated is considerable, while the nervous system disposes of a relatively small number of execution systems, or servo-systems, by means of which it can act upon the vast field of physiological parameters. Therefore, it seems possible that a certain servo-system may act simultaneously upon several variables or parameters. In other words, a servo-system may be common to several biocybernetical regulation systems; it is what we have called the principle of convergence. If we accept this point of view it becomes clear that effector mechanisms represent areas of critical strategy for the organism. On their functional capacity depend all the convergent bioregulation systems. It ought, therefore, follow that from affected effector mechanisms may result serious homeostatic disorders which can be at the origin of certain disease states.

The classical physiology data show that the sympathetic and parasympathetic portion of the nervous vegetative system exert an antagonistic action upon a series of humoral parameters (coagulability of the blood, glycaemia, lipaemia). The stimulation of the sympathicus brings about a hypercoagulability, an increase in glycaemia and in lipaemia; the stimulation of the parasympathicus has reverse effects. Transposing these data in the cybernetical way of speaking, we can say that the vegetative nervous system represents, in alarm situations, the effector mechanism common to the bioregulating systems involved in maintaining coagulability, glycaemia, lipaemia, etc., constant. A disturbance of the nervous system may thus be reflected on all the biocybernetic systems involved in the adjustment of the above mentioned humoral parameters.

THE REACTIVITY OF THE SYMPATHETIC–ADRENERGIC SYSTEM

In order to prove the strategic part played by the vegetative nervous system in the regulation of certain biological processes, it seems necessary: (1) to individualize a certain pathological state characterized by a disfunction of this system; (2) to show that such a disturbance occurs concomittently within the frame of two or several biocybernetic regulation systems (for instance, that of glycaemia and that of the blood coagulability).

For the first object of this research we have investigated the reactivity of the sympathetic–adrenergic system (which is a component of the nervous vegetative system) in several categories of patients suffering from brain vascular accidents, as well as in a group of normal subjects. We know that the sympathetic-adrenergic system exerts its action upon various biological processes by means of its messengers, adrenalin and noradrenalin.

These mediators may be compared with the output messages of a regulation system. Their determination gives indication on the way the respective system operates. We know that, as a result of their inactivation by the enzymatic apparatus, vanilmandelic acid (VMA) results as the main metabolite. The dosage of the latter's urinary elimination thus represents the traces of the sympathetic–adrenergic system output messages, offering the basis for the quantitative estimate of the way it operates. In order to test the functional state of the sympathetic–adrenergic system the basic VMA elimination has been compared with the elimination observed in the course of an induced hypoglycaemia state, which evidently brings about a mobilization of this system. In other words, the basic output messages have been compared with those observed in the conditions of a state of stress. We have used as a hypoglycaemiant substance, 1 g orally administered tolbutamid or 10 U, insulin administered subcutaneously. The dynamic determination of VMA elimination has been carried out on 21 controls, 38 patients suffering from brain haemorrhage, 21 patients suffering from brain circulatory insufficiency, and 37 patients suffering from brain infarction. The VMA urinary estimate has been carried out with the Pisano method, the results being expressed in micrograms of VMA/mg creatinine.

(a) In the control group an increase of urinary VMA elimination occurred both in tolbutamidic hypoglycaemia conditions and in the insulinic conditions; the mean increase was 0.33 ± 0.16 g/mg creatinine after administration of tolbutamide and 0.46 ± 0.15 g/mg creatinine after administration of insuline.

(b) In patients suffering from brain haemorrhage, the sympathetic–adrenergic system behaved in the same way, i.e. it was adequately mobilized after administration of hypoglycaemiant substances, as a result of the increase, of VMA elimination after tolbutamide or insulin; the mean increase was 1.01 ± 0.33 μg for the first condition and 0.87 ± 0.15 μg for the second condition.

(c) In patients with brain circulatory insufficiency the same modification in the urinary VMA elimination has been observed: an increase both after tolbutamide (mean value 0.41 ± 0.32 μg) and after insulin (mean value 0.71 ± 0.28 μg).

(d) Patients with occlusive brain infarction behaved differently from all categories of subjects above mentioned. Tolbutamidic and insulinic hypoglycaemia induced in most cases a reduction of VMA urinary elimination, not an increase as in the previous groups. The mean reduction has been 0.76 ± 0.33 μg after tolbutamide and 0.65 ± 0.22 μg after insulin. All modifications presented above have been statistically significant. We may, therefore, assert that after administration of hypoglycaemiant substances patients suffering from brain infarction present a reduction, and the other subjects belonging to other categories present an increase, both modifications being statistically significant. The dynamic study of VMA elimination, therfore of output messages, shows that the sympathetic–adrenergic system adequately corresponds in the control group and in patients with brain haemorrhage or circulatory insufficiency and it is active in those suffering of brain infarction.

If we accept the assertion that the sympathetic–adrenergic system is a component of the effector mechanism both for the biocybernetic glycoregulation system and for the system of keeping constant the blood coagulability, then we may expect that in patients with brain infarction, in which we have pointed out an inertia of this mechanism, a defective regulation of glycaemia as well as of the blood coagulability may occur. In order to verify this supposition we have studied the modifications induced in the blood coagulability by the processes which take place during the transient regime of the glycoregulation system. This regime has been induced by the use of 1 g orally administered tolbutamide as a disturbing factor.

We have determined the initial glycaemia, then 60, 120, and 180 minutes after administration of the substance (Nelson method). At the same time we have studied the blood coagulability by the heparine tolerance test in vitro (Soulier method) and the recalcified plasma coagulation test (Quick method). The research has been carried out on 17 normal persons, 10 patients suffering from brain haemorrhage, 10 patients with brain circulatory insufficiency, and 26 patients with brain infarction.

Our investigations have shown that the evolution of the transient regime at the level of the glycoregulation system is different in patients with brain infarction as compared with the control group and with patients suffering of brain haemorrhage or circulatory insufficiency. In the first group of patients the glycoregulation system is not apt to restore hypoglycaemia induced by tolbutamide, so that we witness the maintenance of low glycaemia levels even 180 minutes after administering the substance. In the other investigated groups we observe a tendency of restoration of glycaemia after administration of tolbutamide, i.e. of returning to normal values (the prescribed value of the system).

Parallel with the variations of glycaemia we observe variations of the blood coagulability. In the control group, in patients with brain haemorrhage and in those with brain circulatory insufficiency, we observe the first 120 minutes a relative hypocoagulability, but after 180 minutes, in all the 3 groups, blood coagulability is restored by its regulation system. On the other hand, in patients with brain infarction, we observe a hypocoagulability as in the previous groups, which is mainteined or even increased 180 minutes after administering tolbutamide. We, therefore find, the same incapacity of the homeostatic system to restore coagulability, i.e. of returning it to the prescribed value of the system.

CONCLUSIONS

It follows, therefore, that in the group of patients with brain infarction, both the system of maintaining coagulability constant and the glycoregulation system are defective.

This confirms the hypothesis supported by us that the two systems use for their alarm regulation a common execution mechanism, the sympathetic and parasympathetic portion of the vegetative nervous system.

The stationary error which occurred at the level of both regulation systems may be explained by the inertia of the effector mechanism, a fact proved also by the analysis of its output messages. The mediator of this system, adrenalin, intervenes in all hypoglycaemic states as a mechanism of quick recovery of the latter, due to the glycogenolysis it induces. At the time adrenalin brings about a state of hypercoagulability, as the works of Takata, Ingram, and Rowsell have demonstrated. The researches carried out by our laboratory prove that the mobilization of hypercoagulant factors by adrenalin has a lower threshold than glycogenolysis.

The lack of recovery of glycaemia, coexisting with a sustained coagulability observed in patients suffering from brain infarction, may be explained by a reduced secretion of adrenalin, incapable of initiating glycogenolysis, but apt to mobilize the hypercoagulant factors. The functional inertia of the effector system may thus determine a prolonged stress of the hypercoagulant factors, which can play its part in the start of the brain occlusive processes within this condition. The model suggested, also explains why brain occlusive accidents are more frequent during the night when the glycoregulating system is maximally stressed.

We are convinced that our model may be extended so as to include several homeostatic regulation systems concentrated on the same execution servosystem (the sympathicus–parasympathicus system, or others). The cybernetic mechanism referred to, represents a principle of organization which explains and will explain a whole series of pathological manifestations.

On the extraction of uncorrelated parameters from human electrocardiograms for use in automatic classification and diagnosis

L.F. FAVELLA, P.C. GIOLITO, and G. MARAZIO

Istituto di Fisica Teorica della Università di Torino, Torino, Italy

Summary

After a brief illustration of a general method for the extraction of significant parameters from random functions, attention is paid to the application of this method to human electrocardiograms. The second part of the paper deals with the results which have been obtained from a practical application of the method to a collection of 650 E.C.G. by means of an I.B.M. 360/44 computer. The grouping may be shown by means of suitable maps, and a possible connection between the classification with the computer and the usual pathology criteria in cardiology is given.

INTRODUCTION

The subject of this paper are some preliminary results of a research which is still in progress, and will require a good deal of work to be completed.

This research is essentially connected with the idea that electrocardiograms may be considered as a special class of random functions, and as a consequence the methods for analysis and classification should not be very different from those which could be profitably used for other kinds of random functions.

We have been, moreover, encouraged to proceed along this way by previous work on the recognition of the seven italian vowels. In that work we started with a set of recorded spoken vowels, which were expressed in terms of twelve parameters (energy contents of some frequency bands) and we wanted the ensemble of vowels to break into classes. We succeeded in building automatically out of those parameters three functions of them which were enough to classify 97% of our samples of vowels. The theory of such a method is connected with the Loewe-Karhunen expansion of random functions. Further details may be found on a paper which has been recently published in *Kybernetik*[1].

Suppose we have an ensemble of realizations $\{\xi_n\}$ of random functions in an interval of time T. In order to get from these functions a set of parameters with a marked content of information we must build a set of functions $\{\varphi_k\}$ such that the approximation, which can be obtained by representing every ξ_n by means of a sum of a preassigned number of φ's, is the best possible in the mean. From the mathematical standpoint we must determine a set of functions $\{\varphi_k\}$ such that the functional

$$\varepsilon^2 = \int_0^T M \left\{ \left| \xi(x) - \sum_1^N a_n \varphi_n(x) \right|^2 \right\} \mathrm{d}t$$

has a minimum. (The average is taken with respect to the ξ's.) Minimizing ε^2 with respect to the $\{\varphi_n\}$ together with the condition

$$\int_0^T |\varphi_n(t)|^2 \,\mathrm{d}t = 1$$

leads to the following integral equation of Fredholm type:

$$\int_0^T K(t, t')\, \varphi(t')\, \mathrm{d}t' = \lambda \varphi(t)$$

which must be solved for φ and λ. K is the covariance kernel and is defined as follows

$$K(t, t') = M\{\eta(t)\, \eta(t')\}$$

$$\eta(t) = \xi(t) - M(\xi(t))$$

In the case of the electrocardiograms we have a vector random function, which does not introduce difficulties; the previous integral equation is replaced by a set of coupled integral equations whose number is exactly that of the electrocardiographic leads which are used. The coefficients a_n are parameters which are uncorrelated with one another, and this makes them useful for classification purposes, because they provide pieces of information which are almost independent from one another.

There is another reason because we consider such parameters well suited for classifying electrocardiograms. Let us suppose that the currents inside the cardiac muscle are expressible as a sum of certain elementary currents in terms of which we may reconstruct any cardiac distribution of currents. From the mathematical point of view this hypothesis is just that of supposing an orthogonal expansion for the cardiac currents. Any current is thus expressed in terms of its expansion coefficients as a sum of these elementary currents. Any one of this elementary currents generates an electromagnetic field and the electromotive force between any two points on the body surface is given as a weighted sum of these electromagnetic fields, and because of the linearity of Maxwell equations the weights which give the total electromotive force are just the same weights as for the elementary currents. Such a structure implies that these weights, which are the useful parameters, must be expressible as linear functions of the electrocardiographic leads. This seems a good justification for considering parameters which are linearly expressible in terms of the electrocardiographic leads, and which may be easily constructed by means of a computer.

ANALYSIS OF ELECTROCARDIOGRAMS

Our first step in this field of research has been of an explorative character. We wanted to be sure that division into classes took place with this method also for electrocardiograms. So we have collected 691 electrocardiograms*, of which 25 are certainly pathological. These E.C.G.'s have been collected by means of a four-traces magnetic tape recorder and the leads which have been used are D1, AVF, V2, V6. (In a paper by Rosettani and others[2] it is proved that in most cases the four leads we used are sufficient for diagnosis. The E.C.G.'s have been converted in digital form by sampling at a rate of 300 points/second.)

The processing of these E.C.G.'s consists of three phases. The first one consists in a filtering of E.C.G.'s to eliminate alternate current noise, resetting of the base line and choosing a good period beginning with the *P* wave.

* With the collaboration of the staff of the Medical Clinic of the University of Torino, Italy.

The second phase consists in extracting the parameters for class grouping, while the third phase consists in detecting the groupings. There is also a fourth phase in which a comparison is made between the results of the computer and medical response. The first phase is conceptually very simple, although building the program has been a hard task. The average error in finding the P-wave origin is 0.02 seconds. In the second phase all the periods which have been selected are reduced to the same length with a time scale compression. This is a rather questionable operation and I think that it is responsible of some troubles. At that time this compression seemed necessary but recently we realized that it was not; on the contrary it was misleading. By means of the periods we had selected, we have built the function

$$K_\xi(t, t') = \frac{1}{N} \sum_{1}^{N}{}_K \, [\xi_K(t) - m(t)] \, [\xi_K(t') - m(t')]$$

$$m(t) = M\,[\xi(t)]$$

then we have solved the integral equations

$$\int_0^T K_\xi(t, t')\,\varphi(t')\,\mathrm{d}t' = \lambda\varphi(t); \quad \xi \equiv \begin{cases} \text{D 1} \\ \text{AVF} \\ \text{V2} \\ \text{V6} \end{cases}$$

for φ and λ and we have chosen for each lead and for each E.C.G. 32 parameters

$$a_{n,\alpha}^{(\xi)} = \int_0^T \varphi_n^{(\xi)}(t)\,\xi_\alpha(t)\,\mathrm{d}t$$

the φ'*s* which have been chosen belong to the greatest 32 eigenvalues. We have thus coded each E.C.G. in terms of a set of (32 parameters times 4 leads = 128 parameters). These parameters are furtherly processed to obtain another set of 128 uncorrelated parameters. We have stored only 64 of these parameters and then we have tried to see whether there were groupings by examining both the parameters of every single lead and those of the final set of parameters. We obtained thus some maps; some are shown in Figure 1. These maps have suggested to us the hypothesis that the E.C.G.'s in the space of these parameters are condensed around a region, which is hollow near a central point and has several spikes or protuberances; there seem also to be some empty regions.

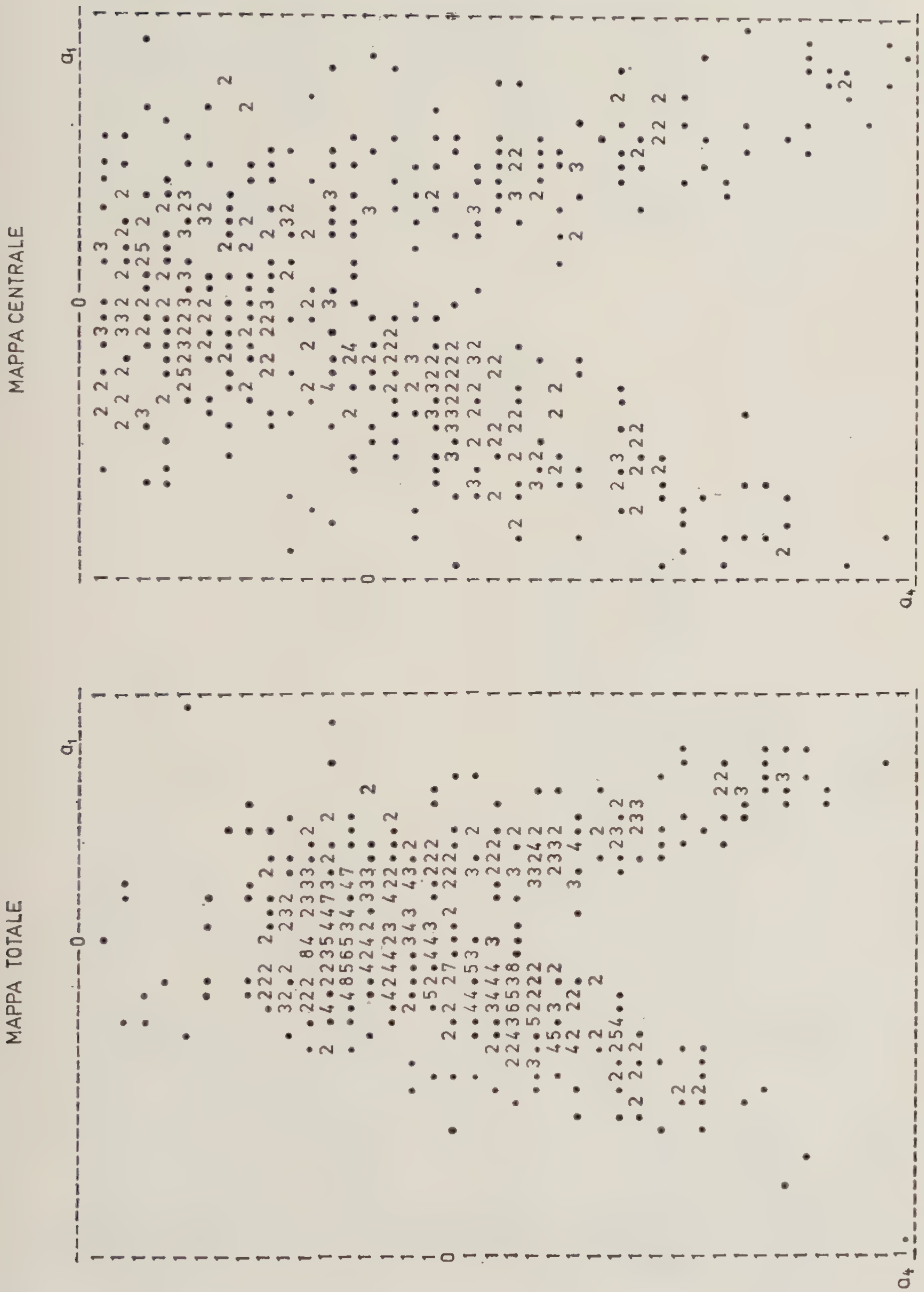


FIGURE 1 Maps of E.C.G.'s

Most spikes are all well distinct from one another and they could be defined as classes. There are only 7 variables in whose space considerable spikes or protuberances appear, but there is a large number of variables in whose space some electrocardiograms receive appreciable contributions. Figure 2 represents a three-dimensional fancy solid which should give an idea of the shape of the region in which E.C.G.'s are distributed. This particular shape has suggested us to cut the region which contains, the E.C.G.'s by means of concentric shells starting from the exterior boundary and going

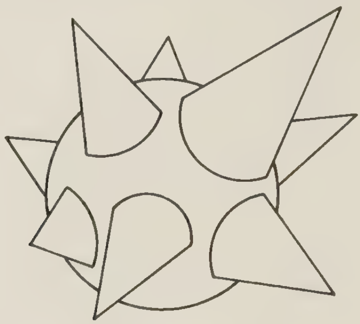


FIGURE 2

towards the interior and pursuing the aforementioned spikes. In this way we have obtained a tree one branch of which is shown in Figure 3. This tree is an attempt of class grouping using only 6 parameters, which are certainly not sufficient for obtaining a good approximation of the E.C.G.'s in the QRS region by means of the orthogonal expansion of which I have spoken because of a broadening of the QRS region for several E.C.G.'s in consequence of our mistake in compressing all the E.C.G.'s to the same length. Perhaps this is the reason because not all of the pathological E.C.G.'s have a distinctive location in this tree. However, in spite of our mistakes and of our six parameters we have seen that more than 50% of the pathological E.C.G.'s lie on the top of some tree branch or in isolated points. If one could prove that by increasing the number of parameters these fact hold true for all the pathological E.C.G.'s then one could make the following statement: "Pathological E.C.G.'s may be interpreted as final stages of gradual evolutions or as anomalous configurations which lie in forbidden regions." At present we are just trying to run the same programs with 16 parameters; moreover, we want to eliminate scale compression and repeat the

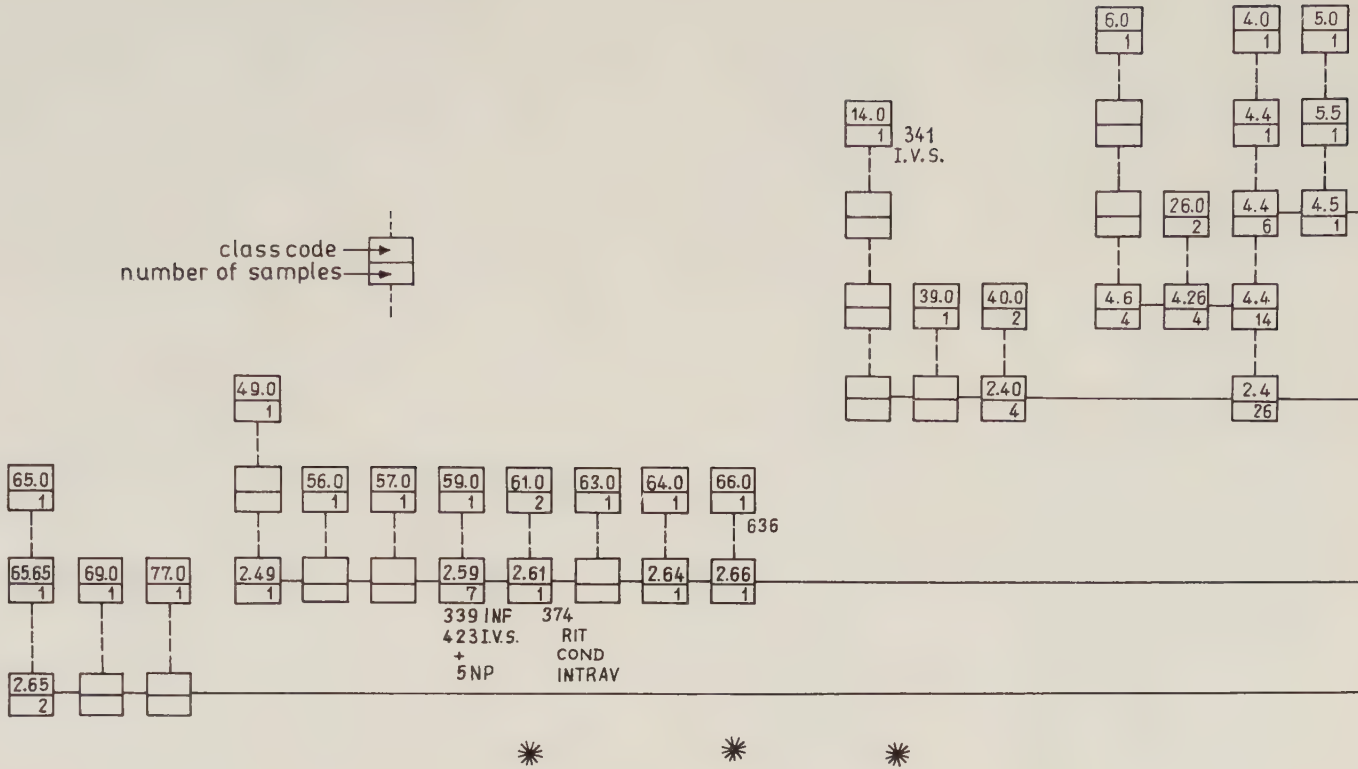

class code
number of samples
65.0 1
65.65 1
2.65 2
69.0 1
77.0 1
49.0 1
2.49 1
56.0 1
57.0 1
59.0 1
2.59 7
61.0 2
2.61 1
63.0 1
64.0 1
2.64 1
66.0 1
2.66 1
636
339 INF
423 I.V.S.
+
5 NP
374
RIT
COND
INTRAV
14.0 1
341
I.V.S.
39.0 1
40.0 2
2.40 4
6.0 1
4.6 4
26.0 2
4.26 4
4.0 1
4.4 1
4.4 6
4.4 14
2.4 26
5.0 1
5.5 1
4.5 1
* * *

FIGURE 3

same work we have done before, by distinguishing three kinds of parameter: (1) geometrical parameters; (2) energetical parameters; (3) morphological parameters. In doing this there are some difficulties connected to the following facts: (1) there are very few chances that the usual vector model is correct; (2) there is a good lot of confusion depending on geometrical representations of points which do not exist; (3) there is not general criterion for choosing a geometrical reference frame for the heart starting from electrocardiograms. We hope to arrive at some constructive results concerning the choice of parameters and class grouping in six or seven months.

References

1. Favella, Reineri, and Righini, "On a mathematical procedure for detecting significant parameters in the classifications of a statistical ensemble of phenomena and its applications", *Kybernetik*, 187–94 (1969),
2. Rosettani, Braguzzi *et al.*, "Proposta di um sistema semplicato di derivazioni convenzionali", *Minerva Med.* (1963).

CHAPTER II–21

The automation of clinical diagnosis

J.H. MITCHELL

Formerly Medical Records Research Group, University of Glasgow, U.K.

Summary

An examination of the possibilities leads to the conclusion that the automation of clinical diagnosis is impossible, and that efforts should be concentrated on other uses of computors in medicine which are already known to be feasible.

"To apply precise methods to vague and ill-defined concepts is likely to perpetuate confusion, by lending to it an air of respectability."

SCADDING, 1968

INTRODUCTION

In the past few decades the patterns of disease have changed dramatically, both quantitatively and qualitatively (Mitchell, 1969), and the total of medical knowledge has become so vast and so complex that the days are long past when any one physician or surgeon could hope to be an expert in all fields. The specialist is truly said to now know more and more about less and less, and most of his personal expertise, which has taken so long to acquire, is lost for ever when he dies.

Again, a great deal of the modern "information explosion" is dissipated in the pages of the multitude of scientific journals which no one has sufficient time, facilities or, indeed, incentive to read, and which are fast becoming too numerous to even index promptly and accurately.

It is not surprising, therefore, that serious consideration is now being given to the storage of medical information in electronic digital computers, and to the possibility, moreover, of these computers collating data already stored within their files and information newly presented to them about the same or other patients, so as to offer physicians and surgeons diagnoses, and perhaps even suggestions regarding treatment.

I propose to examine this attractive concept to try to determine whether it is basically feasible in our present state of knowledge.

TERMINOLOGY

In the most basic and comprehensive terms "illness" in a person may be defined as the reaction of that person to a stressful situation. As the human organism is a biological system, the possible variations in reaction to stress are infinitely numerous, but, in fact, as we all know, certain patterns tend to occur predictably in certain situations, and these patterns are called "diseases". Thus, a high metabolic turnover of uric acid, from whatever cause, will probably result in the deposition of urates in the joints, kidneys, and certain areas of skin, giving rise to acute arthritis, renal calculi, and tophi, the characteristic features of the disease of gout.

Pursuing this example, mention of the term "gout" will bring this picture to the mind of a physician, and the presence in a patient of the three features already mentioned will suggest to the attending physician the possibility of a "diagnosis" of gout, which will be confirmed or refuted by the value of the serum uric acid concentration.

A diagnosis is a succinct comprehensive term which is used to describe the disorder present in a patient, and which indicates, where relevant, the tissue involved, the nature of the abnormality of that tissue, and the cause of the abnormality. For example, "tuberculous meningitis" means an inflammation of the meninges (the membranes covering the brain and spinal cord) by tubercle bacilli, and "non-specific gastro-enteritis" means the presence of vomiting and diarrhoea without any cause being found, either in or outwith the gastro-intestinal tract.

In the end, however, a diagnosis is only useful as a label or headline: "malignant hypertension" is not at all vague in its meaning to a physician, although no actual tissue is specifically mentioned. Moreover (and this must be understood by computer scientists) certain diseases are not what their names might suggest: for example, ulcerative colitis does *not* mean *any* ulcerating inflammatory disorder of the colon, such as bacillary or amoebic dysentery, but a particular and apparently quite distinct disease-entity; and, whilst angina pectoris is a cardiac disorder, agranulocytic angina and herpangina are disorders of the throat. Medical terminology poses a major difficulty to computer programmers (Anderson, 1967; Scadding, 1967).

THE MANAGEMENT OF ILLNESS

A patient presenting with symptoms of illness is questioned and examined by the attending doctor, who, on the basis of the medical history he elicits, and his findings on physical examination, makes a diagnosis and recommends treatment (Figure 1). It may be that special investigations (e.g. bacteriological, biochemical or radiological) are necessary before he can reach a diagnosis of any kind, but very often he forms a tentative opinion even before he begins the physical examination, and he may finally select certain

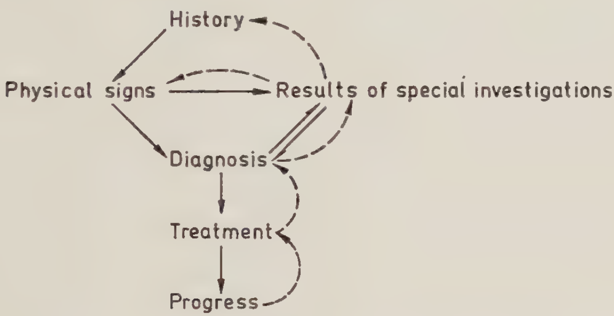


FIGURE 1 The management of illness. The solid lines illustrate the links between the normal stages in the diagnostic and therapeutic processes. The chain lines illustrate the feedback which serves as a validity check to ensure efficient therapy being given to each individual patient

special investigations only to clinch matters beyond reasonable doubt. Thus are diagnoses of renal calculi confirmed by X-rays, and diagnoses of myocardial infarction strongly supported by electrocardiograms and the values of certain serum enzymes.

After treatment has been started, the patient's subsequent progress may be judged by the doctor as unsatisfactory, in the sense that it is contrary in quantity or quality to his (the doctor's) hopes or expectations. He will then review the treatment in case it is actually making matters worse (iatrogenic illness), and, having excluded this possibility, he will reconsider the diagnosis. After that he will, if necessary, review all the data on which the diagnosis was based, i.e. the special investigation results, the physical findings and the history. This feedback in medical management acts as a very efficient validity check to ensure the likelihood of adequate and efficient therapy being given to each individual patient, and it is highly significant that errors and omissions are more often found in the history elicited than in any other part of the diagnostic procedure (Douthwaite, 1956).

DIFFERENTIAL DIAGNOSIS

Very few diagnoses are so obvious and certain that you could, to use a favourite expression of one of my former mentors, "hang your hat on them" (Currie, 1959). The wise clinician, when confronted with a symptom-sign

TABLE I Some examples of common differential diagnoses

Presenting feature	Condition responsible
anterior chest pain	Bornholm disease cervical disk lesion gall stones herpes zoster hiatus hernia lung cancer myocardial ischaemia (with or without infarction) pleurisy pneumothorax rib fracture
severe nervousness	anxiety state brucellosis phaeochromocytoma thyrotoxicosis
excessive thirst and polyuria	chronic nephritis diabetes insipidus diabetes mellitus functional polyuria
severe weight loss	cancer of some kind chronic infection (including tuberculosis) diabetes mellitus malabsorption syndrome thyrotoxicosis

It should be noted that these lists are not exhaustive, and that many of the diagnoses are not final ones: e.g. malabsorption syndrome is a feature of more than 36 diseases.

cluster (syndrome), considers a number of possible diagnoses before he chooses the one he thinks the most likely. Some diseases mimic one another so closely that one is regularly considered in the differential diagnosis of another (Table 1), and some syndromes (e.g. arthritis, anaemia, chronic diarrhoea, jaundice) are both so usual and so complicated that most clinicians tend to clarify them by a fairly standard investigatory approach.

Where definitive diagnostic data commonly remains confusingly incomplete, attempts have been made to employ scoring methods based on probabilities calculated from past experience. A diagnostic formula created for epilepsy (Renfrew, 1958) was entirely pragmatic, but indexes for thyrotoxicosis (Crooks *et al.*, 1959) and arthritis and Hashimoto's thyroiditis (Boyle, 1967) have become increasingly mathematically precise, and a very similar technique has been used by Acheson (1967) for linking medical records, employing weightings both for and against a given conclusion, which are carefully calculated to try to ensure that dangerous mismatching (analogous to misdiagnosis) is particularly unlikely to occur.

The validity of diagnostic scoring has certainly not been generally accepted by practising clinicians, and one obvious objection to the probability factors is that they have been derived from the clinical features of patients in a very small selected population, which may not be at all representative of the community at large. This objection could be met to a great extent by setting up a large computerised data bank, which would serve as an electronic index of differential diagnosis, and which would contain the "profiles" of every known illness (Payne, 1968), but even this would hardly justify the enormous effort and expense involved in its creation, because of the difficulty in keeping it up to date with current medical knowledge and points of controversy, and because it would, as a factual list, prove all too often inferior to an account and discussion in English prose (i.e. a textbook).

PATTERN RECOGNITION

For these very reasons it is already being argued that a computerized data bank should not be fed with information abstracted from the case record as late as when the medical consultation is over and when even some of the results of special investigations may already be to hand. After all, the doctor makes his diagnosis on the basis of pattern recognition. He matches the pattern of what he has been told (the history) and what he has found (the clinical signs) with patterns he has either personally encountered previously or learned about through reading and formal teaching, and the completeness of his match constitutes the certainty of his diagnosis. Suppose initial information about the history of an illness were obtained from a patient by means of a standard questionnaire, and to this were added information derived from a standard physical examination, might not a final and accurate diagnosis be reached by feeding this into a computerized data bank?

This is the concept of automated diagnosis being provided by a "total hospital information system" (Hall, 1967), and the analogy has recently been drawn between identifying the *significance* of a spot cluster on a radar screen in a Ballistic Missile Early Warning Station and that of a symptom-sign cluster on a patient's computerized case record (Boyle, 1967). The computer, it is claimed, would become an increasingly accurate substitute for

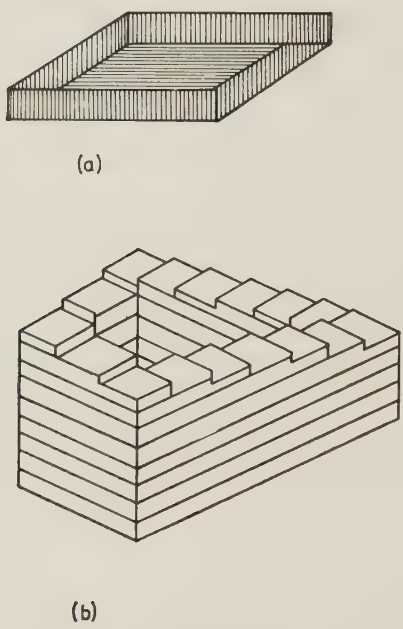


FIGURE 2 Impossible figures: (a) the impossible box; (b) the impossible staircase

human experience as its store (memory) became larger with each input of data, and therefore its diagnoses would become correspondingly more precise.

The basic and highly dangerous assumption in this argument is that, because a computer is superficially similar in a few respects to the human brain, and can perform certain types of calculation and collation very much faster, diagnostic perspicacity can be translated into machine logic, and therefore usefully replaced by it. In other words, the computer program would imitate the processes of the human mind, and this is, I submit, patently absurd.

How can one even begin to explain in concise terms self-evident facts like the difference between right and left, or that certain three-dimensional figures could not physically exist (Figure 2)? Our memories are partly paravisual, so that we can read quickly, and we can instantly recognize a familiar face, even in a photograph, and we can (or so we think) judge a person's age fairly accurately from his appearance. Yet we usually could not *draw* that face from memory on a sheet of paper, nor could we *list* the characteristics of the age pattern we claim to identify so easily.

These faculties are so subtle that in health we take them for granted and undervalue them, yet a patient with a parieto-occipital brain lesion, who loses his right–left discrimination, or his ability to recognize faces, *even his own*, is usually seriously confused as a result, and is very fearful of becoming, or of being thought by others to be insane (*British Medical Journal*, 1962; Kremer, 1958; Mitchell, 1964, 1967).

Diagnostic skill involves the full use of the five natural senses, plus the so-called "sixth sense", by which is meant the reaction aroused by the patient in the attending doctor: this may, indeed, be the most diagnostic of all (Browne and Freeling, 1967). Even the observation that a patient can never answer any question directly can be highly significant (Alvarez, 1967), and a knowledge of games theory is essential in psychiatric practice. The doctor's subjective reactions, empathy not least of all, plus his intuitive skill, are basic requirements in the art of medicine.

DIAGNOSTIC TECHNIQUE

Medicine is not an exact science: not only are emotions important factors in the etiology, course and treatment of an illness, but clinical judgement and ethical considerations are also involved. The medical record of an illness is not, therefore, an objective report on the dysfunction of a biomechanical system, and the use of a standard test procedure, employing branching pattern logic, does not necessarily result in the identification of a lesion (system fault) in medical practice as it *must do* in engineering (Figure 3).

The patient's history is a story which the doctor begins to edit mentally as he hears it, and which he may seek to enlarge in certain respects by further questions in order to clarify the situation. Much of the basis of the final diagnosis lies in the way the patient tells the story, which cannot, therefore, necessarily be reduced to replies to a standard list of questions.

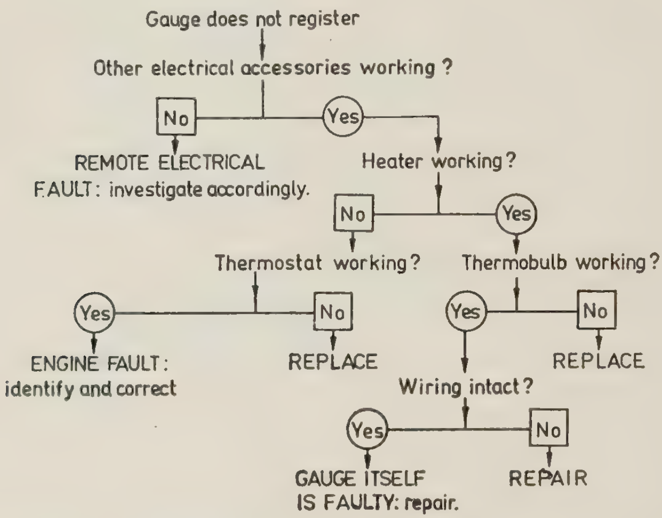


FIGURE 3 A standard test procedure: to determine why the water temperature gauge on the dashboard of an automobile has stopped registering

If the diagnosis proves ultimately to be wrong, it is often because of how the original history was edited by the doctor, or even by the patient, when it was being elicited. Apparently straightforward stories of pain, vomiting, diarrhoea, convulsions, and so on prove to be misleading because of the very way these words are used, so not only may a standard questionnaire be quite insufficient for taking a history, but it may positively encourage semantic confusion.

Similarly, although he conducts a physical examination by a fairly systematic approach, the doctor does not record every possible finding in every patient. He is selective, according to his impression so far gained from the history, and this selection is a measure of his professional skill. He does not, for example, routinely listen for a subclavian bruit on abduction of an arm, although this may be of great diagnostic importance in a case of cervical rib; nor does he routinely assess the visual fields of a patient with a strangulated inguinal hernia.

The facts sought routinely on clinical examination are selected on the basis of probability of return of useful information; beyond that point the yield from a blunderbuss approach is governed by the law of diminishing returns, in terms of both information and time.

The history of an illness and the physical findings in that illness constitute a story, and to try to reduce a story to a series of answers to a list of predetermined questions is to invite the loss of vital information in many instances.

Finally, the medical symptoms and signs of an illness are comprehensively described not in three dimensions but in four (space–time narrative). They cannot, therefore, be thus stored in a computer except in variable field/length format, which makes scoring impossible without preliminary abstraction.

CONCLUSION

It seems to me, therefore, that automated diagnosis is, like the philosopher's stone and perpetual motion, a beguiling but impossible dream. It would be better if this were recognized now, so that our energy, money and time were concentrated on the application of computers to control machine systems in biochemistry, anaesthesia, renal dialysis and intensive-care monitoring, and to analyse wave forms in electrocardiography and electro-encephalography, where feasibility has already been proved and the potential gains are enormous.

References

E.D. Acheson, *Medical Record Linkage*. Oxford University Press, London (1967).

W.C. Alvarez, "Patients who cannot answer a question", An Editorial. In *Modern Medicine*, October, p.823 (1967).

J. Anderson, Private communication (1967).

J.A. Boyle, Private communication (1967).

British Medical Journal "Blank faces", An Editorial. **1**, 1748 (1962).

K. Browne, and P. Freeling, *The Doctor–Patient Relationship*. Livingstone, Edinburgh and London (1967).

J. Crooks, I.P.C. Murray, and E.J. Wayne, "Statistical methods applied to the clinical diagnosis of thyrotoxicosis". *Quart. J. Med.*, **28**, 211 (1959).

J.P. Currie, Private communication (1959).

A.H. Douthwaite, "Pitfalls in medicine." *Brit. Med. J.*, **2**, 895 (1956).

P. Hall, "Automated diagnosis." *Joint B.M.A.–I.M.A. Symp., London* (1967).
M. Kremer, "Sitting, standing and walking." *Brit. Med. J.*, **2**, 63 (1958).
J.H. Mitchell, "Pneumococcal meningo-encephalitis." *Postgrad. Med. J.*, **40**, 97 (1964).
J.H. Mitchell, Unpublished data (1967).
J.H. Mitchell, *A New Look at Hospital Case Records*. H.K. Lewis, London (1969).
L.C. Payne, Private communication (1968).
S. Renfrew, Private communication (1958).
J.G. Scadding, "Diagnosis: the clinician and the computer." *Lancet*, **2**, 877 (1967).
J.G. Scadding, "The clinician and the computer." *Lancet*, **1**, 140 (1968).

CHAPTER II–22

*Psychosynthesis: TV-linked cybernetic system psychotherapy**

H. C. TIEN

Medical Center West, Lansing, Michigan, U.S.A.

"It perhaps explains, too, why there are circumstances where a joint use of shock treatment and psychotherapy is indicated, combining a physical or pharmacological therapy for the phenomena of reverberation in the nervous system, and a psychological therapy for the long-time memories which, without interference, might re-establish from within the vicious circle broken up by the shock treatment."

Cybernetics, WIENER, 1961

Summary

Psychosynthesis is a TV-linked cybernetic system of psychotherapy. Psychosynthesis is a unified theory of energy transformation. In theory, psychosynthesis is based on the unified concept of personality transformation of brain-mind equivalence in the same sense of the mass-energy equation of Einstein. In practice, psychosynthesis is the cybernetic union of psychoanalysis of Freud, behaviour therapy of Pavlov, electrotherapy of Cerletti and information theory of Shannon linked by modern television technology as a cybernetic process of Wiener. And so, the psychosynthesis model unites all schools of psychoanalysis, psychotherapy and electrochemotherapy into a cybernetic system with a goal orientation of the individual seeking his identity in the community.

With the psychosynthesis model, the author has fused the eclectic elements of modern psychiatry into a multiple TV-linked feedback system, composed of interview psychotherapy, couch analysis, family therapy, drug therapy, electrolytic therapy (E.L.T.), group therapy and super-group therapy as

* This paper relates to a scientific exhibit at the Congress (editor).

on-going dynamics of mental health mini-community of 50 families. In this system, the psychiatrist is also being transformed into a family psychiatrist, who could potentially work simultaneously with 100 families or more as the psychosynthesist of a mental health mini-community.

INTRODUCTION

The manner in which psychiatrists received the above theoretical statement may well be summarized by Franz Alexander[1] in his historical survey of psychiatry up to 1968: "The question how far the neurotic machines of Norbert Wiener can be compared with a neurotic personality offers interesting areas for further research. As of the moment, the positive-feedback theory must remain in the realm of speculation." In 1961, the author[2] as a practising psychiatrist, took up this speculation to develop a cybernetic system of psychotherapy. The result is "psychosynthesis"[3].

HISTORIAL BACKGROUND

Psychosynthesis is the meaningful union of the eclectic elements of modern psychiatry into a cybernetic system. The growth of this cybernetic system began with the fusion of psychoanalysis and cybernetics. Freud[4] (1886) tried Erb's electrotherapy and abandoned it for hypnosis, which in turn led to psychoanalysis. However, electrotherapy returned in 1938, when Cerletti introduced electric shock treatment (E.S.T.), which use has been somewhat restricted in the treatment of severe depressions, especially since the advent of drug therapy in midst of mushrooming of psychotherapies. The conflicts and contradictions have been most bewildering for all, professionals, patients and their families alike! To answer all these contending claims, the author gradually built the eclectic elements into a system based on a unified theory of energies which was predicted by an Italian psychiatrist, Assagioli[5], who said of his classical psychosynthesis that: "At this stage, since we do not have scientific instruments which enable us to measure these energies directly, we still have to rely on essentially a phenomenological position, in the sense of insisting on the experience itself, and hoping that sooner or later—maybe not in the lifetime of the author, science will attack this problem on a rigorous 'energy' basis." Zweig[6] (1933) made a similar prediction when he discussed Freud and psychoanalysis: "Freud has done marvels, but there remain other marvels still to do." He also said: "The analysis and enlightenment it effects must have superadded to them a conjoining and fusing technique; psychoanalysis must be supplemented by psychosynthesis. Perhaps this will be achieved by the science of tomorrow." And tomorrow is now today! The basic technology to develop the science of psychosynthesis is here.

CURRENT TECHNOLOGY

Psychosynthesis is still at its newborn state, but the current technology will speed up its growth and development as never before. Today, we view on television that man has reached the Moon; meanwhile, more quietly, we are also witnessing on television the re-making of human personality at will, as desired, within certain limits, of course. The fact that it can be done at all should make us pause. The same science and technology of outer space are giving rise to the third psychiatric revolution of the inner space: the transformation of human mind on demand. The modern techniques of electrolytic therapy (E.L.T.), chemotherapy, and psychotherapy may be linked by television as the recording and feedback instruments to realize the psychosynthesis model, which has fused and integrated the eclectic elements of psychiatry into a new science, as hoped for and predicted by Assagioli[5] and Zweig[6]. The state of current technology has made the model of psychosynthesis practicable anywhere. Let us look at the model.

THE MODEL OF PSYCHOSYNTHESIS

The psychosynthesis model can be simply represented as follows:

$$\underbrace{E_i(E_s) \rightarrow E_o}_{\uparrow\_\_\_\_\_\_\_\_\_|}$$

E_i = input energy pattern. E_s = stored energy pattern, i.e. system function of both genetic and experiential information. E_o = output energy pattern, i.e. thinking or behaviour. And the energy function of E_i, E_s, or E_o is the same E as in the mass–energy equivalence equation of Albert Einstein: $E = mc^2$. The author[3] views mind as mass in relative motion and brain as energy at relative rest. Consciousness is a single frame of electrical charges in motion, like electrons bombarding a television screen. And personality is a time series of the scintillating frames of consciousness. Personality, therefore, becomes a reverberating input–output pattern of self-creation, seeking information or patterns of energy from the environment as well as from its own memories. The stability of any given personality is its identity, which is maintained by feedback upon *the principle of most-similarity*[7]. In other words, the personality never recreates itself, but only a pattern most similar to itself. The phenomenon of unique individuality and personal continuity depends on memory. The psychosynthesis model thus provides us with a clue as to how a personality may be radically transformed. The Cartesian dictum of "I think, therefore I am!" is sharpened to the dictum of psychosynthesis: "I remember, therefore I am!" Here is the clue: if the memory of "I" is loosened or somewhat erased, then the personality of "I" is changed. Consciousness is the most recent memory, which is most subject to erasure and loosening. Personality transformation becomes energy pattern modification of not only scintillating consciousness, but also of recent circulating and old stored memories of childhood.

THE TRANSFORMATION OF PERSONALITY

Hithertofore, all the classical theories of psychoanalysis and psychotherapies and psychosynthesis have one thing in common, the adding of information into the brain in order to modify memories and hopefully to transform personality. However, the psychosynthesis model demands the *substraction* of information from the brain as well the *addition* of information as used by the classical theories of all schools of thought. If personality is a time series of patterns, then its transformation may be achieved by adding and also by taking away information. Addition of information may be accomplished by verbal or visual input; and subtraction of information may be effected by electrical or perhaps by chemical means. It is now clear that the transformation of personality is the differential modification of energy patterns. Energy patterns may be modified verbally, visually, chemically, or electrically, or any combinations thereof, without contradictions of the contending schools of psychiatric thought, whether organic or psychogenic in formulation. The electrolytic therapy, by way of mutation from the electric shock therapy of Cerletti, became a conscious instrument to take information away from the brain in order to permit a new pattern of personality to emerge. Let us describe the electrolytic therapy briefly.

THE ELECTROLYTIC THERAPY (ELT)

The E.L.T. session is divided into three parts. (1) The *prelytic step*, in which the undesirable personality is activated of its memories, whether it be depressive, neurotic, psychopathic, or psychotic in content, to full consciousness either verbally, or with written materials of life history or by videotapes of analytical couch sessions. (2) The *lytic step*, in which the patient receives actual electrical current at the height of such activation for the purpose of loosening consciousness and erasing memories. And (3) the *postlytic step*, in which the patient is immediately reprogrammed with new information for a new personality in this semi-conscious or infant-like state by parents or spouse as the case may be. The reprogramming is usually begun with bottle feeding and sometimes with a new first name. This is *transnomation*. (Later on, the patient's birth certificate is also changed legally with thir new name of rebirth, as if it were.) Here we have the theoretical union of Pavlovian[8] conditioning and Freudian transference. The electrolytic process can best be illustrated by an actual clinical history.

THE PSYCHOSYNTHESIS OF KATE

The psychosynthesis of Kate is the transformation of Lana into Kate. Lana was a 16-year old delinquent girl, who was sexually involved with older boys since age 14. She came under the probate judge's recommendation because of antisocial acting out, skipping school, and other legal problems. She was given 10 E.L.T. sessions. In each she was reactivated of her childhood fears and conflicts, parental divorce, father's alcoholic and sexual episodes, mother's neglect of the family and her sexual involvement with an irresponsible and probably an antisocial man for selective loosening and erasure, so that she could be transformed into Kate as she desired. In sum,

(Electrolysis)
↑ ↑ ↑ ↑ ↑
(psychoanalysis) Lana → 10 E.L.T.'s → Kate (psychosynthesis)
↓ ↓ ↓ ↓ ↓
(Catalysers: parents)

The divorced parents served as postlytic programmers (catalysers) to reprogram their daughter into Kate, a more responsible, more stable, happier girl of self-respect. Having achieved her desirable personality, Kate also joined the TV group for psychosynthesis in which her new personality which is known as Kate was reinforced.

RESULTS

From 1962 to 1967, the author treated 66 patients with this evolving electrolytic therapy, based on the psychosynthesis model. Fifty-six of the original sample (or 85%) emerged as new, happier and more stable personalities like Kate, on follow-up for about three years. Kate represents one of the second series of 50 families, treated with a more elaborate TV-linked cybernetic system of psychotherapy. On preliminary follow-up, 90% of these 50 families report the improvement of the patient's emotional stability, work performance and social functioning. The TV-linked cybernetic system of psychosynthesis groups began in May 1968 because of the need to have a new social group to re-inforce and strengthen the identity of a new personality with a new name (e.g. Kate). Three psychosynthesis groups developed this way in the author's private practice of psychiatry. The three groups also

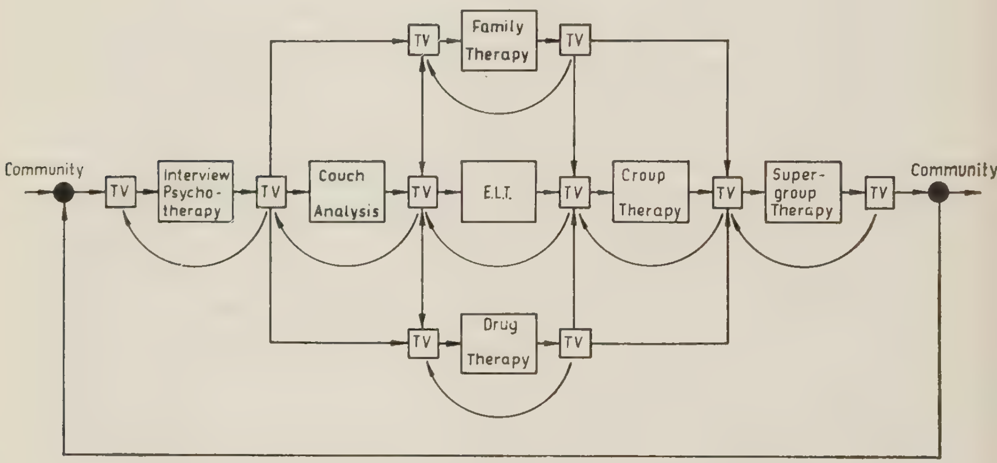


FIGURE 1 Television-linked cybernetic system of psychosynthesis

became linked by television and as a result, a super-group for psychosynthesis was formed. The super-group became a mental health mini-community, in which the newly transformed personalities found themselves happier, more productive, healthier and more meaningful in their everyday life. (See Figure 1.)

References

1. F.G. Alexander and S.P. Selesnick, *The History of Psychiatry*, The New American Library, New York (1968).
2. H.C. Tien, "Transformative hypnodynamics", *J. Nervous Mental Disease*, **133**, 497 (1961).
3. H.C. Tien, "Psychosynthesis: a cybernetic system of psychotherapy", *6th Ann. Rocky Mountain Bioeng. Symp., Laramie, Wyoming* (1969), pp. 78–81.
4. S. Freud, *The Basic Writings of Sigmund Freud*, Random House, New York (1938).
5. R. Assagioli, *Psychosynthesis*, Hobbs, Dorman and Co., *Company, Inc.*, New York (1965), p. 194.
6. G. Lauzun, *Sigmund Freud*, Fawcett World Library, New York (1966), p. 215.
7. H.C. Tien, "Pattern recognition and psychosynthesis", *Am. J. Psychother.*, **23**, 1 (1969).
8. I.P. Pavlov, *Experimental Psychology and Other Essays*, Philosophical Library, New York (1957).
9. N. Wiener, *Cybernetics*, M.I.T. Press and Wiley, New York (1966).

John Rose is the principal of a large and important college of technology. He has spent many years in industry and in education, and has acted as a part-time industrial consultant. His interests range from the history of science to computing and cybernetics. He has written or edited eight books as well as many papers and essays. He founded the Institution of Computer Sciences (U.K.) and was elected the Honorary Secretary of the International Cybernetics Congress Committee (I.C.C.C.) and General-Organiser of the First International Congress of Cybernetics, London, 1969. In addition, he is joint chief editor of the *International Journal of Bio-medical Computing,* and will also edit a new international journal of environmental sciences. Dr. Rose won the first prize in the 1966 International Shell/British Association Competition with an essay entitled 'The problem of technological barriers'. He is also a member of several international committees and societies.

GORDON AND BREACH SCIENCE PUBLISHERS

150 Fifth Avenue, New York, N.Y. 10011
12 Bloomsbury Way, London W.C.1
7-9 rue Emile Dubois, Paris 14e

0 677 14310 9